Optimal Information Disclosure: A Linear Programming Approach*

Anton Kolotilin†

First Version: November, 2011
This Version: July, 2017

Abstract

An uninformed sender designs a mechanism that discloses information about her type to a privately informed receiver, who then decides whether to act. I impose a single-crossing assumption, so that the receiver with a higher type is more willing to act. Using a linear programming approach, I characterize optimal information disclosure and provide conditions under which full and no revelation are optimal. Assuming further that the sender's utility depends only on the sender's expected type, I provide conditions under which interval revelation is optimal. Finally, I show that the expected utilities are not monotonic in the precision of the receiver's private information.

Keywords: Bayesian persuasion, information design, information disclosure, informed receiver

JEL classification: C72, D82, D83

---

*This paper is based on the second chapter of my 2012 Ph.D. dissertation at MIT and was previously circulated under the title "Optimal Information Disclosure: Quantity vs. Quality." I thank Robert Gibbons and Muhamet Yildiz for their invaluable guidance and advice. I thank Hongyi Li for many detailed comments. I also thank Ricardo Alonso, Sandeep Baliga, Abhijit Banerjee, Gabriel Carroll, Denis Chetverikov, Glenn Ellison, Richard Holden, Jin Li, Uliana Loginova, Niko Matouschek, Parag Pathak, Michael Powell, Andrea Prat, Juuso Toikka, Alexander Wolitzky, Juan Xandri, Luis Zermeño, various seminar and conference participants, anonymous referees, and especially the editor, George Mailath, for helpful comments and suggestions. Financial support from the Australian Research Council is acknowledged.

†UNSW Australia, School of Economics. Email: a.kolotilin@unsw.edu.au.

1 Introduction

In the Bayesian persuasion literature (Rayo and Segal 2010 and Kamenica and Gentzkow 2011), an uninformed sender (she) designs an information disclosure mechanism to influence the beliefs of a receiver (he) about the sender's type. I use a linear programming approach to study this problem.

In my model, the receiver privately knows his one-dimensional type and chooses between two actions: to act or not to act. Before observing her type, the sender can commit to any (stochastic) mapping from her types to messages, which I call an information disclosure mechanism. After observing the message generated by the mechanism and his type, the receiver decides whether to act. The sender and receiver's types are drawn from a continuous joint prior distribution. The sender and receiver have continuous utility functions that depend on the sender and receiver's types. I impose a single-crossing assumption which ensures that each message of a mechanism induces the receiver to act if and only if his type exceeds a threshold type.

It turns out that my model is equivalent to an alternative model where the receiver is uninformed, chooses a one-dimensional action, and has utility that is single-peaked in his action for each message of a mechanism. In my model, each message of a mechanism corresponds to a threshold type above which the receiver acts. Likewise, in the alternative model, each message of a mechanism corresponds to an optimal action of the uninformed receiver. That is, the receiver's threshold type in my model is isomorphic to the receiver's optimal action in the alternative model.

I characterize conditions for a candidate mechanism to be optimal, and derive comparative statics on the precision of the receiver's private information. The characterization results apply directly both to my model and to the alternative model. But the comparative statics results do not apply to the alternative model, in which the receiver is uninformed. Hereafter, I discuss my results in the context of my model, in which the receiver is privately informed and chooses between two actions.

For concreteness, consider a school that wishes to persuade a potential employer to hire a student by choosing a grade disclosure policy for the student. The school can freely choose what information about the student's grades appears on the student's transcript. Moreover, the school chooses this disclosure policy before observing anything about the student. The employer observes the student's transcript but also obtains private information, for example, from conducting an employment interview with the student and competing candidates. The single-crossing assumption requires that all possible interview outcomes can be appropriately ranked.

The sender's problem of finding an optimal mechanism reduces to a linear program, because a mechanism is described by the conditional probabilities of messages given the sender's types, and the expected utilities are linear in these probabilities. The linear programming approach gives necessary and sufficient conditions under which a candidate mechanism is optimal. This enables the characterization of conditions that justify many commonly observed grade disclosure policies, such as those reported in Ostrovsky and Schwarz (2010). These conditions imply that to verify that a grade disclosure policy is optimal, it suffices to check that there is no *simple* deviation from this policy that the school prefers. At the one extreme, some schools report all grades and class rank on transcripts. Such a full revelation mechanism is optimal if and only if the sender prefers to reveal any two of her types than to pool them. At the other extreme, some schools release no transcripts. Such a no revelation mechanism is optimal if and only if the sender prefers to pool any three of her types with the uninformative message than to pool two of them and reveal the third one.

Assume further that the sender's utility (under the receiver's optimal action) depends on the message only through the posterior expectation of the sender's type given this message.[1] Under this assumption, the sender can choose any distribution of posterior expectations of the sender's type, subject to the constraint that the prior distribution of the sender's type is a mean-preserving spread of this distribution. As a result, the shape of the optimal mechanism is jointly determined by the convexity properties of the sender's utility function and by the prior distribution of her type. I provide necessary and sufficient conditions under which the sender optimally chooses an interval revelation mechanism that reveals moderate types and hides extreme types.

In general, the sender and receiver's expected utilities under the optimal mechanism are not monotonic in the precision of the receiver's private information. First, as the receiver becomes more informed, his expected utility may decrease despite the fact that he is the only player who takes an action that directly affects his utility. This happens because the optimal mechanism depends on the precision of the receiver's private information, and the sender may prefer to disclose significantly less information if the receiver's information is

[1]The statement of this assumption is silent about the sets of receiver's types and actions. Consequently, the assumption and the corresponding results apply directly both to my model and to the equivalent alternative model with an uninformed receiver. Kamenica and Gentzkow (2011) refer to this assumption as "Sender's payoff depends only on the expected state". Ostrovsky and Schwarz (2010) also impose this assumption and characterize the unique equilibrium (rather than optimal) information disclosure mechanism.

more precise. Returning to the school–employer example, this suggests that low reliability of employment interview procedures (summarized by Arvey and Campion 1982) may be beneficial for employers, as it motivates schools to design more informative disclosure policies. Second, it may be easier for the sender to influence a more informed receiver. This happens because the sender may optimally choose to target only the receiver with favorable private information, and it becomes easier for the sender to persuade such a receiver, as he becomes more informed.

The linear programming approach to Bayesian persuasion complements the standard concavification approach of Kamenica and Gentzkow (2011). They work with the distribution of posterior beliefs induced by a mechanism. They define the sender's indirect utility of posterior beliefs, and derive the optimal mechanism by taking the concave closure of this indirect utility function. In contrast, the linear programming approach solves the dual problem: it derives conditions under which a given mechanism is optimal.

As Gentzkow and Kamenica (2016) point out, the concavification approach has limited applicability when the set of sender's types is an interval, because the set of posterior beliefs becomes infinite-dimensional. The linear programming approach instead works with utilities directly expressed as functions of the one-dimensional sender and receiver's types, and thus yields sharper results for the class of problems I consider.

My model is a special case of Kamenica and Gentzkow (2011), who do not restrict either the sets of receiver's types and actions or the functional form of the receiver's utility. Rayo and Segal (2010), on the other hand, is a special case of my model. They assume that the receiver's type is uniformly distributed and does not affect the sender's utility.[2]

Subsequent to the first version of this paper, some papers on Bayesian persuasion have assumed that the sender's utility depends only on the sender's expected type. Gentzkow and Kamenica (2016) provide an alternative characterization of the set of feasible mechanisms, and use it to find optimal mechanisms in stylized examples. Kolotilin et al. (2017) allow the sender to condition mechanisms on the receiver's reports, and provide simple sufficient conditions for optimality of upper-censorship – a special case of the interval revelation mechanisms characterized in this paper.[3]

[2]In Section 4, I discuss in more details how my paper relates to Rayo and Segal (2010), Kamenica and Gentzkow (2011), and other papers on Bayesian persuasion.

[3]The Bayesian persuasion problem of this paper is mathematically similar to the delegation problem initiated by Holmstrom (1984). Alonso and Matouschek (2008) and Amador and Bagwell (2013) characterize necessary and sufficient conditions under which interval delegation is optimal. Their conditions resemble my conditions under which interval revelation is optimal, but their proofs are more involved.

2 School–Employer Example

A school chooses a grading policy to maximize the probability of an employer hiring a student. The student is either a *peach* or a *lemon*. The school and the employer have a common prior belief that the student is a peach with probability 0.2. The employer hires the student if the employer believes that the student is a peach with probability at least 0.5.

The timing of the game is as follows. First, the school chooses a grading policy Φ described by a finite (ordered) set M of grades and the conditional distribution of grades given the student's type. Second, the student's type is drawn and the grade is generated according to Φ. Third, the employer observes the grade and conducts an employment interview. The interview produces a two-valued signal about the student's type with precision $p \in [1/2, 1]$ in the sense that $\Pr(\text{positive}|\text{peach}) = \Pr(\text{negative}|\text{lemon}) = p$ and $\Pr(\text{negative}|\text{peach}) = \Pr(\text{positive}|\text{lemon}) = 1 - p$. The employer is *positive* (*negative*) if the interview signal is positive (negative). Finally, the employer makes a hiring decision.

I restrict attention to grading policies that generate three possible grades: A, B, or C that convince both, only positive, or no employer to hire. This is without loss of generality because convincing the negative employer also convinces the positive employer.

The school chooses Φ to maximize the probability of hire:

$$1 \cdot \Pr_{\Phi}(A) + \Pr_{\Phi}(\text{positive}|B) \cdot \Pr_{\Phi}(B) + 0 \cdot \Pr_{\Phi}(C),$$

subject to the constraint imposed by the prior distribution of the student's ability:

$$\sum_{m} \Pr_{\Phi}(\text{peach}|m) \cdot \Pr_{\Phi}(m) = \Pr(\text{peach}) = 0.2.$$

Under the optimal grading policy, grades A and B barely persuade the negative and positive employers to hire, whereas grade C makes the employer certain that the student is a lemon; so, after some algebra, $\Pr_\Phi(\text{peach}|A) = p$, $\Pr_\Phi(\text{peach}|B) = 1 - p$, and $\Pr_\Phi(\text{peach}|C) = 0$. Using these conditions, it is easy to show that $\Pr_\Phi(\text{positive}|B) = 2p(1 - p)$.

The school's problem is thus a *linear program*: to maximize the utility function

$$\Pr(A) + 2p(1-p)\Pr(B)$$

over probabilities $\Pr(A)$, $\Pr(B)$, and $\Pr(C)$, subject to the *Bayesian budget constraint*

$$p\Pr(A) + (1-p)\Pr(B) = 0.2.$$

The *marginal utilities* of grades A, B, and C are 1, $2p(1-p)$, and 0; the *prices* of these grades are p, $(1-p)$, and 0. Thus, the school faces a tradeoff: to choose a grading policy that generates A with a small probability and persuades both the negative and positive employers or a grading policy that generates B with a high probability but persuades only the positive employer. The school resolves this tradeoff by choosing a policy that frequently generates grades with the highest *marginal utility-price ratio* ($1/p$ for A and $2p$ for B).

As an aside, this argument that optimal grading policies should frequently generate messages with high marginal utility-price ratios requires the student's type to take only two values, but does not rely on the cardinality of the set of receiver's types, the utility functional forms, or the form of the joint distribution of the sender and receiver's types.

The optimal grading policy can take three forms depending on the interview precision. If the interview is imprecise ($1/2 \leq p < 1/\sqrt{2}$), the marginal utility-price ratio is higher for A than for B; so the optimal grading policy *targets* the negative employer and generates grades A and C. If the interview is precise ($p > 1/\sqrt{2}$), the ratio is higher for B, so the optimal policy targets the positive employer. In this case, if it is impossible to convince the positive employer with probability one ($p < 4/5$), the optimal policy generates grades B and C; otherwise ($p > 4/5$), the optimal policy generates grades A and B.

When the interview is not too precise ($p < 4/5$), the optimal grading policy exhibits *grade inflation*: peaches and some lemons get a good grade but only lemons get a bad grade. Moreover, when the interview is imprecise ($p < 1/\sqrt{2}$), grade inflation is moderate and a good grade impresses all employers. When the interview is relatively precise ($1/\sqrt{2} < p < 4/5$), grade inflation is severe and a good grade convinces only the positive employer. Finally, when the interview is too precise ($p > 4/5$), the optimal grading policy is *noisy*: with a positive probability, a peach gets a bad grade and a lemon gets a good grade.

