Optimal Information Disclosure: Quantity vs. Quality*

Anton Kolotilin†

First Version: November, 2011
This Version: March, 2014

Abstract

A sender chooses ex ante how her information will be disclosed to a privately informed receiver who then takes one of two actions. The sender wishes to maximize the probability that the receiver takes the desired action. The sender faces an ex ante quantity-quality tradeoff: sending positive messages more often (in terms of the sender's information) makes it less likely that the receiver will take the desired action (in terms of the receiver's information). Interestingly, the sender's and receiver's welfare is not monotonic in the precision of the receiver's private information: the sender may find it easier to influence a more informed receiver, and the receiver may suffer from having more precise private information. Necessary and sufficient conditions are derived for full and no information revelation to be optimal.

Key words: information disclosure, persuasion, informed decision maker

JEL Codes: C72, D81, D82, D83

*This paper is a revised version of Chapter 2 of my dissertation at MIT. I thank Robert Gibbons and Muhamet Yildiz for their invaluable guidance and advice. I also thank Ricardo Alonso, Sandeep Baliga, Abhijit Banerjee, Gabriel Carroll, Denis Chetverikov, Glenn Ellison, Richard Holden, Hongyi Li, Jin Li, Uliana Loginova, Niko Matouschek, Parag Pathak, Michael Powell, Andrea Prat, Jeroen Swinkels, Juuso Toikka, Alexander Wolitzky, Juan Xandri, Luis Zermeño, and the seminar participants at CUHK, HKU, HKUST, MIT (Economics and Sloan), NES, Northwestern University (Kellogg M&S), University of Melbourne, UNSW, UQ as well as the participants at the Latin American Meeting of the Econometric Society for helpful comments and suggestions.

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

1 Introduction

A decision maker, the receiver, often relies on information obtained from an interested party, the sender. In contrast to most of the literature on communication, I allow the receiver to obtain information not only from the sender but also from other sources. The main goal is to understand economic aspects of optimal information disclosure from a sender to a privately informed receiver and provide a linear programming approach to this problem.

In my model, the receiver decides whether to act or not to act. The sender's utility depends only on the action taken by the receiver, and she prefers the receiver to act. The receiver's utility depends both on his action and on information. The receiver takes an action that maximizes his expected utility, given his private information and information disclosed by the sender. Before observing her private information, the sender can commit to how her private information will be disclosed to the receiver. Formally, the sender can choose any (stochastic) mapping from her information to messages, which I call a *mechanism*. The sender chooses the mechanism that maximizes the ex ante probability that the receiver will act. I impose a single-crossing assumption on the receiver's preferences and information requiring that receiver's types can be ordered according to their willingness to act.

For example, consider a school that chooses a disclosure policy for a student in order to persuade a potential employer to hire him. The school has a lot of freedom in choosing which part of available information about the student will appear on his transcript. Moreover, the school commits to its disclosure policy before it learns 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. In addition, the school uses the same disclosure policy for all students, who apply to different employers. This also contributes to the receiver's private information in terms of my model.

Since the receiver has private information, he acts or does not act depending not only on a message received from the sender but also on his private information. Thus, from the sender's perspective, each message generates a probability distribution over receiver's actions. Therefore, when the sender chooses a mechanism, she faces an important quantity-quality tradeoff of messages that she will later send. On the one hand, the sender wants to send the positive message, corresponding to higher types of the sender, more often. On the other hand, the sender wants the positive message to persuade a larger set of receiver's types. The optimal mechanism balances these two conflicting objectives. For example, when the school chooses lower standards for getting good grades, more students get good-looking transcripts, but employers rationally account for this and each student with a good-looking transcript will find it harder to get a job. This ex ante tradeoff does not appear in cheap talk and verifiable message games where the sender chooses a report at the interim stage when she already has her private information.

Interestingly, under the optimal mechanism, the sender's and receiver's expected utilities 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 structure of the receiver's private information, and the sender may prefer to disclose significantly less information if the receiver's information is more precise.[1] 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. In this case, it becomes easier for the sender to persuade the receiver with more precise favorable information, so the sender may be able to persuade the receiver with a higher total probability.

The sender's problem of finding an optimal mechanism reduces to a linear program. Using duality theory, I show how to obtain primitive necessary and sufficient conditions for a candidate mechanism to be optimal. This is the main technical contribution of the paper, which can be applied to other models of information disclosure because the sender's and receiver's expected utilities are always linear in probabilities that constitute a mechanism. In reality, schools choose various disclosure policies and duality theory allows us to find primitive conditions on the environment that justify each choice. At the one extreme, schools report all grades and class rank on transcripts. The 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, schools release no transcripts. The no revelation mechanism is optimal if and only if the sender prefers to pool any three of her types than to pool two of them and reveal the third one. Under an additional assumption that the receiver's utility is linear in the sender's and receiver's types that are independent of each other, I show that the shape of the optimal mechanism is determined by the convexity properties of the distribution of the receiver's type and by the expectation of the sender's type.

The most related literature is the one in which the sender can commit to an information disclosure mechanism. Kamenica and Gentzkow (2011) study a much more general model

[1] Continuing the school-employer application, Arvey and Campion (1982) summarize research on employment interviews and report low reliability for interview-based assessments, which may actually be beneficial for employers because it motivates schools to design more informative disclosure policies as shown in my model.

but focus on the case of an uninformed receiver. They also show that some results generalize to the case of a privately informed receiver. In my companion paper Kolotilin (2014), I derive monotone comparative statics results with respect to the probability distribution of information for the case of an uninformed receiver. In contrast, in this paper I focus on the case in which the receiver does have private information, where both the results and analytical techniques are very different. Similar to this paper, Rayo and Segal (2010) assume that the receiver has a binary action choice, but they allow the sender's utility to depend not only on the action but also on information. To make the analysis tractable, they assume that the receiver's type is uniformly distributed. This assumption would make my model trivial in that the sender's expected utility would be the same under any mechanism, as follows from part 1 of Theorem 1 below. Ostrovsky and Schwarz (2010) study information disclosure in matching markets with private information. The main conceptual difference is that they study equilibrium rather than optimal information disclosure. Finally, Au (2012), Horner and Skrzypacz (2012), and Ely et al. (2014) study optimal dynamic information disclosure.

A few papers study cheap talk with a privately informed receiver. In the cheap talk version of my model, the unique equilibrium outcome involves no information revelation because the sender's utility depends only on the receiver's action and the information structure satisfies the single-crossing assumption. If either of these two assumptions fails, a fully revealing equilibrium may exist (Seidmann (1990), Watson (1996), and Olszewski (2004)). Chen (2009), de Barreda (2013), and Lai (2013) study cheap talk with an informed receiver under standard Crawford and Sobel (1982)'s assumptions. They all show that the receiver's expected utility may be not monotonic in the precision of his private information. The mechanics of these results, however, is different from that of my non-monotone comparative statics results: In cheap talk games with an uninformed receiver, the sender's and receiver's expected utilities are non-monotonic even in the precision of the sender's information (Green and Stokey (2007) and Ivanov (2010)) and in the precision of public information (Chen (2012)); but in optimal information disclosure games, the sender's and receiver's expected utilities are monotonic both in the precision of the sender's information and in the precision of public information (Blackwell (1953), Bergemann and Morris (2013), and Kolotilin (2014)).

Less related to this paper, Glazer and Rubinstein (2004), Glazer and Rubinstein (2006), and Sher (2014) study optimal persuasion mechanisms chosen by the receiver, rather than the sender, when certain exogenous constraints are imposed on the set of feasible reports of the sender conditional on the state. Similar to this paper, they use a linear programming approach and duality theory to characterize optimal mechanisms.

The rest of the paper is organized as follows. Section 2 develops a general model. Section 3 presents two examples that illustrate the quantity-quality tradeoff of the sender and non-monotone comparative statics. Section 4 analyzes the model under a fairly general information structure of the sender and receiver. This section partially characterizes the optimal mechanism and derives primitive necessary and sufficient conditions for optimality of the full revelation and no revelation mechanisms. Section 5 concludes. All proofs and technical details are relegated to the appendices.

2 Model

Consider a communication game between a female sender and a male receiver. The receiver takes a binary action: *to act* ($a = 1$) or *not to act* ($a = 0$). The sender's utility depends only on a, but the receiver's utility depends both on a and on (r, s) where components r and s denote the receiver's and sender's types, respectively. Specifically, the sender's utility is a, and the receiver's utility is $au(r, s)$ where u is continuous in s and continuously differentiable in r. (The assumption that the sender's preferences are state independent is not crucial and is relaxed in Section 4.) Before s is realized, the sender can commit to a mechanism that sends a message m to the receiver as a (stochastic) function of her type s; specifically, the sender can choose any conditional distribution $\phi(m|s)$ of m given s. With a slight abuse of notation, the joint distribution of (m, s) is denoted by $\phi(m, s)$. (Appendix B discusses modelling of information disclosure mechanisms in more detail.)

The set M of messages is $\mathbb{R}$, the set R of receiver's types is $[\underline{r}, \overline{r}] \subset \mathbb{R}$, and the set S of sender's types is $[\underline{s}, \overline{s}] \subset \mathbb{R}$. The pair (r, s) has some joint distribution. Except for the binary-signal examples of Section 3 and Appendix D, I assume that 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 timing of the communication game is as follows:

1. The sender publicly chooses a mechanism $\phi(m|s)$.
2. A triple (m, r, s) is drawn according to ϕ, F, and G.
3. The receiver observes (m, r) and takes an action a.
4. Utilities of the sender and receiver are realized.

The solution concept used is Perfect Bayesian Equilibrium. At the third stage, the receiver forms a belief about s and acts if and only if the conditional expectation $\mathbb{E}_\phi\left[u\left(r,s\right)|m,r\right]$ of u given (m,r) is at least 0. At the first stage, the sender chooses an *optimal mechanism* that maximizes her expected utility, the probability that the receiver acts. The main assumption, formally imposed later, is the single-crossing assumption: each message m induces types $r \geq r^*\left(m\right)$ to act, for some function r^*.

Hereafter, use the following definitions and conventions. All notions are in the weak sense, unless stated otherwise. For example, increasing means non-decreasing and higher means not lower. Two mechanisms are *equivalent* if they result in the same probability that the receiver acts. One mechanism *dominates* another mechanism if the former results in a higher probability that the receiver acts than the latter. *The full revelation mechanism* (denoted by ϕ_{full}) is a mechanism that sends a different message for each s. *The no revelation mechanism* (denoted by ϕ_{no}) is a mechanism that sends the same message regardless of s.

3 Examples

In this section, I discuss two complementary examples. In the first example, the sender's and receiver's types are binary and the receiver's type is a noisy signal about the state, known by the sender. In the second example, the sender's and receiver's types are continuous and the receiver's type is independent of the sender's type. For these examples, I derive the optimal mechanism and illustrate the sender's quantity-quality tradeoff. Further, I show that the sender's and receiver's expected utilities are non-monotonic in information. Finally, I discuss what determines the form of an optimal mechanism and how much of information is optimally disclosed.

3.1 Binary Example

In this example, the sender is perfectly informed, but the receiver is partially informed. That is, the sender knows the receiver's utility from acting, but the receiver only gets a signal about his utility. Specifically, the receiver's utility from acting is equal to the sender's type s that takes two values: $s = 1$ with probability $1/5$ and $s = -1$ with probability $4/5$. The receiver's type (equivalently signal) r also takes two values $r = 1$ and $r = -1$ according to the following conditional probabilities:

$$\Pr\left(r = 1|s = 1\right) = \Pr\left(r = -1|s = -1\right) = p.$$

The parameter p captures the precision of the receiver's private signal. Without loss of generality, assume that $p \in [1/2, 1]$. For a given mechanism, the receiver $r = 1$ assigns a higher probability that s is 1, than the receiver $r = -1$. Moreover, the difference in their assessments of the probability that s is 1 increases with p. Thus, p can be alternatively viewed as the measure of *polarization* between the *optimistic receiver* ($r = 1$) and the *pessimistic receiver* ($r = -1$).

In the school-employer application, $s = 1$ and $s = -1$ correspond to good and bad students, p to the quality of an employment interview, $r = 1$ and $r = -1$ to good and bad interview outcomes, a to the hiring decision, and ϕ to the grading system that describes how students' performance is measured.

A message m under a mechanism ϕ generates a posterior probability $\Pr_\phi(s|m)$ of s given m for each value s. Upon receiving m, the receiver r acts if

$$\begin{aligned}
& \Pr(r|s=1)\Pr{}_\phi(s=1|m) - \Pr(r|s=-1)\Pr{}_\phi(s=-1|m) \\
= \ & \Pr(r|s=1)\Pr{}_\phi(s=1|m) - \Pr(r|s=-1)(1-\Pr{}_\phi(s=1|m)) \geq 0.
\end{aligned}$$

Thus, the optimistic receiver acts if $\Pr_\phi(s=1|m) \geq 1-p$, and the pessimistic receiver acts if $\Pr_\phi(s=1|m) \geq p$. Clearly, if m induces the pessimistic receiver to act, it also induces the optimistic receiver to act. Thus, using a similar argument to the revelation principle, we can restrict attention to mechanisms with three messages: (i) m_0 that induces the receiver not to act regardless of his signal ($\Pr_\phi(s=1|m_0) \in [0, 1-p)$), (ii) m_1 that induces only the optimistic receiver to act ($\Pr_\phi(s=1|m_1) \in [1-p, p)$), and (iii) m_2 that induces the receiver to act regardless of his signal ($\Pr_\phi(s=1|m_2) \in [p, 1]$). Because the sender's expected utility is equal to the probability that the receiver acts, she would strictly prefer to send m_2 over m_1 and m_1 over m_0 if there were no constraints on how often she can send various messages.

The prior distribution of s, however, imposes a constraint on how often the sender can send various messages:

$$\sum_{i=0}^{2} \Pr{}_\phi(s=1|m_i)\Pr{}_\phi(m_i) = \Pr(s=1) = \frac{1}{5}, \tag{1}$$

where $\Pr_\phi(m_i)$ denotes the probability that m_i is sent under a mechanism ϕ. Constraint (1) implies that to maximize the probability of the messages m_2 and m_1, the sender should choose a mechanism that satisfies: $\Pr_\phi(s=1|m_0) = 0$, $\Pr_\phi(s=1|m_1) = 1-p$, and $\Pr_\phi(s=1|m_2) = p$.[2] That is, m_0 gives the maximal possible evidence against acting; m_1 gives the minimal

[2]Formally, the optimal mechanism is derived in Appendix D for a setting that nests this example.

possible evidence to make the optimistic receiver act; and m_2 gives the minimal possible evidence to make the pessimistic receiver act. These observations imply that the sender's expected utility simplifies to:[3]

$$2p(1-p)\Pr(m_1) + \Pr(m_2), \tag{2}$$

and constraint (1) simplifies to:

$$(1-p)\Pr(m_1) + p\Pr(m_2) = \frac{1}{5}. \tag{3}$$

The sender's problem of finding the optimal mechanism can be viewed as a problem of maximizing the linear utility function (2) over probabilities $\Pr(m_0)$, $\Pr(m_1)$, and $\Pr(m_2)$ subject to the budget constraint (3). That is, the marginal utilities of the messages m_0, m_1, and m_2 are 0, $2p(1-p)$, and 1; and the prices of these messages are 0, $1-p$, and p. Thus, the sender faces a quantity-quality tradeoff: to send m_1 with a high probability and persuade only the optimistic receiver or to send m_2 with a small probability and persuade both the pessimistic and optimistic receivers. This tradeoff is resolved by a choice of a mechanism that sends messages with the highest *marginal utility-price ratio*. Before discussing the optimal mechanism in a greater detail, I highlight non-monotone comparative statics.

Figure 1 shows the sender's and receiver's expected utilities under the optimal mechanism. Naive intuition may suggest that (i) the sender's expected utility should decrease with p because it is harder to influence a better informed receiver and (ii) the receiver's expected utility should increase with p because a better informed receiver takes a more appropriate action. This naive intuition, however, does not take into account that the optimal mechanism changes with p, and the sender may choose to disclose significantly less information if the receiver is more informed. This effect may overturn the results. In fact, the sender's expected utility strictly increases with p for $p \in \left(1/\sqrt{2}, 4/5\right)$, and the receiver's expected utility jumps down to zero as p exceeds $1/\sqrt{2}$.[4] Thus, a more informative employment interview may help

[3]Equation (2) is obtained using the fact that m_2 induces the receiver to act with probability 1, m_1 induces the receiver to act with probability

$$\Pr(r=1|s=1)\Pr_\phi(s=1|m_1) + \Pr(r=1|s=-1)\Pr_\phi(s=-1|m_1) = p(1-p) + (1-p)p,$$

and m_0 induces the receiver to act with probability 0.

[4]Consistent with the naive intuition, even in a general model, the sender's expected utility decreases and the receiver's expected utility increases with the precision of the receiver's information if this precision is either low or high. Indeed, the sender is best off and the receiver is worst off when the receiver is uniformed; and the sender is worst off and the receiver is best off when the receiver is perfectly informed.

