Scoring Strategic Agents*

Ian Ball†

31 January 2020

[Click for latest version]

Abstract

I introduce a model of predictive scoring. A receiver wants to predict a sender's quality. An intermediary observes multiple features of the sender and aggregates them into a score. Based on the score, the receiver takes a decision. The sender wants the most favorable decision, and she can distort each feature at a privately known cost. I characterize the most accurate scoring rule. This rule underweights some features to deter sender distortion, and overweights other features so that the score is correct on average. The receiver prefers this score to full disclosure because the aggregated information mitigates his commitment problem.

Keywords: scoring, multidimensional signaling, screening, intermediation
JEL Codes: C72, D82, D83, D86.

*This paper was previously circulated with the title "Incentive-Compatible Prediction."

†Yale University, Department of Economics, ian.ball@yale.edu. For continual guidance, I thank my advisor Dirk Bergemann and my committee members Larry Samuelson and Johannes Hörner. For helpful discussions, I thank Nageeb Ali, Isiah Andrews, Simon Board, Alessandro Bonatti, Tilman Börgers, Hector Chade, Gonzalo Cisternas, Joyee Deb, José-Antonio Espín-Sánchez, Mira Frick, Marina Halac, Rick Harbaugh, Tibor Heumann, Ryota Iijima, Deniz Kattwinkel, Jan Knoepfle, Soonwoo Kwon, Patrick Lahr, Stephan Lauermann, Ro'ee Levy, Xiangliang Li, Heng Liu, Niccolò Lomys, Erik Madsen, George Mailath, Chiara Margaria, Idione Meneghel, Meg Meyer, Weicheng Min, Xiaosheng Mu, Miquel Oliu-Barton, Juan Ortner, Yujie Qian, Tim Roughgarden, Anna Rubinchik, Grant Schoenebeck, Eran Shmaya, Nicholas Snashall-Woodhams, Suk Joon Son, Rani Spiegler, Philipp Strack, Juuso Välimäki, Nisheeth Vishnoi, Allen Vong, Conor Walsh, Xinyang Wang, and Weijie Zhong.

1 Introduction

1.1 Motivation and results

As data sources proliferate, predictive scores are increasingly used to guide decisions. Banks use credit scores to set the terms of loans; judges use defendant risk scores to set bail; and online platforms score sellers, businesses, and job-seekers. We now live in a "scored society" (Citron and Pasquale, 2014). These scores have a common structure: An intermediary gathers data about an agent from different sources and then aggregates the agent's features into a score. For example, a FICO credit score predicts a consumer's creditworthiness from multiple features including credit utilization rate and length of credit history.[1]

Predictive scoring is not only a statistical problem: strategic manipulation poses an additional challenge. An agent who understands that she is being scored can distort her features to improve her score, without changing her quality. For example, a consumer can spread her spending across multiple credit cards to lower her credit utilization rate, without reducing her risk of default. "The scoring models may not be telling us the same thing that they have historically," according to Mark Zandi, chief economist at Moody's, "because people are so focused on their scores and working hard to get them up."[2] In general, when scores are introduced to guide high-stakes decisions, people learn to manipulate them.[3] In the presence of such strategic behavior, what scoring rule induces the most accurate decisions?

To answer this question, I build a model of predictive scoring. There are three players: sender, intermediary, and receiver. The receiver wants to predict the *quality* of the sender. The intermediary observes multiple manipulable *features* of the sender, and commits to a rule for aggregating these features into a score (from an unrestricted score set). The receiver sees this score and takes a decision. He wants his decision to match the sender's quality. The sender wants the most favorable decision, and she can distort each of her features at a cost. For each feature, the sender has two dimensions of private information: Her *intrinsic level*, which correlates with her quality, is the value of the feature if she does not distort it; her *distortion ability* parameterizes her

[1] FICO claims that its scores are used in 90% of U.S. lending decisions (`https://ficoscore.com`).

[2] "How More Americans Are Getting a Perfect Credit Score," *Bloomberg*, August 14, 2017.

[3] For instance, hospitals admit healthy patients to improve their scorecards, and law schools hire their own graduates to boost their US News rankings. For a discussion of these examples and others, see Ederer et al. (2018).

cost of distortion.

The intermediary designs the scoring rule to minimize the mean squared error between the receiver's decision and the sender's quality. If the sender's features were exogenous, then predicting her quality would be a purely statistical problem. Instead, the intermediary must consider how the scoring rule motivates the sender to distort her features. Formally, each scoring rule induces a different game between the sender and receiver.

I first consider a *signaling* setting in which the receiver observes the sender's features. In the resulting game between the sender and the receiver, the sender's features serve as signals of her quality. The interpretation of this benchmark is that the intermediary fully discloses the sender's features. Therefore, this setting gives a lower bound on the performance of optimal scoring.

In the signaling game, the sender's distortion can interfere with the receiver's prediction of the sender's quality. I first show that the signaling game has exactly one equilibrium in linear strategies (Theorem 1). In this equilibrium, the amount that the sender distorts each feature depends on her distortion ability. In the special case in which distortion ability is homogeneous in the population, every sender type distorts her features by the same amount. The receiver anticipates the sender's distortion on each feature and subtracts it to determine the sender's intrinsic level. But in general, with heterogeneous distortion ability, each feature confounds the sender's intrinsic level with her distortion ability. The receiver cannot distinguish a sender with a high intrinsic level and low distortion ability from a sender with a lower intrinsic level but higher distortion ability.

Next, I return to the main *scoring* setting in which the receiver learns about the sender's quality only through the intermediary's score. While more information always helps a decisionmaker acting in isolation, here the coarsening of information improves the receiver's predictions by mitigating a commitment problem. To illustrate this problem, suppose that the receiver tried to discourage distortion by making his decision less sensitive to the sender's features. As the sender distorts less, her features become more informative about her quality, and hence the receiver would want to react to them fully.

To isolate the effect of the intermediary, I focus on linear scoring rules. I give a necessary and sufficient condition on the sender's type distribution under which the optimal linear scoring rule yields strictly more accurate decisions than the signaling

equilibrium (Theorem 2). My condition holds for generic covariance parameters. It only rules out the symmetric case in which the sender's distortion in the signaling equilibrium equally reduces the informativeness of each feature. The optimal scoring rule underweights some features to deter sender distortion. A feature is underweighted if on that feature, distortion ability is more heterogeneous or the intrinsic level is more informative of quality (Theorem 3). On these features, dampening distortion has the greatest informational benefit. The optimal rule overweights other features so that the score remains correct on average. If the receiver could observe the sender's features ex post, he would change his decision, but from the score alone he cannot disentangle the value of each feature.

Finally, I consider a *screening* setting in which the receiver can commit to his decision. The receiver can in particular ignore information, so it is optimal for the intermediary to fully disclose the sender's features. Thus, the receiver commits to a decision as a function of these features. To isolate the effect of commitment, I focus on linear rules. Unlike in the scoring setting, the receiver can reduce the weight on every feature simultaneously. As commitment increases—from no commitment (signaling) to information commitment (scoring) to decision commitment (screening)—the receiver's decision becomes less sensitive to the sender's features, thus reducing the informational loss from distortion (Theorem 4). Naturally, screening improves the receiver's payoff relative to the scoring, but it also improves the sender's payoff, on average, by lessening the sender's burden to distort her features.

Outline Section 1.2 reviews related literature. Section 2 presents the model of predictive scoring. Section 3 analyzes the signaling setting without the intermediary. In Section 4, I study the intermediary's scoring problem. I characterize which decisions the intermediary can induce, and I optimize over this set. In Section 5, I analyze the screening setting and then I compare the solutions in the three settings—signaling, scoring, and screening. Section 6 extends the baseline model to consider stochastic scores and a more general social welfare objective. The conclusion is in Section 7. Proofs are in Appendix A. Additional results are in Appendix B.

1.2 Related literature

In my model, the sender distorts her features to influence the receiver's beliefs, and the intermediary designs the receiver's information about the sender's features. Here

I relate my model to the two closest literatures—signaling and information design.

Signaling In Spence (1973) and many subsequent models of signaling, the sender's type satisfies a single-crossing condition so there exists a fully separating equilibrium. Thus, there is no opportunity for an intermediary to improve information transmission. The starting point for my analysis is a signaling environment without a separating equilibrium. Specifically, I consider a multi-feature extension of the elliptical setting in Frankel and Kartik (2019a), which in turn builds upon the Gaussian settings in Bénabou and Tirole (2006), Fischer and Verrecchia (2000), and Prendergast and Topel (1996). Incorporating multiple features allows me to study the intermediary's signal aggregation problem, which is my primary contribution. What makes intermediation interesting in my setting is the double multi-dimensionality—multiple features and multiple type dimensions per feature.[4]

In my model, the three commitment settings—signaling, scoring, and screening—generally yield different solutions. Frankel and Kartik (2019b) compare signaling and screening in the single-feature setting from Frankel and Kartik (2019a). They illustrate that the receiver benefits from committing to under-react to the sender's feature.[5] In my multi-feature model, I show that when the receiver cannot commit, an informational intermediary can provide partial commitment power by aggregating these features into a score.[6] A computer science literature studies a related screening problem termed strategic classification (Dalvi et al., 2004; Brückner and Scheffer, 2009; Hardt et al., 2016). The receiver is a Stackelberg leader and takes a binary decision. Into the strategic classification setting, Hu et al. (2019) introduce heterogeneous distortion ability. They study how this heterogeneity affects the distribution of welfare across different sender types, while I focus on the informational loss for the receiver.

Scoring itself has received little theoretical analysis. The closest paper on scoring is Bonatti and Cisternas (forthcoming). There, a sequence of monopolists price-discriminate using a score that aggregates noisy measurements of past purchase quantities. While the focus is on consumer welfare, they also show that the most

[4]The classical work on signaling with multiple signals (Engers, 1987; Quinzii and Rochet, 1985) focuses on fully separating equilibrium.

[5]In a different setting with binary actions, Cunningham and Moreno de Barreda (2019) show that a receiver can benefit by committing to a more demanding threshold for the desired action.

[6]Crémer (1995) shows in a moral hazard setting that a principal may prefer a less accurate monitoring technology so that she finds it sequentially rational to refuse to renegotiate.

informative score is more persistent than the ex-post optimal prediction. Persistence discourages strategic quantity reduction, thus increasing the signal-to-noise ratio. My result holds in a noiseless static setting and is driven by the different informativeness and manipulability of different features.[7]

While I study signal aggregation, other papers focus on adding noise to a single signal. Rick (2013) studies the effect of garbling a costly signal. He gives conditions under which transmitting the signal over a noisy channel can improve social welfare. Whitmeyer (2019) shows that in a binary signaling game, the receiver-optimal garbling coincides with the receiver's full commitment solution. In particular, this implies that in strategic classification, an intermediary can achieve the screening outcome.

The intermediary in my model privately observes payoff-relevant choices by the sender, unlike a classical mediator who elicits cheap talk reports (Aumann, 1974; Myerson, 1982, 1986; Forges, 1986). While a classical mediator can be consulted in any strategic environment, my intermediary has a private monitoring technology and hence must be modeled as a primitive of the environment.

Information design In the growing literature on information design (for surveys, see Kamenica, 2019; Bergemann and Morris, 2019), the designer controls information about an exogenous state. The designer's information policy influences the beliefs of a single agent (Kamenica and Gentzkow, 2011) or of multiple agents playing a simultaneous-move game (Bergemann and Morris, 2013, 2016). But the information policy does not generally affect the state distribution. In my model, the intermediary's scoring rule changes the distribution of the sender's features. This channel is crucial: The intermediary has the same preferences as the receiver, so if the sender's features were exogenous, full disclosure would be trivially optimal.

The papers on information design with an endogeneous state focus on moral hazard, without private information. In Boleslavsky and Kim (2018), Rodina and Farragut (2016), Mekerishvili (2018), and Zapechelnyuk (2019), the designer simultaneously persuades the receiver and motivates the sender to exert effort.[8] In a dynamic setting, Rodina (2016) and Hörner and Lambert (forthcoming) solve for

[7] Duffie and Dworczak (2018) study the design of financial benchmarks that are robust to manipulation. In their model, different banks hold different positions, so they manipulate in different directions. Unlike my paper, Duffie and Dworczak (2018) do not model how the receiver strategically responds to the benchmark.

[8] Perez-Richet and Skreta (2018) study persuasion within a different model of distortion. Distortion is costless, but the receiver observes the rate of distortion.

effort-maximizing feedback in Holmström's (1999) career concerns model. In all these papers, the objective is to increase the sender's return to productive effort; in my model, the objective is to dampen the sender's return from unproductive distortion.

2 A model of predictive scoring

There are three players. The agent being scored is called the sender (she). An intermediary commits to a rule that maps the sender's features into a score. A receiver (he) observes this score and takes a decision.

The receiver wants to predict the sender's *quality* $\theta \in \mathbf{R}$. The receiver takes a decision $y \in \mathbf{R}$, and his utility u_R is given by

$$u_R = -(y - \theta)^2.$$

Hence the receiver matches his decision with his posterior expectation of θ.

The sender has k manipulable *features*, labeled $j = 1, \ldots, k$, where $k \geq 1$. For each feature j, the sender privately knows her *intrinsic level* $\eta_j \in \mathbf{R}$ and her *distortion ability* $\delta_j \in \mathbf{R}_+$. As a mnemonic, η is for *in*trinsic and δ is for *d*istortion. The distributions are formally specified below, but think of the intrinsic levels as being correlated with quality θ. The sender does not observe her quality directly, though her type $(\eta, \delta) = (\eta_1, \ldots, \eta_k, \delta_1, \ldots, \delta_k)$ may reveal it.[9]

For each feature j, the sender chooses distortion $d_j \in \mathbf{R}$; feature j takes the value

$$x_j = \eta_j + d_j.$$

The sender's utility u_S is given by

$$u_S = y - (1/2) \sum_{j=1}^{k} d_j^2/\delta_j.$$

The sender wants the decision y to be high, and she experiences a quadratic cost from distorting each feature. Her marginal cost of distorting feature j decreases in her distortion ability δ_j. If $\delta_j = 0$, she cannot distort feature j, so $x_j = \eta_j$.[10]

[9] The equilibria I construct would remain equilibria if the receiver observed her own quality.

[10] For $\delta_j = 0$, the sender's utility is defined by the limit as δ_j converges to 0 from above.

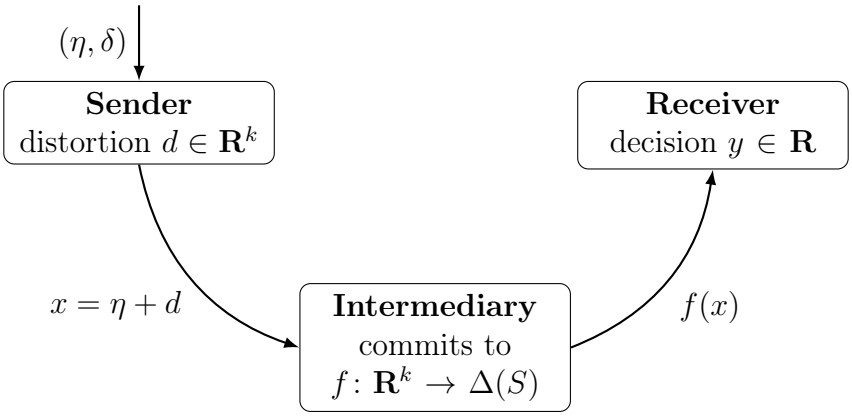


Figure 1. Flow of information

The random $(1+2k)$-vector (θ, η, δ) has an elliptical distribution with finite second moments. The elliptical distribution generalizes the multivariate Gaussian. It is flexible enough to accommodate the sign restriction on δ, yet it retains the convenient property that conditional expectations are linear. This property is stated formally in Section 3.1. I make further covariance assumptions in Section 3.1 and in Section 3.4.

Unlike the sender and receiver, the intermediary has commitment power. Before observing the sender's features, the intermediary commits to a score set S and a scoring rule

$$f\colon \mathbf{R}^k \to \Delta(S),$$

which assigns to each feature vector x a (stochastic) score $f(x)$ in S. Here $\Delta(S)$ is the set of probability measures on S, but I use $f(x)$ to denote the random score itself.[11] The score set S is not restricted. I will show, however, that there is no loss in taking S equal to $\mathbf{R}$.

The intermediary has the same utility function as the receiver. Therefore, the intermediary designs the score to make the receiver's decision most accurate (in the sense of minimizing mean squared error). An interpretation is that the intermediary is a monopolist who sells a scoring service to the receiver and secures a share of the surplus, but this is not modeled.

The intermediary's scoring rule induces a game between the sender and receiver. Figure 1 illustrates the flow of information in this game. The sender observes her

[11] All measurability issues are handled in Appendix B.9.

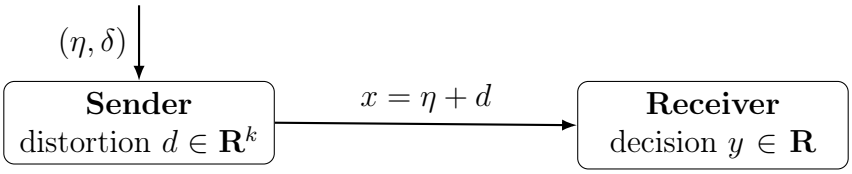


Figure 2. Signaling setting—no intermediary

private type (η, δ) and then chooses how much to distort each feature. Her distorted feature vector x is observed by the intermediary. The intermediary assigns the score $f(x)$ and passes it to the receiver. The receiver updates his beliefs about the sender's quality θ and takes a decision y. Finally, payoffs are realized.

3 Signaling without the intermediary

For this section, suppose that the receiver observes the sender's features, without the intermediary. Thus, the sender and receiver play a signaling game, with features as signals. Figure 2 shows the flow of information in this game. This signaling setting gives a lower bound on the receiver's utility under optimal scoring: The intermediary can replicate the signaling game by fully disclosing the sender's features.

In this section, I first discuss the linear conditional expectation property of elliptical distributions. Then I use this property to construct equilibria.