Figure 1 shows that the school and employer's expected utilities under the optimal grading policy are not monotonic in the interview precision.[4] Naive intuition may suggest that (i) the school's expected utility should decrease with p because it is harder to influence a better informed employer and (ii) the employer's expected utility should increase with p because a better informed employer takes a more appropriate hiring decision. This naive intuition, however, ignores that the optimal mechanism changes with p, and the school may choose to disclose significantly less information when the employer is more informed. This effect may overturn the naive intuition. In equilibrium, a more informative interview may help the school because the employer hires *more* students; it may also hurt the employer because the employer hires *worse* students.

[4]Figure 1 normalizes the utility functions as follows. Both the school and the employer get utility 0 if the student is not hired. The school gets utility 1 if the student is hired. The employer gets utility 1 from hiring a peach and utility -1 from hiring a lemon.

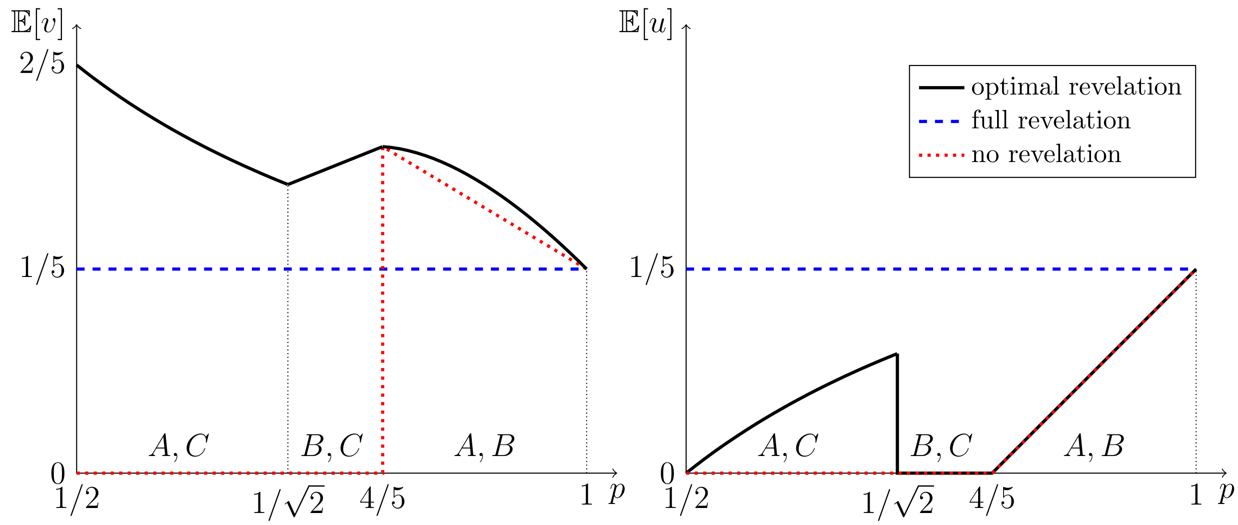

Figure 1: The sender and receiver's expected utilities in the school–employer example

In fact, the school's expected utility strictly increases with the interview precision for $p \in \left(1/\sqrt{2}, 4/5\right)$, where the optimal grading policy targets the positive employer. As the interview precision increases, the positive employer becomes more positive, so it becomes easier for the school to persuade the employer.

Moreover, the employer's expected utility drops down to 0 as the interview precision exceeds $1/\sqrt{2}$. At p around $1/\sqrt{2}$, neither the positive nor the negative employer would hire if the grading policy was completely uninformative. For p slightly above $1/\sqrt{2}$, the optimal grading policy targets the positive employer and thus extracts all rent from the positive employer. In contrast, for p slightly below $1/\sqrt{2}$, the optimal grading policy targets the negative employer, and thus leaves some rent to the positive employer.

3 Model

3.1 Setup

Consider a communication game between a sender and a receiver. The sender chooses an information disclosure mechanism (described below) and the receiver takes one of two actions: *to act* $(a = 1)$ or *not to act* $(a = 0)$.

The set of receiver's types is $R = [\underline{r}, \overline{r}]$, and the set of sender's types is $S = [\underline{s}, \overline{s}]$. The pair (r, s) has some joint prior distribution. For this joint distribution, the marginal distribution $F(s)$ of s and the conditional distribution $G(r|s)$ of r given s admit strictly positive densities $f(s)$ and $g(r|s)$ that are continuous in s and continuously differentiable in r.

The sender and receiver's utilities from $a = 0$ are normalized to zero. The sender and receiver's utilities from $a = 1$ are $v(r,s)$ and $u(r,s)$, respectively, where functions v and u are continuous in s and continuously differentiable in r.

Before s is realized, the sender chooses a *mechanism* that sends a message $m \in \mathbb{R}$ to the receiver as a (stochastic) function of the sender's type s. Specifically, the sender chooses a joint distribution $\Phi(m,s)$ of m and s such that the marginal distribution of s under Φ equals the prior marginal distribution F.

The timing of the communication game is as follows. First, the sender publicly chooses a mechanism Φ. Second, a triple (m,s,r) is drawn according to distributions Φ and G. Third, the receiver observes (m,r) and takes an action a. Finally, the sender and receiver's utilities are realized.

Let $P_\Phi(s|m)$ denote the distribution of s given m under Φ. The distribution of s given m and r is then

$$P_\Phi(s|m,r) = \frac{\int_{\underline{s}}^{s} g(r|\tilde{s}) dP_\Phi(\tilde{s}|m)}{\int_{\underline{s}}^{\overline{s}} g(r|\tilde{s}) dP_\Phi(\tilde{s}|m)}.$$

The receiver's expected utility from $a = 1$ given m and r is

$$\int_{\underline{s}}^{\overline{s}} u(r,s) dP_\Phi(s|m,r) = \frac{\int_{\underline{s}}^{\overline{s}} u(r,s) g(r|s) dP_\Phi(s|m)}{\int_{\underline{s}}^{\overline{s}} g(r|s) dP_\Phi(s|m)}.$$

Therefore, the receiver strictly prefers to act if $\int_S \widetilde{u}(r,s)\, dP_\Phi(s|m) > 0$ and strictly prefers not to act if $\int_S \widetilde{u}(r,s)\, dP_\Phi(s|m) < 0$, where

$$\widetilde{u}(r,s) \equiv u(r,s)\, g(r|s).$$

I impose a single-crossing assumption which ensures that each message of a mechanism induces the receiver to act if and only if his type exceeds a threshold type.

Assumption 1 (Single crossing) *For each distribution Q on S, there exists $r_Q \in R$ such that $\int_S \widetilde{u}(r,s)\, dQ(s) \gtreqless 0$ if $r \gtreqless r_Q$. Moreover, there exists a strictly decreasing function r^* that satisfies $u(r^*(s), s) = 0$ for all $s \in S$.*

A message m of a mechanism Φ induces a distribution Q of s. By Assumption 1, after observing m, the receiver of type r_Q is indifferent between the two actions, and the receiver of type $r > r_Q$ strictly prefers to act. Without loss of generality, I restrict attention to mechanisms Φ such that each message m of Φ induces the receiver to act if and only if

$r \geq m$.[5] Therefore, the set of feasible messages is the image $R^* \equiv r^*(S)$ of S under r^*, and the sender's expected utility from message $m \in R^*$ is

$$V(m,s) \equiv \int_m^{\overline{r}} v(r,s)\, g(r|s)\, dr.$$

Remark 1 *Assumption 1 is stronger than a standard single-crossing assumption, which requires that for each distribution Q on S, inequality $\int_S \widetilde{u}(r_1,s)\, dQ(s) \geq (>) 0$ implies $\int_S \widetilde{u}(r_2,s)\, dQ(s) \geq (>) 0$ whenever $r_2 > r_1$. This standard single-crossing assumption holds if and only if, for all $s_1, s_2 \in S$, functions $\widetilde{u}(r,s_1)$ and $\widetilde{u}(r,s_2)$ of $r \in R$ satisfy signed-ratio monotonicity (Theorem 1 of Quah and Strulovici 2012). In particular, it holds if $u(r,s)$ increases with (r,s) and types (r,s) are affiliated.*

Section 4.3 imposes a stronger linearity assumption which ensures that the sender's expected utility depends only on the sender's expected type, $\mathbb{E}_\Phi[s|m]$.

Assumption 2 (Linearity) *For all $(r,s) \in R \times S$, $u(r,s) = s - r$, $v(r,s) = v(r)$, $G(r|s) = G(r)$, and $S \subset R$.*[6]

Under Assumption 2, the receiver acts if and only if $r \leq \mathbb{E}_\Phi[s|m]$. Therefore a message m of a mechanism Φ satisfies $m = \mathbb{E}_\Phi[s|m]$, and the sender's expected utility from a message m is $V(m) \equiv \int_{\underline{r}}^m v(r)\, g(r)\, dr$ (which depends only on $\mathbb{E}_\Phi[s|m]$).

Remark 2 *Equivalent to Assumption 2, I could directly assume that the sender's expected utility V depends only on $\mathbb{E}_\Phi[s|m]$. Kamenica and Gentzkow (2011) refer to this assumption as "Sender's payoff depends only on the expected state". This assumption may hold even if the receiver has more than two actions. In particular, it holds if the set of actions is compact, the sender and receiver's types are independent, and the sender and receiver's utility functions are linear in the sender's type and continuous in the receiver's type and action.*

[5] Although type $r = m$ is indifferent between the two actions, I assume that he acts. This assumption is innocuous, because, for any $r \in R$, the receiver has type r with probability 0, since G admits a density.

[6] Notice that if r is replaced with $-r$ in Assumption 2, then higher types of the receiver are more willing to act, and Assumption 1 holds. But exposition is easier without this replacement. Notice also that Assumption 2 requires r to be independent of s. In the context of the school–employer example, s may correspond to the student's type privately known by the school, and r to the opportunity cost from hiring privately known by the employer.

3.2 Equivalent Alternative Model

Consider an alternative model where an uninformed receiver takes an action r from set $R = [\underline{r}, \overline{r}]$. If the receiver takes action r and the sender's type is s, then the sender and receiver's utilities are $V(r,s)$ and $U(r,s)$, where V and U are continuous in s and twice continuously differentiable in r. The set of sender's types remains $S = [\underline{s}, \overline{s}]$, and the prior distribution of s remains F.

Parallel to Assumption 1, I impose an assumption which ensures that for each message of a mechanism the receiver's utility is single-peaked in his action.

Assumption 1' (Single crossing) *For each distribution Q on S, there exists $r_Q \in R$ such that $\int_S [-\partial U(r,s)/\partial r]\, dQ(s) \gtreqless 0$ if $r \gtreqless r_Q$. Moreover, there exists a strictly decreasing function r^* that satisfies $\partial U(r^*(s), s)/\partial r = 0$ for all $s \in S$.*

It turns out that the sender's problem of choosing an optimal mechanism in this alternative model under Assumption 1' is the same as in the original model from Section 3.1 under Assumption 1. Given v, u, and g from the original model, set $V(r,s) = \int_r^{\overline{r}} v(\tilde{r}, s)\, g(\tilde{r}|s)\, d\tilde{r}$ and $U(r,s) = \int_r^{\overline{r}} u(\tilde{r}, s)\, g(\tilde{r}|s)\, d\tilde{r}$.[7] Notice that in both models, a message m under mechanism Φ induces some distribution Q of s. In the original model, Q induces the receiver to act if and only if $r \geq r_Q$, so the sender's utility is $\int_{r_Q}^{\overline{r}} v(\tilde{r}, s)\, g(\tilde{r}|s)\, d\tilde{r} = V(r_Q, s)$; in this alternative model, Q induces the receiver to take action r_Q, so the sender's utility is again $V(r_Q, s)$. The receiver's threshold type r_Q in the original model is thus isomorphic to the receiver's optimal action r_Q in this alternative model.

Similar to the equivalence between Assumptions 1 and 1', the following assumption is equivalent to Assumption 2.