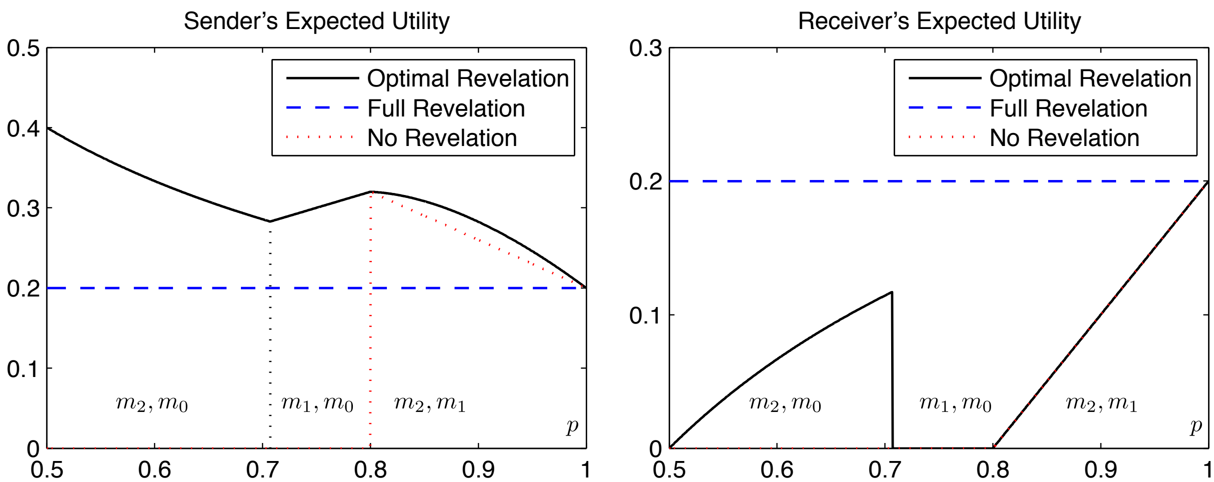

Figure 1: The sender's and receiver's expected utilities in the binary example.

the school in influencing the employer's decision and may hurt the employer in making him hire worse students on average.

I stress that these non-monotone comparative statics results with respect to the precision of information arise only when the receiver is privately informed. If the receiver's signal was public, then the sender's and receiver's expected utilities would be monotonic both in the precision of the sender's private information and in the precision of public information (Kolotilin (2014)).

Figure 1 also sheds light on the extent to which information disclosure can affect the receiver's action and on the informativeness of the optimal mechanism. As the left panel shows, for a wide range of p, the probability that the receiver acts is considerably higher under the optimal mechanism than under the two benchmark mechanisms: the full revelation and no revelation mechanisms. As the right panel shows, from the receiver's perspective, the optimal mechanism is maximally uninformative if $p = 1/2$ or $p \in \left[1/\sqrt{2}, 1\right]$, and its informativeness gradually increases with p for $p \in \left(1/2, 1/\sqrt{2}\right)$. I now explain the three forms that the optimal mechanism can take as p increases from $1/2$ to 1.

First, if the receiver's signal is imprecise in that p is close to $1/2$, then it is almost as cheap to persuade the pessimistic receiver to act as it is to persuade the optimistic receiver to act, because the prices p and $1-p$ are close. Thus, the sender prefers to target the pessimistic receiver, so the optimal mechanism sends the messages m_2 and m_0. As p increases, it becomes harder to persuade the pessimistic receiver to act and, thus, sending m_2 becomes more expensive. As a result, the sender's expected utility decreases with p. Since the optimal mechanism gives no rent to the pessimistic receiver, the optimistic receiver gets a strictly positive rent, which increases with p.

Second, as p exceeds $1/\sqrt{2}$ (but falls behind $4/5$), the polarization between the optimistic and pessimistic receivers becomes so high that it becomes much more expensive to persuade the pessimistic receiver to act than to persuade the optimistic receiver to act. Thus, the sender prefers to target the optimistic receiver, so the optimal mechanism sends the messages m_1 and m_0. In other words, the sender switches from the more expensive and more persuasive message m_2 to the less expensive and less persuasive message m_1. As p increases, the price $1-p$ of sending m_1 decreases and it becomes easier to persuade the optimistic receiver to act. As a result, the sender's expected utility increases with p. The receiver's expected utility jumps down to 0 as p exceeds $1/\sqrt{2}$, and it stays at 0 as p increases, because the optimal mechanism makes the receiver indifferent to act whenever he acts.

Third, as p exceeds $4/5$, the receiver's signal becomes so precise that the sender can persuade the optimistic receiver to act by disclosing no information. Thus, the sender prefers to target the optimistic receiver with certainty and the pessimistic receiver with some probability, so the optimal mechanism sends m_1 and m_2. As p increases further, the sender can persuade the pessimistic receiver to act more often, so the optimal mechanism sends m_2 with a higher probability. But the probability of the receiver being optimistic decreases, so m_1 induces the receiver to act with a lower probability. In this example, the latter effect dominates the former, so the sender's expected utility decreases with p. The receiver's expected utility increases with p because the optimal mechanism gives no rent to the receiver, so a better informed receiver takes a more appropriate action.

Continuing the school-employer application, when the employment interview is not too informative ($p < 4/5$), the optimal grading policy exhibits grade inflation: a good transcript is given to some good and some bad students whereas a bad transcript is given only to bad students. Moreover, when the employment interview is not very informative ($p < 1/\sqrt{2}$), the optimal policy has low grade inflation such that a good transcript impresses all employers; but when the employment interview is moderately informative ($1/\sqrt{2} < p < 4/5$), the optimal policy has high grade inflation such that a good trascript persuades only employers with a positive interview experience. Finally, when the employment interview is too informative ($p > 4/5$), the school optimally uses a noisy grading policy: a good student has a higher chance of getting a good trascript than a bad student does, but there is a chance that a good student gets a bad transcript and a bad student gets a good transcript.

The sender's quantity-quality tradeoff illustrated in this example carries on to a general version of the model. If the sender's signal is binary, this tradeoff is resolved by the choice of messages with the highest marginal utility-price ratio (Section 4 and Appendix D), otherwise the tradeoff becomes more intricate because the budget constraint becomes multidimensional (Sections 3.2 and 4).

3.2 Continuous Example

In this example, the receiver's utility is additive in sender's and receiver's types that are continuous and independent of each other. Specifically, $u(r,s) = s - r$ where s and r are independently distributed with distributions F and G. The supports are such that the receiver $\underline{r}$ always acts ($\underline{r} < \underline{s}$) and the receiver $\overline{r}$ never acts ($\overline{r} > \overline{s}$). For example, s may correspond to the student's ability privately known by the school, and r to the opportunity cost from hiring, privately known by the employer. A message m under a mechanism ϕ induces the receiver to act if and only if $r \leq \mathbb{E}_{\phi}[s|m]$; so without affecting the sender's expected utility we can merge all messages that have the same $\mathbb{E}_{\phi}[s|m]$ into one message $\widetilde{m} = \mathbb{E}_{\phi}[s|m]$. Therefore, without loss of generality, we can focus on mechanisms ϕ for which each message m induces the receiver to act if and only if $r \leq m$.

Proposition 1 simplifies the sender's problem of finding an optimal mechanism to a problem of finding an optimal distribution of messages.

Proposition 1 *Let H denote the marginal distribution of m under the optimal mechanism. Then H*

$$\begin{array}{c} \textit{maximizes } \int_{\underline{r}}^{\overline{r}} G(m)\, dH(m) \\ \textit{subject to } F \textit{ is a mean-preserving spread of } H. \end{array} \tag{4}$$

The objective function in (4) corresponds to the probability that the receiver acts; the constraint in (4) describes the set of feasible distributions of m. The intuition for the constraint is as follows. If F is a mean-preserving spread of H, then F is more informative about the underlying (hypothetical) state than H (Blackwell (1953)). Since the sender has full commitment, she can garble her information to achieve any less informative distribution H than her prior F. If she then fully reveals this garbled information to the receiver, then the distribution of m will be H. Conversely, because the sender can only garble her information, F must be a mean-preserving spread of H for any feasible mechanism.

This example is more general than it may seem. First, the example includes the case in which $u(r,s) = b(r)c(s) + d(r)$ for some functions b, c, and d where b is positive and all functions satisfy certain regularity conditions. Indeed, the receiver acts whenever $-d(r)/b(r) \leq \mathbb{E}_{\phi}[c(s)|m]$, so redefining the receiver's type as $-d(r)/b(r)$ and the sender's type as $c(s)$ gives the required result.

Second, the example includes the special case of Kamenica and Gentzkow (2011) in which "the sender's payoff depends only on the expected state". Specifically, suppose that the receiver does not have private information and the set of receiver's actions is R rather than $\{0, 1\}$. Let the receiver's optimal action r^* given m depend only on $\mathbb{E}_\phi [s|m]$, rather than on the entire distribution of s given m; and let the sender's utility u_S depend only on the receiver's action. Again, we can focus on mechanisms for which $m = \mathbb{E}_\phi [s|m]$; so the sender's expected utility under m is $G(m) \equiv u_S(r^*(m))$.[5] Kamenica and Gentzkow (2011) note that all feasilbe mechanisms satisfy $\mathbb{E}[\mathbb{E}_\phi [s|m]] = \mathbb{E}[s]$ but not all mechanisms that satisfy the equality are feasible; Proposition 1 gives a complete characterization of the set of feasible mechansims: a mechanism ϕ is feasible if and only if the prior distribution of s is a mean-preserving spread of the distribution of $\mathbb{E}_\phi [s|m]$.

Third, the example includes the setting of Ostrovsky and Schwarz (2010) who study information disclosure in matching markets. Specifically, a student with ability s receives a transcript m according to a distribution $\phi(m|s)$ and then he is matched to an employer of quality $G(\mathbb{E}_\phi [s|m])$. The main difference is that in Ostrovsky and Schwarz (2010), the function G is endogenous rather than exogenous.

Proposition 1 suggests that the sender faces a similar tradeoff to that of the binary example. The sender's marginal utility from sending m is $G(m)$. But besides requiring the expectation of m and s to be the same, the budget constraint also requires the distribution of m to be less variable than the prior distribution of s. A corollary of this proposition is that the curvature of G determines the form of the optimal mechanism.

Corollary 1 *In this example:*

1. *All mechanisms are equivalent if and only if G is linear on S.*

2. *ϕ_{full} is optimal if and only if G is convex on S.*

3. *ϕ_{no} is optimal if and only if the concave closure $\mathbf{G}$ of G on S is equal to G at $r_{no} \equiv \mathbb{E}_F [s]$ in that*

$$G(r) - G(r_{no}) \leq g(r_{no})(r - r_{no}) \text{ for all } r \in S.^6 \quad (5)$$

All three parts of Corollary 1 are straightforward because the optimal mechanism is the solution to problem (4). First, if G is linear, then the sender is risk neutral, so all mechanisms

[5] G, being a distribution function, is increasing, but all results of this section continue to hold if G is nonmonotonic; only convexity properties of G matter.

[6] 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 (see Figure 2).

are equivalent. Second, if G is convex, then the sender is risk loving, so the full revelation mechanism is optimal. Third, if G is concave, then the sender is risk averse, so the no revelation mechanism is optimal.[7]

Note that the optimal mechanism may be very sensitive to primitives of the model. For example, if G is almost uniform but strictly convex, then ϕ_{full} is uniquely optimal. However, if G is almost uniform but concave, then ϕ_{no} is uniquely optimal. This observation gives an explanation for why many similar-looking schools may choose very different disclosure policies regarding what information (if any) to report on transcripts (grading scale, class rank, distinctions).

I now discuss comparative statics in this example. By Proposition 1, as F becomes more informative in the mean-preserving spread sense, the set of feasible mechanisms expands, so the sender's expected utility increases.[8] That is, an additional information about a student can only help a school if the school can commit to a disclosure policy in advance. Moreover, Proposition 1 implies that as the sender's and receiver's priors become more favorable for acting (F increases and G decreases in the first-order stochastic dominance sense), the sender's expected utility increases. That is, a school should find it easier to place its students if their ability is higher and if employment opportunities are abundant. These monotone comparative statics results are similar in spirit to the results in Kolotilin (2014).

However, similarly to the binary example, the sender's and receiver's expected utilities are not monotonic in the precision of the receiver's private information. In particular, the sender's expected utility may decrease as the receiver's private information G becomes more precise in the mean-preserving spread sense. To see this, consider F that puts probability one on some s and note that the sender's expected utility $G(s)$ changes ambiguously. Moreover, the receiver's expected utility may decrease with the precision of his private information. To see this, suppose that G_1 is almost uniform but convex, and G_2 is concave and slightly more informative than G_1. By Corollary 1, ϕ_{full} is optimal under G_1 and ϕ_{no} is optimal under G_2. Thus, from the receiver's perspective a small gain from having more precise private information under G_2 is outweighed by a large loss from getting less precise information from the sender.

Finally, I discuss possible forms of an optimal mechanism under the assumption that

[7] Corollary 1 can alternatively be derived using tools developed in Kamenica and Gentzkow (2011). Moreover, Corollary 1 is similar to the results obtained in Section VIII B of Rayo and Segal (2010).

[8] In the two extreme cases, if F were to put probability one on some s, then the only feasible H would put probability one on $m = s$, but if F were to put strictly positive probabilities only on $\underline{s}$ and $\overline{s}$, then any H supported on S with $\mathbb{E}_H[m] = \mathbb{E}_F[s]$ would be feasible.

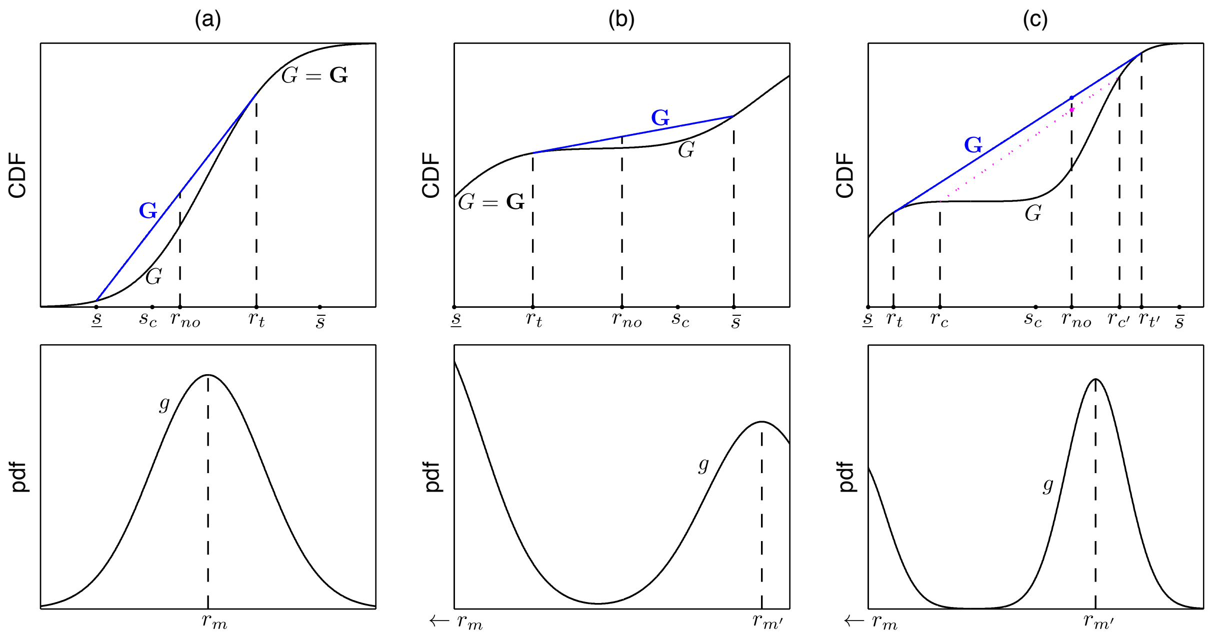

Figure 2: The optimal mechanism in the continuous example.

the distribution G is either unimodal or bimodal (see Appendix C for formal results and corresponding proofs based on duality theory introduced in Secion 4). A distribution G is *unimodal* if its density g has a unique local (and therefore global) maximum at $r_m \in R$; the maximum point r_m is called a *mode*. If the mode $r_m \notin S$, then G is either convex or concave on S; so either ϕ_{full} or ϕ_{no} is optimal by Corollary 1. Similarly, if $r_m \in S$ and $r_{no} \geq r_t$ (see Figure 2 (a)); then $\mathbf{G}(r_{no}) = G(r_{no})$; so ϕ_{no} is still optimal.