3.1 Linear conditional expectations

The elliptical distribution generalizes the multivariate Gaussian. Recall that a (non-degenerate) ellipse in $\mathbf{R}^p$ can be expressed as

$$E(\mu, \Sigma, r) = \{x \in \mathbf{R}^p : (x - \mu)^T \Sigma^{-1} (x - \mu) = r^2\},$$

for some center μ, radius r, and full-rank matrix Σ. A Gaussian distribution with mean μ and full rank covariance matrix Σ has two properties: (i) for each radius r, the density is constant around the ellipse $E(\mu, \Sigma, r)$; and (ii) on rays extending from μ, the density decays exponentially in the square of the radius.

Elliptical distributions retain property (i) but relax property (ii) to allow for any

normalized radial density function.[12] In particular, if the radial density function vanishes beyond a certain value, then the distribution is supported within some ellipse, as in the example of a uniform distribution on the interior of an ellipse. Other common elliptical distributions include the Student and Laplace distributions and the stable distributions.[13]

Denote the mean and variance of (θ, η, δ) by

$$\mu = \begin{bmatrix} \mu_\theta \\ \mu_\eta \\ \mu_\delta \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_\theta^2 & \Sigma_{\theta\eta} & \Sigma_{\theta\delta} \\ \Sigma_{\eta\theta} & \Sigma_{\eta\eta} & \Sigma_{\eta\delta} \\ \Sigma_{\delta\theta} & \Sigma_{\delta\eta} & \Sigma_{\delta\delta} \end{bmatrix}.$$

I do not specify the radial density function, as it will not play a role in my analysis. The radial density is implicitly restricted, however, by the nonnegativity of δ. Next I impose nondegeracy conditions on the covariance. Denote the Moore–Penrose inverse of a matrix M by $M^\dagger$.[14] Random vectors W and Z are *uncorrelated* if $\operatorname{cov}(W, Z)$ equals the zero matrix; otherwise, W and Z are *correlated*.

Assumption A (Nondegeneracy)

(i) $\Sigma_{\eta\eta} - \Sigma_{\eta\delta}\Sigma_{\delta\delta}^\dagger\Sigma_{\delta\eta}$ has full rank.

(ii) (η, δ) and θ are correlated.

Assumption A holds for a generic choice of the covariance matrix Σ. Assumption A.i ensures that the conditional variance $\operatorname{var}(\eta|\delta)$ has full rank almost surely.[15] Assumption A.ii rules out the trivial case in which the sender's type (η, δ) is uncorrelated with her quality θ.

To state the linear conditional expectation property, I introduce notation for linear regression. Given a random variable Y and a random p-vector X, the regression

[12] In fact, elliptical distributions permit distributions without density. For the general definition in terms of quality functions, see Appendix B.4.

[13] These distributions have full support, but the mean and covariance parameters can be chosen so that δ is nonnegative with arbitrarily high probability. If we truncate the distribution by setting the density to zero whenever it is below a fixed tolerance, we get a new elliptical distribution with the property that δ is nonnegative.

[14] The Moore–Penrose inverse extends the usual matrix inverse from the domain of invertible square matrices to all (possibly nonsquare) matrices. See Meyer (2000, pp. 422–424).

[15] If (η, δ) had a Gaussian distribution with the same covariance matrix Σ, then the conditional variance $\operatorname{var}(\eta|\delta)$ would be nonrandom and equal to $\Sigma_{\eta\eta} - \Sigma_{\eta\delta}\Sigma_{\delta\delta}^\dagger\Sigma_{\delta\eta}$. For elliptical distributions, the conditional variance matrix is random, but every realization is a scalar multiple of $\Sigma_{\eta\eta} - \Sigma_{\eta\delta}\Sigma_{\delta\delta}^\dagger\Sigma_{\delta\eta}$ (Cambanis et al., 1981, Corollary 5).

problem is to choose a coefficient vector b in $\mathbf{R}^p$ and an intercept b_0 in $\mathbf{R}$ to minimize the expected square error

$$\mathbf{E}\big[(Y - b_0 - b^T X)^2\big].$$

Provided that X and Y are square integrable and $\mathrm{var}(X)$ has full rank, the minimizers are unique and given by

$$b^\star = \mathrm{var}^{-1}(X)\,\mathrm{cov}(X, Y), \qquad b_0^\star = \mathbf{E}[Y] - (b^\star)^T\,\mathbf{E}[X].$$

Denote the p-vector of regression coefficients by

$$\mathrm{reg}(Y|X) = \mathrm{var}^{-1}(X)\,\mathrm{cov}(X, Y).$$

This is a column vector; its transpose is denoted $\mathrm{reg}^T(Y|X)$.

Lemma 1 (Linear conditional expectations)
Let $X = A\eta + B\delta$ for some matrices A and B, where A has full row rank. We have

$$\mathbf{E}[\theta|X] = \mu_\theta + \mathrm{reg}^T(\theta|X)(X - \mathbf{E}[X]).$$

The right side is the regression of θ on X, so it is the best *linear* prediction of θ from X. The left side is the conditional expectation of θ given X, so it is the best prediction of θ from X, without any linearity constraint. The elliptical family is exactly the class of distributions for which this conditional expectation is linear; for a precise statement of this characterization, see Appendix B.4.

Let β_0 and $\beta = (\beta_1, \ldots, \beta_k)$ denote the coefficients from regressing θ on η:

$$\beta = \Sigma_{\eta\eta}^{-1}\Sigma_{\eta\theta}, \qquad \beta_0 = \mu_\theta - \beta^T\mu_\eta. \tag{1}$$

By Lemma 1, we have $\mathbf{E}[\theta|\eta] = \beta_0 + \beta^T\eta$. Below I compare these regression coefficients with the weights in the linear equilibrium.

3.2 Linear equilibrium characterization

In the signaling game, (pure) strategies are defined as follows. The sender's type space T is the support of (η, δ). A distortion strategy for the sender is a map

$$d\colon T \to \mathbf{R}^k,$$

which assigns a distortion vector to each sender type. A decision strategy for the receiver is a map

$$y\colon \mathbf{R}^k \to \mathbf{R},$$

which assigns a decision to each feature vector.

I focus on Bayesian Nash equilibria in linear strategies, which I call *linear equilibria*.[16] With multi-dimensional types, signaling equilibria are generally difficult to construct, unless we focus on equilibrium in linear strategies or on discrete (often binary) action spaces. Discrete action spaces suffer from a ceiling effect: When types pool on a high action, there is a mechanical loss of information. To avoid this effect, I focus on a continuous model with linear equilibria. Restricting to linear equilibria disciplines the receiver's off-path decisions. In particular, linearity rules out discontinuous jumps in the receiver's decisions off-path.[17]

I now use the linear conditional expectation property of the elliptical distribution to construct linear equilibria. Suppose that the receiver uses the linear strategy

$$y(x) = b_0 + b^T x,$$

for some intercept b_0 and coefficient vector $b = (b_1, \ldots, b_k)$. I call $b_1, \ldots, b_k$ the feature weights, though these weights can be negative. Plugging this strategy into the sender's utility gives

$$b_0 + b^T(\eta + d) - (1/2)\sum_{j=1}^{k} d_j^2/\delta_j.$$

[16] Technically, these functions are affine, but I use the more common term linear throughout. To be sure, nonlinear deviations are feasible, but in equilibrium they are not optimal.

[17] Such jumps are permitted by perfect Bayesian equilibrium. With Gaussian uncertainty, linear strategies have full support so nothing is off-path. To avoid negative cost functions, I follow Frankel and Kartik (2019a) in using elliptical distributions with compact support. As a consequence, linear strategies do not have full support so some feature vectors are off-path.

Given the receiver's linear strategy, the sender's utility is concave and additively separable in the distortion amounts $d_1, \ldots, d_k$. The marginal benefit of distorting feature j equals the feature weight b_j. The marginal cost of distorting feature j is d_j/δ_j. Equating these expressions gives

$$d_j(\eta, \delta) = b_j \delta_j.$$

On each feature j, the sender's best response is increasing in her distortion ability and in the weight on feature j. Because the receiver's strategy is linear, the return to distortion is constant. Therefore, the sender's distortion choice does not depend on her intrinsic type. Denoting the componentwise (Hadamard) product of vectors with with the symbol $\circ$, the sender's best response can be expressed as

$$d(\eta, \delta) = b \circ \delta.$$

With this best response, the sender's feature vector is $\eta + b \circ \delta$. The equilibrium condition is

$$b_0 + b^T(\eta + b \circ \delta) = \mathbf{E}[\theta | \eta + b \circ \delta]. \tag{2}$$

This is an equality between random variables. The left side is the receiver's linear strategy, evaluated at the feature vector $\eta + b \circ \delta$. The right side is the receiver's posterior expectation of θ, conditional upon seeing the sender's feature vector. By the linear conditional expectation property (Lemma 1), the conditional expectation on the right side equals the population regression of θ on the random feature vector $\eta + b \circ \delta$. Therefore, this equality holds if and only if the coefficients on the left are the correct regression coefficients.[18]

Taking expectations of each side of (2) gives

$$b_0 + b^T(\mu_\eta + b \circ \mu_\delta) = \mu_\theta.$$

Hence, the intercept b_0 is pinned down by the vector b. Hereafter I refer to linear equilibria by the coefficient vector b alone, with the understanding that b_0 is chosen to satisfy this equation. The equilibrium condition for the vector b is

$$b = \text{reg}(\theta | \eta + b \circ \delta),$$

[18] Assumption A.i, the variance of $\eta + b \circ \delta$ has full rank, for every vector b.

which can be expressed in terms of the covariance as

$$\operatorname{var}(\eta + b \circ \delta)b = \operatorname{cov}(\eta + b \circ \delta, \theta). \tag{3}$$

This condition is a cubic polynomial system of k equations in the k unknowns $b_1, \ldots, b_k$. The coefficients in this system are determined by the covariance matrix Σ.[19]

The equilibrium condition is cubic even though both players have quadratic utilities. The receiver's quadratic utility makes his best response linear in his expectation of θ. If the receiver updated his belief about θ from an exogenous signal, then the equilibrium condition would be linear, as in simultaneous-move global games (Morris and Shin, 2002; Lambert et al., 2018). Here the receiver updates his beliefs from the sender's features, which are endogenous. In particular, the variance of the sender's feature vector is a quadratic function of b. This quadratic term increases the degree of the system from linear to cubic.

3.3 Homogeneous distortion

As a special case, suppose that the distortion type δ is nonrandom. Mathematically, this means that δ equals the mean vector μ_δ. The components of μ_δ need not be equal, so distortion ability can vary across features. On each feature, however, it is homogeneous in the population. The sender's private information is only her intrinsic type η.

Proposition 1 (Full separation)
If the distortion type δ is nonrandom, then the signaling game has exactly one linear equilibrium. This equilibrium is fully separating.

This result follows directly from the equilibrium condition (3). With δ nonrandom, (3) reduces to the linear equation $\Sigma_{\eta\eta} b = \Sigma_{\eta\theta}$. The matrix $\Sigma_{\eta\eta}$ has full rank by Assumption A.i, so this linear system has a unique solution. In terms of the regression

[19] The equation can be expanded as follows. Let $\operatorname{diag}(b)$ denote the diagonal matrix with the vector b along the diagonal. Using the identity $b \circ \delta = \operatorname{diag}(b)\delta$, (3) becomes

$$\big[\Sigma_{\eta\eta} + \Sigma_{\eta\delta} \operatorname{diag}(b) + \operatorname{diag}(b)\Sigma_{\delta\eta} + \operatorname{diag}(b)\Sigma_{\delta\delta} \operatorname{diag}(b)\big]b = \Sigma_{\eta\theta} + b \circ \Sigma_{\delta\theta}.$$

coefficients $\beta_0, \ldots, \beta_k$ defined in (1), we have

$$b = \beta, \qquad b_0 = \beta_0 - \beta^T(\beta \circ \mu_\delta).$$

The receiver's coefficient vector is the regression vector β. The sender chooses distortion $\beta \circ \mu_\delta$, which increases the the score by $\beta^T(\beta \circ \mu_\delta)$. The receiver subtracts this amount from the intercept to offset the effect of distortion, and no information is lost.

In this equilibrium, the receiver's decision equals the conditional expectation $\mathbf{E}[\theta|\eta]$. This is the receiver's first-best outcome, so full disclosure is an optimal scoring rule. Returning to the example of credit scoring, if every consumer distorts her features by the same amount, the effect can be offset by subtracting a constant from everyone's credit score. In reality, different consumers experience different costs and benefits form distortion. I turn to this general case next.

3.4 Equilibrium existence and uniqueness

If the distortion vector δ is not constant, then the equilibrium condition (3) is cubic. For the rest of the paper, unless specified otherwise, I make the following covariance assumption.

Assumption B (Distortion abilities)

(i) δ is uncorrelated with (θ, η).

(ii) $\text{cov}(\delta_i, \delta_j) \geq 0$ for all features i and j.

Assumption B.i says that the sender's distortion type is not correlated with her quality or with her intrinsic type. This is a stylized way to capture the motivation that the sender's intrinsic level is what reveals information about her latent quality. In the presence of Assumption B.i, Assumption A reduces to (i) $\Sigma_{\eta\eta}$ has full rank, and (ii) $\Sigma_{\eta\theta} \neq 0$.

Assumption B.ii says that the sender's distortion ability on different features is nonnegatively correlated. In the sender's utility function, the distortion ability δ_j parameterizes the sender's cost of distortion, but it can also be interpreted as capturing the intensity of the sender's preferences for a high decision. The following example formalizes this interpretation and shows that the preference component of δ makes the covariance nonnegative.

Example 1 (Distortion type as decision preference)
Suppose that the sender's utility is given by

$$u_S = \delta_0 y - (1/2) \sum_{j=1}^{k} d_j^2 / \bar{\delta}_j,$$

where $\bar{\delta}_1, \ldots, \bar{\delta}_k$ are constants and δ_0 is a nonnegative random variable with positive variance σ_0^2. Without changing the receiver's preferences, we can divide by δ_0 to obtain a utility function that fits the baseline model with $\delta_j = \delta_0 \bar{\delta}_j$. Here $\operatorname{cov}(\delta_i, \delta_j)$ equals $\sigma_0^2 \bar{\delta}_i \bar{\delta}_j$, which is nonnegative. Hence Assumption B.ii holds.

With these standing assumptions, I obtain the main result for the signaling game.

Theorem 1 (Existence and uniqueness)
The signaling game has exactly one linear equilibrium.

The proof of existence uses only Assumption B.i. Define the chained best-response function $\mathrm{BR}\colon \mathbf{R}^k \to \mathbf{R}^k$ by

$$\mathrm{BR}(b) = \operatorname{reg}(\theta | \eta + b \circ \delta).$$

If the receiver uses a linear strategy with coefficient vector b, then as shown in Section 3.2, the sender's best response induces the feature vector $\eta + b \circ \delta$. The receiver's best response to this feature vector is a linear function with coefficient vector $\mathrm{BR}(b)$.

The function BR is continuous. To apply Brouwer's fixed-point theorem, I show that BR is bounded. For every vector b, the variance matrix $\operatorname{var}(\eta + b \circ \delta)$ is uniformly bounded below (in the positive semidefinite order) by a positive definite matrix.[20] This gives a uniform upper bound on the magnitude of BR. The intuition is that if the coefficient vector is too large, then the receiver's decision will pick up too much noise from the sender's intrinsic levels.

Uniqueness is more subtle. In the single-dimensional case, the linear equilibria are the roots of a univariate cubic polynomial. In general, such a polynomial can have up to three real roots. Without the covariance assumptions there can indeed by three equilibria. With k features, this bound becomes 3^k; see Appendix B.7. Nevertheless, the covariance assumptions ensure that the equilibrium is unique.

[20] For two symmetric matrices A and B, matrix A is weakly greater than matrix B in the positive semidefinite order, denoted $A \succeq B$, if the difference $A - B$ is positive semidefinite.

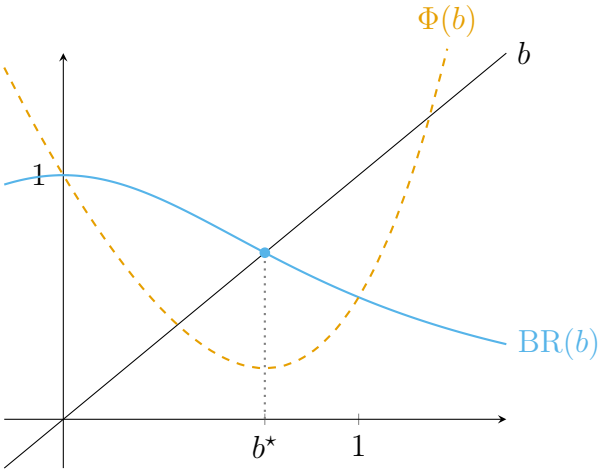


Figure 3. Equilibrium and convex representation

For uniqueness, I express the equilibrium condition as a stationary point of a strictly convex function Φ defined by

$$\Phi(z) = \operatorname{var}(z^T\eta - \theta) + (1/2)\operatorname{var}((z \circ z)^T\delta).$$

The coefficient vector b is an equilibrium if and only if $\nabla\Phi(b) = 0$. Since Φ is strictly convex, there can be at most one such point.

Example 2 (Single feature)
Suppose $k = 1$. Now η, δ, and b are scalars instead of vectors. If the receiver uses a linear decision strategy with slope b, the sender's distortion best response is $d(\eta, \delta) = b\delta$, and we have

$$\mathrm{BR}(b) = \frac{\sigma_{\theta\eta}}{\sigma_\eta^2 + b^2\sigma_\delta^2}.$$

Suppose that $\theta = \eta$ and all variances equal 1. Then $\mathrm{BR}(b) = 1/(1+b^2)$. Figure 3 plots this best-response function against the 45-degree line. The best response achieves its maximum of 1 at $b = 0$. As $|b|$ increases, the sender has a stronger incentive to distort her feature, so the receiver's best response is less sensitive to this noisier feature. The equilibrium $b^\star$ is the unique root of the cubic polynomial $b^3 + b - 1$, hence $b^\star \approx 0.68$.

This function Φ resembles a potential function (Rosenthal, 1973; Monderer and Shapley, 1996), but it is not. A potential function acts on an entire strategy profile, not on one player's strategy. I can show, however, that the game is best-response equivalent to a zero-sum game, in the same way that a potential game is best-response

equivalent to an identical interest game (Voorneveld, 2000).[21] In this zero-sum game, each player's utility function is concave. Therefore I can show that that the equilibrium is stable in the sense that the continuous best-response dynamics converge to the unique linear equilibrium. See Appendix B.2.