Assumption 2' (Linearity) *For all $(r,s) \in R \times S$, $U(r,s) = -(r-s)^2$, $V(r,s) = V(r)$, and $S \subset R$.*

Under Assumption 2', the receiver takes action $r = \mathbb{E}_\Phi[s|m]$. Therefore a message m of a mechanism satisfies $m = \mathbb{E}_\Phi[s|m]$, and the sender's expected utility from a message m is $V(m)$, as in the original model.

To sum up, this alternative model with Assumptions 1' and 2' is equivalent to the original model with Assumptions 1 and 2. Consequently, all the results in Section 4 hold verbatim in this alternative model.

[7] Equivalently, given V and U from this alternative model, set v, u, and g such that $v(r,s)g(r|s) = -\partial V(r,s)/\partial r$ and $u(r,s)g(r|s) = -\partial U(r,s)/\partial r$.

4 Optimal Mechanisms

Section 4.1 sets up the sender's problem as a linear program and presents some basic duality results. Under Assumption 1, Section 4.2 characterizes necessary and sufficient conditions under which the full and no revelation mechanisms are optimal. Under Assumption 2, Section 4.3 characterizes necessary and sufficient conditions under which an interval revelation mechanism is optimal.

A mechanism is called an *interval revelation mechanism with bounds* s_L *and* s_H if $s_L, s_H \in S$, $s_L \leq s_H$, and it generates one message for all $s \in [\underline{s}, s_L)$, another message for all $s \in (s_H, \overline{s}]$, and a different message for each $s \in (s_L, s_H)$.[8] In particular, the *full revelation mechanism* (denoted by Φ_{full}) is an interval revelation mechanism with bounds $s_L = \underline{s}$ and $s_H = \overline{s}$; and the *no revelation mechanism* (denoted by Φ_{no}) is an interval revelation mechanism with bounds $s_L = s_H = \underline{s}$ (or equivalently $s_L = s_H = \overline{s}$).

4.1 Linear Programming Characterization

Under Assumption 1, an optimal mechanism is a distribution Φ that solves the following *primal* linear program,

$$\text{maximize} \int_{R^* \times S} V(r, s)\, d\Phi(r, s) \tag{P}$$

$$\text{subject to} \int_{R^* \times \widetilde{S}} d\Phi(r, s) = \int_{\widetilde{S}} f(s)\, ds \text{ for any measurable set } \widetilde{S} \subset S, \tag{P1}$$

$$\int_{\widetilde{R} \times S} \widetilde{u}(r, s)\, d\Phi(r, s) = 0 \text{ for any measurable set } \widetilde{R} \subset R^*. \tag{P2}$$

The objective function is the sender's expected utility under Φ. The first constraint (P1) is the feasibility requirement that the marginal distribution of s under Φ is F. The second constraint (P2) is the consistency requirement that message $m = r$ makes the receiver r indifferent between the two actions.

The *dual* problem is to find bounded measurable functions η and ν that

$$\text{minimize} \int_S \eta(s) f(s)\, ds \tag{D}$$

$$\text{subject to } \eta(s) + \widetilde{u}(r, s)\, \nu(r) \geq V(r, s) \text{ for all } (r, s) \in R^* \times S. \tag{D1}$$

The variables $\eta(s)$ and $\nu(r)$ are multipliers for constraints (P1) and (P2).

[8] Since the distribution F of s admits a density, it is not necessary to specify what message is sent for a finite number of types, such as s_L and s_H.

Say that Φ is *feasible* for (P) if it is a distribution that satisfies (P1) and (P2). Similarly, say that η and ν are *feasible* for (D) if they are bounded measurable functions that satisfy (D1). Feasible Φ and (η,ν) that solve their respective problems (P) and (D) are called *optimal solutions*.

Lemma 1 gives sufficient conditions under which candidate feasible solutions Φ and (η,ν) are optimal:

Lemma 1 *Suppose Assumption 1 holds. If Φ is feasible for (P), (η,ν) is feasible for (D), and*

$$\int_{R^*\times S} \left(\eta(s)+\widetilde{u}(r,s)\nu(r)-V(r,s)\right)d\Phi(r,s)=0, \tag{C}$$

then Φ and (η,ν) are optimal solutions, and the values of (P) and (D) are the same.

Lemma 2 establishes the existence of optimal solutions and shows that *complementarity condition* (C) is not only sufficient but also necessary for optimality of Φ and (η,ν):

Lemma 2 *Suppose Assumption 1 holds. There exists an optimal mechanism Φ, an optimal solution to the primal problem (P). There exists an optimal solution to the dual problem (D) in which η is continuous. Moreover, (C) holds for these optimal Φ and (η,ν).*

Lemmas 1 and 2 yield necessary and sufficient conditions under which a given mechanism is optimal. Specifically, a candidate mechanism Φ is optimal if and only if there exists (η,ν) that satisfies feasibility condition (D1) and complementarity condition (C). Given (D1), Condition (C) holds if and only if $\eta(s)=V(r,s)-\widetilde{u}(r,s)\nu(r)$ for all (r,s) in the support of Φ. For this $\eta(s)$, we can find conditions on the primitives $\widetilde{u}$, V, and F which are equivalent to the existence of function $\nu(r)$ which satisfies (D1).[9] Lemma 1 (Lemma 2) implies that these conditions are sufficient (necessary) for Φ to be optimal.

4.2 Full and No Revelation under Single Crossing

Besides their simplicity and widespread use, the full revelation mechanism Φ_{full} and the no revelation mechanism Φ_{no} satisfy two important properties under Assumption 1. First, Φ_{full} and Φ_{no} are extremal in the following strong sense: (i) Φ_{full} uniquely maximizes the receiver's expected utility, and (ii) Φ_{no} uniquely minimizes the receiver's expected utility (see Proposition 6 in the Appendix). Second, if the sender privately knew s, did not have commitment power, and always preferred to act in that $v(r,s)>0$ for all (r,s), then: (i)

[9]This step is known as Fourier-Motzkin elimination of $\nu(r)$.

Φ_{full} would be the unique equilibrium outcome of a *persuasion game* (Milgrom 1981) in which the sender can withhold information but cannot misrepresent information; and (ii) Φ_{no} would be the unique equilibrium outcome of a *cheap-talk game* (Crawford and Sobel 1982) in which the sender can say anything.[10]

The first main result derives necessary and sufficient conditions under which Φ_{full} and Φ_{no} are optimal. Note that Φ_{full} generates message $r^*(s)$ for each $s \in S$, and Φ_{no} generates the same message r_{no} for all $s \in S$, where r_{no} is a unique r that solves $\int_S \widetilde{u}(r,s) f(s)\, ds = 0$. Let s_{no} be a unique s that solves $u(r_{no}, s) = 0$.

Proposition 1 *Suppose Assumption 1 holds.*

1. *All mechanisms are optimal if and only if, for all $s_1, s_2 \in S$ and $r \in R$ such that $s_2 > s_1$ and $r \in (r^*(s_2), r^*(s_1))$*
$$\frac{V(r^*(s_2), s_2) - V(r, s_2)}{\widetilde{u}(r, s_2)} = \frac{V(r^*(s_1), s_1) - V(r, s_1)}{\widetilde{u}(r, s_1)}. \tag{1}$$

2. *The full revelation mechanism is optimal if and only if, for all $s_1, s_2 \in S$ and $r \in R$ such that $s_2 > s_1$ and $r \in (r^*(s_2), r^*(s_1))$,*
$$\frac{V(r^*(s_2), s_2) - V(r, s_2)}{\widetilde{u}(r, s_2)} \geq \frac{V(r^*(s_1), s_1) - V(r, s_1)}{\widetilde{u}(r, s_1)}. \tag{2}$$

3. *The no revelation mechanism is optimal if and only if, for all $s_1, s_2 \in S$ and $r \in R$ such that $s_2 > s_1$ and $r \in (r^*(s_2), r^*(s_1))$,*
$$\frac{V(r,s_2)-V(r_{no},s_2)}{\widetilde{u}(r,s_2)} + \frac{\widetilde{u}(r_{no},s_2)}{\widetilde{u}(r,s_2)} \frac{\partial V(r_{no},s_{no})/\partial r}{\partial \widetilde{u}(r_{no},s_{no})/\partial r} \leq \frac{V(r,s_1)-V(r_{no},s_1)}{\widetilde{u}(r,s_1)} + \frac{\widetilde{u}(r_{no},s_1)}{\widetilde{u}(r,s_1)} \frac{\partial V(r_{no},s_{no})/\partial r}{\partial \widetilde{u}(r_{no},s_{no})/\partial r}. \tag{3}$$

To verify optimality of a mechanism Φ, one needs to check that no deviation from Φ to any feasible mechanism increases the sender's expected utility, which requires a lot of checks. It turns out that Proposition 1 can be interpreted as follows: for optimality of Φ_{full} and Φ_{no}, it is necessary and sufficient to check that only certain deviations from these mechanisms do not increase the sender's expected utility. I now define these deviations.

For any $s_1, s_2 \in S$ and $r \in R$ such that $s_2 > s_1$ and $r \in (r^*(s_2), r^*(s_1))$, say that the sender prefers to reveal s_1 and s_2 than to pool them at r if, for the prior distribution that

[10] In the persuasion game, if the sender sent the same message r for two or more different s in equilibrium, then there would exist $\widetilde{s}$ such that the sender $\widetilde{s}$ sent r but $u(r, \widetilde{s}) > 0$, which leads to a contradiction because the sender $\widetilde{s}$ would strictly prefer to reveal her type. In the cheap-talk game, if the sender sent two different messages r_1 and r_2 in equilibrium, then she would strictly prefer to send $\min\{r_1, r_2\}$ regardless of s, which leads to a contradiction.

assigns probabilities p_1 and $p_2 = 1 - p_1$ on states s_1 and s_2 that make type r indifferent between the two actions, $\sum_{i=1}^{2} p_i \widetilde{u}(r, s_i) = 0$, the sender's expected utility is higher under the full revelation mechanism than under the no revelation mechanism,

$$\sum_{i=1}^{2} p_i V(r^*(s_i), s_i) \geq \sum_{i=1}^{2} p_i V(r, s_i). \tag{4}$$

Similarly, say that the sender is indifferent between revealing s_1 and s_2 and pooling them at r if (4) holds with equality.

For any $s_1, s_2, s_3 \in S$ and $r \in R$ such that $s_2 > s_1$, $r \in (r^*(s_2), r^*(s_1))$, and $\operatorname{sgn}(r_{no} - r) = \operatorname{sgn}(r^*(s_3) - r_{no})$, say that the sender prefers to pool s_1, s_2, and s_3 at r_{no} than to pool s_1 and s_2 at r and to reveal s_3 if, for the prior distribution that assigns probabilities p_1, p_2, and $p_3 = 1 - p_1 - p_2$ on states s_1, s_2, and s_3 that make type r_{no} indifferent between the two actions, $\sum_{i=1}^{3} p_i \widetilde{u}(r_{no}, s_i) = 0$, and make type r indifferent between the two actions given that $s \neq s_3$, $\sum_{i=1}^{2} p_i \widetilde{u}(r, s_i) = 0$, the sender's expected utility is higher under the no revelation mechanism than under the mechanism that generates message r for s_1 and s_2 and message $r^*(s_3)$ for s_3,

$$\sum_{i=1}^{3} p_i V(r_{no}, s_i) \geq \sum_{i=1}^{2} p_i V(r, s_i) + p_3 V(r^*(s_3), s_3). \tag{5}$$

Finally, say that the sender prefers to pool s_1, s_2, and s_3 at r_{no} than to pool s_1 and s_2 at r and to reveal s_3 for s_3 approaching s_{no} if (5) holds in the limit as $s_3 \to s_{no}$. Using these definitions, I can now restate Proposition 1 as follows.