Consider the remaining case of unimodal G in which $r_m \in S$ and $r_{no} < r_t$. If F were to put 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 receiver would act with probability $\mathbf{G}(r_{no})$. This mechanism, however, is not feasible when F admits a density because s is equal to $\underline{s}$ with probability 0. Thus, the optimal mechanism reveals s for $s < s_c$ and sends the same message for all $s > s_c$ where the cutoff s_c is such that the sender is indifferent between revealing s_c or pooling it with $s > s_c$.[9]

I now interpret these results with the school-employer application. Suppose that the distribution of the employer's opportunity cost of hiring is unimodal. If the students' ability is

[9] In the extreme case when G is a step function with $G(r) = 0$ for $r < r_t$ and $G(r) = 1$ for $r \geq r_t$, the optimal mechanism reveals s for $s < s_c$ and sends the same message r_t for $s > s_c$ where s_c solves $\mathbb{E}[s|s \geq s_c] = r_t$. This is the case of uninformed receiver ($r = r_t$ with probability 1) studied in Kolotilin (2014).

high or the employment opportunities are abundant, then the school should optimally reveal no information about its students, in which case most students will find a job. Otherwise the school should fully separate bad students but pool good and very good students together. Under this disclosure policy, bad students will find it progressively harder to find a job, but all good students will find it equally easy to find a job. As the students' ability deteriorates or the employment opportunities become scarce, the school should optimally reveal more information about its students by using stricter grading policy (smaller set of good students should be pooled together).

The distribution G is *bimodal* if its density g has two local maxima at $r_m, r'_m \in R$. If the modes $r_m, r'_m \notin S$, then g has a unique minimum on S and the analysis is symmetric to that in the case of a unimodal distribution. For example, in the most interesting case (the only case with partial information disclosure) of $r_m < \underline{s} < r_t < r_{no} < \overline{s} < r'_m$ (see Figure 2 (b)), the optimal mechanism reveals s for $s > s_c$ and sends the same message for all $s < s_c$.

Finally, consider the case in which the modes r_m and r'_m satisfy $r_m < \underline{s} < r'_m < \overline{s}$ (see Figure 2 (c)).[10] If $r_{no} \notin (r_t, r'_t)$, then ϕ_{no} is optimal by Corollary 1. If $r_{no} \in (r_t, r'_t)$, then the optimal mechanism takes one of the three forms depending on the distribution F. Informally, the three cases correspond to the degree of the "variability" of s under F. Recall that Proposition 1 implies that the sender's expected utility increases as the distribution F of s becomes more spread out. First, if the variability of s is high, then the optimal mechanism sends the two messages r_t and r'_t and the receiver acts with probability $\mathbf{G}(r_{no})$ (the sender's types for which each of the two messages are sent are not uniquely determined). Second, if the variability of s is medium, then the optimal mechanism sends a message $r_c > r_t$ for $s < s_c$ and a message $r'_c < r'_t$ for $s > s_c$, where the cutoff s_c is such that the sender is indifferent between pooling s_c with $s < s_c$ or with $s > s_c$. Third (this case is not illustrated on Figure 2 (c)), if the variability of s is low, then the optimal mechanism sends a message $r_f > r_t$ for $s < s_f$, reveals s for $s \in (s_f, s'_f)$, and sends a message $r'_f < r'_t$ for $s > s'_f$, where the cutoffs s_f and s'_f are such that the sender is indifferent between pooling and revealing.[11]

[10] It is straightforward to characterize the optimal mechanism under other cases of bimodal and multimodal distributions G, but it involves consideration of many cases and does not generate novel insights.

[11] A related work of Ginzburg (2013) studies the same continuous example as in this section. He finds that in the optimal mechanism, the types of the sender that are not completely revealed are pooled into just one message. Unfortunately, this result is incorrect as this paragraph shows: low types of the sender are pooled into one message and high types of the sender are pooled into another message.

4 General Case

This section presents a linear programming approach to general problems of optimal information disclosure. The key assumption maintained throughout this section is that the receiver with a higher type is always more willing to act.[12] If the sender's type is binary, then, similarly to the binary example of Section 3.1, the optimal mechanism maximizes a linear utility function subject to a linear budget constraint. However, if the sender's type is not binary, then the budget constraint becomes multidimensional and it becomes hard to solve for an optimal mechanism directly. Nevertheless, an optimal mechanism always solves a linear program; so duality theory applies. Duality theory gives a relatively simple solution to the reverse problem of finding necessary and sufficient conditions on the primitives of the model that ensure that a candidate mechanism is optimal. That is, we can derive conditions of an environment under which an actual disclosure policy chosen by a given school is optimal.[13] In the continuous example of Section 3.2, I use duality theory to find an optimal mechanism by a guess and verify method. In Section 4.2, I use duality theory to characterize general necessary and sufficient conditions for optimality of the two most important mechanisms: the full revelation and no revelation mechanisms.

4.1 Characterization of Optimal Mechanism

Section 4 maintains all assumptions imposed in Section 2 except for the state-independent preferences for the sender. Instead, I allow the sender's utility from acting to be any continuous function w of r and s (the utility from not acting is still normalized to 0). Under mild regularity conditions, all results of this section could be extended to the case of multidimensional types of the sender in that S can be a closed cube in $\mathbb{R}^n$. But, for simplicity, I maintain the assumption that $S = [\underline{s}, \overline{s}]$. The key assumption imposed in this section is the following *single-crossing assumption*:

Assumption 1 (single-crossing) *The function $v_H(r) \equiv \int_S \widetilde{u}(r,s)\, dH(s)$ crosses the horizontal axis once and from below for all distributions H on S where $\widetilde{u}(r,s) \equiv u(r,s)\, g(r|s)$. Moreover, the function $r^*(s)$ is continuous and strictly decreasing in s where $r^*(s)$ is the unique r that solves $u(r,s) = 0$.*

[12] In the binary example, the optimistic receiver is more willing to act. In the continuous example, the receiver with a lower opportunity cost (a higher type $-r$) is more willing to act.

[13] Since all mechanisms are equivalent in the continuous example from Section 3.2 if G is linear, we know that every disclosure policy is optimal under certain conditions.

The single-crossing assumption allows us to restrict attention to mechanisms ϕ in which a message m induces the receiver to act if and only if $r \geq m$ (Lemma 1).[14] Therefore, the utility of the sender s from message r is $c(r,s) \equiv \int_r^{\overline{r}} w(\widetilde{r},s)\, g(\widetilde{r}|s)\, d\widetilde{r}$, where c is continuous in s and continuously differentiable in r. In particular, if the sender's preferences are state-independent as in Section 3, then $c(r,s) = 1 - G(r|s)$.

Lemma 1 *For each mechanism $\widetilde{\phi}$, there is a mechanism ϕ that induces the same mapping from (r,s) to the receiver's action a and each message m sent by ϕ induces the receiver to act if and only if $r \geq m$.*

The essence of the single-crossing assumption is that $v_H(r)$ crosses the horizontal axis at most once and in the same direction, the remaining requirements are just technical conditions.[15] The continuous example satisfies the single-crossing assumption and the binary example satisfies this weak version of the single-crossing assumption.[16] To illustrate broad applicability of this assumption, Proposition 2 provides an alternate representation and primitive sufficient conditions for the weak version of the single-crossing assumption.

Proposition 2 *Let all assumptions imposed in Section 2 hold.*

1. *The function $v_H(r)$ crosses the horizontal axis at most once and from below for all distributions H if and only if for each $r_2 \geq r_1$ there exists a constant $b \geq 0$ such that $\widetilde{u}(r_2,s) \geq b\widetilde{u}(r_1,s)$ for all s.*

2. *If $u(r,s)$ is increasing in both r and s, and the density $g(r|s)$ has the monotone likelihood ratio property in that $g(r_2|s_2)\, g(r_1|s_1) - g(r_2|s_1)\, g(r_1|s_2) \geq 0$ for all $s_2 \geq s_1$ and $r_2 \geq r_1$, then $v_H(r)$ crosses the horizontal axis at most once and from below for all distributions H.*

[14] Without the single-crossing assumption, a message could induce an arbitrary set of receiver's types to act.

[15] Indeed, extending $u(r,s)$ to $\widetilde{R} \supset R$ for all s and making $g(r|s)$ infinitesimally small for all s and $r \notin R$ yields that $v_H(r)$ crosses the horizontal axis exactly once on $\widetilde{R}$, not just at most once. Reordering R yields that $v_H(r)$ crosses the horizontal axis from below, not just in the same direction. Considering H that puts probability one on s yields that $u(r,s)$ crosses the horizontal axis once for all s, so $r^*(s)$ is well defined. Finally, reordering S yields that $r^*(s)$ is decreasing.

[16] To see this, note that for the continuous example (after replacing r with $-r$), $v_H(r) = \int_S (s+r)\, g(r)\, dH(s) \geq 0$ if and only if $r \geq -\mathbb{E}_H[s]$ and for the binary example, $v_H(-1) = (1-p)\Pr_H(s=1) - p\Pr_H(s=-1) \geq 0$ implies $v_H(1) = p\Pr_H(s=1) - (1-p)\Pr_H(s=-1) \geq 0$ because $p \geq 1/2$. The reader interested in an example that does not satisfy even the weak version of the single-crossing assumption is referred to Appendix D.

Before turning to the general problem where both the sender's and receiver's types are continuous, it is instructive to consider the case where the receiver's type is continuous but the sender's type is binary in that $G(r|s)$ admits a density $g(r|s)$ but F is supported on $\underline{s}$ and $\overline{s}$. For all $r \in [r^*(\overline{s}), r^*(\underline{s})]$, let $p(r)$ denote the probability of $\overline{s}$ at which the receiver r is indifferent to act. In the optimal mechanism, the distribution H of messages

$$\text{maximizes} \int_{r^*(\overline{s})}^{r^*(\underline{s})} \mathbb{E}\left[c(m,s)|m\right] dH(m)$$

$$\text{subject to} \int_{r^*(\overline{s})}^{r^*(\underline{s})} p(m)\, dH(m) = \Pr(\overline{s}).$$ [17]

The objective function is the sender's expected utility and the constraint is the feasibility constraint that requires that posterior probabilities $\Pr(\overline{s}|m)$ average out to the prior probability $\Pr(\overline{s})$. Again, the objective function can be interpreted as a linear utility function and the constraint as a Bayesian budget constraint. As a result, the sender faces the same quantity-quality tradeoff as in the binary example of Section 3.1. To resolve this tradeoff, the optimal mechanism sends at most two messages with the highest marginal utility-price ratio $\mathbb{E}\left[c(m,s)|m\right]/p(m)$.[18]

In general (if both the sender's and receiver's types are continuous), the optimal mechanism is a distribution ϕ that

$$\text{maximizes} \int_{R\times S} c(r,s)\, d\phi(r,s) \tag{6}$$

$$\text{subject to} \int_{R\times \widetilde{S}} d\phi(r,s) = \int_{\widetilde{S}} f(s)\, ds \text{ for any measurable set } \widetilde{S}\subset S, \tag{7}$$

$$\int_{\widetilde{R}\times S} \widetilde{u}(r,s)\, d\phi(r,s) = 0 \text{ for any measurable set } \widetilde{R}\subset R. \tag{8}$$

The objective function is the sender's expected utility under a mechanism ϕ. The first constraint (7) is the requirement that the marginal distribution of s for ϕ is F. Intuitively, (7) is a multidimensional Bayesian budget constraint. The second constraint (8) is the requirement that a message r makes the receiver r indifferent to act. Intuitively, (8) determines multidimensional prices of various messages.

[17] Explicitly, $p(m) = \widetilde{u}(m,\underline{s})/(\widetilde{u}(m,\underline{s}) - \widetilde{u}(m,\overline{s}))$ and $\mathbb{E}\left[c(m,s)|m\right] = p(m)\, c(m,\overline{s}) + (1-p(m))\, c(m,\underline{s})$.

[18] The optimal mechanism is a solution to a linear program, so it is an extreme point of the constraint set. If s is binary, then the constraint is one dimensional, so the optimal mechanism sends at most two messages.

A general problem of information disclosure with uninformed receiver studied in Kamenica and Gentzkow (2011) can be described by (6)-(8) under mild regularity conditions; so all results of Section 4 apply to that setting as well. To see this, suppose that the receiver does not have private information and the set of receiver's actions is R rather than $\{0,1\}$. The sender's utility is $c(r,s)$ and the receiver's marginal utility of r is $-\widetilde{u}(r,s)$. The single-crossing assumption ensures that there is a unique and interior optimal action of the receiver for any posterior H of s; so we can restrict attention to direct mechanisms ϕ in which a message m induces the receiver to take action $r=m$. Then (6) is the sender's expected utility under ϕ; (7) is the requirement that the marginal distribution of s for ϕ is F; and (8) is the necessary and sufficient first-order condition for the receiver to be willing to follow any recommendation m of ϕ.

The problem (6) is called the *primal problem*. This primal problem is an infinite dimensional linear program, because the objective function and both constraints are linear in a probability distribution ϕ.[19] The *dual problem* is to find bounded functions η and ν that

$$\text{minimize} \int_S \eta(s) f(s)\, ds \tag{9}$$

$$\text{subject to } \eta(s) + \widetilde{u}(r,s)\nu(r) \geq c(r,s) \text{ for all } (r,s) \in R \times S. \tag{10}$$

Intuitively, the variables $\eta(s)$ and $\nu(r)$ are multipliers for constraints (7) and (8).

Say that ϕ is *feasible* for (6) if it is a distribution that satisfies (7) and (8). Similarly, say that η and ν are *feasible* for (9) if they are bounded functions that satisfy (10). Feasible ϕ and (η,ν) that solve their respective problems (6) and (9) are called *optimal solutions*.

The reader should not be concerned about how the dual problem is derived; what is important is the linkage between the primal and dual problems stated in Lemmas 2 and 3. Weak duality gives an easy way to check that candidate feasible solutions ϕ and (η,ν) are optimal:

Lemma 2 *If ϕ is feasible for (6), (η,ν) is feasible for (9), and*

$$\int_{R\times S} \left(\eta(s) + \widetilde{u}(r,s)\nu(r) - c(r,s)\right) d\phi(r,s) = 0, \tag{11}$$

then ϕ and (η,ν) are optimal solutions.

[19] The primal problem would be a finite dimensional linear program if the sets R and S were finite. I impose enough smoothness to guarnatee that standard results for finite dimensional linear programs extend to an infinite dimensional case. If R and S are finite, neither full revelation mechanism nor no revelation mechanism is generically optimal, because G and F are step functions. For this reason, I assume that the sets S and R are intervals in which case both full revelation and no revelation mechanisms can be generically optimal.

Strong duality establishes the existence of optimal solutions and shows that complementarity condition (11) is not only sufficient but also necessary for optimality of ϕ and (η, ν).

Lemma 3 *There exists an optimal mechanism ϕ, an optimal solution to the primal problem (6). There exists an optimal solution to the dual problem (9) in which η is continuous. Moreover, (11) holds for these optimal ϕ and (η, ν).*

The key benefit of this linear programming approach to problems of optimal information disclosure is that duality theory gives an easy way to solve the reverse problem (that is, to find conditions on the primitives that guarantee that a given mechanism is optimal). Specifically, by Lemmas 2 and 3, a candidate mechanism ϕ is optimal if and only if there exists (η, ν) that satisfies feasibility condition (10) and complementarity condition (11). (11) holds if and only if the integrand is zero ($\eta(s) = c(r,s) - \widetilde{u}(r,s)\,\nu(r)$) for all (r,s) in the support of ϕ. For this $\eta(s)$, we can find conditions on $\widetilde{u}$, c, and F that are equivalent to the existence of the function $\nu(r)$ that satisfies (10) for all $(r,s) \in R \times S$ (this step is known as Fourier-Motzkin elimination of $\nu(r)$). Weak (strong) duality implies that these conditions are sufficient (necessary) for ϕ to be optimal.

In the continuous example, I use this approach to check optimality of various mechanisms, which include different combinations of pooling and revelation of the sender's types. Although this approach can be used to derive necessary and sufficient conditions for optimality of any of these mechanism in a general problem, in the next section, I focus on the full revelation mechanism ϕ_{full} and the no revelation mechanism ϕ_{no}.

There are at least two reasons that make mechanisms ϕ_{no} and ϕ_{full} prominent, besides their widespread use. First, if the sender did not have commitment power and her preferences were state-independent, then ϕ_{no} would be the unique equilibrium outcome under unverifiable information of the sender in the sense of Crawford and Sobel (1982), and ϕ_{full} would be the unique equilibrium outcome under verifiable information of the sender in the sense of Milgrom (1981).[20] The second reason is that these two mechanisms are extremal in a strong sense:

Proposition 3 *Let the single-crossing assumption hold.*

1. *The receiver's expected utility under ϕ_{no} is strictly lower than under any other ϕ.*

[20] Under unverifiable communication, if the sender sent two different messages r_1 and r_2 in equilibrium, then she would strongly prefer to send $\min\{r_1, r_2\}$ regardless of s, which leads to a contradiction. Under verifiable communication, 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 strongly prefer to reveal $\widetilde{s}$ instead.

2. *The receiver's expected utility under ϕ_{full} is strictly higher than under any other ϕ.*

A more informed receiver is better at maximizing his expected utility by taking a more appropriate action, so a weak version of Proposition 3 is immediate. The single-crossing assumption together with the smoothness assumption guarantees that the strict version of Proposition 3 holds. Note that the strict version does not hold in the binary example: the optimal mechanism is different from ϕ_{no}, yet the receiver's expected utility under the optimal mechanism is the same as under ϕ_{no} for $p = 1/2$ and for $p \geq 1/\sqrt{2}$.