3.5 Information loss from distortion

Now I analyze the receiver's loss of information from distortion. If the receiver could directly observe the sender's type, then his posterior expectation would be

$$\mathbf{E}[\theta|\eta,\delta] = \beta_0 + \beta^T\eta.$$

An equilibrium is *fully revealing* if the receiver's decision coincides with $\mathbf{E}[\theta|\eta,\delta]$. If δ is homogeneous, then 1 implies that the linear equilibrium is fully revealing. If δ is heterogeneous, it is straightforward to check that the unique linear equilibrium is not fully revealing. With a more involved argument, I can completely rule out fully revealing, linear or otherwise.

Proposition 2 (No fully revealing equilibrium)
If $\text{var}(\delta)$ *has full rank, then no Bayesian Nash equilibrium is fully revealing.*

The details of the proof are involved, but the intuition is straightforward. A type with a higher distortion abilities will prefer to mimick another type with lower distortion ability but higher intrinsic levels (and hence higher expected quality).

How does the receiver's utility in the unique linear equilibrium depend on the utility weight that the sender places on the decision? Suppose that the receiver's utility equals

$$\alpha y - (1/2)\sum_{j=1}^{k} d_j^2/\delta_j,$$

where $\alpha > 0$. In the language of Frankel and Kartik (2019a), the parameter s controls the sender's *stake* in the decision. Taking $\alpha = 1$ gives the baseline model. Without changing the sender's preferences, we can divide by α to get

$$y - (1/2)\sum_{j=1}^{k} d_j^2/(\alpha\delta_j).$$

[21] Hwang and Rey-Bellet (2018a) and Hwang and Rey-Bellet (2018b) study these games.

This is the utility function from the baseline model, except that the sender's distortion type is $\alpha\delta$ rather than δ. In the variance matrix Σ, the submatrix $\Sigma_{\delta\delta}$ is scaled up by α^2 and all other components are unchanged because the covariance between δ and (θ, η) is zero by Assumption B.i. Therefore, the stakes α can be represented by scaling the variance matrix by α^2. I study the effect of these stakes on receiver's utility in the unique linear equilibrium.

Proposition 3 (Stakes and information loss)
If $\text{var}(\delta)$ *is positive definite, the receiver's utility in the unique linear equilibrium is strictly decreasing in the sender's stakes* α*.*

For a given strategy for the receiver, the sender distorts her features more as the sender's stakes increase. This result is more subtle because the the receiver's equilibrium strategy changes as well. I show that the equilibrium change in the receiver's strategy is not large enough to offset the direct effect of the increased stakes.

One might expect that the equilibrium utility is decreasing in $\Sigma_{\delta\delta}$, with respect to the positive semidefinite matrix order. Fixing the receiver's strategy, the informativeness of the sender's features are indeed decreasing in $\Sigma_{\delta\delta}$, but the equilibrium effect on the receiver's strategy can outweigh this direct effect, as shown in the following counterexample.

Example 3 (Receiver's payoff is not monotone in $\Sigma_{\delta\delta}$)
Suppose $k = 2$. Consider the covariance matrices[22]

$$\Sigma_{\theta\eta} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \qquad \Sigma_{\eta\eta} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \qquad \Sigma_{\delta\delta} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

With these parameters, the regression coefficient β equals $(7/3, -2/3)$. Even though η_1 and η_2 are both positively correlated with the quality θ, the regression coefficient on η_2 is negative because η_1 and η_2 have positively correlated errors and η_2 has a weaker correlation with θ. The signaling equilibrium is $b^\star = (1.20, -0.10)$. As $\text{var}(\delta_2)$ increases, both components of $b^\star$ shrink in magnitude, and the receiver's payoff increases, though the effect is small.[23]

[22] Set $\sigma_\theta^2 = 9$. This variance does not affect the analysis, as long as the induced variance matrix is positive semidefinite.

[23] Increasing $\text{var}(\delta_2)$ from 1 to 2 shifts $b^\star$ from $(1.1951, -0.0971)$ to $(1.1950, -0.0966)$. The receiver's utility increases from -4.3167 to -4.3165.

This result extends the single-dimensional result in Frankel and Kartik (2019a). With a single feature, increasing the stakes and increasing the variance are equivalent. With multiple features, the variance matrix can be increased in different directions that do not correspond to increasing the stakes. My result pinpoints the increasing stakes as the source of reduced equilibrium information transmission.

4 Scoring

Now I return to the scoring model with the intermediary. To simplify the intermediary's problem, I first establish a revelation principle: There is no loss in restricting scores to direct decision recommendations that the receiver obeys. This is not a standard revelation principle because the intermediary observes costly actions, not cheap-talk reports. I use this revelation principle to characterize the linear outcome rules that the intermediary can induce. Then I maximize the intermediary's objective over this set.

4.1 Revelation principle and obedience

Each scoring rule $f\colon \mathbf{R}^k \to \Delta(S)$ induces a different game between the sender and receiver. To state the revelation principle in full generality, I allow for mixed strategies in this game. A distortion strategy for the sender is a map $d\colon T \to \Delta(\mathbf{R}^k)$, which assigns to each sender type a distribution over distortion vectors. A decision strategy for the receiver is a map $y\colon S \to \Delta(\mathbf{R})$, which assigns to each score a distribution over decisions. The solution concept is Bayesian Nash equilibrium, just as in the signaling setting without the intermediary.

An *outcome rule* $\sigma = (\sigma_S, \sigma_R)$ consists of a map $\sigma_S\colon T \to \Delta(\mathbf{R}^k)$ specifying the distribution of features for each sender type, and a map $\sigma_R\colon \mathbf{R}^k \to \Delta(\mathbf{R})$ specifying the distribution of decisions for each feature vector. This is a different object than the social choice function used in the usual revelation principle. Since the sender and receiver play a sequential-move game, I must specify off-path play as part of the outcome.

Together a scoring rule f and a strategy profile (d, y) in the resulting game *induce* the outcome rule σ given by $\sigma_S(\eta, \delta) = \eta + d(\eta, \delta)$ and $\sigma_R(x) = y(f(x))$. Here the composition of functions is extended in the natural way to the composition of mixed

strategies. An outcome rule σ is *implementable* if there exists a scoring rule f and a Bayesian Nash equilibrium (d, y) in the associated game that together induce σ. An outcome rule is *directly implementable* if it can be implemented by a scoring rule with $S = \mathbf{R}$ and a Bayesian Nash equilibrium with $y = \mathrm{id}$. In words, direct implementation means that the intermediary makes a decision recommendation, which the receiver follows.

Proposition 4 (Revelation principle)
Every implementable outcome rule is directly implementable.

Given an indirect scoring rule, construct an equivalent direct scoring rule by pooling all the scores that induce the receiver to take the same decision. This way, the intermediary gives the receiver the Blackwell minimal information that he needs to follow the decision rule. The intuition resembles the revelation principle from Myerson (1986), but the formal statement is different because the intermediary in my model observes the sender's features (which are payoff-relevant actions) rather than cheap-talk messages. With cheap-talk messages, the mediator can, without loss, restrict the sender to a fixed set of messages. The mediator distinguishes one default message from the message set and commits to treat any message outside the message space as if the sender had sent that default message. The same approach does not work in my model because different distortion vectors give the sender different payoffs, even if they result in the same decision by the receiver.[24]

The revelation principle allows me to focus my analysis on the outcome rule σ. An outcome rule σ is directly implementable if it satisfies the following obedience conditions. The sender's obedience condition is simply her best-response condition for the strategy profile (σ_S, σ_R). For the receiver, the obedience condition is

$$\sigma_R(\sigma_S(\eta, \delta)) = \mathbf{E}\big[\theta | \sigma_R(\sigma_S(\eta, \delta))\big].$$

Both sides of this equation are random variables. The randomness comes from the random variable (θ, η, δ) and also from the mixed decision rule.

Hereafter, I suppose that the intermediary uses a direct recommendation scoring rule f, and I denote the induced random feature vector by X. With this notation,

[24] Doval and Ely (2016) and Makris and Renou (2018) establish a revelation principle for dynamic games in which the designer can observe past actions.

the obedience condition takes the simple form

$$f(X) = \mathbf{E}[\theta | f(X)].$$

This condition says upon receiving the recommendation $f(X)$, the receiver finds it optimal to follow it.

4.2 Linear outcome rules

For the rest of the analysis, I restrict to linear outcome rules. Scoring rules often take a simple form so that they are explainable to scored agents. For example, FICO scores aggregate five different credit factors with associated weights. The linearity restriction also allows me to isolate the effect of the intermediary when comparing the scoring solution to the signaling equilibrium.

Suppose that the intermediary uses the direct scoring rule

$$f(x) = b_0 + b^T x,$$

for some intercept b_0 and coefficient vector $b = (b_1, \ldots, b_k)$. As in the signaling game, the sender's best response is to choose distortion vector $b \circ \delta$. The receiver's decision must match his conditional expectation of θ, given the intermediary's decision recommendation:

$$b_0 + b^T(\eta + b \circ \delta) = \mathbf{E}[\theta | b_0 + b^T(\eta + b \circ \delta)]. \tag{4}$$

By the linear conditional expectations property (Lemma 1), this equality between random variables reduces to an equality between regression coefficients. Taking the expectation of each side of (4) gives

$$b_0 = \mu_\theta - b^T(\mu_\eta + b \circ \mu_\delta).$$

Hence, the intercept b_0 is pinned down by the vector b. The vector b must satisfy the regression equation

$$1 = \operatorname{reg}\big(\theta \big| b^T(\eta + b \circ \delta)\big).$$

In terms of the covariance, this condition is

$$b^T \operatorname{var}(\eta + b \circ \delta) b = b^T \operatorname{cov}(\eta + b \circ \delta, \theta). \tag{5}$$

This is a single quartic polynomial equation in the coefficients $b_1, \ldots, b_k$. Compare this scalar equality with the vector equality from the signaling equilibrium

$$\operatorname{var}(\eta + b \circ \delta) b = \operatorname{cov}(\eta + b \circ \delta, \theta).$$

If b satisfies the equilibrium condition, then it automatically satisfies the scoring condition. If there are multiple features ($k > 2$), the scoring condition is more permissive. If there is a single feature ($k = 1$), then the signaling and scoring conditions are the same, except that scoring also allows $b = 0$, which means that the intermediary provides no information.

4.3 Optimal scoring

So far, I have characterized the set of linear obedient decision rules that the intermediary can induce. The analysis so far holds for any objective of the intermediary. Now I maximize the receiver's utility over this set, and I will identify when the intermediary can strictly improve the receiver's payoff.

By the standard bias–variance decomposition, the intermediary's problem becomes[25]

$$\begin{aligned} &\text{minimize} && \operatorname{var}\big(b^T(\eta + b \circ \delta) - \theta\big) \\ &\text{subject to} && b^T \operatorname{var}(\eta + b \circ \delta) b = b^T \operatorname{cov}(\eta + b \circ \delta, \theta). \end{aligned} \tag{6}$$

The feasible set in particular contains the signaling equilibrium b^{signal}. The loss function is continuous and convex, but the constraint set is not necessarily compact or convex.

Proposition 5 (Scoring existence and uniqueness)
The scoring problem (6) has exactly one solution.

[25] The bias–variance decomposition gives

$$\mathbf{E}^2\big[b_0 + b^T(\eta + b \circ \delta) - \theta\big] + \operatorname{var}\big(b_0 + b^T(\eta + b \circ \delta) - \theta\big).$$

For obedient decision rules, the first term vanishes. In the second term, b_0 does not affect the variance so it can be dropped.

I prove existence by showing that the loss function is lower semi-compact, and I prove uniqueness by analyzing the Lagrangian. Denote the scoring solution b^{score}. I characterize whether this scoring solution coincides with the signaling equilibrium.

Theorem 2 (Scoring versus signaling)
The following are equivalent.

(i) $b^{\text{signal}} = b^{\text{score}}$.

(ii) b^{signal} *is a scalar multiple of* β.

(iii) $\operatorname{diag}(\beta)\Sigma_{\eta\eta}(\beta \circ \beta)$ *is a scalar multiple of* $\Sigma_{\eta\theta}$.

From the definition of b^{signal} and β, condition (ii) says for some scalar λ,

$$\operatorname{reg}(\theta|\eta + b^{\text{signal}} \circ \delta) = \lambda \operatorname{reg}(\theta|\eta).$$

This means that in the signaling equilibrium, the sender's distortion dampens the informativeness of every feature by the same factor. The obedience constraint prevents the intermediary from systematically reducing every feature weight, but the intermediary can still control the relative feature weights. As long as distortion has a different effect on different features, the intermediary can improve information transmission by decreasing the weight on features that are most garbled by equilibrium distortion.

Condition (iii) gives an explicit condition for $b^{\text{score}} = b^{\text{signal}}$ in terms of the covariance primitives. Provided that $k > 1$, this condition is non-generic. If $k = 1$, however, the vectors become scalars, so the condition always holds, hence scoring and signaling coincide.

I now turn to comparative statics for the receiver's utility under the optimal scoring rule. A real-valued function h act on symmetric square matrices is *increasing* if $A \succeq B$ implies $h(A) \geq h(B)$, and $A \succ B$ implies $h(A) > h(B)$.

Proposition 6 (Scoring comparative statics)
The receiver's scoring utility is decreasing in the variance $\Sigma_{\delta\delta}$.

This is the intuitive result that failed for the signaling setting. If the variance $\Sigma_{\delta\delta}$ decreases, the receiver can secure the same payoff by adding uncorrelated noise to the score. The optimal score is strictly better.

I will illustrate the theorem with examples in a particularly tractable setting that I introduce now.

4.4 Setting with uncorrelated errors

Consider the following *setting with uncorrelated errors*, which I return to throughout the paper. The sender's quality is given by

$$\theta = (\theta_1 + \cdots + \theta_k)/k.$$

The intrinsic level on each feature j is given by

$$\eta_j = \theta_j + \varepsilon_j.$$

Thus, the vector (θ, η, δ) is determined by the $3k$ random variables

$$\theta_1, \ldots, \theta_k, \varepsilon_1, \ldots, \varepsilon_k, \delta_1, \ldots, \delta_k.$$

These random are jointly elliptically distributed. After normalization, we may assume without loss that each θ_j has unit variance. For all i and j, the correlation between θ_i and θ_j equals ρ, but all other correlations are zero. The remaining variances are parameterized as $\sigma^2_{\varepsilon,i} = \text{var}(\varepsilon_i)$ and $\sigma^2_{\delta,i} = \text{var}(\delta_i)$, for each feature j. These parameters control the precision the intrinsic level and the heterogeneity of distortion ability on each feature.

I will illustrate many of the general theorems through the following numerical example.

Example 4 (Two features)
Consider the setting of uncorrelated errors with $k = 2$, $\rho = 1$, and $\sigma^2_{\varepsilon,1} = \sigma^2_{\varepsilon,2} = 1$. The intrinsic levels are equally precise signals about θ. For the distortion heterogeneity, let $\sigma^2_{\delta,1} = 1$. I will vary the parameter $\sigma^2_{\delta,2}$. If $\sigma^2_{\delta,2} = 1$, then this example is symmetric. As $\sigma^2_{\delta,2}$ increases above 1, distortion ability is more heterogeneous on the second feature than on the first.

The left panel of Figure 4 plots the set of obedient coefficient vectors $b = (b_1, b_2)$ in Example 4 for the symmetric case $s^2_{\delta,2} = 1$ (solid blue) and the asymmetric case $s^2_{\delta,2} = 4$ (dashed orange). In both cases, the interior of the curve is convex. If the receiver uses a scoring rule inside the curve, then the receiver's best response is a linear function of the recommendation with slope strictly greater than 1. Conversely, if the intermediary uses a rule lying outside the curve, then the receiver's best response is

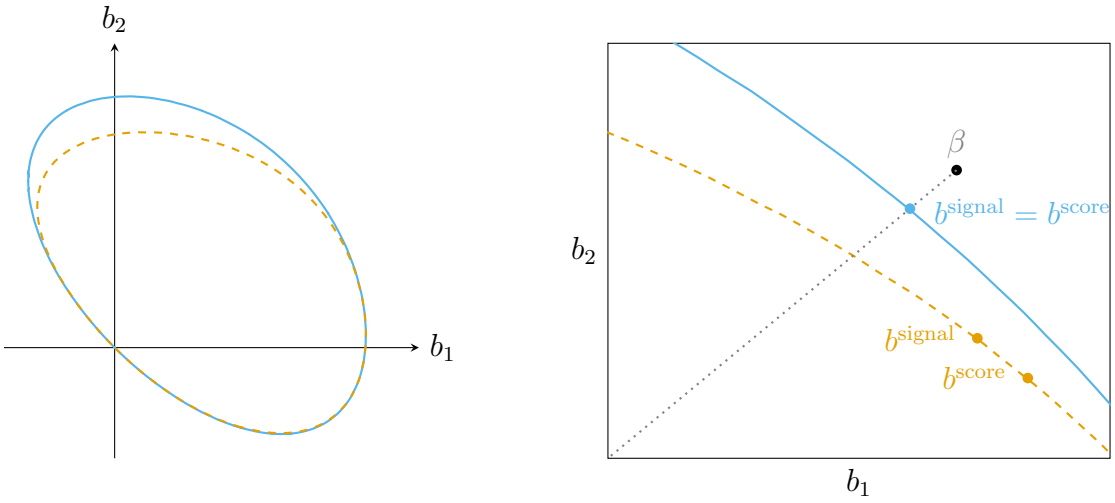


Figure 4. *Left*: Obedient coefficient vectors for two different parameter values. *Right*: Optimal scoring rule and signaling equilibrium.

linear in the recommendation, but with a coefficient strictly less than 1 (and possibly negative). Both curves pass through the origin, which represents no information provision. As $\sigma^2_{\delta,2}$ increases from the symmetric case to the asymmetric case, the curve shrinks toward the origin.[26]

The right panel of Figure 4 shows a magnified view of the obedient scoring rules in the left panel. I label the regression vector β as well as the vectors b^{signal} and b^{score} for both parameter values. As before, the solid blue curve is $(\sigma^2_{\delta,1}, \sigma^2_{\delta,2}) = (1, 1)$ and the dashed blue curve is $(\sigma^2_{\delta,1}, \sigma^2_{\delta,2}) = (1, 4)$. In the symmetric case, the signaling equilibrium vector is a scalar multiple of the regression coefficient β. In this case, the signaling equilibrium maximizes the receiver's utility over all obedient decisions rules, so the scoring solution coincides with the signaling equilibrium.