Corollary 1 *Suppose Assumption 1 holds.*

1. *All mechanisms are optimal if and only if, for all $s_1, s_2 \in S$ and $r \in R$ such that $s_2 > s_1$ and $r \in (r^*(s_2), r^*(s_1))$, the sender is indifferent between revealing s_1 and s_2 and pooling them at r.*

2. *The full revelation mechanism is optimal if and only if, for all $s_1, s_2 \in S$ and $r \in R$ such that $s_2 > s_1$ and $r \in (r^*(s_2), r^*(s_1))$, the sender prefers to reveal s_1 and s_2 than to pool them at r.*

3. *The no revelation mechanism is optimal if and only if, for all $s_1, s_2 \in S$ and $r \in R$ such that $s_2 > s_1$ and $r \in (r^*(s_2), r^*(s_1))$, the sender prefers to pool s_1, s_2, and s_3 at r_{no} than to pool s_1 and s_2 at r and to reveal s_3 for s_3 approaching s_{no}.*

Conditions (1)–(3) in Proposition 1 are weaker than the corresponding conditions (i) – (iii) below from Kamenica and Gentzkow (2011).[11] Following the notation of Assumption 1, suppose a message m of a mechanism Φ generates posterior distribution $Q(s) = P_\Phi(s|m)$ of s, and let r_Q be the receiver's type who is indifferent between the two actions. The sender's indirect expected utility under Q is $\widehat{V}(Q) = \int_S V(r_Q, s)\, dQ(s)$. Kamenica and Gentzkow (2011) show the following:

(i) All mechanisms are optimal if $\widehat{V}(Q)$ is linear in Q, so that the sender is indifferent between separating posteriors Q_1 and Q_2 and pooling them at $\alpha Q_1 + (1-\alpha) Q_2$;

(ii) Full revelation Φ_{full} is optimal if $\widehat{V}(Q)$ is convex in Q, so that the sender prefers to separate Q_1 and Q_2 than to pool them at $\alpha Q_1 + (1-\alpha) Q_2$;

(iii) No revelation Φ_{no} is optimal if the concave closure of $\widehat{V}$ evaluated at the prior F is equal to $\widehat{V}(F)$,[12] so that (after a moment of reflection), for Q_F whose mean is arbitrarily close to $\mathbb{E}_F[s]$, the sender prefers to pool Q and Q_F at F than to separate them.

Proposition 1 shows that it is sufficient to check (i) and (ii) only for degenerate distributions Q_1 and Q_2 whose supports are s_1 and s_2, respectively, and to check (iii) only for discrete Q whose support is $\{s_1, s_2\}$ and degenerate Q_F whose support is s_3 where s_3 is arbitrarily close to s_{no}.

Conditions (1)–(3) are necessary because, for optimality of a candidate mechanism, one needs to check all deviations from the mechanism, including those described in (1)–(3).

The proof of sufficiency of conditions (1)–(3) relies on Lemmas 1 and 2, but we can build the intuition by focusing on *decomposed* mechanisms in which each message is sent by at most two types of the sender. To justify this focus, I construct a decomposed version of Φ_{no} for the case in which $u(r,s)$ is linear in s, s is uniformly distributed on $S = [-1, 1]$, and r is independent of s. Consider a mechanism that sends a message m_0 for $s = 0$ and a different message m_e for each pair $\{-e, e\}$ of S, where $e \in (0, 1]$. Noting that $\mathbb{E}[s|m_e] = \mathbb{E}[s] = 0$ for all $e \in [0, 1]$ implies that this (decomposed) mechanism induces the same mapping from

[11] Rayo and Segal (2010) study a special case of my model with $u(r,s) = s - r$, $v(r,s) = v(s)$, and $g(r|s) = (\overline{r} - \underline{r})^{-1}$ for all $(r,s) \in R \times S$. Proposition 1 can be used to establish their key lemma (Lemma 1), which shows that pooling two types $s_2 > s_1$ yields a higher (lower) expected utility to the sender than separating them if $v(s_2) \leq v(s_1)$ (if $v(s_2) \geq v(s_1)$). If, in addition, $v(s) = bs + c$ for all s, then Assumption 2 holds and Proposition 1 implies: (i) all mechanisms are optimal if $b = 0$, (ii) Φ_{full} is optimal if $b > 0$, and (iii) Φ_{no} is optimal if $b < 0$ (see Corollary 2 below).

[12] Intuitively, a concave closure of a function (defined on a convex set) is the smallest concave function that is everywhere greater than the original function.

(r, s) to the receiver's action as Φ_{no}. This argument can be generalized to show that any mechanism can be decomposed in this way.

I now discuss the intuition for sufficiency conditions of Proposition 1, starting from part 1. Consider any non-trivial message of a decomposed mechanism. This message is sent by some two types of the sender. By (1), the sender is indifferent between revealing these two types and pooling them, so the sender is indifferent between the original mechanism and the mechanism that differs only in that it reveals these two types. Continuously modifying the original mechanism for each message until all types are revealed implies that the sender is indifferent between the original mechanism and Φ_{full}, so part 1 follows.

I now turn to part 2 of Proposition 1. Again, consider any non-trivial message of a decomposed mechanism. This message is sent by some two types of the sender. By (2), the sender prefers to reveal these two types than to pool them, so the sender prefers the mechanism that differs from the original one only in that it reveals these two types. Continuously modifying the original mechanism for each message until all types are revealed implies that the sender prefers Φ_{full} to the original mechanism, so part 2 follows.

Finally, I provide the intuition for a weaker version of part 3 of Proposition 1. Namely, if the sender prefers to pool s_1, s_1', s_2, s_2' at r_{no} than to pool s_1, s_1' at r_1 and to pool s_2, s_2' at r_2 for all feasible s_1, s_1', s_2, s_2', r_1, r_2, then Φ_{no} is optimal. Consider two non-trivial messages of a decomposed mechanism. Suppose that the first message m_1 is sent by s_1 and s_1' and makes the receiver $r_1 \leq r_{no}$ indifferent. Similarly, suppose that the second message m_2 is sent by s_2 and s_2' and makes the receiver $r_2 \geq r_{no}$ indifferent. The sender prefers the mechanism that differs only in that it sends one message that makes the receiver r_{no} indifferent instead of sending both m_1 and m_2. Continuously applying this argument for pairs of messages until all types are pooled implies that the sender prefers Φ_{no} to the original mechanism, so this weaker version of part 3 follows.

4.3 Interval Revelation under Linearity

Under Assumption 2, Proposition 2 simplifies the sender's problem of finding an optimal mechanism to a problem of finding an optimal distribution of messages.

Proposition 2 *Suppose Assumption 2 holds. Let H denote the marginal distribution of m under the optimal mechanism. Then,*

$$\begin{array}{c} H \text{ maximizes } \int_{\underline{s}}^{\overline{s}} V(m)\, dH(m) \\ \text{subject to } F \text{ is a mean-preserving spread of } H. \end{array} \tag{6}$$

The objective function in (6) represents the sender's expected utility; and the constraint in (6) describes the set of feasible distributions of m.[13] The intuition for the constraint is as follows. If F is a mean-preserving spread of H, then F is more informative about the sender's type than H in the sense of Blackwell (1953). A mechanism can garble the sender's information to achieve any less informative distribution H of m than the prior F. Conversely, because a mechanism can only garble the sender's information, F must be a mean-preserving spread of H for any feasible mechanism.

By Proposition 2, the curvature of V determines the form of the optimal mechanism.

Corollary 2 *Suppose Assumption 2 holds and let $r_{no} = \mathbb{E}_F[s]$.*

1. *All mechanisms are optimal if and only if V is linear on S.*

2. *Φ_{full} is optimal if and only if V is convex on S.*

3. *Φ_{no} is optimal if and only if $V(r) \leq V(r_{no}) + V'(r_{no})(r - r_{no})$ for all $r \in S$.*

All three parts of Corollary 2 are straightforward implications of (6).[14] First, if V is linear, then the sender is risk neutral, so all mechanisms are equivalent. Second, if V is convex, then the sender is risk loving, so the full revelation mechanism is optimal. Third, if V is concave, then the sender is risk averse, so the no revelation mechanism is optimal. More precisely, part 3 requires that the concave closure $\mathbf{V}$ of V on S is equal to V at r_{no}.

The second main result derives necessary and sufficient conditions under which an interval revelation mechanism with bounds s_L and s_H is optimal. This mechanism generates message

[13] Kamenica and Gentzkow (2011) note that all feasible H have the same mean as F, but that not all such H are feasible. Proposition 2 shows that H is feasible if and only if F is a mean-preserving spread of H. Using Proposition 2, Gentzkow and Kamenica (2016) provide an alternative characterization of feasible mechanisms.

[14] Rayo and Segal (2010) and Kamenica and Gentzkow (2011) also obtain versions of Corollary 2.

$r_L = \mathbb{E}_F[s|s < s_L]$ for all $s \in [\underline{s}, s_L)$, message $r_H = \mathbb{E}_F[s|s > s_H]$ for all $s \in (s_H, \overline{s}]$, and message s for each $s \in (s_L, s_H)$. In a special case of $s_L = s_H$, the revelation interval (s_L, s_H) is empty, so the mechanism generates only two messages, r_L and r_H.

Proposition 3 *Suppose Assumption 2 holds.*

1. *An interval revelation mechanism with bounds $s_L < s_H$ is optimal if and only if*

$$V(r) \leq V(r_L) + V'(r_L)(r - r_L) \text{ for all } r \in [\underline{s}, s_L] \text{ with equality at } s_L,$$

$$V(r) \leq V(r_H) + V'(r_H)(r - r_H) \text{ for all } r \in [s_H, \overline{s}] \text{ with equality at } s_H,$$

$$V(r) \text{ is convex for all } r \in (s_L, s_H).$$

2. *An interval revelation mechanism with bounds $s_L = s_H$ is optimal if and only if*

$$V(r) \leq V(r_L) + V'(r_L)(r - r_L) \text{ for all } r \in [\underline{s}, s_L],$$

$$V(r) \leq V(r_H) + V'(r_H)(r - r_H) \text{ for all } r \in [s_H, \overline{s}],$$

$$V(r_L) + V'(r_L)(s_L - r_L) = V(r_H) + V'(r_H)(s_H - r_H),$$

$$V'(r_L) \leq V'(r_H).$$

I now discuss implications of Proposition 3 for the case when the derivative $V'(r)$ of the sender's expected utility is either unimodal or bimodal. The derivative V' is *unimodal* if it has a unique local (and therefore global) maximum at $r_m \in R$; the maximum point r_m is called a *mode*. Consider the case of unimodal V' in which $r_m \in S$ and $r_{no} < r_t$, where r_t is the point of tangency illustrated in Figure 2 (a).[15] If F were to assign strictly positive probabilities only on $\underline{s}$ and $\overline{s}$, then the optimal mechanism would send two messages $\underline{s}$ and r_t and the sender's expected utility would achieve the concave closure $\mathbf{V}(r_{no})$. This mechanism, however, is not feasible when F admits a density because s is equal to $\underline{s}$ with probability 0. By part 1 of Proposition 3, $s_L = \underline{s}$ and $s_H \in (\underline{s}, \overline{s})$, so the optimal mechanism reveals s for $s < s_H$ and sends the same message r_H for all $s > s_H$, where the bound s_H is determined by the condition that the sender is indifferent between revealing s_H and pooling it with r_H.[16]

[15] In the remaining cases of unimodal V', either Φ_{full} or Φ_{no} is optimal by Corollary 2.

[16] In an extreme case, when V is a step function with $V(r) = 0$ for $r < r_t$ and $V(r) = 1$ for $r \geq r_t$, the optimal mechanism reveals s for $s < s_H$ and sends the same message r_t for $s > s_H$, where s_H is a unique solution to $\mathbb{E}_F[s|s > s_H] = r_t$. This is the case of an uninformed receiver ($r = r_t$ with probability 1) studied in Kolotilin (2015).