4.2 Optimality of Specific Mechanisms

By definition, a mechanism ϕ is optimal if and only if it dominates all feasible mechanisms. This observation gives trivial necessary and sufficient conditions for optimality of ϕ. However, to check these conditions, one needs to compare ϕ with all feasible mechanisms, which requires a lot of comparisons. It turns out that for the optimality of ϕ, it is necessary and sufficient to check that only certain deviations from ϕ do not increase the probability that the receiver acts.

I now present the main result of this section, which shows what deviations it is necessary and sufficient to check for optimality of ϕ_{full} and ϕ_{no}. Note that ϕ_{full} sends the message $r^*(s)$ for each $s \in S$, and ϕ_{no} sends the same message r_{no} for each $s \in S$, where r_{no} is the unique r that solves $\int_S \widetilde{u}(r,s) f(s)\,ds = 0$.

Theorem 1 *Let the single-crossing assumption hold. Then:*

1. *All mechanisms are equivalent if and only if, for all $s_1, s_2 \in S$ and $r \in (r^*(s_2), r^*(s_1))$,*

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

2. *ϕ_{full} is optimal if and only if, for all $s_1, s_2 \in S$ and $r \in (r^*(s_2), r^*(s_1))$,*

$$\frac{c(r^*(s_2), s_2) - c(r, s_2)}{\widetilde{u}(r, s_2)} \geq \frac{c(r^*(s_1), s_1) - c(r, s_1)}{\widetilde{u}(r, s_1)}. \tag{13}$$

3. *ϕ_{no} is optimal if and only if, for all $s_1, s_2 \in S$ and $r \in (r^*(s_2), r^*(s_1))$,*

$$\frac{c(r,s_2)-c(r_{no},s_2)}{\widetilde{u}(r,s_2)} + \frac{\widetilde{u}(r_{no},s_2)}{\widetilde{u}(r,s_2)} \frac{\partial c(r_{no},s_{no})/\partial r}{\partial \widetilde{u}(r_{no},s_{no})/\partial r} \leq \frac{c(r,s_1)-c(r_{no},s_1)}{\widetilde{u}(r,s_1)} + \frac{\widetilde{u}(r_{no},s_1)}{\widetilde{u}(r,s_1)} \frac{\partial c(r_{no},s_{no})/\partial r}{\partial \widetilde{u}(r_{no},s_{no})/\partial r}, \tag{14}$$

where s_{no} is the unique s that solves $u(r_{no}, s) = 0$.

The proof relies on duality theory as explained in the previous section. As discussed in the proof, an interpretation of Theorem 1 is that (i) all mechanisms are equivalent if and only if the sender is indifferent to reveal s_1 and s_2 or to pool them at r for all feasible $s_1, s_2 \in S$ and $r \in R$; (ii) ϕ_{full} is optimal if and only if the sender prefers to reveal s_1 and s_2 than to pool them at r for all feasible $s_1, s_2 \in S$ and $r \in R$; (iii) and ϕ_{no} is optimal if and only if the sender prefers to pool s_1, s_2, s_3 at r_{no} than to pool s_1, s_2 at r and to reveal s_3 for all feasible $s_1, s_2 \in S$, $r \in R$, and for s_3 arbitrarily close to s_{no}.

It is straightforward to see that conditions (12)-(14) are necessary because, for optimality of a candidate mechanism, we need to check all deviations from the candidate mechanism, including those described in (12)-(14).

Before discussing the intuition for sufficiency of conditions (12)–(14), I note that these conditions are more primitive than sufficient conditions based on Kamenica and Gentzkow (2011). Following their approach, each message m generates the posterior distribution H of s, which is equal to $\phi(s|m)$. Let $r^*(H)$ be the receiver who is indifferent to act; specifically, $r^*(H)$ is the unique solution to $\int_S \widetilde{u}(r,s)\,dH(s) = 0$. Then the sender's expected utility under H is $\widehat{v}(H) \equiv \int_S c(r^*(H), s)\,dH(s)$. Kamenica and Gentzkow (2011) show that (i) all mechanisms are equivalent if $\widehat{v}$ is linear in H, (ii) ϕ_{full} is optimal if $\widehat{v}$ is convex in H, and (iii) ϕ_{no} is optimal if the concave closure of $\widehat{v}$ evaluated at the prior F is equal to $\widehat{v}(F)$. These conditions can be written as:

$$\widehat{v}(\alpha H_1 + (1-\alpha) H_2) = \alpha \widehat{v}(H_1) + (1-\alpha)\widehat{v}(H_2) \text{ for all } \alpha \in (0,1) \text{ and } H_1, H_2 \in \Delta S, \quad \text{(i)}$$

$$\widehat{v}(\alpha H_1 + (1-\alpha) H_2) \leq \alpha \widehat{v}(H_1) + (1-\alpha)\widehat{v}(H_2) \text{ for all } \alpha \in (0,1) \text{ and } H_1, H_2 \in \Delta S, \quad \text{(ii)}$$

$$\int_S \nabla\widehat{v}(F)\, d(H(s) - F(s)) \geq \widehat{v}(H) - \widehat{v}(F) \text{ for all } H \in \Delta S, \quad \text{(iii)}$$

where ΔS denotes the set of distributions on S and $\nabla\widehat{v}(F)$ denotes the gradient of $\widehat{v}$ evaluated at F. These conditions can be interpreted as (i) the sender is indifferent to separate posteriors H_1 and H_2 or to pool them at $\alpha H_1 + (1-\alpha) H_2$, (ii) the sender prefers to separate H_1 and H_2 than to pool them at $\alpha H_1 + (1-\alpha) H_2$, and (iii) the sender prefers to pool H and H_F at F than to separate them, where H_F is aribitrarily close to F. Theorem 1 shows that it is sufficient to check (i) and (ii) only for degenerate distributions H_1 and H_2 whose supports are s_1 and s_2, respectively, and to check (iii) only for discrete H whose support is $\{s_1, s_2\}$ and degenerate H_F whose support is s_3 where s_3 is arbitrarily close to s_{no}.

The intuition for sufficiency of conditions (12)–(14) relies on Lemma 4. Consider a message r of a mechanism ϕ. This message r generates a lottery $\phi(s|r)$ that makes the receiver r indifferent to act. Lemma 4 shows that this lottery can be decomposed into simpler lotteries indexed by e in such a way that (i) the support of each lottery e contains at most two elements, and (ii) each lottery e makes the receiver r indifferent to act.[21]

Lemma 4 *Let the single-crossing assumption hold. For each mechanism $\phi(r,s)$, there exists a mechanism $\varphi(m,s)$ that (i) induces the same mapping from (r,s) to the receiver's action a and (ii) has two dimensional messages $m=(r,e)\in R\times[0,1]$ such that for each $m=(r,e)$, the support of $\varphi(.|m)$ contains at most two elements of S and the receiver r is indifferent to act.*

I now discuss each sufficiency condition of Theorem 1 in turn. Suppose that the sender is indifferent to reveal s_1 and s_2 or to pool them at r for all feasible s_1, s_2, and r. By Lemma 4, without loss of generality, consider a mechanism in which each message is sent only by some two types s_1 and s_2. Since the sender is indifferent to reveal s_1 and s_2 or to pool them, this mechanism is equivalent to the mechanism that differs only in that it reveals s_1 and s_2. Sequentially modifying the mechanism for each message gives that this arbitrarily chosen mechanism is equivalent to ϕ_{full}; so part 1 follows. Rewriting (12) gives that all mechanisms are equivalent in the knife-edge case when there exists a function $b(r)$ such that, for all $s\in S$ and $r\in(r^*(\overline{s}),r^*(\underline{s}))$,

$$c(r^*(s),s)-c(r,s)=b(r)\widetilde{u}(r,s). \tag{15}$$

The model of this section nests Rayo and Segal (2010). To get tractable results, they assume that $u(r,s)=r+s$, where r and s are independent, r is uniformly distributed on $[-1,0]$, and the support of s is contained in the interval $[0,1]$. They allow the sender's utility to depend on a and s. However, if the sender's utility depends only on a, then all mechanisms are equivalent by part 1 of Theorem 1, because $c(r,s)=1-G(r)=-r$ and $r^*(s)=-s$ so that (15) holds with $b(r)=1$.[22]

I now turn to part 2 of Theorem 1. Again, without loss of generality, consider a mechanism in which each message is sent only by two types s_1 and s_2. Since the sender prefers to reveal s_1 and s_2 than to pool them, this mechanism is dominated by the mechanism that differs only in that it reveals s_1 and s_2. Sequentially modifying the mechanism for each message, we get that ϕ_{full} dominates this arbitrarily chosen mechanism; so part 2 follows.

Finally, I provide the intuition for a weaker version of part 3 of Theorem 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

[21] Golosov et al. (2014) use a similar result to construct a fully revealing equilibrium in a dynamic cheap talk game.

[22] Note that the continuous example of Section 3 has the same functional form of the receiver's utility (after replacing r with $-r$), but it does not assume that r is uniformly distributed.

all feasible s_1, s_1', s_2, s_2', r_1, r_2, then ϕ_{no} is optimal. Again, without loss of generality, consider a mechanism in which each message is sent only by two types; say $r_1 \leq r_{no}$ is sent by s_1 and s_1' and $r_2 \geq r_{no}$ is sent by s_2 and s_2'. This mechanism is dominated by the mechanism that differs only in that it sends the message r_{no} instead of r_1 and r_2. Sequentially applying this argument for pairs of messages, we get that ϕ_{no} dominates this arbitrarily chosen mechanism; so the weaker version of part 3 follows.

5 Conclusions

In this paper, I have studied optimal information disclosure mechanisms with two-sided asymmetric information. The receiver bases his action not only on the information disclosed by the sender but also on his private information. Thus, from the sender's perspective, each message results in a stochastic action by the receiver. The analysis reveals an important quantity-quality tradeoff of messages. The optimal mechanism finds a balance between quantity and quality. This balance is easiest to explain when the sender's type is binary. In this case, the prior distribution of the sender's information imposes a budget constraint on the frequencies of various messages, whereas the distribution of the receiver's information determines the sender's expected utility, which is linear in the frequencies of various messages. The optimal mechanism sends messages with the highest marginal utility-price ratio.

I also derive interesting non-monotone comparative statics results with respect to the receiver's private information for the binary-signal example in which the sender is perfectly informed but the receiver is partially informed. If the receiver's private information is either very precise or very imprecise, then the sender's expected utility decreases and the receiver's expected utility increases with the precision of the receiver's information. However, if the precision of the receiver's information is intermediate, then these results can be overturned. Surprisingly, the receiver may become worse off as his private information becomes more precise. Thus, if there is an earlier stage when the receiver can publicly choose how informed he will be, he may not want to be as informed as possible.

The paper also makes several technical contributions, which can be applied to other models of information disclosure. First, it identifies and characterizes the single-crossing assumption that is crucial for tractable results and for the quantity-quality tradeoff. Second, it provides a simple guess and verify method based on duality theory that allows to check that a candidate mechanism is optimal. Third, it provides primitive necessary and sufficient conditions for full revelation and no revelation mechanisms to be optimal.

In this paper, the receiver is not allowed to communicate with the sender. This assumption fits many real-life examples. In particular, the school uses the same grading system for all students regardless of where they apply for a job and before they get interviewed by employers. However, this assumption is not without loss of generality because the sender can potentially increase the probability that the receiver acts by conditioning a mechanism on receiver's reports about his private information. I leave the analysis of optimal two-way communication for future work.

Appendix A: Proofs

Proof of Proposition 1. For any mechanism ϕ and any message m, $m = \mathbb{E}_\phi [s|m]$, which implies that the distribution F of s is a mean-preserving spread of the distribution H of m generated by ϕ. Conversely, if F is a mean-preserving spread of H, then 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. That is, the set of distributions H generated by all feasible mechanisms is given by the constraint of (4), whereas the objective of (4) is simply the probability that the receiver acts under a mechanism that generates H. ■

Proof of Corollary 1. The proof relies on Theorem 1 from Section 4 and, therefore, should be read after Section 4. Clearly, this example satisfies the single-crossing assumption of Section 4 if r is replaced with $-r$. Thus, all results of Section 4 apply. With this change of variables, $r^*(s) = s$, $c(r, s) = G(r)$, and $\widetilde{u}(r, s) = (s - r) g(r)$.

By part 1 of Theorem 1, all mechanisms are equivalent if and only if (12):

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

By part 2 of Theorem 1, ϕ_{full} is optimal if and only if (13):

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

By part 3 of Theorem 1, ϕ_{no} is optimal if and only if (14), which simplifies to (5). ■

Proof of Lemma 1. Let $\widetilde{\phi}(\widetilde{m}, s)$ be the joint distribution of the message $\widetilde{m}$ and the sender's type s under a mechanism $\widetilde{\phi}$. There exists a version $\widetilde{\phi}(s|\widetilde{m})$ of the conditional distribution of s given $\widetilde{m}$ by Theorem 33.3 of Billingsley (1995). The receiver's expected utility given $\widetilde{m}$ and r is given by:

$$\mathbb{E}_{\widetilde{\phi}}\left[u(r,s)\,|\widetilde{m},r\right]=\frac{\int_S u(r,s)\,g(r|s)\,d\widetilde{\phi}(s|\widetilde{m})}{\int_S g(r|s)\,d\widetilde{\phi}(s|\widetilde{m})}.$$

Therefore, by the single-crossing assumption, there exists unique $r(\widetilde{m})$ such that the receiver r acts (that is, $\mathbb{E}_{\widetilde{\phi}}\left[u(r,s)\,|\widetilde{m},r\right]\geq 0$) if and only if $r\geq r(\widetilde{m})$. Let $\phi(m,s)$ be the joint distribution of $m\equiv r(\widetilde{m})$ and s under the mechanism $\widetilde{\phi}$. Since $\int_S \widetilde{u}(r(\widetilde{m}),s)\,d\widetilde{\phi}(s|\widetilde{m})=0$ for all $\widetilde{m}$ sent by $\widetilde{\phi}$, we have that for each measurable set $N\subset M$,

$$\int_{N\times S}\widetilde{u}(m,s)\,d\phi(m,s)=\int_{r^{-1}(N)\times S}\widetilde{u}(r(\widetilde{m}),s)\,d\widetilde{\phi}(\widetilde{m},s)=0,$$

where $r^{-1}(N)$ is the preimage of N. Therefore, there exists a version $\phi(s|m)$ such that $\int_S \widetilde{u}(m,s)\,d\phi(s|m)=0$ for all m; so ϕ is the required mechanism. ■

Proof of Proposition 2. The function $v_H(r)$ crosses the horizontal axis at most once and from below if and only if $v_H(r_1)\geq 0$ and $r_2\geq r_1$ imply $v_H(r_2)\geq 0$. Denote $\widetilde{u}_i(s)$ as the function $\widetilde{u}(r_i,s)$ of s. If $\widetilde{u}_2(s)\geq b\widetilde{u}_1(s)$ for all s and some constant $b\geq 0$, then $v_H(r)$ crosses the horizontal axis at most once and from below for all distributions H because $\int_S \widetilde{u}_1(s)\,dH(s)\geq 0$ and $r_2\geq r_1$ imply:

$$\int_S \widetilde{u}_2(s)\,dH(s)\geq b\int_S \widetilde{u}_1(s)\,dH(s)\geq 0.$$

Conversely, suppose that $v_H(r_1)\geq 0$ and $r_2\geq r_1$ imply $v_H(r_2)\geq 0$ and let us show that there exists $b\geq 0$ such that $\widetilde{u}_2(s)\geq b\widetilde{u}_1(s)$ for all s. This result is obvious if $\widetilde{u}_1(s)<0$ for all s. Consider now the case in which $\widetilde{u}_1(s)\geq 0$ for some s. Suppose to get a contradiction that there does not exist the required $b\geq 0$. Then the function $\widetilde{u}_2(s)$ does not belong to the closed convex cone C defined as the set of functions which can be represented as $d\widetilde{u}_1(s)+v(s)$ for some constant $d\geq 0$ and some continuous positive function $v(s)$. By the Separating Hyperplane Theorem (Corollary 5.84 of Aliprantis and Border (2006)), there exists a continuous linear functional ψ satisfying $\psi(\widetilde{u}_2)<0$ and $\psi(c)\geq 0$ for all $c\in C$. By the Riesz Representation Theorem (Theorem 6 in Section 36 of Kolmogorov and Fomin (1975)), ψ can be represented in the form $\psi(c)=\int_S c(s)\,d\Psi(s)$, where Ψ is a function of bounded variation on S. Define the function H as $H(s)=\Psi(s)/V(\Psi)$, where $V(\Psi)>0$ denotes the total variation of Ψ on S. Recall that the set C contains all positive continuous functions and $\int_S c(s)\,d\Psi(s)\geq 0$ for all $c\in C$. Applying the Dominated Convergence Theorem (Theorem 11.21 of Aliprantis and Border (2006)) to an appropriate sequence of positive continuous functions converging to the indicator function $\mathbf{1}_{[s_1,s_2]}$ yields that $\Psi(s_2) - \Psi(s_1) \geq 0$ for all $s_2 > s_1$, which in turn implies that H is a distribution function on S. Recalling that $V(\Psi) > 0$, $\widetilde{u}_1 \in C$, $\psi(\widetilde{u}_2) < 0$ and $\psi(c) \geq 0$ for all $c \in C$ yields $v_H(r_1) \geq 0$ and $v_H(r_2) < 0$, which is a contradiction.