The dashed, orange curve illustrates the (generic) asymmetric case. Since $\sigma^2_{\delta,1} < \sigma^2_{\delta,2}$, the receiver's equilibrium decision is more sensitive to feature 1 than to feature 2. The intermediary can improve the accuracy of the receiver's decision by sliding along the obedient curve further away from the 45-degree line, to shift weight from feature 2 to feature 1. In general, there exists a local improvement away from the

[26] To see that these regions are nested, observe that the region inside the curve is given by the inequality

$$b^T \Sigma_{\eta\eta} b + (b \circ b)^T \Sigma_{\delta\delta} (b \circ b) \leq b^T \Sigma_{\eta\theta}.$$

As $\Sigma_{\delta\delta}$ decreases with respect to the positive semidefinite order, this inequality becomes more permissive.

signaling equilibrium, unless the gradient of the receiver's objective is orthogonal to the surface of obedient decision rules.

I will be interested in underweighting and overweighting. To quantify underweighting and overweighting, define for each vector b in $\mathbf{R}^k$ and feature j, the weighting ratio

$$w_j(b) = \frac{b_j}{\text{reg}(\theta|\eta + b \circ \delta)}.$$

Thus, $w_j(b)$ is the ratio between the weight b_j placed on feature j and the ex-post optimal weight $\text{reg}(\theta|\eta+b\circ\delta)$. A coefficient vector b *underweights* feature j if $w_j(b) < 1$ and *overweights* feature j if $w_j(b) > 1$.

For each feature j, define

$$\tau_j = \frac{(1 - \rho + \sigma^2_{\varepsilon,j})^3}{\sigma^2_{\delta,j}}.$$

I show that underweighting and overweighting are captured by the expression

Theorem 3 (Underweighting and overweighting)
Consider the setting of uncorrelated errors.

(i) If the τ_j are all equal, then $b^{\text{signal}} = b^{\text{score}}$.

(ii) If the τ_j are not all equal, then at least one feature is overweighted and at least one feature is underweighted. Moreover, for all i and j,

$$w_i(b^{\text{score}}) \geq w_j(b^{\text{score}}) \iff \tau_i \geq \tau_j.$$

The general condition from Theorem 2 takes a simple form in the setting of uncorrelated errors. Part (i) says that scoring and screening coincide if and only if all the τ_j are equal. Part (ii) identifies which features tend to be underweighted and which features tend to be overweighted. Underweighting is most productive when the direct effect of dampened distortion is large (high $\sigma^2_{\delta,j}$) and this additional information about the intrinsic level helps the receiver to predict quality (low $\sigma^2_{\varepsilon,j}$).

Figure 5 plots the signaling equilibrium and the scoring solution for Example 4, as the variance $\sigma^2_{\delta,2}$ interpolates between the symmetric case and the asymmetric case. As distortion ability on the second feature becomes more heterogeneous, weight shifts from the second feature to first feature for both signaling and scoring. But the effect is more dramatic for the scoring setting. To see that the first feature is overweighted

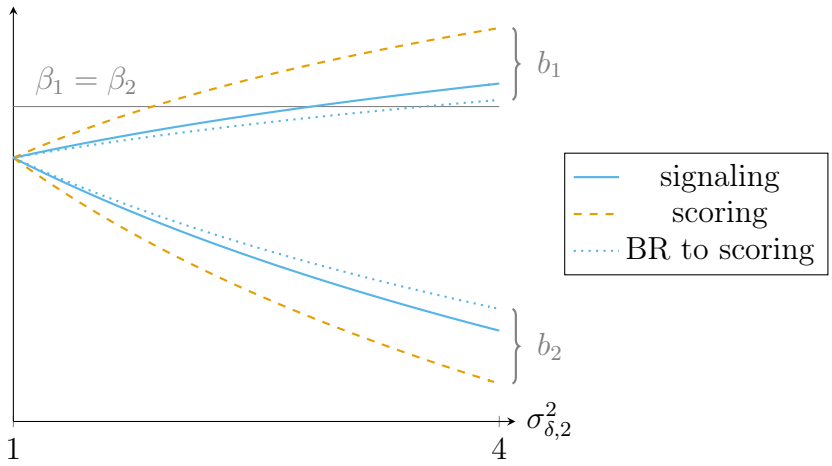


Figure 5. Scoring—underweighting and overweighting

and the second is underweighted by the scoring solution, I also plot the receiver's best response if he could observe the sender's features ex post. He would put less weight on the first feature, even less than in the signaling equilibrium. Conversely, he would put more weight on the second feature, even more than in the signaling equilibrium.

5 Screening

Finally, I consider a *screening* setting, in which the receiver commits to a decision as a function of the sender's features. I then compare the three settings—signaling, scoring, and screening.

5.1 Decision commitment

The intermediary and the receiver have the same preferences, so if the receiver has decision commitment power, we can assume without loss that the intermediary fully discloses the sender's features to the receiver. The receiver therefore commits to a decision rule as a function of the sender's features.

To isolate the effect of commitment, I again restrict to linear decision rules. The problem is to minimize the expected loss

$$\mathbf{E}\Big[\big(b_0 + b^T(\eta + b \circ \delta) - \theta\big)^2\Big].$$

This objective incorporates the sender's best response to the receiver's decision rule.

Since there is no longer an obedience constraint, the receiver can commit to a decision rule whose expectation differs from μ_θ. But starting from such a decision rule, the receiver can strictly increase her payoff by adjusting the intercept to eliminate the bias without changing the sender's incentives. In the bias–variance decomposition, the bias vanishes, so the receiver equivalently minimizes the variance

$$\operatorname{var}\big(b^T(\eta + b \circ \delta) - \theta\big),$$

over all vectors $b \in \mathbf{R}^k$. Unlike the signaling setting, there is no obedience constraint on the vector b. The same argument as the proof of Proposition 5 establishes that the screening solution exists and is unique. Denote the solution by b^{screen}.

5.2 Commitment reduces distortion

By Assumption B.i, δ and (θ, η) are uncorrelated, so the variance can be decomposed as

$$\operatorname{var}(b^T\eta - \theta) + \operatorname{var}((b \circ b)^T\delta) = \operatorname{var}(b^T\eta - \theta) + (b \circ b)^T\Sigma_{\delta\delta}(b \circ b).$$

The second term is the receiver's loss from the sender's distortion. This loss depends on the size of b. Raising this term to power $1/4$, define the norm $\|\cdot\|_{4,\delta}$ by

$$\|b\|_{4,\delta} = \Big[(b \circ b)^T\Sigma_{\delta\delta}(b \circ b)\Big]^{1/4}.$$

In Appendix B.8, I use the nonnegativity of $\Sigma_{\delta\delta}$ to show that this is in fact a norm. It measures the receiver's feature sensitivity in terms of the receiver's loss from the induced distortion.

Theorem 4 (Commitment reduces distortion)
Suppose $k > 1$. If $\Sigma_{\delta\delta}$ is positive definite, then

$$\|\beta\|_{4,\delta} > \|b^{\text{signal}}\|_{4,\delta} \geq \|b^{\text{score}}\|_{4,\delta} > \|b^{\text{screen}}\|_{4,\delta}.$$

The middle inequality is strict if $b^{\text{signal}} \neq b^{\text{score}}$.

As the receiver's commitment power increases, the receiver makes his decision less sensitive to the sender's features in order to deter the sender's distortion. Starting on the left side of the inequality, the regression coefficient β is the optimal coeffi-

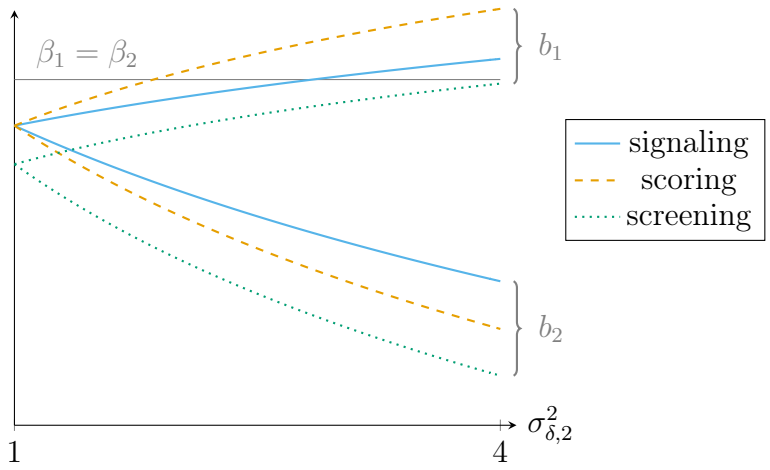


Figure 6. Comparing commitments

cient vector if the sender did not distort her features at all. Next, in the signaling equilibrium, the receiver is playing a best response to the sender's distorted features, so his decision become less sensitive to the sender's features. In the scoring setting, the intermediary rearranges the weights to further reduce distortion, subject to the obedience constraint. Finally, with decision commitment, the receiver can commit to under-react to all the sender's features.

5.3 Comparing feature weights

To compare the feature weights individually, I return to the setting of uncorrelated errors from Section 4.4. Figure 6 plots the features weights in all three settings for Example 4, as $\sigma^2_{\delta,2}$ varies from 1.5 to 6. First, look at the left axis, where $\sigma^2_{\delta,1} = \sigma^2_{\delta,2} = 1.5$. The signaling equilibrium and the scoring solution coincide, but the screening solution puts strictly smaller weights on both features. Scoring allows the intermediary to change the relative feature weights, but when the features are symmetric, there is nothing to be gained. With commitment power, the receiver can reduce the weight on both features. Moving to the right, as $\sigma^2_{\delta,2}$ increases, the weight shifts from feature 2 to feature 1. The weights in the screening solution lie below the scoring solution.

The observations in this example hold more generally.

Proposition 7 (Feature weights)
In the setting of uncorrelated errors, the following hold.

(i) b^{signal}, b^{score}, *and* b^{screen} *are all strictly positive.*

(ii) *If* $(\sigma^2_{\varepsilon,i}, \sigma^2_{\delta,i}) > (\sigma^2_{\varepsilon,j}, \sigma^2_{\delta,j})$, *then* $b_i < b_j$ *for all* b *in* $\{b^{\text{signal}}, b^{\text{score}}, b^{\text{screen}}\}$.
(iii) $b^{\text{score}}_j > b^{\text{screen}}_j$ *for all* j.
(iv) *For all* i *and* j, *we have* $w_i(b^{\text{screen}}) \geq w_j(b^{\text{screen}})$ *if and only if* $\tau_i \geq \tau_j$.

Part (i) says that the coefficient vectors in all three settings are all strictly positive. The covariance vector $\Sigma_{\eta\theta}$ is strictly positive, but this is not a sufficient condition for the regression vector to be nonnegative. In the simple setting, we have nonnegative weights for each of the three settings because the errors ε_i and the distortion abilities δ_i are all uncorrelated. Part (ii) compares the weights on different features within the same level of commitment. Less weight is a placed on a feature if the intrinsic level is a noisier signal of the latent quality and if the distortion ability is more heterogeneous. Part (iii) compares the feature weights across different settings. Relative to scoring, the screening solution puts less weight on every feature. Part (iv) shows that under optimal screening, the amount of underweighting is still ordered by the parameter τ_i, even though all the weights are smaller under screening.

6 Extensions

6.1 Random scoring

In the main model, I restrict to deterministic linear scores. Now I allow for stochastic scores, and I impose the linearity restriction on the conditional expectation. Suppose that there exist b_0 in $\mathbf{R}$ and b in $\mathbf{R}^k$ such that

$$\mathbf{E}[f(x)] = b_0 + b^T x,$$

for all $x \in \mathbf{R}^k$. The sender is risk neutral, so her best response is still $d(\eta, \delta) = b \circ \delta$. Let X denote the induced random feature vector $\eta + b \circ \delta$. Scoring is noise-free if $f(X) = \mathbf{E}[f(X)|X]$, that is, the scoring rule does not add noise on-path.

Proposition 8 (Noise-free scoring)
Optimal scoring is noise-free.

Assumption B.i implies that $\Sigma_{\delta\theta} = 0$, but the proof also goes through if some component $\Sigma_{\delta\theta}$ is strictly positive. Instead of adding uncorrelated noise, the intermediary can add weight to feature with this positive covariance. If $\Sigma_{\theta\eta} < 0$, however, adding noise can be optimal, as illustrated by the following example.

Example 5 (Noisy scoring)
Take $k = 1$. Suppose η and δ are uncorrelated with unit variance and let $\theta = \eta - 2\delta$. Hence $\sigma_{\theta\eta} = 1$ and $\sigma_{\theta\delta} = -2$. In this case the optimal scoring rule uses $b = 0.25$ and $\mathbf{E}[\mathrm{Var}(f(X)|X)] = 0.059$. The added noise makes it obedient to use such a small coefficient b, which dampens the sender's distortion.

6.2 Efficient scoring

In the main model, the intermediary maximizes the receiver's utility u_R. More generally, suppose that the intermediary maximizes the expectation of the social welfare function

$$\pi u_S + (1 - \pi)u_R,$$

where π in $[0, 1)$ is the Pareto weight on the sender. The sender is risk-neutral and the scoring rule cannot change the receiver's expected decision. Thus, the scoring rule affects the sender only through her cost of distortion

$$(1/2)\sum_{i=1}^{k}(b_i\delta_i)^2/\delta_i = (1/2)(b \circ b)^T\delta.$$

If $\pi = 1$, the solution is to provide no information. I rule out that this case in order to focus on scoring rules with positive variance.

For a linear decision rule, the intermediary minimizes

$$\pi(1/2)(b \circ b)^T\mu_\delta + (1 - \pi)\left[\mathrm{var}(b^T\eta - \theta) + (b \circ b)^T\Sigma_{\delta\delta}(b \circ b)\right].$$

The noisiness can be captured by the *noise ratio*

$$\frac{\mathbf{E}[\mathrm{var}(f(X)|X)]}{\mathrm{var}(f(X))}.$$

The denominator is the variance of the score. The numerator is the average variance of the noise that is added to the score.

Proposition 9 (Efficient scoring)
There exists a cutoff $\bar{\pi} \in [0, 1)$ such that optimal scoring is noise-free if and only if $\pi \leq \bar{\pi}$. For $\pi > \bar{\pi}$, the noise ratio is strictly increasing in π.

6.3 Observable features

In the main model, the receiver observes sender's score, but he does not directly observe any of the sender's intrinsic levels. What if instead the receiver can observe some of the sender's features?

Suppose a subset of features is observable. Partition the index set as

$$\{1, \ldots, k\} = I \cup J,$$

where features i in I are directly observable, but features j in J are not. Therefore, the decision rule is parameterized by $(b_0, b) = (b_0, b_I, b_J)$. The obedience requirement is now

$$\begin{aligned} b_I &= \text{reg}(\theta | \eta_I + b_I \circ \delta_i), \\ 1 &= \text{reg}(\theta | b^T(\eta + b \circ \delta)). \end{aligned}$$

Thus, this general problem interpolates between signaling, in which $I = \{1, \ldots, k\}$, and scoring, in which $J = \{1, \ldots, k\}$.

6.4 Correlated distortion ability

The covariance assumptions capture the intuition that distortion impedes the informativeness of the sender's features. Without the covariance assumptions, the variance cannot be decomposed into the uncertainty introduced by the distortion ability and the information from the sender's intrinsic levels. In the scoring benchmark, the equilibrium still exists, but uniqueness is not guaranteed. I can prove that generically scoring strictly improves upon every signaling equilibrium, but there is no easy expression for the generic condition.

6.5 Further extensions

I discuss additional extensions in Appendix B.3. In particular, I allow for the sender's cost function to be nonseparable across the different dimensions of distortion and for multi-dimensional decisions. I also allow the sender's distortion activity to be productive in the sense that it shifts the receiver's bliss point from θ to $\theta + v^T d$, where v is a fixed vector measuring the productivity of distortion on each dimension.

Since the linear structure is preserved, the equilibria can be characterized by a large cubic system, but his system is difficult to analyze.

7 Conclusion

I study the design of predictive scores—like FICO credit scores—when the agents being scored behave strategically. When information is dispersed and costly to acquire, it is common for intermediaries to aggregate information from different sources. Instead of disclosing everything they learn, intermediaries often transmit simple scores. I show that with appropriate choice of feature weights, the intermediary can mitigate a commitment problem for the receiver. In order to induce the most accurate decisions, the intermediary should commit to make predictions that are strictly suboptimal, given the information collected.

There are natural directions for future work. To focus on the weighting of different features, I study a static model. In a dynamic model, I could analyze the relative weights on different features at different times. I also assume that the intermediary knows the distribution of the sender's type and latent quality. It would be interesting to try to estimate the moments of the distribution from observed behavior. This would be a first step towards applying the theory to design more accurate scoring systems.

A Proofs

A.1 Proof of Lemma 1

This result is implied by Lemma 3 (Appendix B.4).

A.2 Proof of Proposition 1

This result is proved in the main text (Section 3.3).

A.3 Proof of Theorem 1

I separately prove existence and uniqueness. Existence does not require Assumption B.

Existence Drop Assumption B and let the receiver's bliss point be $\theta + v^T d$, for some fixed vector v. Define a function $g\colon \mathbf{R}^k \to \mathbf{R}^k$ by

$$g(b) = \operatorname{var}^{-1}(\eta + b \circ \delta) \operatorname{cov}(\eta + b \circ \delta, \theta + v^T(b \circ \delta)).$$

There is a linear equilibrium with coefficient vector b if and only if b is a fixed point of g. This function g is continuous. To apply Brouwer, I check that g is uniformly bounded.