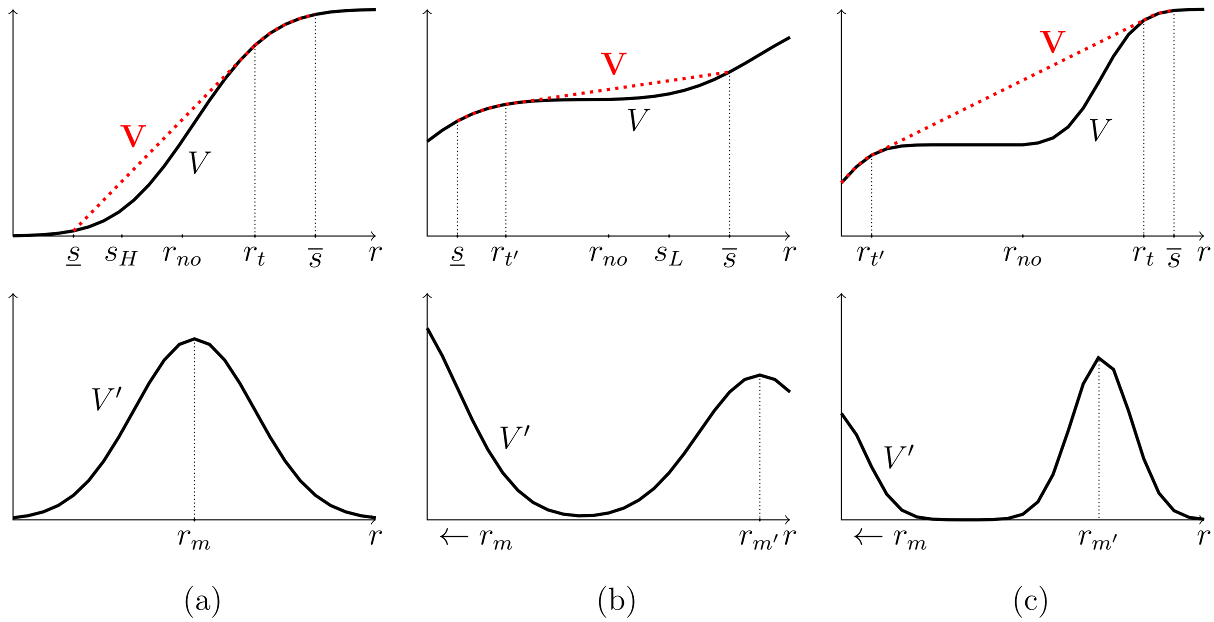

Figure 2: Sender's utility V and concave closure $\mathbf{V}$ when derivative V' is unimodal (a) and bimodal (b,c)

The derivative V' is *bimodal* if it has two local maxima at $r_m, r_{m'} \in R$. If $r_m < \underline{s} < r_{t'} < r_{no} < \overline{s} < r_{m'}$, where $r_{t'}$ is the point of tangency illustrated in Figure 2 (b), then, by part 1 of Proposition 3, $s_H = \overline{s}$ and $s_L \in (\underline{s}, \overline{s})$, so the optimal mechanism reveals s for $s > s_L$ and sends the same message r_L for all $s < s_L$. If $r_m < \underline{s} < r_{t'} < r_{no} < r_{m'} < r_t < \overline{s}$, where r_t and $r_{t'}$ are the points of tangency illustrated in Figure 2 (c), then the optimal mechanism takes one of the following three forms. The first two forms correspond to the interval revelation mechanisms (with interior bounds $s_L, s_H \in (\underline{s}, \overline{s})$) from parts 1 and 2 of Proposition 3. The third form corresponds to the mechanism that sends the two messages r_t and $r_{t'}$, so that the sender's expected utility achieves the concave closure $\mathbf{V}(r_{no})$.[17]

[17]This mechanism does not generally belong to the class of interval revelation mechanisms; so this case is not considered in Proposition 3.

5 Comparative Statics

This section studies the value of the receiver's information. I depart from the assumptions of Section 3 and instead impose the following three assumptions.

Assumption 3 *u and v are increasing in s.*

Assumption 4 *$v(r,s) = v(s)$ and $u(r,s) = u(s)$ for all $(r,s) \in R \times S$.*

Assumption 5 *$v(\bar{s}) > 0$, $u(\bar{s}) > 0$, and $\mathbb{E}_F[u(s) | s : v(s) > 0] < 0$.*

Assumption 3 requires that the sender and receiver are more willing to act for higher types s. Assumption 4 requires that the receiver's type affects the receiver's belief but does not directly affect the sender and receiver's utilities. Assumption 5 requires that the sender can influence the receiver's action but cannot achieve her first-best outcome if the receiver is uninformed. Assumption 3 is mainly for ease of presentation; Assumption 4 is crucial for Proposition 4; and Assumption 5 is crucial for Proposition 5.

Let the set of receiver's types R be finite; so the receiver's *information structure* $\mathcal{G}$ can be described by conditional probabilities $q(r|s)$ of r given s, where $q(r|s)$ is a measurable function of s for each $r \in R$. For each $\mathcal{G}$, the sender and receiver's expected utilities under the optimal mechanism are denoted by $V_{\mathcal{G}}$ and $U_{\mathcal{G}}$, respectively. I use Blackwell (1953)'s ordering of information structures: $\mathcal{G}$ is *more informative* than $\mathcal{G}'$ if there exists a stochastic matrix D such that $q'(r'|s) = \sum_{r \in R} D(r'|r) q(r|s)$ for all $(r', s) \in R' \times S$. An information structure $\mathcal{G}$ is *public* if $q(r|s)$ is either 0 or 1 for all $(r,s) \in R \times S$; that is, the receiver's type is deterministically determined by the sender's type if $\mathcal{G}$ is public. Let $\mathcal{G}_{full}$ and $\mathcal{G}_{no}$ represent fully informative and completely uninformative (public) information structures. That is, $R_{full} = S$ and $q_{full}(s|s) = 1$ for all $s \in S$; $R_{no} = \{r_{no}\}$ and $q_{no}(r_{no}|s) = 1$ for all $s \in S$. Although R_{full} is not finite, it is clear that any information structure $\mathcal{G}$ is less informative than $\mathcal{G}_{full}$ and more informative than $\mathcal{G}_{no}$.

Before discussing non-monotone comparative statics, I present a benchmark result (also found in Kolotilin 2015) that provides sufficient conditions for monotone comparative statics.[18] The receiver's expected utility increases and the sender's expected utility decreases with the precision of the receiver's private information if this precision is either very low or very high. Moreover, this monotonicity holds for all levels of precision if the receiver's information is public.

[18] Relatedly, Bergemann and Morris (2016a) show that the set of implementable outcomes decreases as the information structure becomes more informative, which implies the sender's part of Proposition 4.

Proposition 4 *Suppose Assumptions 3–5 hold and distribution F admits a strictly positive density f on S.*

1. *For any information structure $\mathcal{G}$, we have $U_{\mathcal{G}_{full}} \geq U_{\mathcal{G}} \geq U_{\mathcal{G}_{no}}$ and $V_{\mathcal{G}_{no}} \geq V_{\mathcal{G}} \geq V_{\mathcal{G}_{full}}$.*

2. *For any two public information structures $\mathcal{G}$ and $\mathcal{G}'$, such that $\mathcal{G}$ is more informative than $\mathcal{G}'$, we have $U_{\mathcal{G}} \geq U_{\mathcal{G}'}$ and $V_{\mathcal{G}'} \geq V_{\mathcal{G}}$.*

Intuitively, if $\mathcal{G}$ and $\mathcal{G}'$ are public, and $\mathcal{G}$ is more informative than $\mathcal{G}'$, then, under $\mathcal{G}'$, the sender can first make public information more precise, from $\mathcal{G}'$ to $\mathcal{G}$, and then implement any mechanism Φ available under $\mathcal{G}$, implying $V_{\mathcal{G}'} \geq V_{\mathcal{G}}$. To get the intuition for the receiver's part of Proposition 4, suppose that the sender's utility is type-independent, so that $v(s) = 1$ for all s. In this case, under public $\mathcal{G}$, the optimal mechanism and the no revelation mechanism give the same expected utility to the receiver, implying $U_{\mathcal{G}} \geq U_{\mathcal{G}'}$.

In light of Proposition 4, the utility non-monotonicity presented in Section 2 can only arise when the receiver's information is private and its precision is intermediate. The next proposition shows that, if the sender's type can take only two values, it is always possible to increase the precision of the receiver's private information in such a way that the sender and receiver's expected utilities change non-monotonically.[19]

Proposition 5 *Suppose Assumptions 3–5 hold and the support of F is $\{\underline{s}, \overline{s}\}$. There exist two binary information structures $\mathcal{G}$ and $\mathcal{G}'$, such that $\mathcal{G}$ is more informative than $\mathcal{G}'$, yet $U_{\mathcal{G}'} > U_{\mathcal{G}}$, and $V_{\mathcal{G}} > V_{\mathcal{G}'}$.*

The intuition for Proposition 5 is similar to that in the school–employer example. Since the sender's type can take only two values, without loss of generality, I assume that $r = \Pr(\overline{s}|r)$ for all $r \in R$. By Assumption 5, the sender wants to persuade the receiver to act, but the receiver prefers not to act if he has no information beyond the prior. By continuity, we can find the receiver's type $\overline{r} > \Pr(\overline{s})$ who would still prefer not to act under the no revelation mechanism. A binary information structure of the receiver with $\underline{r} < \overline{r}$ becomes more informative if $\underline{r}$ decreases and $\overline{r}$ stays constant. As $\underline{r}$ decreases, the probability of $\overline{r}$

[19] Bergemann and Morris (2016b) consider an example with $v(\overline{s}) = v(\underline{s}) = 1$, $u(\overline{s}) = 9/10$, $u(\underline{s}) = -1$, $\Pr(\overline{s}) = 1/2$, and a binary information structure with the restriction that $q(\overline{r}|\overline{s}) = q(\underline{r}|\underline{s}) = p$. They show that the set of implementable outcomes (and, thus, the sender's expected utility) decreases with the precision of the receiver's private information p. This monotonicity does not hold without the restriction that $q(\overline{r}|\overline{s}) = q(\underline{r}|\underline{s}) = p$. Since their example satisfies Assumptions 3–5, Proposition 5 implies that there exist two binary information structures $\mathcal{G}$ and $\mathcal{G}'$, such that $\mathcal{G}$ is more informative than $\mathcal{G}'$, yet the sender is strictly better off under $\mathcal{G}$.

increases, because $\Pr(\underline{r})\,\underline{r} + \Pr(\overline{r})\,\overline{r} = \Pr(\overline{s})$, so it becomes relatively more attractive for the sender to target $\overline{r}$ than $\underline{r}$. There exists a critical value $\underline{r}$ at which the sender is exactly indifferent between which of the two types of the receiver to target. Above this value, the sender targets $\underline{r}$, and the receiver's expected utility is strictly positive, because $\overline{r}$ strictly prefers to act whenever $\underline{r}$ acts. Below this value, the sender targets $\overline{r}$, the receiver's expected utility is zero, and the sender's expected utility increases as the receiver's private information becomes more precise ($\underline{r}$ decreases), because the probability of $\overline{r}$ increases.

Appendix: Proofs

Proof of Lemma 1. The lemma can be proved by applying Theorem 2.1 of Anderson and Nash (1987) to my model. But, to make the paper self-contained, I prove this lemma here.

Since η is bounded and measurable on set S, (P1) implies

$$\int_S \eta(s)\, f(s)\, ds = \int_{R^* \times S} \eta(s)\, d\Phi(r,s).$$

Since ν is bounded and measurable on set R, (P2) implies

$$\int_{R^* \times S} \widetilde{u}(r,s)\,\nu(r)\, d\Phi(r,s) = 0.$$

Summing up these two equalities gives

$$\int_S \eta(s)\, f(s)\, ds = \int_{R^* \times S} \left(\eta(s) + \widetilde{u}(r,s)\,\nu(r)\right) d\Phi(r,s). \tag{7}$$

Integrating (D1) over $R^* \times S$ gives

$$\int_{R^* \times S} V(r,s)\, d\Phi(r,s) \le \int_{R^* \times S} \left(\eta(s) + \widetilde{u}(r,s)\,\nu(r)\right) d\Phi(r,s). \tag{8}$$

Suppose that (C) holds for some feasible (η, ν) and Φ. Conditions (7) and (8) yield

$$\int_{R^* \times S} V(r,s)\, d\Phi(r,s) = \int_S \eta(s)\, f(s)\, ds. \tag{9}$$

Consider any other feasible $\widetilde{\Phi}$. Conditions (7) and (8) imply

$$\int_{R^* \times S} V(r,s)\, d\widetilde{\Phi}(r,s) \le \int_S \eta(s)\, f(s)\, ds.$$

Combining this inequality with (9) gives

$$\int_{R^* \times S} V(r,s)\, d\widetilde{\Phi}(r,s) \le \int_{R^* \times S} V(r,s)\, d\Phi(r,s),$$

showing that Φ is an optimal solution to the primal problem (P). An analogous argument proves that (η, ν) is an optimal solution to (D). Finally, (9) shows that the values of (P) and (D) are the same. ■

Proof of Lemma 2. The proof of this lemma is a modification of the proof of Theorem 5.2 in Anderson and Nash (1987), whose notation I closely follow.