Suppose that $v_H(r_1) \geq 0$, $r_2 \geq r_1$, and the functions u and g satisfy the suppositions of the second part. The second part follows from:

$$\begin{aligned} v_H(r_2) &= \int_S u(r_2, s)\, g(r_2|s)\, dH(s) \\ &\geq \int_S u(r_1, s)\, g(r_2|s)\, dH(s) \\ &\geq \frac{\int_S g(r_2|s)\, dH(s)}{\int_S g(r_1|s)\, dH(s)} \int_S u(r_1, s)\, g(r_1|s)\, dH(s) \\ &= \frac{\int_S g(r_2|s)\, dH(s)}{\int_S g(r_1|s)\, dH(s)} v_H(r_1) \\ &\geq 0. \end{aligned}$$

The first inequality holds because u is increasing in r. Since g has the monotone likelihood ratio, the distribution of s given r_2 first-order stochastically dominates the distribution of s given r_1 in that:

$$\frac{\int_{\underline{s}}^{\widetilde{s}} g(r_2|s)\, dH(s)}{\int_S g(r_2|s)\, dH(s)} \leq \frac{\int_{\underline{s}}^{\widetilde{s}} g(r_1|s)\, dH(s)}{\int_S g(r_1|s)\, dH(s)} \text{ for all } \widetilde{s} \in S,$$

as Milgrom (1981) shows. Thus, the second inequality holds because the function $u(r_1, s)$ is increasing in s. ■

Proof of Lemma 2. The proof of similar results can be found in Anderson and Nash (1987). However, to make the paper self-contained, I prove this lemma.

Multiplying (7) by η and integrating over S gives

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

Multiplying (8) by ν and integrating over R gives

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

Summing up these two equalities gives

$$\int_S \eta(s)\, f(s)\, ds = \int_{R\times S} (\eta(s) + \widetilde{u}(r, s)\, \nu(r))\, d\phi(r, s). \tag{16}$$

Integrating (10) over $R \times S$ gives

$$\int_{R\times S} c(r, s)\, d\phi(r, s) \leq \int_{R\times S} (\eta(s) + \widetilde{u}(r, s)\, \nu(r))\, d\phi(r, s). \tag{17}$$

Suppose that (11) holds for some feasible (η, ν) and ϕ. Then conditions (16) and (17) yield

$$\int_{R\times S} c(r,s)\, d\phi(r,s) = \int_S \eta(s) f(s)\, ds. \tag{18}$$

Consider any other feasible $\widetilde{\phi}$. Conditions (16) and (17) imply

$$\int_{R\times S} c(r,s)\, d\widetilde{\phi}(r,s) \le \int_S \eta(s) f(s)\, ds.$$

Combining this inequality with (18) gives

$$\int_{R\times S} c(r,s)\, d\widetilde{\phi}(r,s) \le \int_{R\times S} c(r,s)\, d\phi(r,s),$$

showing that ϕ is an optimal solution to the primal problem (6). An analogous argument proves that (η, ν) is an optimal solution to (9). ■

Proof of Lemma 3. The proof of this lemma is a generalization 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 $c(r,s)$ 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 (8).

Optimal solution to (6). A mechanism ϕ_{full} is feasible for the primal problem (6). 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 (7) is continuous because it is a projection; the constraint map in (8) 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 (9). Since c is continuous on the compact set $R\times S$, there exist finite values $\underline{c} = \min_{r,s} c(r,s)$ and $\overline{c} = \max_{r,s} c(r,s)$. Functions $\eta(s) = \overline{c}$ and $\nu(r) = 0$ are feasible for the dual problem. We can make the set of feasible (η,ν) bounded by adding the constraints:

$$-\frac{\overline{c}-\underline{c}}{\min_s \max_r \widetilde{u}(r,s)} \equiv \underline{K} \le \nu(r) \le \overline{K} \equiv -\frac{\overline{c}-\underline{c}}{\max_s \min_r \widetilde{u}(r,s)} \text{ for all } r,$$

$$\underline{c} \leq \eta(s) \leq \overline{N} = \overline{c} - \min_{(r,s,\nu)\in R\times S\times[\underline{K},\overline{K}]} \widetilde{u}(r,s)\,\nu \text{ for all } s.$$

To see that the value of the dual problem is not affected, note that $\overline{K} > 0$ because $\widetilde{u}(r,s) < 0$ for $r < r^*(s)$ and $\underline{K} < 0$ because $\widetilde{u}(r,s) > 0$ for $r > r^*(s)$. If $\nu(r) > \overline{K}$, then

$$\begin{aligned} \eta(s) &\geq \sup{}_r \{c(r,s) - \widetilde{u}(r,s)\,\nu(r)\} \\ &> \underline{c} - \overline{K}\min{}_r \widetilde{u}(r,s) \\ &\geq \underline{c} - \overline{K}\max{}_s \min{}_r \widetilde{u}(r,s) = \overline{c}; \end{aligned}$$

so the objective in (9) is strictly higher than under $\eta(s) = \overline{c}$ and $\nu(r) = 0$. Similarly, if $\nu(r) < \overline{K}$, then

$$\begin{aligned} \eta(s) &\geq \sup{}_r \{c(r,s) - \widetilde{u}(r,s)\,\nu(r)\} \\ &> \underline{c} - \underline{K}\max{}_r \widetilde{u}(r,s) \\ &\geq \underline{c} - \underline{K}\min{}_s \max{}_r \widetilde{u}(r,s) = \overline{c}; \end{aligned}$$

so the objective in (9) is strictly higher than under $\eta(s) = \overline{c}$ and $\nu(r) = 0$. The constraint (10) implies:

$$\begin{aligned} \eta(s) &\geq \sup{}_r \{c(r,s) - \widetilde{u}(r,s)\,\nu(r)\} \\ &\geq \underline{c} + \sup{}_r \{-\widetilde{u}(r,s)\,\nu(r)\} \geq \underline{c}. \end{aligned}$$

Finally, take any ν such that $\nu(r) \in [\underline{K}, \overline{K}]$ for all r, then (η,ν) with $\eta(s) \geq \overline{N}$ for all s is feasible for (9), because $c(r,s) - \widetilde{u}(r,s)\,\nu(r) \leq \overline{N}$ for all r and s. Thus, if (η,ν) is feasible, then (η^*,ν) with $\eta^*(s) \equiv \min\{\eta(s), \overline{N}\}$ is also feasible and the objective in (9) is smaller under η^*.

The constraint map in (10) 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 (11) under optimal solutions. As can be seen from above, the dual problem has a finite value and functions $\eta(s) = 2\overline{c}$ and $\nu(r) = 0$ are in the interior of the constraint set (10). 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

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

Indeed, η^* is feasible and $\eta^* \leq \eta$, because η satisfies (10) for all r; so the objective in (9) is smaller under η^*. We show that η^* is continuous. Since $R \times S$ is compact, c and $\widetilde{u}$ are uniformly continuous. Thus, since ν is bounded, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|(c(r,s) - \widetilde{u}(r,s)\nu(r)) - (c(r,s') - \widetilde{u}(r,s')\nu(r))| < \varepsilon \tag{19}$$

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

$$\eta^*(s) < c(r,s) - \widetilde{u}(r,s)\nu(r) + \varepsilon. \tag{20}$$

Thus,

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

where the first inequality holds by definition of η^*, the second by (19), and the third by (20). 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 3. *Part 1.* The receiver's expected utilities under any mechanism ϕ and the no revelation mechanism ϕ_{no} are:

$$\begin{aligned} \mathbb{E}_{\phi}[u] &= \int_{R\times S} \left(\int_{r}^{\overline{r}} \widetilde{u}(\widetilde{r},s)\, d\widetilde{r} \right) d\phi(r,s), \\ \mathbb{E}_{\phi_{no}}[u] &= \int_{S} \left(\int_{r_{no}}^{\overline{r}} \widetilde{u}(\widetilde{r},s)\, d\widetilde{r} \right) f(s)\, ds \\ &= \int_{R\times S} \left(\int_{r_{no}}^{\overline{r}} \widetilde{u}(\widetilde{r},s)\, d\widetilde{r} \right) d\phi(r,s). \end{aligned}$$

The first two lines hold because a message m induces the receiver r to act if and only if $r \geq m$ and ϕ_{no} sends the message r_{no} regardless of s. The third line holds because the marginal distribution of s for any mechanism ϕ coincides with the prior distribution of s. For a mechanism ϕ, denote the conditional distribution of s given a message r by $\phi(s|r)$ and the marginal distribution of a message r by $\phi(r)$. Fubini's Theorem (Theorem 11.27 of Aliprantis and Border (2006)) gives

$$\begin{aligned} \mathbb{E}_{\phi}[u] - \mathbb{E}_{\phi_{no}}[u] = &\textstyle\int_{\underline{r}}^{r_{no}} \left[\int_{r}^{r_{no}} \left(\int_S \widetilde{u}(\widetilde{r},s)\, d\phi(s|r) \right) d\widetilde{r} \right] d\phi(r) \\ &- \textstyle\int_{r_{no}}^{\overline{r}} \left[\int_{r_{no}}^{r} \left(\int_S \widetilde{u}(\widetilde{r},s)\, d\phi(s|r) \right) d\widetilde{r} \right] d\phi(r). \end{aligned} \tag{21}$$

By the single-crossing assumption, we have $\int_S \widetilde{u}(\widetilde{r},s)\, d\phi(s|r) > 0$ for $\widetilde{r} > r$. Therefore, $\int_r^{r_{no}} \left(\int_S \widetilde{u}(\widetilde{r},s)\, d\phi(s|r) \right) d\widetilde{r} > 0$ for $r < r_{no}$. Since $\phi(r)$ of any mechanism ϕ that differs from ϕ_{no} puts strictly positive probability on messages in $[\underline{r}, r_{no})$, the first integral in (21) is strictly positive. The analogous argument shows that the second integral in (21) is strictly negative; so $\mathbb{E}_{\phi}[u] - \mathbb{E}_{\phi_{no}}[u] > 0$ for any ϕ that differs from ϕ_{no}.

Part 2. The receiver's expected utility under ϕ_{full} is

$$\begin{aligned}\mathbb{E}_{\phi_{full}}[u] &= \int_S \left(\int_{r^*(s)}^{\overline{r}} \widetilde{u}(\widetilde{r}, s)\, d\widetilde{r} \right) f(s)\, ds \\ &= \int_{R\times S} \left(\int_{r^*(s)}^{\overline{r}} \widetilde{u}(\widetilde{r}, s)\, d\widetilde{r} \right) d\phi(r, s).\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}_{\phi_{full}}[u] - \mathbb{E}_{\phi}[u] = \textstyle\int_S \int_{r>r^*(s)} \left(\int_{r^*(s)}^{r} \widetilde{u}(\widetilde{r}, s)\, d\widetilde{r} \right) d\phi(r, s) \\ - \textstyle\int_S \int_{r<r^*(s)} \left(\int_{r}^{r^*(s)} \widetilde{u}(\widetilde{r}, s)\, d\widetilde{r} \right) d\phi(r, s).\end{aligned} \tag{22}$$

By the single-crossing assumption, 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 (22) is strictly positive. The analogous argument shows that the second integral in (22) is strictly negative; so $\mathbb{E}_{\phi_{full}}[u] - \mathbb{E}_{\phi}[u] > 0$ for any ϕ that differs from ϕ_{full}.
■

Proof of Theorem 1. I prove each part in turn.

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

$$\frac{c(r^*(s_2), s_2) - c(r, s_2)}{\widetilde{u}(r, s_2)} > \frac{c(r^*(s_1), s_1) - c(r, s_1)}{\widetilde{u}(r, s_1)}. \tag{23}$$

(The case in which the left hand side of (23) is strictly smaller than the right hand side is analogous.) Let $h_1(x) = \int_{s_1}^{x} \widetilde{u}(r, s) f(s)\, ds$ and $h_2(x) = \int_{x}^{s_2} \widetilde{u}(r, s) f(s)\, ds$. The functions h_1 and h_2 are continuous, strictly decreasing, and vanishing at $x = s_1$ and $x = s_2$, respectively. Thus, in a neighborhood of s_2, we can define a continuous function $s_1^*(x)$ that satisfies $h_1(s_1^*(x)) + h_2(x) = 0$. By the implicit function theorem,

$$\frac{ds_1^*(x)}{dx} = \frac{\widetilde{u}(r, x) f(x)}{\widetilde{u}(r, s_1^*(x)) f(s_1^*(x))}. \tag{24}$$

By continuity of all functions, there exists x_2 such that (23) 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 these s. That is, the former mechanism sends $r^*(s)$ and the latter sends r, because $h_1(s_1^*(x)) + h_2(x) = 0$. The two mechanisms are not equivalent, because the difference in the sender's payoff between the former and latter mechanisms is:

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

where the inequality holds by (23) and the equality holds by (24) and the change of variables formula. This concludes the proof of "only if" part of part 1.

Taking the limit $x_2 \uparrow s_2$ (and thus $s_1^*(x_1) \downarrow s_1$), this argument suggests that an interpretation of condition (23) is that the sender strictly prefers to reveal s_1 and s_2 than to pool them at r. Similarly, an interpretation of condition (12) is that the sender is indifferent to reveal s_1 and s_2 or to pool them at r.

If part of part 1. Consider any mechanism ϕ. Note that condition (12) is equivalent to (15). Substituting (15) into (8) gives

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

Taking into account (7) gives

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

which implies that the probability that the receiver acts is the same for all mechanisms.

Part 2. By Lemmas 2 and 3, ϕ_{full} is optimal if and only if there exists feasible (η, ν) that satisfies

$$\int_{R \times S} (\eta(s) + \widetilde{u}(r, s)\, \nu(r) - c(r, s))\, d\phi_{full}(r, s) = 0. \tag{25}$$

By (10), the integrand is nonnegative; so (25) holds if and only if

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

Since $\widetilde{u}(r^*(s), s) = 0$, we have $\eta(s) = c(r^*(s), s)$ almost everywhere. Since η is continuous by Lemma 3 and c and r^* are continuous by assumption, $\eta(s) = c(r^*(s), s)$ holds for all $s \in S$. Therefore, ϕ_{full} is optimal if and only if there exists ν that satisfies (10):

$$c(r^*(s), s) + \widetilde{u}(r, s)\, \nu(r) \geq c(r, s) \text{ for all } (r, s) \in R \times S, \tag{26}$$

which is equivalent to

$$\frac{c\left(r,s_{2}\right)-c\left(r^{*}\left(s_{2}\right),s_{2}\right)}{\widetilde{u}\left(r,s_{2}\right)}\leq\nu\left(r\right)\leq\frac{c\left(r^{*}\left(s_{1}\right),s_{1}\right)-c\left(r,s_{1}\right)}{-\widetilde{u}\left(r,s_{1}\right)}$$

for all $r\in\left(r^{*}\left(\overline{s}\right),r^{*}\left(\underline{s}\right)\right)$ and s_1, s_2 such that $r\in\left(r^{*}\left(s_{2}\right),r^{*}\left(s_{1}\right)\right)$. (For $r\notin\left(r^{*}\left(\overline{s}\right),r^{*}\left(\underline{s}\right)\right)$, the existence of ν is obvious because (26) bounds ν only from one side.) There exists such ν if and only if (13) holds. Note that (13) is just (23) with the strict inequality replaced by the weak one; so an interpretation of condition (13) is that the sender (weakly) prefers to reveal s_1 and s_2 than to pool them at r.

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

$$\eta\left(s\right)+\widetilde{u}\left(r_{no},s\right)\nu\left(r_{no}\right)=c\left(r_{no},s\right)\text{ for all }s\in S. \tag{27}$$

Therefore, ϕ_{no} is optimal if and only if there exists ν that satisfies (10):

$$c\left(r_{no},s\right)-\widetilde{u}\left(r_{no},s\right)\nu\left(r_{no}\right)+\widetilde{u}\left(r,s\right)\nu\left(r\right)\geq c\left(r,s\right)\text{ for all }\left(r,s\right)\in R\times S. \tag{28}$$

Since $\widetilde{u}$ is continuous and ν is bounded, inequality (28) holds if and only if:

$$\frac{c\left(r,s_{2}\right)-\left(c\left(r_{no},s_{2}\right)-\widetilde{u}\left(r_{no},s_{2}\right)\nu\left(r_{no}\right)\right)}{\widetilde{u}\left(r,s_{2}\right)}\leq\nu\left(r\right)\leq\frac{\left(c\left(r_{no},s_{1}\right)-\widetilde{u}\left(r_{no},s_{1}\right)\nu\left(r_{no}\right)\right)-c\left(r,s_{1}\right)}{-\widetilde{u}\left(r,s_{1}\right)} \tag{29}$$

for all $r\in\left(r^{*}\left(\overline{s}\right),r^{*}\left(\underline{s}\right)\right)$, and $s_1,s_2\in S$ such that $r\in\left(r^{*}\left(s_{2}\right),r^{*}\left(s_{1}\right)\right)$. (For $r\notin\left(r^{*}\left(\overline{s}\right),r^{*}\left(\underline{s}\right)\right)$, the existence of ν is obvious because (28) bounds ν only from one side.)