For each b, the vector $g(b)$ be the unique minimizer of the function

$$z \mapsto \operatorname{var}\big(z^T(\eta + b \circ \delta) - \theta - v^T(b \circ \delta)\big). \tag{7}$$

Comparing $z = v$ with $z = g(b)$ in (7) gives

$$\begin{aligned}\operatorname{var}(v^T\eta - \theta) &\geq \operatorname{var}\big(g(b)^T(\eta + b \circ \delta) - \theta - v^T(b \circ \delta)\big) \\ &\geq \mathbf{E}[\operatorname{var}(g(b)^T\eta - \theta|\delta)],\end{aligned}$$

where the inequality follows from the law of total variance. Expand the last expression to get

$$g(b)^T\,\mathbf{E}[\operatorname{var}(\eta|\delta)]g(b) - 2g(b)^T\,\mathbf{E}[\operatorname{cov}(\eta, \theta|\delta)] + \mathbf{E}[\operatorname{var}(\theta|\delta)].$$

The matrix $\mathbf{E}[\operatorname{var}(\eta|\delta)]$ is a positive scalar multiple of $\Sigma_{\eta\eta} - \Sigma_{\eta\delta}\Sigma_{\delta\delta}^{\dagger}\Sigma_{\delta\eta}$ (Cambanis et al., 1981, Corollary 5). By Assumption A.i, $\mathbf{E}[\operatorname{var}(\eta|\delta)]$ has full rank. It follows that there is some constant r such that $\|g(b)\| \leq r$ for all b.[27] Now restrict g to the ball $B(0, r)$ and apply Brouwer.

[27] Here are the details. Let λ be the smallest eigenvalue of $\mathbf{E}[\operatorname{var}(\eta|\delta)]$, and let C denote the magnitude of the vector $\mathbf{E}[\operatorname{cov}(\eta, \theta|\delta)]$. For all b, we have $\operatorname{var}(v^T\eta - \theta) \geq \lambda^2\|g(b)\|^2 - 2C\|g(b)\|$.

Uniqueness Now impose Assumption B, except relax the assumption $\Sigma_{\delta\theta} = 0$ to $\Sigma_{\delta\theta} \leq 0$. The equilibrium condition (3) can be written as

$$0 = \Sigma_{\eta\eta} b + \operatorname{diag}(b)\Sigma_{\delta\delta}(b \circ b) - \Sigma_{\eta\theta} - \operatorname{diag}(b)\Sigma_{\delta\theta}. \tag{8}$$

I construct a strictly convex function $\Phi\colon \mathbf{R}^k \to \mathbf{R}^k$ such $\nabla\Phi(b)$ equals (a scalar multiple of) the right side of (8). Uniqueness follows from that fact that a strictly convex function has at most one stationary point.[28]

Define $\Phi\colon \mathbf{R}^k \to \mathbf{R}$ by

$$\begin{aligned}\Phi(b) &= \operatorname{var}\big(b^T\eta + (1/2)(b \circ b)^T\delta - \theta\big) \\ &= b^T\Sigma_{\eta\eta} b + (1/2)(b \circ b)^T\Sigma_{\delta\delta}(b \circ b) + \sigma_\theta^2 - b^T\Sigma_{\eta\theta} - (b \circ b)^T\Sigma_{\delta\theta}.\end{aligned} \tag{9}$$

Differentiating, we see that $\nabla\Phi(b)$ equals the right side of (8).

Now I check that Φ is strictly convex. The first term is strictly convex because $\Sigma_{\eta\eta}$ has full rank. The quadratic term is convex because $-\Sigma_{\delta\theta}$ is nonnegative. Since $\Sigma_{\delta\delta}$ is positive semidefinite and componentwise nonnegative, the quartic term is convex. To see this, observe that the gradient and Hessian of the quadratic term are given by

$$2\operatorname{diag}(z)\Sigma_{\delta\delta}(z \circ z) \quad \text{and} \quad 4\operatorname{diag}(z)\Sigma_{\delta\delta}\operatorname{diag}(z) + 2\operatorname{diag}\big(\Sigma_{\delta\delta}(z \circ z)\big).$$

Each term of the Hessian is positive semidefinite: the first matrix it is congruent to $\Sigma_{\delta\delta}$, and the second matrix is a diagonal matrix with nonnegative entries.

A.4 Proof of Proposition 2

To simplify notation for this proof, define the cost function $c\colon \mathbf{R}^k \times \mathbf{R}^k_{++} \to \mathbf{R}$ by

$$c(d, \delta) = (1/2)\sum_{j=1}^{k} d_j^2/\delta_j.$$

The Hessian can be written as 2×2 block matrix, consisting of $n \times n$ submatrices, which are denoted by the operators $\nabla^2_{11}, \nabla^2_{12}, \nabla^2_{21}, \nabla^2_{22}$.

Suppose for a contradiction that there exists a fully-revealing equilibrium (d, y). The strategy y must be pure since the receiver's utility function is strictly convex. We may assume without loss that d is pure; otherwise, select for each type some distortion vector in the support, and denote this selection by d.

The details are a bit technical, but the intuition is straightforward. I will show that some type with high distortion ability will prefer to mimic a type with a higher intrinsic level.

[28] A strictly convex function may not have any stationary points; consider the map $z \mapsto \exp(z_1 + \cdots + z_k)$. But I have independently established existence.

Lemma 2 (Sequence of types)
Suppose $\Sigma_{\eta\eta}$ and $\Sigma_{\delta\delta}$ have full rank. We can select the following types in T: (η^0, δ^0), another type $(\eta^0, t\delta^0)$ for some $t > 1$, and a sequence of types (η^i, δ^i) converging to (η^0, δ^0) satisfying the following:

1. *there exists $\kappa > 0$ such that $\beta^T(\eta^i - \eta^0) \geq \kappa\|\eta^i - \eta^0\|$ for all i;*
2. *$\|\delta^i - \delta^0\|/\|\eta^i - \eta^0\|$ is bounded.*

Now define

$$\bar{x}^0 = \eta^0 + d(\eta^0, t\delta^0)$$

and for all $i \geq 0$, set

$$x^i = \eta^i + d(\eta^i, \delta^i).$$

Of course we then have $y(x^i) = \beta^T\eta^i$ for all $i > 0$ and for $i = 0$, we have $y(x^0) = y(\bar{x}^0) = \beta^T\eta^0$. By passing to a subsequence, we may assume that x^i is convergent. Denote the limit by x^*.

The equilibrium condition for types (η^i, δ^i) and (η^0, δ^0) together imply that

$$c(x^* - \eta^0, \delta^0) = c(x^0 - \eta^0, \delta^0). \tag{10}$$

To complete the proof, I will show that for some i, type $(\eta^0, t\bar{\delta}^0)$ strictly prefers to mimic type (η^i, δ^0) by choosing distortion $x^i - \eta^0$, that is,

$$\begin{aligned}\beta^T(\eta^i - \eta^0) &> c(x^i - \eta^0, t\delta^0) - c(\bar{x}^0 - \eta^0, t\delta^0) \\ &= t^{-1}[c(x^i - \eta^0, \delta^0) - c(\bar{x}^0 - \eta^0, \delta^0)].\end{aligned}$$

Comparing types $(\eta^0, t\delta^0)$ and (η^0, δ^0), we see that $c(\bar{x}^0 - \eta^0, \delta^0) = c(x^0 - \eta^0, \delta^0)$. Together with (10), this implies that the desired inequality is equivalently

$$\beta^T(\eta^i - \eta^0) > t^{-1}[c(x^i - \eta^0, \delta^0) - c(x^* - \eta^0, \delta^0)]. \tag{11}$$

To establish this inequality, first observe from the equilibrium condition that for all i and j,

$$c(x^i - \eta^i, \delta^i) - c(x^j - \eta^i, \delta^i) \leq \beta^T(\eta^i - \eta^j) \leq c(x^i - \eta^j, \delta^j) - c(x^j - \eta^j, \delta^j).$$

Passing to the limit in j, we see that for all i,

$$c(x^i - \eta^i, \delta^i) - c(x^* - \eta^i, \delta^i) \leq \beta^T(\eta^i - \eta^0) \leq c(x^i - \eta^0, \delta^0) - c(x^* - \eta^0, \delta^0). \tag{12}$$

By the first inequality in (12), to prove (11) it suffices to check that

$$t^{-1}[c(x^i - \eta^0, \delta^0) - c(x^* - \eta^0, \delta^0)] < c(x^i - \eta^i, \delta^i) - c(x^* - \eta^i, \delta^i). \tag{13}$$

Now expand the right side as

$$\begin{aligned}[c(x^i-\eta^i,\delta^i)-c(x^*-\eta^i,\delta^i)]-[c(x^i-\eta^i,\delta^0)-c(x^*-\eta^i,\delta^0)]\\ +[c(x^i-\eta^i,\delta^0)-c(x^*-\eta^i,\delta^0)]-[c(x^i-\eta^0)-c(x^*-\eta^0)]\\ +c(x^i-\eta)-c(x^*-\eta).\end{aligned}$$

Applying the mean value theorem to the second differences on the right, we can find w^i, ξ^i and z^i such that the right side equals

$$\begin{aligned}(x^i-x^*)^T\nabla^2_{12}c(w^i,\xi^i)(\delta^i-\delta^0)+(x^i-x^*)^T\nabla^2_{11}c(z^i,\delta^0)(\eta^0-\eta^i)\\ +c(x^i-\eta^0,\delta^0)-c(x^*-\eta^0,\delta^0),\end{aligned}$$

where w^i is on the line segment between $x^i-\eta^i$ and $x^*-\eta^i$; ξ^i is on the line segment between δ^i and δ^0; and finally $z^i-(x^*-\eta^0)$ is in the parallelogram spanned by x^i-x^* and $\eta^0-\eta^i$.[29] Rearranging, we see that (13) is equivalent to

$$\begin{aligned}0<(1-t^{-1})[c(x^i-\eta)-c(x^*-\eta^0)]\\ +(x^i-x^*)^T\nabla^2_{12}c(w^i,\xi^i)(\delta^i-\delta^0)+(x^i-x)^T\nabla^2_{11}c(z^i,\delta^0)(\eta^i-\eta),\end{aligned}\tag{14}$$

Combining the right side of (12) and Lemma 2.1, we have

$$c(x^i-\eta)-c(x^*-\eta)\geq\beta^T(\eta^i-\eta^0)\geq\kappa\|\eta^i-\eta^0\|.$$

Therefore, after dividing both sides of (14) by $\|\eta^i-\eta^0\|$, the new right side is bounded below by

$$(1-t^{-1})\kappa+(x^i-x^*)^T\nabla^2_{12}c(w^i,\xi^i)\frac{\delta^i-\delta^0}{\|\eta^i-\eta^0\|}+(x^i-x)^T\nabla^2_{11}c(z^i,\delta^0)\frac{\eta^i-\eta^0}{\|\eta^i-\eta^0\|}.$$

Since c is twice continuously differentiable, the last two terms vanish in the limit, so this expression must be strictly positive for all i sufficiently large.

A.5 Proof of Lemma 2

Find some ellipse $E(\mu,\Sigma,r)$ in the support, with $r>0$. Simply choose η^0 such that $(\eta^0-\mu_\eta)\Sigma^{-1}_{\eta\eta}(\eta^0-\mu_\eta)$ is strictly between 0 and r^2. Then pick δ^0 to be the smaller of the two scalar multiples of μ_δ that are on the ellipse, and take $t\delta^0$ to be the larger.

[29] For any twice continuously differentiable function $g\colon \mathbf{R}^2\to\mathbf{R}$, the mean value theorem implies that

$$g(h_1,h_2)-g(h_1,0)-g(0,h_2)+g(0,0)=h_1h_2D_{12}g(\theta_1h_1,\theta_2h_2),$$

for some θ_1 and θ_2 in $[0,1]$. The decomposition in the proof applies this expansion to a particular choice of g.

Then pick η^i to be a scalar multiple of η^0 and let δ^i be a scalar multiple of δ^0 so that (η^i, δ^i) is in $E(\mu, \Sigma, r)$.

A.6 Proof of Proposition 3

To simplify notation, set $t = \alpha^2$. The equilibrium condition is

$$0 = \Sigma_{\eta\eta} b + t \operatorname{diag}(b) \Sigma_{\delta\delta} (b \circ b) - \Sigma_{\eta\theta}. \tag{15}$$

The receiver's loss equals the posterior variance

$$b^T \Sigma_{\eta\eta} b - 2 b^T \Sigma_{\eta\theta} + \sigma_\theta^2 + t (b \circ b)^T \Sigma_{\delta\delta} (b \circ b), \tag{16}$$

where b is a function of t, defined by the equilibrium condition, bu I suppress this dependence. I show that this expression is strictly increasing in t.

I use the implicit function theorem to compute the derivative $\dot{b}$ of b with respect to t. The derivative of the left side with respect to t is $\operatorname{diag}(b)\Sigma_{\delta\delta}(b \circ b)$; the derivative with respect to b is given by

$$D = \Sigma_{\eta\eta} + t \operatorname{diag}(\Sigma_{\delta\delta}(b \circ b)) + 2t \operatorname{diag}(b) \Sigma_{\delta\delta} \operatorname{diag}(b).$$

This matrix D is positive definite and hence has full rank.[30] By the implicit function theorem, b is a differentiable function of t, and the derivative is given by

$$\dot{b} = -D^{-1} \operatorname{diag}(b) \Sigma_{\delta\delta} (b \circ b).$$

Therefore, the total derivative of (16) with respect to t is

$$\Big[2\Sigma_{\eta\eta} b - 2\Sigma_{\eta\theta} + 4t (b \circ b)^T \Sigma_{\delta\delta} \operatorname{diag}(b) \Big] \dot{b} + (b \circ b)^T \bar{\Sigma}_{\delta\delta} (b \circ b).$$

By (15), the term in brackets reduces to $2t(b \circ b)^T \Sigma_{\delta\delta} \operatorname{diag}(b)$. Plug in the expression for $\dot{b}$ to get

$$- 2t (b \circ b)^T \Sigma_{\delta\delta} \operatorname{diag}(b) D^{-1} \operatorname{diag}(b) \Sigma_{\delta\delta} (b \circ b) + (b \circ b)^T \Sigma_{\delta\delta} (b \circ b). \tag{17}$$

To complete the proof, I show that this expression is positive. We have

$$D \succ 2t \operatorname{diag}(b) \Sigma_{\delta\delta} \operatorname{diag}(b).$$

The matrix on the right is not necessarily invertible, so we use the Moore–Penrose

[30] The matrix $\Sigma_{\eta\eta}$ is positive definite and the next two matrices are positive semidefinite; in particular, $\bar{\Sigma}_{\delta\delta}$ is a nonnegative matrix, and hence the diagonal matrix $\operatorname{diag}(t\bar{\Sigma}_{\delta\delta}(b \circ b))$ has nonnegative entries and hence is positive semidefinite.

psuedoinverse. We have

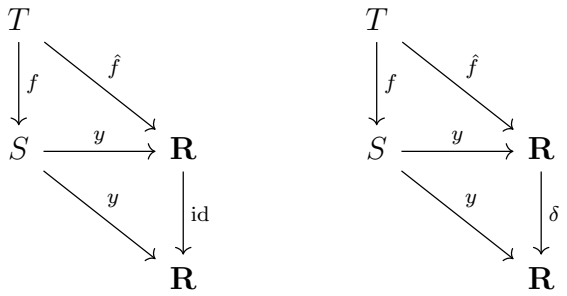
$$D^{-1} \prec \left[2t \operatorname{diag}(b)\Sigma_{\delta\delta} \operatorname{diag}(b)\right]^{\dagger}.$$

Since $\Sigma_{\delta\delta}$ is positive semidefinite and and b is nonzero, we have

$$\begin{aligned}
&2t(b \circ b)^T \Sigma_{\delta\delta} \operatorname{diag}(b) D^{-1} \operatorname{diag}(b) \Sigma_{\delta\delta}(b \circ b) \\
&< 2t(b \circ b)^T \Sigma_{\delta\delta} \operatorname{diag}(b) \left[2t \operatorname{diag}(b)\Sigma_{\delta\delta} \operatorname{diag}(b)\right]^{\dagger} \operatorname{diag}(b) \Sigma_{\delta\delta}(b \circ b) \\
&= b^T \operatorname{diag}(b)\Sigma_{\delta\delta} \operatorname{diag}(b) \left[\operatorname{diag}(b)\Sigma_{\delta\delta} \operatorname{diag}(b)\right]^{\dagger} \operatorname{diag}(b) \Sigma_{\delta\delta} \operatorname{diag}(b) b \\
&= b^T \operatorname{diag}(b)\Sigma_{\delta\delta} \operatorname{diag}(b) b \\
&= (b \circ b)^T \Sigma_{\delta\delta}(b \circ b).
\end{aligned}$$

A.7 Proof of Proposition 4

Let $\sigma = (\sigma_S, \sigma_R)$ be an implementable decision rule. There exists a scoring rule f and an strategy profile (d, y) that together induce σ. Consider the direct implementation with $\hat{d} = d$, $\hat{f} = y \circ f$ and $\hat{y} = \text{id}$. These compositions are shown in the first commutative diagram below. This diagram shows that this strategy profile replicates the decision rule. To see that there are no profitable deviations, observe that the sender has the same set of deviations in the new game. For the receiver, the second diagram shows that any deviation δ by the receiver can be replicated in the original game by the deviation $y' = \delta \circ y$, so it cannot be profitable.



A.8 Proof of Proposition 5

I separately prove existence and uniqueness. Existence does not require Assumption B.

Existence Drop Assumption B and let the receiver's bliss point be $\theta + v^T d$, for some fixed vector v. By Theorem 1, there exists a linear equilibrium, $\hat{b}$. Every optimal vector must in particular outperform $\hat{b}$, so we can equivalently optimize over

only the vectors b satisfying

$$\text{var}(\bar{b}^T(\eta + \bar{b} \circ \delta) - \theta - v^T(\bar{b} \circ \delta)) \geq \text{var}(b^T(\eta + b \circ \delta) - \theta - v^T(b \circ \delta)).$$

Now follow the argument in the proof of Theorem 1 in Appendix A.3 to show that this set is bounded. The set of obedient vectors b satisfying this inequality is therefore closed and bounded, and hence compact. The objective is continuous, so a solution exists.