Conventions. The primal variable Φ is in $M_r(R^* \times S)$, the space of finite signed measures on $R^* \times S$ with the total variation norm. The mechanism Φ is chosen from the positive closed convex cone P of finite positive measures on $R^* \times S$. The dual constraint function $V(r, s)$ is in $C(R^* \times S)$, the space of continuous measurable functions on $R^* \times S$ with the uniform norm. The dual variables (η, ν) are in $L_\infty(S) \times L_\infty(R^*)$, the space of bounded measurable functions with the uniform norm. The primal constraint function (f, θ) is in $L_1(S) \times L_1(R^*)$, the space of absolutely integrable functions with the 1-norm, where θ is a zero function on the right hand side of (P2).

Optimal solution to (P). One feasible Φ for the primal problem (P) is the full revelation mechanism. The feasible set of the primal problem is bounded because the total variation of any probability measure Φ is equal to one. The constraint map in (P1) is continuous because it is a projection; the constraint map in (P2) is continuous because $\widetilde{u}$ is continuous. The space M_r is the dual of C by Corollary 14.15 of Aliprantis and Border (2006). Therefore, there exists an optimal solution Φ by Theorem 3.20 in Anderson and Nash (1987).

Optimal solution to (D). Since V is continuous on the compact set $R^* \times S$, there exists a finite value $\overline{V} = \max_{r,s} V(r, s)$. Functions $\eta(s) = \overline{V}$ and $\nu(r) = 0$ are feasible for the dual problem, and the set of feasible (η, ν) can be bounded without affecting the value of the dual problem. The constraint map in (D1) is continuous because $\widetilde{u}$ is continuous. The space L_∞ is the dual of L_1 by Theorem 13.28 of Aliprantis and Border (2006). Therefore, there exists an optimal solution (η, ν) by Theorem 3.20 in Anderson and Nash (1987).

Equality (C) under optimal solutions. As can be seen from above, the dual problem has a finite value and functions $\eta(s) = 2\overline{V}$ and $\nu(r) = 0$ are in the interior of the constraint set (D1). Therefore, there is no duality gap by Theorem 3.13 in Anderson and Nash (1987).

Continuity of η. Observe that if (η, ν) is optimal, then (η^*, ν) is also optimal, where

$$\eta^*(s) = \sup{}_r \{V(r, s) - \widetilde{u}(r, s)\nu(r)\}.$$

Indeed, η^* is feasible and $\eta^* \leq \eta$, because η satisfies (D1) for all r, so the objective in (D) is smaller under η^*. I now show that η^* is continuous. Since $R^* \times S$ is compact, V and $\widetilde{u}$ are uniformly continuous. Thus, since ν is bounded, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|(V(r, s) - \widetilde{u}(r, s)\nu(r)) - (V(r, s') - \widetilde{u}(r, s')\nu(r))| < \varepsilon \tag{10}$$

for all $r \in R^*$ and $s, s' \in S$ such that $|s - s'| < \delta$. By definition of η^*, for any s there exists r such that

$$\eta^* (s) < V (r, s) - \widetilde{u} (r, s) \nu (r) + \varepsilon. \tag{11}$$

Thus,

$$\begin{aligned} \eta^* (s') &\geq V (r, s') - \widetilde{u} (r, s') \nu (r) \\ &> V (r, s) - \widetilde{u} (r, s) \nu (r) - \varepsilon \\ &> \eta^* (s) - 2\varepsilon, \end{aligned}$$

where the first inequality holds by definition of η^*, the second by (10), and the third by (11). Analogously, $\eta^* (s) > \eta^* (s') - 2\varepsilon$, so $|\eta^* (s) - \eta^* (s')| < 2\varepsilon$ whenever $|s - s'| < \delta$, implying that η^* is continuous. ■

Proof of Proposition 1. *If part of part 1.* Consider any mechanism Φ. Note that condition (1) holds if and only if there exists a function $b (r)$, such that, for all $s \in S$ and all $r \in (r^* (\overline{s}), r^* (\underline{s}))$, we have $V (r^* (s), s) - V (r, s) = b (r) \widetilde{u} (r, s)$. Substituting this equation into (P2) gives

$$\int_{R^* \times S} V (r, s) \, d\Phi (r, s) = \int_{R^* \times S} V (r^* (s), s) \, d\Phi (r, s).$$

Taking into account (P1) gives

$$\int_{R^* \times S} V (r, s) \, d\Phi (r, s) = \int_S V (r^* (s), s) f (s) \, ds,$$

which implies that the sender's expected utility is the same under all mechanisms.

Only if part of part 1. Suppose to get a contradiction that there exist $s_2 > s_1$, and $r \in (r^* (s_2), r^* (s_1))$ such that

$$\frac{V (r^* (s_2), s_2) - V (r, s_2)}{\widetilde{u} (r, s_2)} > \frac{V (r^* (s_1), s_1) - V (r, s_1)}{\widetilde{u} (r, s_1)}. \tag{12}$$

(The case in which the left hand side of (12) is strictly smaller than the right hand side is analogous.) Let $w_1 (x) = \int_{s_1}^{x} \widetilde{u} (r, s) f (s) \, ds$ and $w_2 (x) = \int_{x}^{s_2} \widetilde{u} (r, s) f (s) \, ds$. There exists $\varepsilon_1 > 0$ such that the function w_1 is continuously differentiable, strictly decreasing on $[s_1, s_1 + \varepsilon_1]$, and vanishing at s_1. Likewise, there exists $\varepsilon_2 > 0$ such that the function w_2 is continuously differentiable, strictly decreasing on $[s_2 - \varepsilon_2, s_2]$, and vanishing at s_2. Thus, on $[s_2 - \varepsilon_2, s_2]$, we can define a continuously differentiable and strictly decreasing function $s_1^* (x)$ that satisfies $w_1 (s_1^* (x)) + w_2 (x) = 0$. By the implicit function theorem,

$$\frac{ds_1^* (x)}{dx} = \frac{\widetilde{u} (r, x) f (x)}{\widetilde{u} (r, s_1^* (x)) f (s_1^* (x))}. \tag{13}$$

By continuity, there exists $x_2 < s_2$ such that (12) holds for all $(s_1, s_2) \in [s_1, s_1^*(x_2)] \times [x_2, s_2]$. Consider two mechanisms that differ only in that one reveals s for all $s \in [s_1, s_1^*(x_2)] \cup [x_2, s_2]$ and the other sends the same message for all $s \in [s_1, s_1^*(x_2)] \cup [x_2, s_2]$. That is, the former mechanism sends $r^*(s)$ and the latter sends r, because $w_1(s_1^*(x)) + w_2(x) = 0$. The sender strictly prefers the former mechanism, because the difference in the sender's expected utility between the former and latter mechanisms is:

$$\begin{aligned}
&\int_{[s_1, s_1^*(x_2)] \cup [x_2, s_2]} (V(r^*(s), s) - V(r, s)) f(s)\, ds \\
> \; &\int_{s_1}^{s_1^*(x_2)} (V(r^*(s), s) - V(r, s)) f(s)\, ds \\
&+ \int_{x_2}^{s_2} \frac{\widetilde{u}(r, s)}{\widetilde{u}(r, s_1^*(s))} (V(r^*(s_1^*(s)), s_1^*(s)) - V(r, s_1^*(s))) f(s)\, ds \; = \; 0,
\end{aligned}$$

where the inequality holds by (12) and the equality holds by (13) and the change of variables formula. This concludes the proof of "only if" part of part 1.

Part 2. By Lemmas 1 and 2, Φ_{full} is optimal if and only if there exists feasible (η, ν) that satisfies

$$\int_{R^* \times S} (\eta(s) + \widetilde{u}(r, s)\nu(r) - V(r, s))\, d\Phi_{full}(r, s) = 0. \tag{14}$$

By (D1), the integrand is nonnegative, so (14) holds if and only if

$$\eta(s) + \widetilde{u}(r^*(s), s)\nu(r^*(s)) = V(r^*(s), s) \text{ almost everywhere.}$$

Since $\widetilde{u}(r^*(s), s) = 0$, we have $\eta(s) = V(r^*(s), s)$ almost everywhere. Since η is continuous by Lemma 2, and V and r^* are continuous by assumption, $\eta(s) = V(r^*(s), s)$ holds for all $s \in S$. Therefore, Φ_{full} is optimal if and only if there exists ν that satisfies (D1):

$$V(r^*(s), s) + \widetilde{u}(r, s)\nu(r) \geq V(r, s) \text{ for all } (r, s) \in R^* \times S, \tag{15}$$

which is equivalent to

$$\frac{V(r, s_2) - V(r^*(s_2), s_2)}{\widetilde{u}(r, s_2)} \leq \nu(r) \leq \frac{V(r^*(s_1), s_1) - V(r, s_1)}{-\widetilde{u}(r, s_1)}$$

for all $r \in (r^*(\overline{s}), r^*(\underline{s}))$ and s_1, s_2 such that $r \in (r^*(s_2), r^*(s_1))$. (For $r \in \{r^*(\overline{s}), r^*(\underline{s})\}$, the existence of ν is obvious because (15) bounds ν only from one side.) There exists such ν if and only if (2) holds.

Part 3. Analogously to part 2, Φ_{no} is optimal if and only if there exists feasible (η, ν) that satisfies

$$\eta(s) + \widetilde{u}(r_{no}, s)\nu(r_{no}) = V(r_{no}, s) \text{ for all } s \in S. \tag{16}$$

Therefore, Φ_{no} is optimal if and only if there exists ν that satisfies (D1):

$$V(r_{no}, s) - \widetilde{u}(r_{no}, s)\,\nu(r_{no}) + \widetilde{u}(r, s)\,\nu(r) \geq V(r, s) \text{ for all } (r, s) \in R^* \times S, \tag{17}$$

which is equivalent to

$$\frac{V(r,s_2)-(V(r_{no},s_2)-\widetilde{u}(r_{no},s_2)\nu(r_{no}))}{\widetilde{u}(r,s_2)} \leq \nu(r) \leq \frac{(V(r_{no},s_1)-\widetilde{u}(r_{no},s_1)\nu(r_{no}))-V(r,s_1)}{-\widetilde{u}(r,s_1)} \tag{18}$$

for all $r \in (r^*(\overline{s}), r^*(\underline{s}))$, and $s_1, s_2 \in S$ such that $r \in (r^*(s_2), r^*(s_1))$. (For $r \in \{r^*(\overline{s}), r^*(\underline{s})\}$, the existence of ν is obvious because (17) bounds ν only from one side.)

At $r = r_{no}$, both sides of (18) become $\nu(r_{no})$. Thus, for (18) to be satisfied everywhere, the derivatives of both sides of (18) with respect to r evaluated at $r = r_{no}$ must coincide, which gives

$$\nu(r_{no}) = \frac{\frac{\partial V(r_{no},s_1)/\partial r}{\widetilde{u}(r_{no},s_1)} - \frac{\partial V(r_{no},s_2)/\partial r}{\widetilde{u}(r_{no},s_2)}}{\frac{\partial \widetilde{u}(r_{no},s_1)/\partial r}{\widetilde{u}(r_{no},s_1)} - \frac{\partial \widetilde{u}(r_{no},s_2)/\partial r}{\widetilde{u}(r_{no},s_2)}}. \tag{19}$$

Taking the limit of (19) as $s_2 \downarrow s_{no}$ gives

$$\nu(r_{no}) = \frac{\partial V(r_{no}, s_{no})/\partial r}{\partial \widetilde{u}(r_{no}, s_{no})/\partial r}. \tag{20}$$

Substituting $\nu(r_{no})$ from (20) into (18) completes the proof of Proposition 1. ■

Proof of Corollary 1. *Parts 1 and 2.* Since $p_2 = 1 - p_1$ and $\sum_{i=1}^{2} p_i \widetilde{u}(r, s_i) = 0$,

$$p_1 = \frac{\widetilde{u}(r, s_2)}{\widetilde{u}(r, s_2) - \widetilde{u}(r, s_1)} \text{ and } p_2 = -\frac{\widetilde{u}(r, s_1)}{\widetilde{u}(r, s_2) - \widetilde{u}(r, s_1)}.$$

By Assumption 1, $\widetilde{u}(r, s_1) < 0 < \widetilde{u}(r, s_2)$ because $r \in (r^*(s_2), r^*(s_1))$; so $p_1, p_2 \in (0, 1)$. Substituting p_1 and p_2 in (4) gives (2). Finally, (4) with equality is equivalent to (1).