At $r=r_{no}$, both sides of (29) become $\nu\left(r_{no}\right)$. Thus, for (29) to be satisfied everywhere, the derivatives of both sides of (29) with respect to r evaluated at $r=r_{no}$ must coincide, which gives

$$\nu\left(r_{no}\right)=\frac{\frac{\partial c\left(r_{no},s_{1}\right)/\partial r}{\widetilde{u}\left(r_{no},s_{1}\right)}-\frac{\partial c\left(r_{no},s_{2}\right)/\partial r}{\widetilde{u}\left(r_{no},s_{2}\right)}}{\frac{\partial\widetilde{u}\left(r_{no},s_{1}\right)/\partial r}{\widetilde{u}\left(r_{no},s_{1}\right)}-\frac{\partial\widetilde{u}\left(r_{no},s_{2}\right)/\partial r}{\widetilde{u}\left(r_{no},s_{2}\right)}}. \tag{30}$$

Taking the limit of (30) as $s_2\downarrow s_{no}$ gives

$$\nu\left(r_{no}\right)=\frac{\partial c\left(r_{no},s_{no}\right)/\partial r}{\partial\widetilde{u}\left(r_{no},s_{no}\right)/\partial r}. \tag{31}$$

Substituting $\nu\left(r_{no}\right)$ from (31) into (29) completes the proof of Theorem 1.

We now interpret condition (14). Let $s_1,s_2\in S$, $r\in\left(r^{*}\left(s_{2}\right),r^{*}\left(s_{1}\right)\right)$ and s_3 be such that there exists the prior distribution that puts positive probabilities p_1, p_2, p_3 only on s_1, s_2, s_3 such that $\sum_{i=1}^{3}p_i\widetilde{u}\left(r_{no},s_i\right)=0$, and $\sum_{i=1}^{2}p_i\widetilde{u}\left(r,s_i\right)=0$. The requirement that for this prior distribution the no revelation mechanism, which sends r_{no} for s_1, s_2, s_3, dominates the mechanism that sends r for s_1, s_2 and $r^*(s_3)$ for s_3 can be written as:

$$\frac{c(r,s_2)-c(r_{no},s_2)}{\widetilde{u}(r,s_2)} + \frac{\widetilde{u}(r_{no},s_2)}{\widetilde{u}(r_{no},s_3)}\frac{c(r_{no},s_3)-c(r^*(s_3),s_3)}{\widetilde{u}(r,s_2)} \leq \frac{c(r,s_1)-c(r_{no},s_1)}{\widetilde{u}(r,s_1)} + \frac{\widetilde{u}(r_{no},s_1)}{\widetilde{u}(r_{no},s_3)}\frac{c(r_{no},s_3)-c(r^*(s_3),s_3)}{\widetilde{u}(r,s_1)}. \tag{32}$$

Taking the limit of (32) as $s_3 \to s_{no}$ gives (14), because

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

where the first equality holds by L'Hospital's rule and the second by the implicit function theorem applied to $\widetilde{u}(r^*(s), s) = 0$. Thus, an interpretation of condition (14) is that the sender prefers to pool s_1, s_2, s_3 at r_{no} than to pool s_1, s_2 at r and to reveal s_3 for s_3 arbitrarily close to s_{no}. ■

Proof of Lemma 4. Consider any $\widetilde{r}$ in the support of ϕ. For a moment assume that $\phi(s|\widetilde{r})$ admits a density. Because $\mathbb{E}_\phi[u(\widetilde{r}, s)|r = \widetilde{r}] = 0$, we can construct a decreasing function $v_1(e)$ and an increasing function $v_2(e)$ defined on $[0,1]$ such that $\Pr_\phi(s \in [v_1(e), v_2(e)]|r = \widetilde{r}) = e$ and $\mathbb{E}_\phi[u(\widetilde{r}, s)|r = \widetilde{r}, s \in [v_1(e), v_2(e)]] = 0$. If $\phi(s|\widetilde{r})$ does not admit a density, then a similar result holds but with possible randomization at the boundaries $v_1(e)$ and $v_2(e)$. Formally, there exists a quadruple function (v_1, v_2, q_1, q_2) from $R \times [0,1]$ to $[\min_{s\in S}\widetilde{u}(r,s), 0] \times [0, \max_{s\in S}\widetilde{u}(r,s)] \times [0,1] \times [0,1]$ such that

$$\begin{aligned}
&\textstyle\int_{v_1(\widetilde{r},e)<\widetilde{u}(\widetilde{r},s)<v_2(\widetilde{r},e)} \widetilde{u}(\widetilde{r}, s)\, d\phi(s|\widetilde{r}) \\
&+ \textstyle\sum_{i=1,2} v_i(\widetilde{r}, e)\, q_i(\widetilde{r}, e) \Pr{}_\phi(\widetilde{u}(\widetilde{r}, s) = v_i(\widetilde{r}, e)|r = \widetilde{r}) = 0, \\
&\Pr{}_\phi(v_1(\widetilde{r}, e) < \widetilde{u}(\widetilde{r}, s) < v_2(\widetilde{r}, e)|r = \widetilde{r}) \\
&+ \textstyle\sum_{i=1,2} q_i(\widetilde{r}, e) \Pr{}_\phi(\widetilde{u}(\widetilde{r}, s) = v_i(\widetilde{r}, e)|r = \widetilde{r}) = e
\end{aligned}$$

for all $(\widetilde{r}, e) \in R\times[0,1]$. Define distribution φ of $(\widetilde{r}, e, s)$ as follows. The marginal distribution of $\widetilde{r}$ for φ coincides with the marginal distribution of $\widetilde{r}$ for ϕ. The conditional distribution of e given $\widetilde{r}$ is uniform on the unit interval $[0,1]$. The conditional distribution of s given $\widetilde{r}$ and e puts probabilities p_1 and $1-p_1$ on s_1 and s_2, where s_1 and s_2 satisfy $\widetilde{u}(\widetilde{r}, s_1) = v_1(\widetilde{r}, e)$ and $\widetilde{u}(\widetilde{r}, s_2) = v_2(\widetilde{r}, e)$, and p_1 solves $p_1 v_1(\widetilde{r}, e) + (1-p_1) v_2(\widetilde{r}, e) = 0$. Clearly, this φ satisfies the required properties of Lemma 4. ■

Appendix B: Modelling of Information Disclosure

This appendix formalizes my modelling of information disclosure and discusses other possible alternatives.

Let all subsets of the real line be equipped with the Borel sigma algebra. Let $(\Omega, \mathcal{F}, P)$ be a (sufficiently rich) probability space and let the state v, the receiver's type r, and the sender's type s be random variables on this space. The receiver's von Neumann-Morgenstern utility is va; the conditional expectation of v given (r, s) is $\mathbb{E}[v|r,s] \equiv u(r,s)$; the conditional distribution of r given s is $G(r|s)$; and the marginal distribution of s is F.

The sender's and receiver's mixed strategies are defined using the distributional approach of Milgrom and Weber (1985). The sender's strategy is any random variable m such that (a version of) the conditional distribution of m given (v, r, s) is independent of (v, r) and the marginal distribution ϕ of (m, s) is any element (richness of the probability space is used) of the set of all probability distributions on $M \times S$ whose marginal distribution on S is F. Similarly, the receiver's strategy is any random variable a such that (a version of) the conditional distribution of a given (m, v, r, s) is independent of (v, s) and the marginal distribution of (a, m, r) is any element of the set of all probability distributions on $\{0, 1\} \times M \times R$ whose marginal distribution on $M \times R$ coincides with the marginal distribution of (m, r), given by $\int_{S\times[\underline{m},m]} G(r|s)\, d\phi(\widetilde{m}, s)$.

To simplify the discussion, in the remainder of this appendix, restrict attention to pure strategies of the sender and receiver: $m(\omega) = h_m(s(\omega))$ and $a(\omega) = h_a(m(\omega), r(\omega))$ for all $\omega \in \Omega$ where h_m and h_a are measurable functions that take values in M and $\{0, 1\}$, respectively. Equivalently, by Theorem 20.1 of Billingsley (1995), a pure strategy $m : \Omega \to M$ is $\mathcal{F}_s$-measurable random variable and a pure strategy $a : \Omega \to \{0, 1\}$ is $\mathcal{F}_{m,r}$-measurable random variable, where $\mathcal{F}_{x_1,\ldots,x_k}$ denotes the sigma algebra generated by random variables $x_1, \ldots, x_k$. In the paper, I restrict attention to pure strategies of the receiver in which he takes the sender's preferred action $a = 1$ whenever he is indifferent to act. Under certain regularity conditions, it may be possible to restrict attention to pure strategies of the sender by a purification theorem similar to Dvoretzky et al. (1951) and Ambrosio (2003). But I do not explore this possibility in the paper; instead, I study a general problem when mixed strategies are allowed.

Since the outcome of the game can depend only on joint information of the receiver and sender (r, s), the paper focuses on the reduced form game where the state v is implicit. It turns out that for any reduced form game we can construct an underlying binary state v; so the analysis cannot be simplified under the binary state case. To construct this binary state, it suffices for any utility function u and for any probability measure p of (r, s), to find

v_1, v_2, and probability measure q of (v, r, s) that satisfy

$$\begin{aligned} p(T) &= q(v_1, T) + q(v_2, T), \\ \int_T u(t)\, dp &= v_1 q(v_1, T) + v_2 q(v_2, T) \end{aligned}$$

for each measurable set $T \subset R \times S$. Take any v_1 and v_2 that satisfy $v_1 \leq u(r, s) \leq v_2$ for all $(r, s) \in R \times S$. Then $q(v_1, T) = \int_T \frac{v_2 - u(r,s)}{v_2 - v_1} dp$ and $q(v_2, T) = \int_T \frac{u(r,s) - v_1}{v_2 - v_1} dp$ constitute the required measure q.

The set of feasible mechanisms essentially corresponds to all garblings of the sender's information. For example, in the school-employer application, the school can choose any grading policy that determines how the student's performance maps to a transcript; the employer observes the chosen grading policy, the realized transcript, his private signal (for example, outside option and inteview results) and forms a Bayesian belief about the value of hiring the student. One may consider more general mechanisms whose messages conditional on the sender's information can be correlated (to some extent) with the receiver's information or more restrictive mechanisms in which certain information of the sender cannot be revealed. For example, in the first situation, the school can issue different transcripts depending on what kind of jobs the student applies to; in the second situation, the school cannot disclose information about private life of the student. Both situations can be modelled by allowing the sender to choose any $\mathcal{F}_z$-measurable random variable, where $\mathcal{F}_s \subset \mathcal{F}_z \subseteq \mathcal{F}_{v,r,s}$ in the first situation and $\mathcal{F}_z \subset \mathcal{F}_s$ in the second situation. Renaming z as s, we get the same model of information disclosure.[23] At the extreme case of $\mathcal{F}_z = \mathcal{F}_{v,r,s}$, message m can depend on v, r, and s; so r is akin to public information. Therefore, under certain regularity conditions (see Kolotilin (2014)), in the optimal mechanism, $m = m_1$ if $v \geq v^*(r)$ and $m = m_0$ otherwise, where $v^*(r)$ is the unique solution to $\mathbb{E}[v \geq v^*(r) | r] = 0$. (Since the optimal mechanism does not depend on s, the optimal mechanism remains the same if m is $\mathcal{F}_{v,r}$-measurable rather than $\mathcal{F}_{v,r,s}$-measurable.) Similarly, if $\mathcal{F}_z = \mathcal{F}_{r,s}$, then in the optimal mechanism, $m = m_1$ if $u(r, s) \geq u^*(r)$ and $m = m_0$ otherwise, where $u^*(r)$ is the unique solution to $\mathbb{E}[u(r, s) \geq u^*(r) | r] = 0$.

[23]If the set Ω is finite, then for any sigma algebra $\mathcal{G}$, there exists a random variable z with $\mathcal{F}_z = \mathcal{G}$, but if the set Ω is infinite, there exist sigma algebras $\mathcal{G}$ that are not generated by any random variable z (for example, $\Omega = [0, 1]$, $\mathcal{F}$ is the Borel sigma algebra, and $\mathcal{G}$ is the sigma algebra consisting of countable sets and sets with countable compliments). If on top of choosing any $\mathcal{F}_z$-measurable random variable m, the sender can choose a sigma algebra $\mathcal{F}_z$ from some set that contains the finest sigma algebra $\mathcal{F}_{z^*}$, which satisfies $\mathcal{F}_z \subseteq \mathcal{F}_{z^*}$ for all feasible$\mathcal{F}_z$, then it is without loss of generality to assume that the sender chooses $\mathcal{F}_{z^*}$-measurable random variable m. Under the restriction to pure strategies, Gentzkow and Kamenica (2012)'s model of multi-sender information disclosure can be viewed in this way.

More generally, one can study my communication game using the existing notions of correlated equilibrium in games with incomplete information, surveyed in Forges (1993). Strategic form correlated equilibrium corresponds to $\mathcal{F}_s$-measurable m. Bayesian solution corresponds to $\mathcal{F}_{r,s}$-measurable m. Bayes correlated equilibrium of Bergemann and Morris (2013) corresponds to $\mathcal{F}_{v,r}$-measurable m. Finally, communication equilibrium corresponds to $\mathcal{F}_{s,n}$-measurable m where n is $\mathcal{F}_r$-measurable report rule chosen by the receiver. I leave the analysis of optimal communication mechanisms for future work.

Appendix C: Results for Continuous Example

This appendix characterizes the optimal mechanism under various distributions G discussed in Section 3.2. Lemma 2 gives a technique for proving all results in this appendix. Lemma 2 implies that a candidate mechanism is optimal if there exists feasible (η, ν) for (9),

$$\eta(s) + (s - r)\nu(r) \geq G(r) \text{ for all } (r, s) \in R \times S, \tag{33}$$

such that the complementarity condition (11) holds,

$$\int_{R\times S} (\eta(s) + (s - r)\nu(r) - G(r))\, d\phi(r, s) = 0. \tag{34}$$

To prove that each of the mechanisms described in Propositions 4 and 5 is optimal, I construct (η, ν) that satisfies (33) and (34).

Proposition 4 *Let G be convex on $[\underline{s}, r_m]$, concave on $[r_m, \overline{s}]$, and $G(r_{no}) < \mathbf{G}(r_{no})$ (see Figure 2 (a)). Then the optimal mechanism reveals s for $s < s_c$ and sends the same message $r_c \equiv \mathbb{E}[s|s \geq s_c]$ for $s \geq s_c$ where $s_c \in (\underline{s}, r_m)$ is uniquely determined by*

$$G(s_c) = G(r_c) + g(r_c)(s_c - r_c).$$

Proof. We can ignore verification of conditions (33) and (34) for (r, s) such that $r \notin S$, because no mechanism can send a message $r \notin S$. Technically, condition (33) bounds ν only from one side for $r \notin S$: $\nu(r) \geq (G(r) - \eta(s)) / (s - r)$ for $r < \underline{s}$ and $\nu(r) \leq (\eta(s) - G(r)) / (r - s)$ for $r > \overline{s}$. Thus, to satisfy (33), we can set $\nu(r) = K$ for $r < \underline{s}$ and $\nu(r) = -K$ for $r > \overline{s}$ where K is sufficiently large. The values of ν for $r \notin S$ do not affect (34) because the support of ϕ is included in $S \times S$.

For $(r,s) \in S \times S$, the required (η, ν), which satisfies (33) and (34), is

$$\eta(s) = \begin{cases} G(s) \text{ for } s \in [\underline{s}, s_c], \\ G(r_c) + g(r_c)(s - r_c) \text{ for } s \in (s_c, \overline{s}], \end{cases}$$
$$\nu(r) = \begin{cases} -g(r) \text{ for } r \in [\underline{s}, s_c], \\ -g(r_c) \text{ for } r \in (s_c, \overline{s}]. \end{cases}$$

The pair (η, ν) satisfies (33) for all $(r,s) \in S \times S$ because:

$$\eta(s) \geq \eta(r) - \nu(r)(s - r) \geq G(r) - \nu(r)(s - r),$$

The first inequality holds because η is convex on S and $-\nu(r)$ is a subderivative of $\eta(r)$ for all $r \in S$. The second inequality holds because $\eta(r) \geq G(r)$ for all $r \in S$. Further, (η, ν) satisfies (34) because (33) holds with equality for (r,s) in the support of the mechanism. ■

The case when G has the form illustrated on Figure 2 (b), rather than Figure 2 (a), is completely analagous; so I omit the formal result and the corresponding proof.