Uniqueness First I establish an alternative characterization of the solution. If b is a solution, then by the Lagrange multiplier theorem, there exists a Lagrange multiplier λ such that

$$\Sigma_{\eta\eta} b - 2\Sigma_{\eta\theta} + 4\,\text{diag}(b)\Sigma_{\delta\delta}(b \circ b) = \lambda\left[\Sigma_{\eta\eta} b - \Sigma_{\eta\theta} + 4\,\text{diag}(b)\Sigma_{\delta\delta}(b \circ b)\right].$$

Rearranging gives

$$(1-\lambda)\Sigma_{\eta\eta} b - 2\Sigma_{\eta\theta} + 4\,\text{diag}(b)\Sigma_{\delta\delta}(b \circ b) = \lambda\Sigma_{\eta\theta}.$$

The solution is not $b = 0$, so we must have $\lambda \neq 1$. Therefore,

$$0 = (1-\lambda)\left[\Sigma_{\eta\eta} b - (2-\lambda)/(2-2\lambda)\Sigma_{\eta\theta} + 4\,\text{diag}(b)\Sigma_{\delta\delta}(b \circ b)\right].$$

Taking $\bar{\lambda} = (2-\lambda)/(2-2\lambda)$, this is implies that b minimizes the function

$$\text{var}(b^T(\eta + b \circ \delta) - \bar{\lambda}\theta).$$

for some $\bar{\lambda}$. This function is strictly convex, so for each $\bar{\lambda}$, there is exactly one solution. It remains to check that the solutions for two different values of $\bar{\lambda}$ cannot simultaneously satisfy the obedience constraint. For b satisfying the obedience constraint, we have

$$\text{var}(b^T(\eta + b \circ \delta) - \bar{\lambda}\theta) = \bar{\lambda}\sigma_\theta^2 - \bar{\lambda}\Sigma_{\eta\theta},$$

so the minimizer does not depend on $\bar{\lambda}$.

A.9 Proof of Theorem 2

I prove (i) $\implies$ (ii) $\implies$ (iii) $\implies$ (i).

Suppose (i). Let b denote the common value of b^{signal} and b^{score}. Then there exists $\bar{\lambda}$ such that

$$\begin{aligned}
\Sigma_{\eta\theta} - \Sigma_{\eta\eta} b &= \text{diag}(b)\Sigma_{\delta\delta}(b \circ b), \\
\bar{\lambda}\Sigma_{\eta\theta} - \Sigma_{\eta\eta} b &= 2\,\text{diag}(b)\Sigma_{\delta\delta}(b \circ b).
\end{aligned}$$

Subtracting gives

$$\bar{\lambda}\Sigma_{\eta\theta} = \operatorname{diag}(b)\Sigma_{\delta\delta}(b \circ b). \tag{18}$$

Hence $(1-\bar{\lambda})\Sigma_{\eta\theta} = \Sigma_{\eta\eta}b$, and we conclude that $b = (1-\bar{\lambda})\beta$, which implies (ii).

Now suppose (ii). There exists λ such that $b^{\text{signal}} = \lambda\beta$, so

$$(1-\lambda)\Sigma_{\eta\theta} = \lambda^3 \operatorname{diag}(\beta)\Sigma_{\delta\delta}(\beta \circ \beta).$$

We cannot have $\lambda = 0$ or $\lambda = 1$ because exactly one side of this equality would vanish. Hence (iii) follows.

Finally, suppose (iii). Write $\operatorname{diag}(\beta)\Sigma_{\delta\delta}(\beta \circ \beta) = \nu\Sigma_{\eta\theta}$. Left multiplying by b^T shows that $\nu > 0$. If $b = t\beta$, then we have

$$\Sigma_{\eta\theta} - \Sigma_{\eta\eta}b = (1-t)\Sigma_{\eta\theta},$$

and

$$\operatorname{diag}(b)\Sigma_{\delta\delta}(b \circ b) = t^3\nu.$$

The equilibrium system comes $(1-t) = t^3\nu$ gives $0 = \nu t^3 + t - 1$. The right side is strictly increasing, so this equation has a unique solution, denoted $\bar{t}$. Therefore, $b = t^\star\beta$ also satisfies the Lagrangian condition for the scoring solution, where $\bar{\lambda} = 1 + \bar{t}^3\nu$, as needed.

A.10 Proof of Proposition 6

Consider two values $\tilde{\Sigma}_{\delta\delta}$ and $\Sigma_{\delta\delta}$ with $\tilde{\Sigma}_{\delta\delta} \succeq \Sigma_{\delta\delta}$. Let $\tilde{b}^{\text{score}}$ and b^{score} be the corresponding scoring solutions. If $\operatorname{var}(\delta) = \Sigma_{\delta\delta}$, the intermediary can secure the optimal payoff when $\operatorname{var}(\delta) = \tilde{\Sigma}_{\delta\delta}$ by using coefficient vector $\tilde{b}^{\text{score}}$ together with noise with expected variance $(\tilde{b}^{\text{score}})^T(\tilde{\Sigma}_{\delta\delta} - \Sigma_{\delta\delta})\tilde{b}^{\text{score}}$. But noise is strictly suboptimal by Proposition 8, so the receiver's payoff from $\Sigma_{\delta\delta}$ must be strictly higher.

A.11 Proof of Theorem 3

First we need some notation. In the setting of uncorrelated errors, for any vector b, let

$$\hat{b}_j = (1 - \rho + \sigma^2_{\varepsilon,j})b_j$$

for each j, and denote the corresponding vector by $\hat{b}$.

We claim that, in the setting of uncorrelated errors, each of the following expressions is independent of j:

$$\hat{b}_j^{\text{score}} + 2(\hat{b}_j^{\text{score}})^3/\tau_j, \qquad \frac{\hat{b}_j^{\text{score}} + (\hat{b}_j^{\text{score}})^3/\tau_j}{w_j(b^{\text{score}})}. \tag{19}$$

The result then follows by observing that for all i and j,

$$\begin{aligned} & w_i(b^{\text{score}}) \geq w_j(b^{\text{score}}) \\ \iff & \hat{b}_i^{\text{score}} + (\hat{b}_i^{\text{score}})^3/\tau_i \geq \hat{b}_j^{\text{score}} + (\hat{b}_j^{\text{score}})^3/\tau_j \\ \iff & \hat{b}_i^{\text{score}} \geq \hat{b}_j^{\text{score}} \\ \iff & \tau_i \geq \tau_j. \end{aligned}$$

The middle biconditional follows from writing

$$\hat{b}_i^{\text{score}} + (\hat{b}_i^{\text{score}})^3/\tau_i = (1/2)\hat{b}_i^{\text{score}} + (1/2)\big[\hat{b}_i^{\text{score}} + 2(\hat{b}_i^{\text{score}})^3/\tau_i.$$

Now we establish the claim. Define the variance vectors

$$v_\varepsilon = (\sigma^2_{\varepsilon,1}, \ldots, \sigma^2_{\varepsilon,k}), \qquad v_\delta = (\sigma^2_{\delta,1}, \ldots, \sigma^2_{\delta,k}).$$

Now compute the covariances in the setting with uncorrelated errors:

$$\begin{aligned} \Sigma_{\eta\theta} &= [\rho + (1-\rho)/k]\mathbf{1} \\ \Sigma_{\eta\eta} &= \rho\mathbf{1}\mathbf{1}^T + (1-\rho)I + \operatorname{diag}(v_\varepsilon), \\ \Sigma_{\delta\delta} &= \operatorname{diag}(v_\delta). \end{aligned} \tag{20}$$

Consider the first expression in (19). The scoring solution is given by

$$\lambda\Sigma_{\eta\theta} - \Sigma_{\eta\eta}b = 2\operatorname{diag}(b)\Sigma_{\delta\delta}(b \circ b)$$

for some positive λ. Plugging in the expressions in (20) this condition becomes

$$[\lambda\rho + \lambda(1-\rho)/k - \rho\mathbf{1}^T b]\mathbf{1} = (1-\rho)b + v_\varepsilon \circ b + 2v_\delta \circ (b^{\circ 3}).$$

Therefore, for all j, we have

$$\lambda\rho + \lambda(1-\rho)/k - \rho\mathbf{1}^T b^{\text{score}} = (1 - \rho + \sigma^2_{\varepsilon,j})b_j^{\text{score}} + 2\sigma^2_{\delta,j}(b_j^{\text{score}})^3. \tag{21}$$

Thus, the first expression in (19) is independent of j, and moreover the solution b^{score} has strictly positive entries.

Now consider the second expression in (19). For any vector b, the best response $\operatorname{BR}(b)$ satisfies

$$\operatorname{var}(\eta + b \circ \delta)\operatorname{BR}(b) = \operatorname{cov}(\eta + b \circ \delta, \theta),$$

he

$$(\Sigma_{\eta\eta} + \operatorname{diag}(b)\Sigma_{\delta\delta}\operatorname{diag}(b))\operatorname{BR}(b) = \Sigma_{\eta\theta},$$

hence

$$\Sigma_{\eta\theta} - \Sigma_{\eta\eta}\operatorname{BR}(b) = \operatorname{diag}(b)\Sigma_{\delta\delta}\operatorname{diag}(b)\operatorname{BR}(b)$$

lugging in the expressions in (20) this condition becomes

$$[\rho + (1-\rho)/k - \rho \mathbf{1}^T \operatorname{BR}(b)]\mathbf{1} = (1-\rho)\operatorname{BR}(b) + v_\varepsilon \circ \operatorname{BR}(b) + b^{\circ 2} \circ v_\delta \circ \operatorname{BR}(b).$$

Therefore, for all j, we have

$$[\rho + (1-\rho)/k - \rho \mathbf{1}^T \operatorname{BR}(b)] = (1 - \rho + \sigma_{e,j}^2 + \sigma_{\delta,j}^2 b_j^2) \operatorname{BR}_j(b).$$

For b with nonzero entries we can multiply and divide by b_j to express the right side as

$$\frac{(1-\rho+\sigma_{\varepsilon,j}^2)b_j + b_j^3 \sigma_{\delta,j}^2}{w_j(b)} = \frac{\hat{b}_j + \hat{b}_j^3/\tau_j}{w_j(b)}.$$

Finally, we check that at least one feature is underweighted and at least one is overweighted. For any b satisfying the scoring obedience constraint,

$$b^T \operatorname{var}((\eta + b \circ \delta))b = b^T \operatorname{cov}(\eta + b \circ \delta, \theta).$$

On the other hand, the best response condition is

$$\operatorname{var}(\eta + b \circ \delta) \operatorname{BR}(b) = \operatorname{cov}(\eta + b \circ \delta, \theta),$$

Left multiply to get

$$b^T \operatorname{var}((\eta + b \circ \delta))b = b^T \operatorname{cov}(\eta + b \circ \delta, \theta).$$

comparing, we conclude that

$$b^T \operatorname{var}(\eta + b \circ \delta)b = b^T \operatorname{var}(\eta + b \circ \delta) \operatorname{BR}(b)$$

hence

$$b^T(\Sigma_{\eta\eta} + \operatorname{diag}(b)\Sigma_{\delta\delta} \operatorname{diag}(b))(b - \operatorname{BR}(b)) = 0.$$

But this is the inner product of a strictly positive vector and $(b - \operatorname{BR}(b))$. Hence $b - \operatorname{BR}(b)$ is not comparable with zero vector, so b and c are incomparable.

A.12 Proof of Theorem 4

First I establish the following inequalities:

$$\|\beta\|_{4,\delta}^4 > \|b^{\text{signal}}\|_{4,\delta}^4 \begin{cases} \geq \|b^{\text{score}}\|_{4,\delta}^4, \\ > \|b^{\text{screen}}\|_{4,\delta}^4. \end{cases}$$

To compare β and b^{signal}, recall that they respectively minimize

$$\text{var}(b^T\eta - \theta) \quad \text{and} \quad \text{var}(b^T\eta - \theta) + (1/2)\|b\|_{4,\delta}^4.$$

To compare b^{signal} and b^{score}, recall that they respectively minimize the functions

$$\text{var}(b^T\eta - \theta) + (1/2)\|b\|_{4,\delta}^4 \quad \text{and} \quad \text{var}(b^T\eta - \theta) + \|b\|_{4,\delta}^4,$$

over the set of obedient scoring rules. In fact, b^{signal} is a global minimizer over all or $\mathbf{R}^k$.

Finally, to compare b^{signal} and b^{screen}, recall that they minimize the functions

$$\text{var}(b^T\eta - \theta) + (1/2)\|b\|_{4,\delta}^4 \quad \text{and} \quad \text{var}(b^T\eta - \theta) + \|b\|_{4,\delta}^4.$$

It remains to prove that $\|b^{\text{score}}\|_{4,\delta}^4 > \|b^{\text{screen}}\|_{4,\delta}^4$. There exists λ such that b^{score} satisfies the first-order condition

$$\Sigma_{\eta\eta} b + 2\,\text{diag}(b)\Sigma_{\delta\delta}(b \circ b) - \lambda\Sigma_{\eta\theta} = 0. \tag{22}$$

I apply the implicit function theorem in λ. If $\lambda = 1$, this condition gives the screening solution. To compare the scoring and the screening solution, I continuously interpolate between scoring and screening, through the parameter λ, and I study the local effect on the two components of the receiver's losses. The derivative of this left side with respect to λ is $-\Sigma_{\eta\theta}$. The derivative of the left side with respect to b is given by

$$D = \Sigma_{\eta\eta} + 2\,\text{diag}(\Sigma_{\delta\delta}(b \circ b)) + 4\,\text{diag}(b)\Sigma_{\delta\delta}\,\text{diag}(b).$$

By the implicit function theorem, b is locally differentiable in λ and

$$\dot{b} = D^{-1}\Sigma_{\eta\theta}.$$

From (22), we have

$$\lambda\Sigma_{\eta\theta} - \Sigma_{\eta\eta} b = 2\,\text{diag}(b)\Sigma_{\delta\delta}(b \circ b).$$

Left multiplying by b^T gives

$$\lambda b^T\Sigma_{\eta\theta} - b^T\Sigma_{\eta\eta} b = 2\|b\|_{\delta,4}^4.$$

Therefore, it suffices to prove that the left side is strictly increasing in λ. Differentiating with respect to λ gives

$$b^T\Sigma_{\eta\theta} + \lambda\Sigma_{\eta\theta}^T\dot{b} - 2b^T\Sigma_{\eta\eta}\dot{b}.$$

Plug in the expression for $\dot{b}$ from the implicit function theorem to get

$$b^T\Sigma_{\eta\theta} + \lambda\Sigma_{\eta\theta}^T D^{-1}\Sigma_{\eta\theta} - 2b^T\Sigma_{\eta\eta}D^{-1}\Sigma_{\eta\theta}. \tag{23}$$

To bound the last term, I use a simple inequality for quadratic forms. If A is positive semidefinite, and x and y are vectors, then expanding $(x-y)^T A(x-y)$ shows that $2x^TAy \leq x^TAx + y^TAy$. Applying this here gives

$$\begin{aligned} 2b^T\Sigma_{\eta\eta}D^{-1}\Sigma_{\eta\theta} &= 2\lambda^{-1}b^T\Sigma_{\eta\eta}D^{-1}(\lambda\Sigma_{\eta\theta}) \\ &\leq \lambda^{-1}\left[b^T\Sigma_{\eta\eta}D^{-1}\Sigma_{\eta\eta}b + (\lambda\Sigma_{\eta\theta})^T D^{-1}(\lambda\Sigma_{\eta\theta})\right] \\ &= \lambda^{-1}b^T\Sigma_{\eta\eta}D^{-1}\Sigma_{\eta\eta}b + \lambda\Sigma_{\eta\theta}D^{-1}\Sigma_{\eta\theta}. \end{aligned}$$

Since $D \succeq \Sigma_{\eta\eta}$, the right side is bounded above by

$$\lambda^{-1}b^T\Sigma_{\eta\eta}b + \lambda\Sigma_{\eta\theta}D^{-1}\Sigma_{\eta\theta}.$$

Plugging this inequality into (23), the last term cancels and we get the lower bound

$$b^T\Sigma_{\eta\theta}b - \lambda^{-1}b^T\Sigma_{\eta\eta}b.$$

Rearrange this expression and use (22) to get

$$\lambda^{-1}b^T\left(\lambda\Sigma_{\eta\theta} - \Sigma_{\eta\eta}b\right) = \lambda^{-1}(b\circ b)^T\Sigma_{\delta\delta}(b\circ b).$$

This final expression is strictly positive.

A.13 Proof of Proposition 7

For signaling, scoring, and screening, the respective first-order conditions are given as follows, for some $\lambda > 1$:

$$\begin{aligned} \Sigma_{\eta\theta} - \Sigma_{\eta\eta}b &= \operatorname{diag}(b)\Sigma_{\delta\delta}(b\circ b), \\ \lambda\Sigma_{\eta\theta} - \Sigma_{\eta\eta}b &= 2\operatorname{diag}(b)\Sigma_{\delta\delta}(b\circ b), \\ \Sigma_{\eta\theta} - \Sigma_{\eta\eta}b &= 2\operatorname{diag}(b)\Sigma_{\delta\delta}(b\circ b). \end{aligned}$$

Adjusting the argument in the proof of Theorem 3 in Appendix A.11, we the corresponding conditions for the three cases. For all j:

$$\begin{aligned} \rho + (1-\rho)/k - \rho\mathbf{1}^Tb &= (1-\rho+\sigma_{\varepsilon,j}^2)b_j + \sigma_{\delta,j}^2 b_j^3, \\ \lambda[\rho + (1-\rho)/k] - \rho\mathbf{1}^Tb &= (1-\rho+\sigma_{\varepsilon,j}^2)b_j + 2\sigma_{\delta,j}^2 b_j^3, \\ \rho + (1-\rho)/k - \rho\mathbf{1}^Tb &= (1-\rho+\sigma_{\varepsilon,j}^2)b_j + 2\sigma_{\delta,j}^2 b_j^3. \end{aligned}$$

With this system, I prove the parts in turn.

(i) In each of the three systems, the right side has the same sign as b_j, so all b_j have the same sign. If $b \leq 0$, then the left side is strictly positive, which yields a contradiction.

(ii) This follows immediately from the system of equations since the left side does not depend on j.

(iii) It suffices to show that

$$\lambda[\rho + (1-\rho)/k] - \mathbf{1}^T b^{\text{score}} > \rho + (1-\rho)/ - \mathbf{1}^T b^{\text{screen}}.$$

If the reverse inequality holds, then the first-order condition implies that $b^{\text{score}} \leq b^{\text{screen}}$, which then yields a contradiction.

(iv) The argument is the same as the proof of Theorem 3 Appendix A.11.