Part 3. Since $\operatorname{sgn}(r_{no} - r) = \operatorname{sgn}(r^*(s_3) - r_{no})$, either $r_{no} \in (r, r^*(s_3))$ or $r_{no} \in (r^*(s_3), r)$ is true. Suppose that $r_{no} \in (r, r^*(s_3))$ (the other case is analogous). By Assumption 1, $\widetilde{u}(r_{no}, s_3) < 0$ because $r_{no} < r^*(s_3)$, $\widetilde{u}(r, s_1) < 0 < \widetilde{u}(r, s_2)$ because $r \in (r^*(s_2), r^*(s_1))$, and $\sum_{i=1}^{2} p_i \widetilde{u}(r_{no}, s_i) > 0$ because $r_{no} > r$ and $\sum_{i=1}^{2} p_i \widetilde{u}(r, s_i) = 0$. Therefore, the system of equations $p_1 + p_2 + p_3 = 1$, $\sum_{i=1}^{3} p_i \widetilde{u}(r_{no}, s_i) = 0$, and $\sum_{i=1}^{2} p_i \widetilde{u}(r, s_i) = 0$ has a unique solution $p_1, p_2, p_3 \in (0, 1)$. Substituting these p_1, p_2, and p_3 in (5) and rearranging gives

$$\frac{V(r,s_2)-V(r_{no},s_2)}{\widetilde{u}(r,s_2)} + \frac{\widetilde{u}(r_{no},s_2)}{\widetilde{u}(r_{no},s_3)} \frac{V(r_{no},s_3)-V(r^*(s_3),s_3)}{\widetilde{u}(r,s_2)} \leq \frac{V(r,s_1)-V(r_{no},s_1)}{\widetilde{u}(r,s_1)} + \frac{\widetilde{u}(r_{no},s_1)}{\widetilde{u}(r_{no},s_3)} \frac{V(r_{no},s_3)-V(r^*(s_3),s_3)}{\widetilde{u}(r,s_1)}.$$

Taking the limit of this inequality as $s_3 \to s_{no}$ gives (3), because

$$\lim_{s_3 \to s_{no}} \frac{V(r_{no}, s_3) - V(r^*(s_3), s_3)}{\widetilde{u}(r_{no}, s_3)} = -\frac{\partial V(r_{no}, s_{no})/\partial r}{\partial \widetilde{u}(r_{no}, s_{no})/\partial s} \frac{dr^*(s_{no})}{ds} = \frac{\partial V(r_{no}, s_{no})/\partial r}{\partial \widetilde{u}(r_{no}, s_{no})/\partial r},$$

where the first equality holds by L'Hospital's rule and the second by the implicit function theorem applied to $\widetilde{u}(r^*(s), s) = 0$. ■

Proof of Proposition 2. Any mechanism Φ, whose messages m satisfy $m = \mathbb{E}_\Phi[s|m]$, generates messages with a distribution H having the property that distribution F is a mean-preserving spread of H. It remains to verify that any distribution H having this property can be generated by a feasible mechanism. If F is a mean-preserving spread of H, then, by definition, s has the same distribution as $m + z$ for some z such that $\mathbb{E}[z|m] = 0$. Define $\Phi(\widetilde{m}, \widetilde{s}) = \Pr(m \leq \widetilde{m}, m + z \leq \widetilde{s})$ for all $(\widetilde{m}, \widetilde{s}) \in S \times S$. For this Φ, the marginal distribution of s is F and $\mathbb{E}_\Phi[s|m] = \mathbb{E}_\Phi[m + z|m] = m$. Therefore, Φ is a feasible mechanism whose messages are distributed according to H. ■

Proof of Corollary 2. Assumption 1 and Proposition 1 hold after replacing r with $-r$. With this change of variables, $r^*(s) = s$, $V(r, s) = V(r)$, and $\widetilde{u}(r, s) = (s - r)g(r)$.

By part 1 of Proposition 1, all mechanisms are optimal if and only if (1):

$$V(r) = \frac{s_2 - r}{s_2 - s_1} V(s_1) + \frac{r - s_1}{s_2 - s_1} V(s_2) \text{ for all } s_1, s_2, r \in S.$$

By part 2 of Proposition 1, Φ_{full} is optimal if and only if (2):

$$V(r) \leq \frac{s_2 - r}{s_2 - s_1} V(s_1) + \frac{r - s_1}{s_2 - s_1} V(s_2) \text{ for all } s_1, s_2, r \in S.$$

By part 3 of Proposition 1, Φ_{no} is optimal if and only if (3), which simplifies to the condition of part 3 of Corollary 2. ■

Proof of Proposition 3. *Only if part of part 1.* By Lemma 2, if the described mechanism Φ is optimal, then there exists feasible (η, ν) that satisfies

$$\int_{R \times S} (\eta(s) + (s - r)\nu(r) - V(r))\, d\Phi(r, s) = 0.$$

By the feasibility condition (D1), the integrand is nonnegative. Moreover, by Lemma 2, η is continuous, so

$$\eta(s) = \begin{cases} V(r_L) - (s - r_L)\nu(r_L) \text{ for all } s \in [\underline{s}, s_L], \\ V(s) \text{ for all } s \in (s_L, s_H), \\ V(r_H) - (s - r_H)\nu(r_H) \text{ for all } s \in [s_H, \overline{s}]. \end{cases}$$

The feasibility condition (D1) implies

$$V(s) + (s - r)\nu(r) \geq V(r) \text{ for all } s, r \in (s_L, s_H).$$

Taking the limits $s \uparrow r$ and $s \downarrow r$ yields $\nu(r) = -V'(r)$ for all $r \in (s_L, s_H)$. Substituting back gives

$$V(s) \geq V(r) + V'(r)(s - r) \text{ for all } s, r \in (s_L, s_H),$$

which implies that $V(r)$ is convex for all $r \in (s_L, s_H)$.

The feasibility condition (D1) also implies

$$V(r_L) - (s - r_L)\nu(r_L) + (s - r)\nu(r) \geq V(r) \text{ for all } s, r \in [\underline{s}, s_L].$$

Writing these inequalities for $s = s_L$ and $s = \underline{s}$ gives

$$\begin{aligned} V(r_L) - (s_L - r_L)\nu(r_L) + (s_L - r)\nu(r) &\geq V(r), \\ V(r_L) - (\underline{s} - r_L)\nu(r_L) + (\underline{s} - r)\nu(r) &\geq V(r). \end{aligned}$$

Multiplying the first inequality by $(r - \underline{s})$, the second by $(s_L - r)$, and adding up yields:

$$(s_L - \underline{s})(V(r_L) + (r_L - r)\nu(r_L) - V(r)) \geq 0.$$

Taking the limits $r \uparrow r_L$ and $r \downarrow r_L$ yields $\nu(r_L) = -V'(r_L)$. Substituting back gives

$$V(r) \leq V(r_L) + V'(r_L)(r - r_L) \text{ for all } r \in [\underline{s}, s_L] \text{ with equality at } s_L,$$

where the equality holds by continuity of η.

To complete the proof, we can use the same argument to get $\nu(r_H) = -V'(r_H)$ and

$$V(r) \leq V(r_H) + V'(r_H)(r - r_H) \text{ for all } r \in [s_H, \overline{s}] \text{ with equality at } s_H.$$

If part of part 1. Consider the described mechanism Φ and the constructed pair (η, ν):

$$\begin{aligned} \eta(s) &= \begin{cases} V(r_L) + V'(r_L)(s - r_L) \text{ for } s \in [\underline{s}, s_L], \\ V(s) \text{ for } s \in (s_L, s_H), \\ V(r_H) + V'(r_H)(s - r_H) \text{ for } s \in [s_H, \overline{s}], \end{cases} \\ \nu(r) &= \begin{cases} -V'(r_L) \text{ for } r \in [\underline{s}, s_L], \\ -V'(r) \text{ for } r \in (s_L, s_H), \\ -V'(r_H) \text{ for } r \in [s_H, \overline{s}]. \end{cases} \end{aligned}$$

The complementarity condition (C) holds by construction. Moreover, (η, ν) is feasible for (D), because, for all $(r, s) \in S \times S$, we have:

$$\eta(s) + (s - r)\nu(r) \geq \eta(r) \geq V(r),$$

where the first inequality holds because $\eta(s)$ is convex for all $s \in S$ and $-\nu(r)$ is a subderivative of $\eta(r)$ for all $r \in S$. Therefore, by Lemma 1, Φ is optimal.

Only if part of part 2. The proof of the first three conditions is the same as in part 1. To prove $V'(r_L) \leq V'(r_H)$, write the feasibility condition (D1) for $s \geq s_H$ and $r = r_L$:

$$V(r_H) + V'(r_H)(s - r_H) \geq V(r_L) + V'(r_L)(s - r_L),$$

and notice that both sides are equal at $s = s_H = s_L$ and linear in s for $s \geq s_H$.

If part of part 2. Consider the described mechanism Φ and the pair (η, ν) from part 1. By the same argument as in part 1, Φ is optimal. ■

Proof of Proposition 4. Although this proposition follows almost immediately from Kolotilin (2015), below I provide a simpler proof adapted to this setting.

Part 1. $V_{\mathcal{G}} \geq V_{\mathcal{G}_{full}}$, because the sender can always achieve $V_{\mathcal{G}_{full}}$ by choosing the full revelation mechanism. Similarly, $V_{\mathcal{G}_{no}} \geq V_{\mathcal{G}}$, because an outcome produced under $\mathcal{G}$ by Φ can also be achieved under $\mathcal{G}_{no}$ by a mechanism Φ' that generates r according to q and m according to Φ. Trivially, $U_{\mathcal{G}_{full}} \geq U_{\mathcal{G}}$, because the receiver is best off when he knows the sender's type.

Finally, $U_{\mathcal{G}} \geq U_{\mathcal{G}_{no}}$, because the optimal mechanism under $\mathcal{G}_{no}$ gives the same expected utility to the receiver as the no revelation mechanism; that is, $U_{\mathcal{G}_{no}} = \max\{\mathbb{E}_F[u(s)], 0\}$. Indeed, since $\mathbb{E}_F[u(s)|s : v(s) > 0] < 0$ by Assumption 5, the optimal mechanism induces the receiver to act if and only if $s \geq s^*$, where s^* is the unique solution to $\mathbb{E}_F[u(s)|s \geq s^*] = 0$ (Kolotilin 2015).

Part 2. Public $\mathcal{G}$ partitions S into disjoint subsets $S_r = \{s : q(r|s) = 1\}$. Moreover, since public $\mathcal{G}$ is more informative than public $\mathcal{G}'$, $\mathcal{G}$ is a *refinement* of $\mathcal{G}'$; that is, for each $r \in R$, there exists $r' \in R'$ such that $S_r \subset S_{r'}$. Therefore, an outcome produced under $\mathcal{G}$ by Φ can also be achieved under $\mathcal{G}'$ by a mechanism Φ' that refines $\mathcal{G}'$ to $\mathcal{G}$ and generates m according to Φ, which implies that $V_{\mathcal{G}'} \geq V_{\mathcal{G}}$.

Under public $\mathcal{G}$, r is deterministically determined by s, so we can allow mechanisms to be conditioned on r. By Kolotilin (2015), the optimal mechanism induces the receiver to act if and only if $s \geq s^*(r)$, where $s^*(r)$ is the minimum $\widetilde{s} \in S$ such that $v(\widetilde{s}) \geq 0$ and $\mathbb{E}_F[u(s)|s \in S_r : s \geq \widetilde{s}] \geq 0$. Therefore,

$$U_{\mathcal{G}} = \sum_{r \in R} \max\left\{\int_{s \in S_r} u(s) f(s)\, ds, \int_{s \in S_r : v(s) \geq 0} u(s) f(s)\, ds, 0\right\},$$

which implies that $U_{\mathcal{G}} \geq U_{\mathcal{G}'}$, because $\mathcal{G}$ is a refinement of $\mathcal{G}'$. ■

Proof of Proposition 5. If s can take only two values, Assumption 5 implies that $u(\underline{s}) < 0 < v(\underline{s})$. Therefore, $\mathbb{E}_F[u(s)|s : v(s) > 0] < 0$ simplifies to $\mathbb{E}_F[u(s)] < 0$, which holds if and only if $\mu = \Pr(\overline{s}) < 1/(1+x)$, where $x = -u(\overline{s})/u(\underline{s})$.