Proposition 5 *Let G be concave on $[\underline{s}, r_m]$ and $[r'_m, \overline{s}]$, convex on $[r_m, r_a]$, and let $G(r_{no}) < \mathbf{G}(r_{no})$ (see Figure 2 (c)). Let $r_t, r'_t \in S$ be uniquely determined by*

$$\mathbf{G}(r_{no}) = \frac{r'_t - r_{no}}{r'_t - r_t} G(r_t) + \frac{r_{no} - r_t}{r'_t - r_t} G(r'_t)$$

and let $r_c \equiv \mathbb{E}[s|s < s_c]$ and $r'_c \equiv \mathbb{E}[s|s \geq s_c]$ where the cutoff $s_c \in S$ is uniquely determined by

$$G(r_c) + g(r_c)(s_c - r_c) = G(r'_c) + g(r'_c)(s_c - r'_c).$$

1. *If $G(r_c) + g(r_c)(s_c - r_c) \geq \mathbf{G}(s_c)$, then an optimal mechanism sends the two messages r_t and r'_t (the sets of s for which each of the two messages are sent are not uniquely determined).*

2. *If $G(r_c) \leq G(r_c) + g(r_c)(s_c - r_c) < \mathbf{G}(s_c)$, then the optimal mechanism sends the message $r_c > r_t$ for $s < s_c$ and the message $r'_c < r'_t$ for $s \geq s_c$.*

3. *If $G(r_c) + g(r_c)(s_c - r_c) > G(r_c)$, then the optimal mechanism sends the message $r_f \equiv \mathbb{E}[s|s \leq s_f] > r_t$ for $s \leq s_f$, reveals s for $s \in (s_f, s'_f)$, and sends the message $r'_f \equiv \mathbb{E}[s|s \geq s'_f] < r'_t$ for $s \geq s'_f$ where the cutoffs $s_f, s'_f \in S$ are uniquely determined by*

$$G(s_f) = G(r_f) + g(r_f)(s_f - r_f),$$
$$G(s'_f) = G(r'_f) + g(r'_f)(s'_f - r'_f).$$

Proof. Similarly to the proof of Proposition 4, the pair (η, ν) satisfies (33) for all $(r, s) \in S \times S$ if for all $r \in S$, η is convex, $-\nu(r)$ is a subderivative of $\eta(r)$, and $\eta(r) \geq G(r)$. Further, this (η, ν) satisfies (34) if (33) holds with equality for (r, s) in the support of the proposed mechanism.

Part 1. The required (η, ν), which satisfies (33) and (34), is given by

$$\begin{aligned} \eta(s) &= G(r_t) + g(r_t)(s - r_t) \text{ for all } s \in S, \\ \nu(r) &= -g(r_t) \text{ for all } r \in S. \end{aligned}$$

Finally, I show that there exists a mechanism that sends the two messages r_t and r_t'. Note that $r_c \leq r_t$ and $r_c' \geq r_t'$ because

$$G(r_c) + g(r_c)(s_c - r_c) = G(r_c') + g(r_c')(s_c - r_c') \geq \mathbf{G}(s_c).$$

Thus, there exist $s_t, s_t' \in S$ such that

$$\begin{aligned} F(s_t') - F(s_t) &= \frac{r_{no} - r_t}{r_t' - r_t}, \\ \mathbb{E}[s | s \in (s_t, s_t')] &= r_t'; \end{aligned}$$

so one required mechanism sends the message r_t for $s \in [\underline{s}, s_t] \cup [s_t', \overline{s}]$ and the message r_t' for $s \in (s_t, s_t')$.

Part 2. The required (η, ν), which satisfies (33) and (34), is given by

$$\begin{aligned} \eta(s) &= \begin{cases} G(r_c) + g(r_c)(s - r_c) \text{ for } s \in [\underline{s}, s_c], \\ G(r_c') + g(r_c')(s - r_c') \text{ for } s \in (s_c, s], \end{cases} \\ \nu(r) &= \begin{cases} -g(r_c) \text{ for } r \in [\underline{s}, s_c], \\ -g(r_c') \text{ for } r \in (s_c, \overline{s}]. \end{cases} \end{aligned}$$

Part 3. The required (η, ν), which satisfies (33) and (34), is given by

$$\begin{aligned} \eta(s) &= \begin{cases} G(r_f) + g(r_f)(s - r_f) \text{ for } s \in [\underline{s}, s_f], \\ G(s) \text{ for } s \in \left(s_f, s_f'\right), \\ G\left(r_f'\right) + g\left(r_f'\right)\left(s - r_f'\right) \text{ for } s \in \left[s_f', \overline{s}\right], \end{cases} \\ \nu(r) &= \begin{cases} -g(r_f) \text{ for } r \in [\underline{s}, s_f], \\ -g(r) \text{ for } r \in \left(s_f, s_f'\right), \\ -g\left(r_f'\right) \text{ for } r \in \left[s_f', \overline{s}\right]. \end{cases} \end{aligned}$$

■

Appendix D: Results for Binary Case

When the receiver has private information, in general, the problem of finding the optimal mechanism becomes complicated, as Section 4 suggests. This appendix fully characterizes the optimal mechanism when s and r are binary. More formally, assume that F puts strictly positive probabilities only on s_1 and s_2 and that $G(.|s_1)$ and $G(.|s_2)$ put strictly positive probabilities only on r_1 and r_2.

The binary case splits into two subcases. In the first subcase, one sender's signal is more favorable for acting than the other, regardless of r. The analysis of this subcase is two-fold. First, it provides formal proofs for the motivating example of Section 3.1. Second, it shows that the quantity-quality tradeoff of the motivating example carries on to a more general setting. In the second subcase, different sender's signals are favorable for acting depending on r. In this subcase, the single-crossing assumption does not hold, so the analysis and results are very different from those in the paper.

Using the revelation principle, for any mechanism, we can find an equivalent mechanism that sends at most four messages: (i) $m_\emptyset$ that induces the receiver not to act for all r, (ii) m_1 that induces the receiver to act only if $r = r_1$, (iii) m_2 that induces the receiver to act only if $r = r_2$, and (iv) $m_{1,2}$ that induces the receiver to act for all r.

For notational simplicity, this appendix uses different notation. In particular, denote $p_j \equiv \Pr(s_j)$, $p_{i|j} \equiv \Pr(r_i|s_j)$, $u_{ij} \equiv u(r_i, s_j)$, $\widetilde{u}_{ij} \equiv u_{ij} p_{i|j}$, $k_i = \widetilde{u}_{i1}/(\widetilde{u}_{i1} - \widetilde{u}_{i2})$, and $\phi_K^j \equiv \Pr_\phi(m = m_K, s = s_j)$ for $i, j = 1, 2$ and $K = \{\emptyset\}, \{1\}, \{2\}, \{1, 2\}$. Indexes i and j are reserved for r and s, respectively. Note that k_i is the cutoff posterior belief $\Pr_\phi(s_2|m)$ at which the receiver r_i is indifferent to act because

$$\mathbb{E}[u|r_i] = \frac{\widetilde{u}_{i1}(1 - \Pr_\phi(s_2|m)) + \widetilde{u}_{i2} \Pr_\phi(s_2|m)}{\Pr(r_i)} = 0.$$

Aligned Preferences

If one sender's signal is more favorable for acting than the other for all r, then the analysis is analogous to that of the binary example in Section 3.1. In particular, the sender faces the quantity-quality tradeoff of messages, which is resolved by the choice of a mechanism that sends messages with the highest marginal utility-price ratio.

To make the analysis non-redundant, assume that $u_{i1} < 0 < u_{i2}$ for $i = 1, 2$, $k_2 < k_1$, and $p_2 < k_1$. Strict inequalities rule out non-generic cases. Inequalities $u_{i2} > u_{i1}$ and $k_2 < k_1$ can be obtained by relabelling elements of S and R, respectively. If u_{i1} and u_{i2} had the same sign for some i, then the receiver r_i would take the same action regardless of the mechanism and the analysis would be as if the receiver was uninformed (Kolotilin (2014)). Finally, if $p_2 \geq k_1$, the no revelation mechanism would induce the receiver to act for all r, and, thus, it would be optimal.

Under these assumptions, the optimal mechanism can take the three forms identified in Section 3, as follows from:

Proposition 6 *If $u_{i1} < 0 < u_{i2}$, $k_2 < k_1$, and $p_2 < k_1$, then the optimal mechanism sends two messages.*

1. *If $p_{1|2} + p_{2|1}\widetilde{u}_{22}/\widetilde{u}_{21} \geq \widetilde{u}_{12}/\widetilde{u}_{11}$, it sends $m_{1,2}$ and $m_{\emptyset}$: $m_{1,2}$ with certainty if $s = s_2$ and with a non-trivial probability if $s = s_1$.*

2. *If $p_{1|2} + p_{2|1}\widetilde{u}_{22}/\widetilde{u}_{21} < \widetilde{u}_{12}/\widetilde{u}_{11}$ and $p_2 < k_2$, it sends m_2 and $m_{\emptyset}$: m_2 with certainty if $s = s_2$ and with a non-trivial probability if $s = s_1$.*

3. *If $p_{1|2} + p_{2|1}\widetilde{u}_{22}/\widetilde{u}_{21} < \widetilde{u}_{12}/\widetilde{u}_{11}$ and $p_2 \geq k_2$, it sends m_2 and $m_{1,2}$: m_2 with a non-trivial probability both if $s = s_2$ and if $s = s_1$.*

In all cases, $m_{\emptyset}$ reveals s_1 in that $\Pr_{\phi}(s_2|m_{\emptyset}) = 0$; m_2 makes the receiver r_2 indifferent to act in that $\Pr_{\phi}(s_2|m_2) = k_2$; and $m_{1,2}$ makes the receiver r_1 indifferent to act in that $\Pr_{\phi}(s_2|m_{1,2}) = k_1$. The receiver's expected utility under the optimal mechanism is strictly greater than that under the no revelation mechanism only in case 1.[24]

The intuition for Proposition 6 is analogous to that of the binary example in Section 3.1. The receiver r_i acts upon receiving a message m under a mechanism ϕ if $\Pr_{\phi}(s_2|m) \geq k_i$. If the message m persuades the receiver r_1 to act, it also persuades the receiver r_2 to act because $k_2 < k_1$ by assumption. Thus, we can restrict attention to mechanisms with the three messages $m_{\emptyset}$, m_2, and $m_{1,2}$. To maximize the probability that the receiver acts, each message of the optimal mechanism either makes the receiver exactly indifferent to act for some r ($\Pr_{\phi}(s_2|m_{1,2}) = k_1$ and $\Pr_{\phi}(s_2|m_2) = k_2$) or makes the receiver certain that $s = s_1$ so that it is optimal not to act ($\Pr_{\phi}(s_2|m_{\emptyset}) = 0$).

Thus, the sender's problem is to maximize the probability that the receiver acts:

$$\left(k_2 p_{2|2} + (1 - k_2)\, p_{2|1}\right) q_2 + q_{1,2}$$

[24]If s and r are independent, then $p_{i|2} = p_{i|1} = \Pr(r_i)$, so $\widetilde{u}_{ij}$ can be replaced with u_{ij} for all $i, j = 1, 2$ in all expressions.

over probabilities $q_\emptyset$, q_2, and $q_{1,2}$ of the messages $m_\emptyset$, m_2, and $m_{1,2}$ subject to the constraint imposed by the prior distribution of s:

$$k_2 q_2 + k_1 q_{1,2} = p_2.$$

Similar to Section 3, we can interpret k_2 and k_1 as unit prices of sending m_2 and $m_{1,2}$, and the probabilities $\left(k_2 p_{2|2} + (1-k_2)\, p_{2|1}\right)$ and 1 as the marginal utilities of sending m_2 and $m_{1,2}$. If $p_{1|2} + p_{2|1}\widetilde{u}_{22}/\widetilde{u}_{21} \geq \widetilde{u}_{12}/\widetilde{u}_{11}$, then the marginal utility-price ratio is highest for $m_{1,2}$, and the sender prefers to send $m_{1,2}$ than m_2, so the optimal mechanism sends $m_{1,2}$ and $m_\emptyset$. If $p_{1|2} + p_{2|1}\widetilde{u}_{22}/\widetilde{u}_{21} < \widetilde{u}_{12}/\widetilde{u}_{11}$, then the ratio is highest for m_2, and the sender prefers to send m_2 than $m_{1,2}$. The optimal mechanism then depends on whether the no revelation mechanism induces the receiver r_2 to act or not. If so ($p_2 \geq k_2$), then it sends the messages m_2 and $m_{1,2}$, otherwise it sends the messages m_2 and $m_\emptyset$.

Proof of Proposition 6. The optimal mechanism ϕ maximizes

$$\Pr{}_\phi\,(a=1) = p_{2|1}\phi_2^1 + p_{2|2}\phi_2^2 + \phi_{1,2}^1 + \phi_{1,2}^2$$

subject to

$$\begin{aligned}
&\phi_K^j \geq 0 \text{ for } j = 1,2 \text{ and } K = \{\emptyset\}, \{2\}, \{1,2\}, \\
&\phi_\emptyset^j + \phi_2^j + \phi_{1,2}^j = p_j \text{ for } j = 1,2, \\
&\widetilde{u}_{21}\phi_2^1 + \widetilde{u}_{22}\phi_2^2 \geq 0, \\
&\widetilde{u}_{11}\phi_{1,2}^1 + \widetilde{u}_{12}\phi_{1,2}^2 \geq 0, \\
&\widetilde{u}_{21}\phi_\emptyset^1 + \widetilde{u}_{22}\phi_\emptyset^2 < 0 \text{ or } \phi_\emptyset^1 = \phi_\emptyset^2 = 0, \\
&\widetilde{u}_{11}\phi_2^1 + \widetilde{u}_{12}\phi_2^2 < 0 \text{ or } \phi_2^1 = \phi_2^2 = 0.
\end{aligned} \tag{35}$$

Consider the relaxed problem that omits the last two constraints with strict inequalities. The solution to the relaxed problem satisfies $\phi_\emptyset^2 = 0$, $\widetilde{u}_{21}\phi_2^1 + \widetilde{u}_{22}\phi_2^2 = 0$, and $\widetilde{u}_{11}\phi_{1,2}^1 + \widetilde{u}_{12}\phi_{1,2}^2 = 0$, otherwise we can increase $\Pr_\phi(a=1)$ by the following changes to the mechanism. If $\phi_\emptyset^2 \neq 0$, change $\widetilde{\phi}_{1,2}^2 = \phi_{1,2}^2 + \phi_\emptyset^2$ and $\widetilde{\phi}_\emptyset^2 = 0$; if $\widetilde{u}_{11}\phi_{1,2}^1 + \widetilde{u}_{12}\phi_{1,2}^2 > 0$, change $\widetilde{\phi}_{1,2}^1 = \phi_{1,2}^1 + \varepsilon$ and either $\widetilde{\phi}_2^1 = \phi_2^1 - \varepsilon$ or $\widetilde{\phi}_\emptyset^1 = \phi_\emptyset^1 - \varepsilon$; if $\widetilde{u}_{21}\phi_2^1 + \widetilde{u}_{22}\phi_2^2 > 0$, change $\widetilde{\phi}_{1,2}^2 = \phi_{1,2}^2 + \varepsilon$ and $\widetilde{\phi}_2^2 = \phi_2^2 - \varepsilon$ where ε is a small positive number. These observations together with $k_2 < k_1$ imply that the solution to the relaxed problem satisfies the last two constraints and, therefore, it also solves the original problem. The original problem simplifies to the maximization of

$$\Pr{}_\phi\,(a=1) = \left(1 - \frac{\widetilde{u}_{12}}{\widetilde{u}_{11}}\right) p_2 - \left(p_{1|2} + \frac{\widetilde{u}_{22}}{\widetilde{u}_{21}} p_{2|1} - \frac{\widetilde{u}_{12}}{\widetilde{u}_{11}}\right)\phi_2^2$$

over ϕ_2^2 subject to

$$\begin{aligned}
&\left(\tfrac{\widetilde{u}_{12}}{\widetilde{u}_{11}} - \tfrac{\widetilde{u}_{22}}{\widetilde{u}_{21}}\right)\phi_2^2 \leq p_1 + \tfrac{\widetilde{u}_{12}}{\widetilde{u}_{11}} p_2. \\
&0 \leq \phi_2^2 \leq p_2.
\end{aligned}$$

The solution to this problem is:

$$\phi_2^2 = \begin{cases} 0 \text{ if } p_{1|2} + \frac{\widetilde{u}_{22}}{\widetilde{u}_{21}} p_{2|1} \geq \frac{\widetilde{u}_{12}}{\widetilde{u}_{11}}; \\ p_2 \text{ if } p_{1|2} + \frac{\widetilde{u}_{22}}{\widetilde{u}_{21}} p_{2|1} < \frac{\widetilde{u}_{12}}{\widetilde{u}_{11}} \text{ and } \widetilde{u}_{21} p_1 + \widetilde{u}_{22} p_2 < 0; \\ \widetilde{u}_{21} \frac{\widetilde{u}_{11} p_1 + \widetilde{u}_{12} p_2}{\widetilde{u}_{12} \widetilde{u}_{21} - \widetilde{u}_{11} \widetilde{u}_{22}} \text{ if } p_{1|2} + \frac{\widetilde{u}_{22}}{\widetilde{u}_{21}} p_{2|1} < \frac{\widetilde{u}_{12}}{\widetilde{u}_{11}} \text{ and } \widetilde{u}_{21} p_1 + \widetilde{u}_{22} p_2 \geq 0. \end{cases}$$

Finally, $\phi_{1,2}^2 = p_2 - \phi_2^2$, $\phi_2^1 = -\phi_2^2 \widetilde{u}_{22} / \widetilde{u}_{21}$, $\phi_{1,2}^1 = -\phi_{1,2}^2 \widetilde{u}_{12} / \widetilde{u}_{11}$, $\phi_\emptyset^2 = 0$, $\phi_\emptyset^1 = p_1 - \phi_2^1 - \phi_{1,2}^1$.