A.14 Proof of Proposition 8

I prove that if the scoring solution has noise, then $\Sigma_{\delta\theta} < 0$. Substituting the constraint into the intermediary's problem, we get the equivalent problem

$$\begin{aligned} &\text{maximize} && \operatorname{cov}\big(b^T\eta + (b \circ b)^T\delta, \theta\big) \\ &\text{subject to} && \operatorname{var}\big(b^T\eta + (b \circ b)^T\delta\big) + t^2 = \operatorname{cov}\big(b^T\eta + (b \circ b)^T\delta, \theta\big). \end{aligned}$$

Suppose for a contradiction that the maximum is achieved at $(\bar{b}, \bar{t})$, where $\bar{t} > 0$. By adjusting t, it is possible to change b in any direction, so a necessary condition is that there is local maximum in the objective. Write out the objective as

$$g(b) = b^T\Sigma_{\eta\theta} + (b \circ b)^T\Sigma_{\delta\theta}.$$

The necessary conditions for a local maximum are

$$\begin{aligned} \nabla g(b) &= \Sigma_{\eta\theta} + 2b \circ \Sigma_{\delta\theta} = 0. \\ \nabla^2 g(b) &= 2\operatorname{diag}(\Sigma_{\delta\theta}) \preceq 0. \end{aligned}$$

It follows that $\Sigma_{\delta\theta} \leq 0$. To rule out $\Sigma_{\delta\theta} = 0$, observe that if that case $\Sigma_{\eta\theta} = 0$, then the unique solution is $(b, t) = 0$.

A.15 Proof of Proposition 9

The intermediary's problem is

$$\begin{aligned} &\text{minimize} && \pi(1/2)(b \circ b)^T\mu_\delta + (1-\pi)\left[b^T\Sigma_{\eta\eta}b - 2b^T\Sigma_{\eta\theta} + (b \circ b)^T\Sigma_{\delta\delta} + t^2\right] \\ &\text{subject to} && b^T\Sigma_{\eta\theta} = b^T\Sigma_{\eta\eta}b + (b \circ b)^T\Sigma_{\delta\delta}(b \circ b). \end{aligned}$$

Substituting the constraint, we get an alternative program with the same set of maximizers:

$$\begin{aligned} \text{minimize} \quad & \pi(1/2)(b \circ b)^T \mu_\delta - (1-\pi) b^T \Sigma_{\eta\theta} \\ \text{subject to} \quad & b^T \Sigma_{\eta\theta} = b^T \Sigma_{\eta\eta} b + (b \circ b)^T \Sigma_{\delta\delta} (b \circ b) + t^2. \end{aligned}$$

Consider the unconstrained problem. The objective is convex. The first-order condition gives

$$\pi(b \circ \mu_g) = (1-\pi)\Sigma_{\eta\theta}.$$

Hence

$$\hat{b} = \pi^{-1}(1-\pi)\Sigma_{\eta\theta} \circ \mu_\delta^{-1},$$

where μ_δ^{-1} is the componentwise reciprocal of μ_δ. As $\hat{b}$ is scaled up, the term in brackets is strictly increasing, as is the noise ratio

$$\frac{\hat{b}^T \Sigma_{\eta\eta} \hat{b} + (\hat{b} \circ \hat{b})^T \Sigma_{\delta\delta} (\hat{b} \circ \hat{b}) - \hat{b}^T \Sigma_{\eta\theta}}{b^T \Sigma_{\eta\theta}} = \frac{\hat{b}^T \Sigma_{\eta\eta} \hat{b} + (\hat{b} \circ \hat{b})^T \Sigma_{\delta\delta} (\hat{b} \circ \hat{b})}{b^T \Sigma_{\eta\theta}} - 1.$$

B Additional results

B.1 Homogeneous intrinsic level

When distortion ability is homogeneous, there is a fully informative equilibrium (Proposition 1). Here I consider the signaling equilibrium when the intrinsic type is homogeneous. For this result, I drop Assumptions i and ii. There is always a trivial equilibrium with $b = 0$.

Proposition 10 (Homogeneous intrinsic level)
Suppose $\Sigma_{\delta\delta}$ is positive definite. If the intrinsic type η is nonrandom, then the nontrivial linear signaling equilibria are given as follows. For each subset J of $\{1, \dots, k\}$ with $\operatorname{reg}_J(\theta|\delta_j) > 0$,

$$b_{-J} = 0, \qquad b_J = \operatorname{reg}^{1/2}(\theta|\delta_j),$$

where the square root is evaluated componentwise.

Proof. It is immediate that these are equilibria. Even if the receiver observed δ_j, he would not have a deviation, but in these equilibria the receiver observes a garbling of δ_j because some components of δ_j are censored.

For the converse, suppose there is a linear equilibrium b. Set $J = \operatorname{supp} b$. The equibrilibrium condition requires that $b_j^2 = \operatorname{reg}_j(\theta|\delta_j)$ for each $j \in J$, so this equilibrium takes the desired form. □

B.2 Equilibrium stability

In the case where the equilibrium is unique, best-response dynamics will converge towards the equilibrium. If one player uses a linear strategy, then the best response

of the other player is linear, so best responses stay within the class of linear strategies, which can be represented by two vectors a and b. The linear strategies are represented by

$$d(\eta, \delta) = a \circ \delta, \qquad y(x) = b_0 + b^T \delta,$$

Initially each player chooses a linear strategy, denoted $a(0)$ and $b(0)$. Then the agents continuously adjust their strategies that that at each time t, the derivative $a'(t)$ is in the direction of the best response. The sender's best response is to match the vector used by the receiver, so

$$\dot{a} = b - a, \qquad \dot{b} = \operatorname{reg}(\theta | \eta + a \circ \delta) - b.$$

The intercept b_0 is adjusted similarly, but this evolution is pinned down by the dynamics of b.[31]

Theorem 5 (Convergence to equilibrium)
Starting from any linear strategy profile, continuous best-response dynamics converge to the unique linear equilibrium. Moreover, the rate of convergence is exponential.

Proof. There are two steps. First I construct a potential function, and then I apply convergence results for zero-sum games.

Construct zero-sum potential I construct a potential function Ψ as follows. We denote the linear strategies of the agents by a for the sender and b for the receiver. Define $\Psi \colon \mathbf{R}^k \times \mathbf{R}^k \to \mathbf{R}$ by

$$\Psi(a, b) = \operatorname{var}\big(b^T(\eta + a \circ \delta) - \theta\big) - (1/2) \operatorname{var}\big((a \circ a)^T \delta - \theta\big).$$

I claim that the function Ψ is a zero-sum potential. The sender chooses the first argument to maximize Ψ and the receiver chooses the second argument to minimize Ψ. For the receiver, this is a cardinal potential function. For the sender, this is an ordinal potential function, and it suffices to check that for all $b \in \mathbf{R}^k$, we have

$$b \in \operatorname*{argmax}_a \Psi(a, b).$$

To prove this, write

$$\begin{aligned} \Psi(a, b) = (a \circ b)^T \Sigma_{\delta\delta} (a \circ b) - (1/2)(a \circ a) \Sigma_{\delta\delta} (a \circ a) \\ - (a \circ b)^T \Sigma_{\delta\theta} + (1/2)(a \circ a)^T \Sigma_{\delta\theta} + h(b), \end{aligned}$$

[31] That is,

$$\dot{b}_0 = \mu_\theta - \operatorname{reg}^T(\theta | \eta + a \circ \delta)(\mu_\eta + a \circ \mu_\delta) - b_0.$$

where the function h is defined to collect the terms that do not depend on a. Expanding componentwise, we have

$$\Psi(a,b) = \sum_{i,j}[a_i b_i a_j b_j - a_i^2 a_j^2/2]\operatorname{cov}(\delta_i,\delta_j) - \sum_i [a_i b_i - a_i^2/2]\operatorname{cov}(\delta_i,\theta) + h(b).$$

For each i and j, we have $\operatorname{cov}(\delta_i,\delta_j) \geq 0$ and $\operatorname{cov}(\delta_i,\theta) \leq 0$. Hence, it suffices to check that each term in brackets is maximized with $a = b$. For the second term this is immediate. For the first, I check that

$$a_i b_i a_j b_j - a_i^2 a_j^2/2 \leq b_i^2 b_j^2 - b_i^2 b_j^2/2 = b_i^2 b_j^2/2,$$

which follows from the inequality $(a_i a_j - b_i b_j)^2 \geq 0$. This shows that the receiver's best response minimizes this function. This is the unique best response if for each i, at least one of the terms $\operatorname{cov}(\delta_i,\theta)$ or one entry of $\operatorname{cov}(\delta_i,\eta)$ is strictly positive. But we do not need this for the result.

Apply results from zero-sum dynamics As long as the sender's best response is contained in the best response for this zero-sum potential function, the result holds. In fact, we get a stronger result because the dynamics are more permissive. We apply the convergence result of Barron et al. (2010), which generalizes Hofbauer and Sorin (2006). To apply this theorem, it suffices to check that the correspondences BR_S and BR_R, defined by

$$\mathrm{BR}_S(b) = \operatorname*{argmax}_{a'} \Psi(a',b), \quad \text{and} \quad \mathrm{BR}_R(a) = \operatorname*{argmax}_{b'} \Psi(a,b').$$

are convex valued and uppersemicontinuous. Upper semicontinuity is immediate from Berge's theorem. The receiver's best-response correspondence is singleton-valued and hence convex-valued. For the sender, notice that

$$\mathrm{BR}_S(b) = \{b_J\} \times \mathbf{R}_{-J},$$

where J is the support described above. The theorem requires that we work in a convex set, but this is without loss becuause of the uniform bound on the sender's best response established above.

□

B.3 Further extensions

Distortion Suppose that negative distortion is free for features in a subset J of $\{1,\dots k\}$. For features in $\{1,\dots,k\}\setminus J$, negative distortion is still costly. We analyze the model as follows. First, analyze the model as before. Each equilibrium with coefficient vector b remains an equilibrium in the alternative model if and only if

$b_J \geq 0$. But this does not necessarily exhaust the set of equilibria. For each subset J' of J, consider the truncated model with dimensions in J' removed. Solve for the resulting equilibria and check whether $b_{J \setminus J'}$ is nonnegative. If so, appending $b_{J'} = 0$ gives an equilibrium.[32]

Noisy features I can add noisy features as long as the elliptical distribution is preserved.

Nonseparable costs To generalize to nonseparable cost functions, suppose that the sender's distortion ability is represented not by a random k-vector δ but rather by a random $k \times k$ matrix Δ. Now suppose that the vector

$$(\theta, \eta, \text{vech}(\Delta)) \in \mathbf{R}^{1+k+k(k+1)/2}$$

is jointly elliptically distributed. Moreover, assume that Δ is almost surely symmetric and positive semidefinite. The sender's utility becomes $e^T \Delta^\dagger e$. This reduces to the baseline model with $\Delta = \text{diag}(\delta)$. For non-diagonal matrices, it becomes more difficult to simultaneously satisfy the elliptical assumptions and the nonnegative definiteness assumptions, but one particular case is if δ is diagonally dominant: $\delta_{ii} > \sum_{j \neq i} |\delta_{ij}|$ for each i. If the diagonal entries of δ are bounded away from 0, then we can always choose a sufficiently small off-diagonal elements to satisfy this condition. Preferences are given by

$$u_S = y - (1/2) d^T \Delta^\dagger d, \qquad u_R = -(y - \theta - v^T d)^2.$$

If the receiver uses a strategy with coefficient vector b, the sender chooses $d \in \mathbf{R}^k$ to maximize

$$b^T(\eta + d) - (1/2) d^T \Delta^\dagger d.$$

This is a concave maximization. The first-order condition is

$$0 = b - \Delta^\dagger d,$$

so $d = \Delta b$. The solution to the first-order condition is not unique, but this is the unique solution with finite distortion cost. The equilibrium condition becomes

$$\text{var}(\eta + \delta b) b = \text{cov}(\eta + \delta b, \theta + v^T \Delta b).$$

Multi-dimensional decisions We can also extend to the model to allow for multidimensional decisions. The sender's latent type θ is an m-vector, and the receiver takes a decision $y \in \mathbf{R}^m$. Decisions are normalized so that each contributes equally to

[32] Indeed, the same model immediately permits lower dimensional distortion, without any change. Suppose distortion has dimension ℓ, with $\ell < k$, and that the feature vector takes the above form with $\tilde{B}$ a $k \times \ell$ matrix with linearly independent columns. Add additional columns to extend $\tilde{B}$ to a $k \times k$ matrix, and take $\delta_i = 0$ for $i = \ell + 1, \ldots, k$. Then apply the inverse transformation as before.

the sender's utility. The productivity of distortion is now an $k \times m$ matrix V, and the receiver's utility is weighted by a symmetric positive semidefinite $m \times m$ weighting matrix. Preferences are given by

$$u_S = \sum_{j=1}^{m} y_j - (1/2) \sum_{i=1}^{k} d_i^2/\delta_i, \qquad u_R = -(y - \theta - V^T d)^T W (y - \theta - V^T d),$$

Now a linear decision function is parametrized by a $k \times m$ coefficient matrix B. Given such a strategy, the sender chooses $d \in \mathbf{R}^k$ to maximize

$$\mathbf{1}^T B^T (\eta + d) - (1/2) \sum_{j=1}^{k} d_j^2/\delta_j,$$

where $\mathbf{1}$ is the m-vector of ones. The first-order condition gives

$$d = B\mathbf{1} \circ \delta,$$

so the equilibrium condition on the $k \times m$ matrix b becomes the $k \times m$ matrix equation

$$\operatorname{var}(\eta + B\mathbf{1} \circ \delta) B = \operatorname{cov}\big(\eta + B\mathbf{1} \circ \delta, \theta + v^T (B\mathbf{1} \circ \delta)\big).$$

Therefore, there are km equilibrium constraints, and there will be m constraints in the intermediary's problem. Relative to the baseline model, the number of constraints is multiplied by m, the number of dimensions of the decision.

B.4 Elliptical distributions

Multivariate elliptical distributions generalize the multivariate Gaussian distribution. This family retains the elliptical symmetry of the multivariate Gaussian but allows for more flexible tail behavior. For a detailed discussion of symmetric multivariate distributions and the properties of elliptical distributions, see the paper Cambanis et al. (1981) or the monograph Fang et al. (1989). Deimen and Szalay (2019) work with a specific subfamily of univariate elliptical distributions with the further property that the tail conditional distributions are linear. Recently, Frankel and Kartik (2019a) uses a linear–quadratic–elliptical specification in their model of signaling. Elliptical distributions were first introduced into economics, however, by Owen and Rabinovitch (1983) and Chamberlain (1983).[33]

[33]Elliptical distributions have a long history: Frankel and Kartik (2019a) cites Gesche (2017), who cites an earlier version of Deimen and Szalay (2019), who in turn cites Mailath and Nöldeke (2008). They cite the finance paper of Foster and Viswanathan (1993), which finally cites Owen and Rabinovitch (1983) and Chamberlain (1983). Independently, Li et al. (1987) first observe the advantage of weakening joint normality assumptions to the linear conditional expectation property. In their Cournot model, each player makes inferences based on an exogenous signal, so they do not

Lemma 3 (Elliptical characterization)
Fix an n-vector a positive semidefinite $n \times n$ matrix Σ. For an integrable random n-vector X, the following are equivalent.

1. *$X =_d \mu + \Sigma^{1/2} Z$ for some integrable spherical random variable Z.*
2. *For each $n \times m$ matrix A and $n \times p$ matrix B, we have*

$$\mathbf{E}[A^T X | B^T X] = A^T \mu + A^T \Sigma B (B^T \Sigma B)^{-1} (X - \mu).$$

3. *For all n-vectors a and b, we have*

$$\mathbf{E}[a^T X | b^T X] = a^T \mu + a^T \Sigma b (b^T \Sigma b)^{-1} (X - \mu).$$

4. *For all n-vectors a and b satisfying $a^T \Sigma b = 0$, we have*

$$\mathbf{E}[a^T X | b^T X] = a^T \mu.$$

Proof. Recall that (1) is the standard definition of an integrable elliptical distribution, and is known to imply (2) (Cambanis et al., 1981). It is immediate that (2) implies (3), which in turn implies (4). So the key step is showing that (4) implies (1). From the characterization, it suffices to show that $\Sigma^{-1/2}(X - \mu)$ is spherical. Suppose $\hat{a}$ and $\hat{b}$ are orthogonal. Using the spherical characterization of Eaton (1986), It suffices to show that the following expectation vanishes:

$$\begin{aligned} &\mathbf{E}\big[\hat{a}^T \Sigma^{-1/2}(X - \mu) \mid \hat{b}^T \Sigma^{-1/2}(X - \mu)\big] \\ &= \mathbf{E}\big[\hat{a}^T \Sigma^{-1/2} X \mid \hat{b}^T \Sigma^{-1/2} X\big] - \hat{a}^T \Sigma^{-1/2} \mu, \end{aligned}$$

but this follows immediately by putting $a = \Sigma^{-1/2}\hat{a}$ and $b = \Sigma^{-1/2}\hat{b}$. □

The observation about linearity of regression was made in Nimmo-Smith (1979) and strengthened in Hardin (1982). A slightly different characterization involving only orthogonal vectors was given by Eaton (1986).

B.5 Equilibrium nonuniqueness examples

I begin with two examples illustrating the at linear equilibria need not be unique.

Example 6 (Non-uniqueness, single feature)
There is a single feature so $k = 1$. The equilibrium condition becomes

$$b = \frac{\sigma_{\theta\eta} + b\sigma_{\theta\delta}}{\sigma_\eta^2 + b^2 \sigma_\delta^2}. \tag{24}$$

need the stronger elliptical assumption, which requires linear conditional expectations conditional on linear combinations of signals.