For any r_1 and r_2 such that $0 < r_1 < \mu < r_2 < 1/(1+x)$, there exists a binary information structure of the receiver with $R = \{r_1, r_2\}$, where $r_1 = \Pr(\overline{s}|r_1)$ and $r_2 = \Pr(\overline{s}|r_2)$. Similarly to the school–employer example, we can restrict attention to mechanisms that generate only three messages m_2, m_1, and m_0, where m_2 makes the receiver r_2 indifferent, m_1 makes the receiver r_1 indifferent, and m_0 convinces both types of the receiver that $s = \underline{s}$.

Since $r_2 < 1/(1+x)$, neither type of the receiver would act if the sender chose the no revelation mechanism. It is easy to show then that the optimal mechanism is either Φ_1 that generates m_1 and m_0 or Φ_2 that generates m_2 and m_0.

Without loss of generality, assume that $m = \Pr(\overline{s}|m)$ for all m. After receiving m_i, where $i \in \{1, 2\}$, the receiver r_i holds the posterior:

$$\Pr(\overline{s}|m_i, r_i) = \frac{\frac{m_i r_i}{\mu}}{\frac{m_i r_i}{\mu} + \frac{(1-m_i)(1-r_i)}{(1-\mu)}}.$$

Since m_i makes r_i indifferent, we have $\Pr(\overline{s}|m_i, r_i) = 1/(1+x)$, which is equivalent to

$$m_i = \frac{\mu(1-r_i)}{\mu(1-r_i) + (1-\mu) r_i x}.$$

Since the posteriors $\Pr(\overline{s}|m)$ must average out to the prior $\Pr(\overline{s})$, the mechanism Φ_i generates m_i with probability μ/m_i. Therefore, the sender's expected utility under Φ_1 is:

$$\begin{aligned} V_1 &= \frac{\mu}{m_1}\left((1-m_1) v(\underline{s}) + m_1 v(\overline{s})\right) \\ &= \frac{(1-\mu) r_1 x v(\underline{s}) + \mu(1-r_1) v(\overline{s})}{1-r_1}. \end{aligned}$$

Similarly, the sender's expected utility under Φ_2 is:

$$\begin{aligned} V_2 &= \frac{\mu}{m_2}\left((1-m_2)\Pr(r_2|\underline{s}) v(\underline{s}) + m_2 \Pr(r_2|\overline{s}) v(\overline{s})\right) \\ &= \frac{(\mu - r_1) r_2 (x v(\underline{s}) + v(\overline{s}))}{r_2 - r_1}. \end{aligned}$$

In the limit $r_1 \downarrow 0$, we have $V_2 = \mu(xv(\underline{s}) + v(\overline{s})) > \mu v(\overline{s}) = V_1$; but in the limit $r_1 \uparrow \mu$, we have $V_1 = \mu(xv(\underline{s}) + v(\overline{s})) > 0 = V_2$. Since both V_1 and V_2 are continuous in r_1, for each $r_2^* \in (\mu, 1/(1+x))$, there exists an $r_1^* \in (0, \mu)$ at which $V_1 = V_2$.

Let $\mathcal{G}$ and $\mathcal{G}'$ be the two information structures with $R = \{r_1^*/2, r_2^*\}$ and $R' = \{r_1^*, r_2^*\}$, respectively. Because $r_1 \le r_1'$ and $r_2 \ge r_2'$, $\mathcal{G}$ is more informative than $\mathcal{G}'$. Under $\mathcal{G}'$, the sender is indifferent between Φ_1 and Φ_2, so Φ_1 is an optimal mechanism. Since V_1 increases with r_1 and V_2 decreases with r_1, the sender's optimal mechanism under $\mathcal{G}$ is Φ_2. The receiver's expected utility is 0 under Φ_2 and is strictly positive under Φ_1; so $U_{\mathcal{G}'} > U_{\mathcal{G}}$. Since V_2 decreases with r_1, the sender is strictly better off under $\mathcal{G}$; so $V_{\mathcal{G}} > V_{\mathcal{G}'}$. ■

Proposition 6 *Suppose Assumption 1 holds.*

1. *The receiver's expected utility under Φ_{full} is strictly higher than under any other Φ.*

2. *The receiver's expected utility under Φ_{no} is strictly lower than under any other Φ.*

Proof of Proposition 6. The receiver's expected utility under Φ, Φ_{full}, and Φ_{no} is:

$$\mathbb{E}_{\Phi}[u] = \int_{R^*\times S}\left(\int_r^{\overline{r}} \widetilde{u}(\widetilde{r},s)\,d\widetilde{r}\right) d\Phi(r,s), \tag{21}$$

$$\mathbb{E}_{\Phi_{full}}[u] = \int_S\left(\int_{r^*(s)}^{\overline{r}} \widetilde{u}(\widetilde{r},s)\,d\widetilde{r}\right) f(s)\,ds = \int_{R^*\times S}\left(\int_{r^*(s)}^{\overline{r}} \widetilde{u}(\widetilde{r},s)\,d\widetilde{r}\right) d\Phi(r,s), \tag{22}$$

$$\mathbb{E}_{\Phi_{no}}[u] = \int_S\left(\int_{r_{no}}^{\overline{r}} \widetilde{u}(\widetilde{r},s)\,d\widetilde{r}\right) f(s)\,ds = \int_{R^*\times S}\left(\int_{r_{no}}^{\overline{r}} \widetilde{u}(\widetilde{r},s)\,d\widetilde{r}\right) d\Phi(r,s). \tag{23}$$

Equation (21) holds because a message m induces the receiver r to act if and only if $r \geq m$. The first equality in (22) holds because Φ_{full} generates message $r^*(s)$ for each $s \in S$. Similarly, the first equality in (23) holds because Φ_{no} generates r_{no} for all $s \in S$. The second equality in (22) and (23) holds because the marginal distribution of s under any mechanism Φ coincides with the prior distribution of s.

Part 1. Fubini's Theorem together with the condition $\widetilde{u}(r^*(s), s) = 0$ gives

$$\begin{aligned}\mathbb{E}_{\Phi_{full}}[u] - \mathbb{E}_{\Phi}[u] = &\int_S \int_{r>r^*(s)} \left(\int_{r^*(s)}^{r} \widetilde{u}(\widetilde{r},s)\,d\widetilde{r}\right) d\Phi(r,s)\\ &- \int_S \int_{r<r^*(s)} \left(\int_{r}^{r^*(s)} \widetilde{u}(\widetilde{r},s)\,d\widetilde{r}\right) d\Phi(r,s).\end{aligned} \tag{24}$$

By Assumption 1, we have $\widetilde{u}(\widetilde{r},s) > 0$ for $\widetilde{r} > r^*(s)$, so $\int_{r^*(s)}^{r} \widetilde{u}(\widetilde{r},s)\,d\widetilde{r} > 0$ for $r > r^*(s)$. Any Φ that differs from Φ_{full} puts strictly positive probability on the event $r > r^*(s)$, otherwise $\int_{R^*\times S} \widetilde{u}(r,s)\,d\Phi(r,s)$ would be strictly negative rather than zero. Therefore, the first integral in (24) is strictly positive. The analogous argument shows that the second integral in (24) is strictly negative, so $\mathbb{E}_{\Phi_{full}}[u] - \mathbb{E}_{\Phi}[u] > 0$ for any Φ that differs from Φ_{full}.

Part 2. For a mechanism Φ, denote the conditional distribution of s given a message r by $P_{\Phi}(s|r)$ and the marginal distribution of message r by $P_{\Phi}(r)$. Fubini's Theorem gives

$$\begin{aligned}\mathbb{E}_{\Phi}[u] - \mathbb{E}_{\Phi_{no}}[u] = &\int_{\underline{r}}^{r_{no}} \left[\int_{r}^{r_{no}} \left(\int_S \widetilde{u}(\widetilde{r},s)\,dP_{\Phi}(s|r)\right) d\widetilde{r}\right] dP_{\Phi}(r)\\ &- \int_{r_{no}}^{\overline{r}} \left[\int_{r_{no}}^{r} \left(\int_S \widetilde{u}(\widetilde{r},s)\,dP_{\Phi}(s|r)\right) d\widetilde{r}\right] dP_{\Phi}(r).\end{aligned} \tag{25}$$

By Assumption 1, we have $\int_S \widetilde{u}(\widetilde{r}, s)\, dP_\Phi(s|r) > 0$ for $\widetilde{r} > r$. Therefore,

$$\int_r^{r_{no}} \left(\int_S \widetilde{u}(\widetilde{r}, s)\, dP_\Phi(s|r) \right) d\widetilde{r} > 0 \text{ for } r < r_{no}.$$

Since $P_\Phi(r)$ of any mechanism Φ that differs from Φ_{no} puts strictly positive probability on messages in $[\underline{r}, r_{no})$, the first integral in (25) is strictly positive. The analogous argument shows that the second integral in (25) is strictly negative, so $\mathbb{E}_\Phi[u] - \mathbb{E}_{\Phi_{no}}[u] > 0$ for any Φ that differs from Φ_{no}. ■

References

Aliprantis, Charalambos D. and Kim C. Border (2006) *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Berlin: Springer-Verlag.

Alonso, Ricardo and Niko Matouschek (2008) "Optimal Delegation," *Review of Economic Studies*, Vol. 75, pp. 259–293.

Amador, Manuel and Kyle Bagwell (2013) "The Theory of Optimal Delegation With an Application to Tariff Caps," *Econometrica*, Vol. 81, pp. 1541–1599.

Anderson, Edward J. and Peter Nash (1987) *Linear Programming in Infinite-Dimensional Spaces: Theory and Applications*, New York: John Wiley and Sons.

Arvey, Richard D. and James E. Campion (1982) "The Employment Interview: A Summary and Review of Recent Research," *Personnel Psychology*, Vol. 35, pp. 281–322.

Bergemann, Dirk and Stephen Morris (2016a) "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, Vol. 11, pp. 487–522.

Bergemann, Dirk and Stephen Morris (2016b) "Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium," *American Economic Review: Papers and Proceedings*, Vol. 106, pp. 586–591.

Blackwell, David (1953) "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, Vol. 24, pp. 265–272.

Crawford, Vincent P. and Joel Sobel (1982) "Strategic Information Transmission," *Econometrica*, Vol. 50, pp. 1431–1451.

Gentzkow, Matthew and Emir Kamenica (2016) “A Rothschild-Stiglitz Approach to Bayesian Persuasion,” *American Economic Review: Papers and Proceedings*, Vol. 106, pp. 597–601.

Holmstrom, Bengt (1984) “On the Theory of Delegation,” in M Boyer and R. Kihlstrom eds. *Bayesian Models in Economic Theory*, New York: North-Holland.

Kamenica, Emir and Matthew Gentzkow (2011) “Bayesian Persuasion,” *American Economic Review*, Vol. 101, pp. 2590–2615.

Kolotilin, Anton (2015) “Experimental Design to Persuade,” *Games and Economic Behavior*, Vol. 90, pp. 215–226.

Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li (2017) “Persuasion of a Privately Informed Receiver,” *Econometrica*, Vol. forthcoming.

Milgrom, Paul (1981) “Good News and Bad News: Representation Theorems and Applications,” *Bell Journal of Economics*, Vol. 12, pp. 380–391.

Ostrovsky, Michael and Michael Schwarz (2010) “Information Disclosure and Unraveling in Matching Markets,” *American Economic Journal: Microeconomics*, Vol. 2, pp. 34–63.

Quah, John K-H and Bruno Strulovici (2012) “Aggregating the Single Crossing Property,” *Econometrica*, Vol. 80, pp. 2333–2348.

Rayo, Luis and Ilya Segal (2010) “Optimal Information Disclosure,” *Journal of Political Economy*, Vol. 118, pp. 949–987.