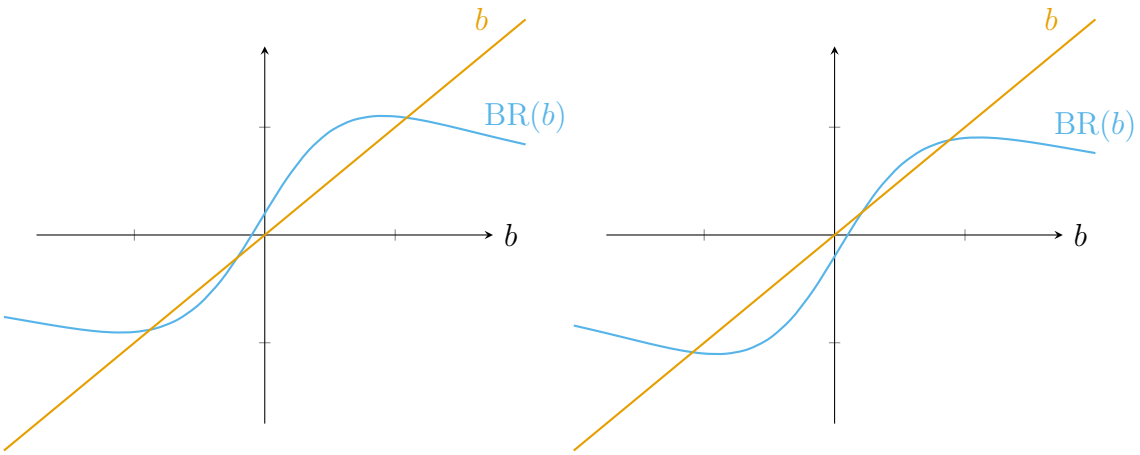


Figure 7. Multiple equilibria with a single feature

At $b = 0$, the right side equals $\beta = \sigma_{\theta\eta}/\sigma_\eta^2$ and has slope $\sigma_{\theta\delta}/\sigma_\eta^2$. The right side converges to 0 as the magnitude of b becomes large.

For this example, take

$$\sigma_\eta^2 = \sigma_\delta^2 = 1, \qquad \sigma_{\theta\eta} = 0.2, \qquad \sigma_{\theta\delta} = 2.$$

Thus, the distortion ability δ is much more informative than the intrinsic level η about he quality θ. The two sides of (24), with these parameter values, are plotted on the left side of Figure 7. There are three equilibria. Start at $b = 0$. Moving the right, the slope of the receiver's best response increases but is eventually attenuated by the variance of the feature. We get an equilibrium at $b = 1.09$. Moving to the left, the best response decreases past 0 and yields the equilibrium at $b = -0.21$. Continuing to the left, the larger gaming ability becomes more dominant, yielding the equilibrium at $b = -0.88$. This equilibrium is smaller in magnitude than the increasing equilibrium because the contributions of the gaming ability and the intrinsic level go in opposite directions.

The right panel of Figure 7 considers the same parameter values except the sign of $\sigma_{\theta\eta}$ is flipped to -0.2. The signs of the equilibria flip as well, so the equilibria are at 0.88, 0.21, and -1.09.

Finally, consider the welfare for the players across equilibria. The expected decision is the same across equilibria, so the sender's utility is determined by the distortion cost, which is increasing in the magnitude of b. The receiver's utility is determined by the magnitude of the correlation coefficient

$$\operatorname{corr}(\theta, \eta + b\delta) = \frac{\sigma_{\theta\eta} + b\sigma_{\theta\delta}}{\sigma_\theta \sqrt{\sigma_\eta^2 + b^2\sigma_\delta^2}}.$$

At each equilibrium, this expression reduces to $b(\sigma_\eta^2 + b^2\sigma_\delta^2)/\sigma_\theta$. Thus, across these three equilibria, the receiver's utility is strictly increasing in the magnitude of b.

With multiple features, there is a further form of nonuniqueness that can arise when the features are correlated.

Example 7 (Non-uniqueness, multiple features)
Let $k = 2$. The nonzero parameters are as follows. The variances are given by

$$\text{var}(\eta) = \text{var}(\delta) = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

The covariances are as follows:

$$\text{cov}(\eta_1, \theta) = \text{cov}(\delta_1, \theta) = 0.4, \qquad \text{cov}(\eta_1, \theta) = 0.2, \qquad \text{cov}(\delta_1, \theta) = 1.2.$$

On the first dimension, the intrinsic level and distortion ability are equally informative about θ. On the second dimension, the gaming ability is much more informative. If the first feature were the only feature, then there would be a unique equilibrium with with coefficient 0.48. If the second feature were the only feature, then there would be a unique equilibrium with coefficient 0.70. If the dimensions were uncorrelated, then the unique equilibrium would be $b = (0.48, 0.70)$. But the covariance changes the equilibrium structure. Now the equilibria are

$$b = (0.60, -0.48), \qquad b = (0.09, 0.66), \qquad b = (0.42, 0.13).$$

Compared to the single-dimensional equilibria, in the equilibria in which both coefficients are positive, the coefficients are attenuated by the positive covariance between the dimensions.

B.6 Nonlinear rules

In the main text, I restrict attention to linear scoring rules. This keeps the analysis tractable, and also allows me to isolate the effect of intermediary when I compare the scoring solution to the equilibrium in linear strategies.

These linear rules are also robust. They remain equilibria for any distribution within the elliptical family, wheres nonlinear equilibria depend on the particular details of the distribution, even within the elliptical family. Elliptical distributions give the linear conditional expectations property, so that if the sender uses a linear strategy, the receiver's best response is linear. These distributions do not, however, ensure that optimality of linear strategies. Here I provide a very simple example.

Lemma 4 (Two-point elliptical distributions)
If a random vector takes two values, each with probability $1/2$*, then it is elliptical.*

Proof. Let X be a random vector that takes two values with equal probability. Denote the mean of X by μ. Then $X - \mu$ takes values $\pm c$ with equal probability, for some nonzero vector c. The characteristic function of $X - \mu$ is given by

$$\varphi_{X-\mu}(t) = (1/2)e^{it^T c} + (1/2)e^{-it^T c} = \cos(t^T c) = g(t^T \Sigma^\dagger t),$$

where $g(x) = \cos(\sqrt{x})$ and $\Sigma = cc^T/(c^T c)^2$. □

Suppose that (η, δ) takes the values $(0, 0.4)$ and $(1, 0.1)$, each with probability $1/2$, and $\theta = \eta$. First I show that there is exactly one equilibrium in linear strategies and that this this equilibrium is completely uninformative. Then I show that there is a fully revealing nonlinear equilibrium.

With two mass points, a pure-strategy equilibrium is either uninformative or completely revealing. I look for an equilibrium with coefficient b. To be fully revealing, a necessary condition is that

$$b(1 + 0.1b) - b(0 + 0.4b) = 1,$$

which has no solution. On the other hand, there is a uninformative equilibrium given by the condition

$$b(1 - 0.3b) = 1,$$

which is solved by $b = 10/3$.

Now I construct a fully revealing nonlinear equilibrium. The receiver chooses $y = 1$ if $x \geq 1$ and $y = 0$ if $x < 1$. Clearly, type $(\eta, \delta) = (1, 0.1)$ does not distort. Type $(0, 0.4)$ does not distort either because the cost of increasing the reciever's decision to 1 is at least $(1/2)/0.4 = 5/4$, and this cost is strictly greater than 1.

B.7 Cubic systems

The linear equilibria are characterized by a cubic polynomial system of k equations in k unknowns $b_1, \ldots, b_k$. For a system of k polynomial equations in k unknowns with generic coefficients, there are finitely many complex solutions. Whenever there are finitely many complex solutions, the Bezout's theorem states that the number of complex solutions is at most the product of the degrees of the equations (Cox et al., 2005, Theorem 5.5, p. 115) Of course, every real solution is a complex solution, so if there are finitely many linear equilibria, then there are at most 3^k equilibria. This upper bound is achieved if each equation is a univariate cubic polynomial in a single variable with three real roots.

B.8 Convexity and norms

If a function $q\colon \mathbf{R}^n \to \mathbf{R}$ is quasiconvex and absolutely homogeneous, then it is a semi-norm.[34] The proof is essentially the same as the proof that the Minkowski functional is a semi-norm. Fix a and b in $\mathbf{R}^n$, and let $\alpha = \|a\|$ and $\beta = \|b\|$. We have

$$q\left(\frac{a+b}{\alpha+\beta}\right) = q\left(\frac{\alpha}{\alpha+\beta}\cdot\frac{a}{\alpha} + \frac{\beta}{\alpha+\beta}\cdot\frac{b}{\beta}\right) \leq 1,$$

where the last line uses the quasiconvexity of q for the level set $[q \leq 1]$. By absolute homogeneity, it follows that

$$q(a+b) \leq \alpha + \beta,$$

which is the desired subaditivity.

B.9 Measurability

The score set S is endowed with a σ-algebra $\mathcal{S}$. The sender's scoring rule f is a Markov transition from $\mathbf{R}^k$ to S, where $\mathbf{R}^k$ is endowed with the usual Borel σ-algebra. That is, f is a function from $\mathbf{R}^k \times \mathcal{S} \to [0,1]$ satisfying the following.

(i) For each $x \in \mathbf{R}^k$, the map $A \mapsto f(x,A)$ is a probability measure on $\mathcal{S}$.

(ii) For each $A \in \mathcal{S}$, the map $x \mapsto f(x,A)$ is measurable.

Behavioral strategies and decision rules are Markov transitions as well.

Markov transitions can be composed in the natural way. Let $(X,\mathcal{X})$, $(Y,\mathcal{Y})$, and $(Z,\mathcal{Z})$ be measurable spaces. Given Markov transitions g from $(X,\mathcal{X})$ to $(Y,\mathcal{Y})$ and h from $(Y,\mathcal{Y})$ to $(Z,\mathcal{Z})$, the composition gh is the Markov transition from $(X,\mathcal{X})$ to $(Z,\mathcal{Z})$ defined by

$$(gh)(x,C) = \int_Y g(x,\mathrm{d}y)h(y,C),$$

for all $x \in X$ and $C \in \mathcal{Z}$. In this integral, the function $y \mapsto h(y,C)$ is integrated over Y against the measure $g(x,\cdot)$. Likewise, the composition of a measure m in $(X,\mathcal{X})$ and a Markov transition g from $(X,\mathcal{X})$ to $(Y,\mathcal{Y})$, the composition mg is the measure on $(Y,\mathcal{Y})$ defined by

$$(mg)(B) = \int_Y m(\mathrm{d}y)g(y,B),$$

for all $B \in \mathcal{Y}$. In this integral, the function $y \mapsto g(y,B)$ is integrated against the measure m.

In the main text, I abuse notation by writign $f(x)$ to denote a random variable with measures $f(x,\cdot)$. Similarly, when I write $\sigma_S(\eta,\delta)$ to denote the sender's distortion, viewed as a random variable. Formally, the distribution is the composition of

[34] A function $q\colon \mathbf{R}^k \to \mathbf{R}$ is absolutely homogenenous if $q(tx) = |t|q(x)$ for all $x \in \mathbf{R}^k$ and $t \in \mathbf{R}$. A function $q\colon \mathbf{R}^k \to \mathbf{R}$ is a seminorm if it is nonnegative, absolutely homogeneous, and subadditive (i.e., satisfies the triangle inequality).

the law of (η, δ) and the Markov transition σ_S. The other compositions are defined similarly.

References

Aumann, R. J. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96. [6]

Barron, E. N., R. Goebel, and R. R. Jensen (2010): "Best Response Dynamics for Continuous Games," *Proceedings of the American Mathematical Society*, 138, 1069–1083. [50]

Bénabou, R. and J. Tirole (2006): "Incentives and Prosocial Behavior," *American Economic Review*, 96, 1652–1678. [5]

Bergemann, D. and S. Morris (2013): "Robust Predictions in Games with Incomplete Information," *Econometrica*, 81, 1251–1308. [6]

——— (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522. [6]

——— (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57, 44–95. [6]

Boleslavsky, R. and K. Kim (2018): "Bayesian Persuasion and Moral Hazard," Working paper. [6]

Bonatti, A. and G. Cisternas (forthcoming): "Consumer Scores and Price Discrimination," *Review of Economic Studies*. [5]

Brückner, M. and T. Scheffer (2009): "Nash Equilibria of Static Prediction Games," in *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, 171–179. [5]

Cambanis, S., S. Huang, and G. Simons (1981): "On the Theory of Elliptically Contoured Distributions," *Journal of Multivariate Analysis*, 11, 368–385. [10, 35, 52, 53]

Chamberlain, G. (1983): "A Characterization of the Distributions that Imply Mean–Variance Utility Functions," *Journal of Economic Theory*, 29, 185–201. [52]

Citron, D. K. and F. Pasquale (2014): "The Scored Society: Due Process for Automated Predictions," *Washington Law Review*, 89. [2]

Cox, D. A., J. Little, and D. O'Shea (2005): *Using Algebraic Geometry*, vol. 185 of *Graduate Texts in Mathematics*, Springer, 2 ed. [56]

Crémer, J. (1995): "Arm's Length Relationships," *Quarterly Journal of Economics*, 110, 275–295. [5]

Cunningham, T. and I. Moreno de Barreda (2019): "Effective Signal-Jamming," Working paper. [5]

Dalvi, N., P. Domingos, Mausam, S. Sanghai, and D. Verma (2004): "Adversarial Classification," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 99–108. [5]

Deimen, I. and D. Szalay (2019): "Delegated Expertise, Authority, and Communication," *American Economic Review*, 109, 1349–1374. [52]

Doval, L. and J. C. Ely (2016): "Sequential Information Design," Working paper. [21]

Duffie, D. and P. Dworczak (2018): "Robust Benchmark Design," Stanford GSB Working Paper 3175. [6]

Eaton, M. L. (1986): "A Characterization of Spherical Distributions," *Journal of Multivariate Analysis*, 20, 272–276. [53]

Ederer, F., R. Holden, and M. Meyer (2018): "Gaming and Strategic Opacity in Incentive Provision," *RAND Journal of Economics*, 49, 819–854. [2]

Engers, M. (1987): "Signalling with Many Signals," *Econometrica*, 55, 663–674. [5]

Fang, K.-T., S. Kotz, and K. W. Ng (1989): *Symmetric Multivariate and Related Distributions*, no. 36 in Monographs on Statistics & Applied Probability, Chapman and Hall. [52]

Fischer, P. E. and R. E. Verrecchia (2000): "Reporting Bias," *Accounting Review*, 75, 229–245. [5]

Forges, F. (1986): "An Approach to Communication Equilibria," *Econometrica*, 54, 1375–1385. [6]

Foster, F. D. and S. Viswanathan (1993): "The Effect of Public Information and Competition on Trading Volume and Price Volatility," *Review of Financial Studies*, 6, 23–56. [52]

Frankel, A. and N. Kartik (2019a): "Muddled Information," *Journal of Political Economy*, 127, 1739–1776. [5, 12, 18, 20, 52]

——— (2019b): "Improving Information from Manipulable Data," Working paper. [5]

Gesche, T. (2017): "De-Biasing Strategic Communication," University of Zurich Working Paper No. 216. [52]

Hardin, Jr, C. D. (1982): "On the Linearity of Regression," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 61, 293–302. [53]

Hardt, M., N. Megiddo, C. Papadimitriou, and M. Wooters (2016): "Strategic Classification," in *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, 111–122. [5]

Hofbauer, J. and S. Sorin (2006): "Best Response Dynamics for Continuous Zero-Sum Games," *Discrete and Continuous Dynamical Systems—Series B*, 6, 215–224. [50]

Holmström, B. (1999): "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66, 169–182. [7]

Hörner, J. and N. Lambert (forthcoming): "Motivational Ratings," *Review of Economic Studies*. [6]

Hu, L., N. Immorlica, and J. W. Vaughan (2019): "The Disparate Effects of Strategic Manipulation," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 259–268. [5]

Hwang, S.-H. and L. Rey-Bellet (2018a): "Simple Characterizations of Potential Games and Zero-sum Games," ArXiv:1602.04410v1. [18]

——— (2018b): "Strategic Decompositions of Normal Form Games: Zero-sum Games and Potential Games," ArXiv:1602.06648v2. [18]

Kamenica, E. (2019): "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11, 249–272. [6]

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615. [6]

Lambert, N. S., G. Martini, and M. Ostrovsky (2018): "Quadratic Games," Working paper. [14]

Li, L., R. D. McKelvey, and T. Page (1987): "Optimal Research for Cournot Oligopolists," *Journal of Economic Theory*, 42, 140–166. [52]

Mailath, G. J. and G. Nöldeke (2008): "Does Competitive Pricing Cause Market Breakdown under Extreme Adverse Selection?" *Journal of Economic Theory*, 140, 97–125. [52]

Makris, M. and L. Renou (2018): "Information Design in Multi-stage games," Working Paper 861, Queen Mary University of London, School of Economics and Finance. [21]

Mekerishvili, G. (2018): "Optimal Disclosure on Crowdfunding Platforms," Working paper. [6]

Meyer, C. D. (2000): *Matrix Analysis and Applied Linear Algebra*, SIAM. [10]

Monderer, D. and L. S. Shapley (1996): "Potential Games," *Games and Economics Behavior*, 14, 124–143. [17]

Morris, S. and H. S. Shin (2002): "Social Value of Public Information," *American Economic Review*, 92, 1521–1534. [14]

Myerson, R. B. (1982): "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematica Economics*, 10, 67–81. [6]

——— (1986): "Multistage Games with Communication," *Econometrica*, 54, 323–358. [6, 21]

Nimmo-Smith, I. (1979): "Linear Regressions and Sphericity," *Biometrika*, 66, 390–392. [53]

Owen, J. and R. Rabinovitch (1983): "On the Class of Elliptical Distributions and their Applications to the Theory of Portfolio Choice," *Journal of Finance*, 38, 745–752. [52]

Perez-Richet, E. and V. Skreta (2018): "Test Design under Falsification," Working paper. [6]

Prendergast, C. and R. Topel (1996): "Favoritism in Organizations," *Journal of Political Economy*, 104, 958–78. [5]

Quinzii, M. and J.-C. Rochet (1985): "Multidimensional Signaling," *Journal of Mathematical Economics*, 14, 261–284. [5]

Rick, A. (2013): "The Benefits of Miscommunication in Communication Games," Working paper. [6]

Rodina, D. (2016): "Information Design and Career Concerns," Working paper. [6]

Rodina, D. and J. Farragut (2016): "Inducing Effort through Grades," Working paper. [6]

Rosenthal, R. W. (1973): "A Class of Games Possessing Pure-strategy Nash Equilibra," *International Journal of Game Theory*, 2, 65–67. [17]

Spence, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–374. [5]

Voorneveld, M. (2000): "Best-response Potential Games," *Economics Letters*, 66, 289–295. [18]

Whitmeyer, M. (2019): "Bayesian Elicitation," Working paper. [6]

Zapechelnyuk, A. (2019): "Optimal Quality Certification," Working paper. [6]