Under ϕ_{no}, the receiver's expected utility is

$$\mathbb{E}\left[\max_a \mathbb{E}_{\phi_{no}} \left[au\left(r,s\right)|r\right]\right] = \max\left\{0, \widetilde{u}_{21} p_1 + \widetilde{u}_{22} p_2\right\}.$$

Under ϕ, the receiver's expected utility is

$$\begin{aligned} \mathbb{E}\left[\max_a \mathbb{E}_{\phi} \left[au\left(r,s\right)|r,m\right]\right] &= \left(\widetilde{u}_{11} \phi_{1,2}^1 + \widetilde{u}_{12} \phi_{1,2}^2\right) + \left(\widetilde{u}_{21} \phi_{1,2}^1 + \widetilde{u}_{22} \phi_{1,2}^2\right) + \left(\widetilde{u}_{21} \phi_2^1 + \widetilde{u}_{22} \phi_2^2\right) \\ &= \widetilde{u}_{21} \phi_{1,2}^1 + \widetilde{u}_{22} \phi_{1,2}^2 \\ &= \begin{cases} \left(\frac{\widetilde{u}_{11} \widetilde{u}_{22} - \widetilde{u}_{12} \widetilde{u}_{21}}{\widetilde{u}_{11}}\right) p_2 \text{ if } p_{1|2} + \frac{\widetilde{u}_{22}}{\widetilde{u}_{21}} p_{2|1} \geq \frac{\widetilde{u}_{12}}{\widetilde{u}_{11}}; \\ 0 \text{ if } p_{1|2} + \frac{\widetilde{u}_{22}}{\widetilde{u}_{21}} p_{2|1} < \frac{\widetilde{u}_{12}}{\widetilde{u}_{11}} \text{ and } \widetilde{u}_{21} p_1 + \widetilde{u}_{22} p_2 < 0; \\ \widetilde{u}_{21} p_1 + \widetilde{u}_{22} p_2 \text{ if } p_{1|2} + \frac{\widetilde{u}_{22}}{\widetilde{u}_{21}} p_{2|1} < \frac{\widetilde{u}_{12}}{\widetilde{u}_{11}} \text{ and } \widetilde{u}_{21} p_1 + \widetilde{u}_{22} p_2 \geq 0. \end{cases} \end{aligned}$$

The second equality holds because $\widetilde{u}_{11} \phi_{1,2}^1 + \widetilde{u}_{12} \phi_{1,2}^2 = \widetilde{u}_{21} \phi_2^1 + \widetilde{u}_{22} \phi_2^2 = 0$. The first case holds because $\phi_{1,2}^2 = p_2$ and $\phi_{1,2}^1 = -p_2 \widetilde{u}_{12} / \widetilde{u}_{11}$. The second case holds because $\phi_{1,2}^1 = \phi_{1,2}^2 = 0$. The third case holds because $\phi_{1,2}^1 = p_1 - \phi_2^1$, $\phi_{1,2}^2 = p_2 - \phi_2^2$, and $\widetilde{u}_{21} \phi_2^1 + \widetilde{u}_{22} \phi_2^2 = 0$. Therefore, the receiver's expected utilities under ϕ and ϕ_{no} differ if and only if $p_{1|2} + p_{2|1} \widetilde{u}_{22} / \widetilde{u}_{21} \geq \widetilde{u}_{12} / \widetilde{u}_{11}$.
■

Misaligned Preferences

The main goal of this section is to illustrate the variety of possible optimal mechanisms in the case where different sender's signals are more favorable for acting depending on the receiver's type. For example, a school may know whether a student is good at natural sciences or liberal arts, but it may be unsure which of these two qualities are valued by the employer. Note that this case violates the single-crossing assumption of Section 4.

All forms that the optimal mechanism can take are characterized by Proposition 7. Similar to the previous subcase, to make the analysis non-redundant, I impose certain assumptions.

Proposition 7 *If $u_{12} < 0 < u_{11}$, $u_{21} < 0 < u_{22}$, and $p_2 > k_1$, then the optimal mechanism sends at most two messages.*

1. *If* $k_2 \leq k_1$*, it sends* m_2 *and* $m_{1,2}$*. The message* m_2 *reveals* s_2 *in that* $\Pr_\phi(s_2|m_2) = 1$ *and the message* $m_{1,2}$ *makes the receiver* r_1 *indifferent to act in that* $\Pr_\phi(s_2|m_{1,2}) = k_1$*.*

2. *If* $k_2 > k_1$*, then depending on parameters, it sends either only* m_2 *or both* m_2 *and* m_1*. If it sends both* m_2 *and* m_1*, there are four cases in which each message* m_i *either reveals* s_i *in that* $\Pr_\phi(s_i|m_i) = 1$*, or makes the receiver* r_i *indifferent to act in that* $\Pr_\phi(s_2|m_i) = k_i$*.*

I only sketch the intuition for this proposition because it is tedious and involves many cases. Note that a message m that assigns a higher probability to s_2 is more persuasive for the receiver r_2, and less persuasive for the receiver r_1. The messages m_1 and m_2 are always feasible because revealing s_1 induces the receiver r_1 to act, and revealing s_2 induces the receiver r_2 to act. However, if $k_2 \leq k_1$ (part 1 of Proposition 7), then the message $m_{1,2}$ is feasible, but the message $m_\emptyset$ is not. In this case, the sender wants to send $m_{1,2}$ as often as possible. As a result, the optimal mechanism sends two types of messages: those that give minimal possible evidence to make the receiver act regardless of his signal, and those that reveal s. In contrast, if $k_1 < k_2$ (part 2 of Proposition 7), then the message $m_\emptyset$ is feasible, but the message $m_{1,2}$ is not. In this case, the optimal mechanism can take five different forms, which, in particular, include the full revelation and no revelation mechanisms.

Proof of Proposition 7. The optimal mechanism ϕ maximizes

$$\Pr{}_\phi(a = 1) = p_{1|1}\phi_1^1 + p_{1|2}\phi_1^2 + p_{2|1}\phi_2^1 + p_{2|2}\phi_2^2 + \phi_{1,2}^1 + \phi_{1,2}^2$$

subject to

$$\begin{aligned}
&\phi_K^j \geq 0 \text{ for } j = 1,2 \text{ and } K = \{\emptyset\}, \{1\}, \{2\}, \{1,2\}, \\
&\phi_\emptyset^j + \phi_1^j + \phi_2^j + \phi_{1,2}^j = p_j \text{ for } j = 1,2, \\
&\widetilde{u}_{i1}\phi_i^1 + \widetilde{u}_{i2}\phi_i^2 \geq 0 \text{ for } i = 1,2, \\
&\widetilde{u}_{i1}\phi_{1,2}^1 + \widetilde{u}_{i2}\phi_{1,2}^2 \geq 0 \text{ for } i = 1,2, \\
&\widetilde{u}_{i1}\phi_{3-i}^1 + \widetilde{u}_{i2}\phi_{3-i}^2 < 0 \text{ or } \phi_{3-i}^1 = \phi_{3-i}^2 = 0 \text{ for } i = 1,2, \\
&\widetilde{u}_{i1}\phi_\emptyset^1 + \widetilde{u}_{i2}\phi_\emptyset^2 < 0 \text{ or } \phi_\emptyset^1 = \phi_\emptyset^2 = 0 \text{ for } i = 1,2.
\end{aligned}$$

Note that $\widetilde{u}_{11}\Pr_\phi(s_1|m) + \widetilde{u}_{12}\Pr_\phi(s_2|m) \geq 0$ is equivalent to $\Pr_\phi(s_2|m) \leq k_1$, and $\widetilde{u}_{21}\Pr_\phi(s_1|m) + \widetilde{u}_{22}\Pr_\phi(s_2|m) \geq 0$ is equivalent to $\Pr_\phi(s_2|m) \geq k_2$. Therefore, the receiver r_1 acts if $\Pr_\phi(s_2|m) \leq k_1$, and the receiver r_2 acts if $\Pr_\phi(s_2|m) \geq k_2$. If $k_2 \leq k_1$, then no mechanism can send the message $m_\emptyset$ because $\Pr_\phi(s_2|m) < k_2$ and $\Pr_\phi(s_2|m) > k_1$ cannot both hold. On the contrary, if $k_2 > k_1$, then no mechanism can send the message $m_{1,2}$ because $\Pr_\phi(s_2|m) \geq k_2$ and $\Pr_\phi(s_2|m) \leq k_1$ cannot both hold. Consider these two cases in turn.

Let $k_2 \leq k_1$ and, thus, $\phi^1_\emptyset = \phi^2_\emptyset = 0$. Consider the relaxed problem with the constraints $\phi^j_K \geq 0$, $\phi^j_1 + \phi^j_2 + \phi^j_{1,2} = p_j$, $\widetilde{u}_{11}\phi^1_1 + \widetilde{u}_{12}\phi^2_1 \geq 0$, and $\widetilde{u}_{11}\phi^1_{1,2} + \widetilde{u}_{12}\phi^2_{1,2} \geq 0$ for all K and j. Note that the last two constraints imply $\widetilde{u}_{11}\left(\phi^1_1 + \phi^1_{1,2}\right) + \widetilde{u}_{12}\left(\phi^2_1 + \phi^2_{1,2}\right) \geq 0$, so the solution to the relaxed problem satisfies $\phi^1_1 = \phi^2_1 = 0$, otherwise we can increase $\Pr_\phi(a=1)$ by the following changes to the mechanism: $\widetilde{\phi}^j_{1,2} = \phi^j_{1,2} + \phi^j_1$ and $\widetilde{\phi}^j_1 = 0$ for $j = 1,2$. Substituting $\phi^j_{1,2} = p_j - \phi^j_2$, the relaxed problem simplifies to: ϕ^1_2 and ϕ^2_2 maximize

$$\Pr{}_\phi(a=1) = 1 - p_{1|1}\phi^1_2 - p_{1|2}\phi^2_2$$

subject to

$$\begin{aligned}&\phi^j_2 \in [0, p_j] \text{ for } j = 1,2,\\ &\widetilde{u}_{11}\phi^1_2 + \widetilde{u}_{12}\phi^2_2 \leq \widetilde{u}_{11}p_1 + \widetilde{u}_{12}p_2.\end{aligned}$$

The solution to this problem is $(\phi^1_2, \phi^2_2) = (0, (\widetilde{u}_{11}p_1 + \widetilde{u}_{12}p_2)/\widetilde{u}_{12})$. It is also the solution to the original problem because it satisfies all constraints of the original problem.

Let $k_2 > k_1$ and, thus, $\phi^1_{1,2} = \phi^2_{1,2} = 0$. In the optimal mechanism, $\phi^1_\emptyset = \phi^2_\emptyset = 0$, otherwise we can increase $\Pr_\phi(a=1)$ by the following changes to the mechanism: $\widetilde{\phi}^i_i = \phi^i_i + \phi^i_\emptyset$ and $\widetilde{\phi}^i_\emptyset = 0$ for $i = 1,2$. Consider the relaxed problem with the constraints $\phi^j_1, \phi^j_2 \geq 0$, $\phi^j_1 + \phi^j_2 = p_j$, $\widetilde{u}_{11}\phi^1_1 + \widetilde{u}_{12}\phi^2_1 \geq 0$, and $\widetilde{u}_{21}\phi^1_2 + \widetilde{u}_{22}\phi^2_2 \geq 0$ for all $j = 1,2$. Substituting $\phi^j_1 = p_j - \phi^j_2$, the relaxed problem simplifies to: ϕ^1_2 and ϕ^2_2 maximize

$$\Pr{}_\phi(a=1) = p_{1|1}p_1 + p_{1|2}p_2 + \left(1 - 2p_{1|1}\right)\phi^1_2 + \left(1 - 2p_{1|2}\right)\phi^2_2$$

subject to

$$\begin{aligned}&\phi^j_2 \in [0, p_j] \text{ for } j = 1,2,\\ &\widetilde{u}_{11}\phi^1_2 + \widetilde{u}_{12}\phi^2_2 \leq \widetilde{u}_{11}p_1 + \widetilde{u}_{12}p_2,\\ &\widetilde{u}_{21}\phi^1_2 + \widetilde{u}_{22}\phi^2_2 \geq 0.\end{aligned}$$

The coefficients $1 - 2p_{1|1}$ and $1 - 2p_{1|2}$ in the objective function can have any sign and, therefore, any extreme point of the constraints can be a solution to this problem. If $p_2 \geq k_2$, the extreme points of (ϕ^1_2, ϕ^2_2) are $(0, p_2)$, $(0, (\widetilde{u}_{11}p_1 + \widetilde{u}_{12}p_2)/\widetilde{u}_{12})$, and (p_1, p_2). If $p_2 < k_2$, the extreme points of (ϕ^1_2, ϕ^2_2) are $(0, p_2)$, $(0, (\widetilde{u}_{11}p_1 + \widetilde{u}_{12}p_2)/\widetilde{u}_{12})$, $(-p_2\widetilde{u}_{22}/\widetilde{u}_{21}, p_2)$, $(\widetilde{u}_{22}, -\widetilde{u}_{21}) \cdot (\widetilde{u}_{11}p_1 + \widetilde{u}_{12}p_2)/(\widetilde{u}_{11}\widetilde{u}_{22} - \widetilde{u}_{12}\widetilde{u}_{21})$. All these extreme points can be a solution to the original problem because they satisfy all constraints of the original problem. ■

References

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

Ambrosio, Luigi (2003) *Lecture Notes on Optimal Transport Problems*: Springer.

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*, **35** (2), pp. 281–322.

Au, Pak Hung (2012) "Dynamic Information Disclosure." Mimeo, NTU.

de Barreda, Ines Moreno (2013) "Cheap Talk with Two-Sided Private Information." Mimeo, University of Oxford.

Bergemann, Dirk and Stephen Morris (2013) "The Comparison of Information Structures in Games: Bayes Correlated Equilibrium and Individual Sufficiency." Mimeo, Yale University.

Billingsley, Patrick (1995) *Probability and Measure*, New York: John Wiley and Sons.

Blackwell, David (1953) "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, **24** (2), pp. 265–272.

Chen, Ying (2009) "Communication with Two-sided Asymmetric Information." Mimeo, Arizona State University.

——— (2012) "Value of Public Information in Sender–Receiver Games," *Economics Letters*, **114** (3), pp. 343–345.

Crawford, Vincent P. and Joel Sobel (1982) "Strategic Information Transmission," *Econometrica*, **50** (6), pp. 1431–1451.

Dvoretzky, Aryeh, Abraham Wald, and Jacob Wolfowitz (1951) "Elimination of Randomization in Certain Statistical Decision Procedures and Zero-Sum Two-Person Games," *Annals of Mathematical Statistics*, **22** (1), pp. 1–21.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica (2014) "Suspense and Surprise," *Journal of Political Economy*, **forthcoming**.

Forges, Françoise (1993) "Five Legitimate Definitions of Correlated Equilibrium in Games with Incomplete Information," *Theory and Decision*, **35** (3), pp. 277–310.

Gentzkow, Matthew and Emir Kamenica (2012) "Competition in Persuasion." Mimeo, University of Chicago.

Ginzburg, Boris (2013) "Optimal News Management." Mimeo, UCL.

Glazer, Jacob and Ariel Rubinstein (2004) "On Optimal Rules of Persuasion," *Econometrica*, **72** (6), pp. 1715–1736.

——— (2006) "A Study in the Pragmatics of Persuasion: a Game Theoretical Approach," *Theoretical Economics*, **1** (4), pp. 395–410.

Golosov, Mikhail, Vasiliki Skreta, Aleh Tsyvinski, and Andrea Wilson (2014) "Dynamic Strategic Information Transmission," *Journal of Economic Theory*, **forthcoming**.

Green, Jerry R. and Nancy L. Stokey (2007) "A Two-Person Game of Information Transmission," *Journal of Economic Theory*, **135** (1), pp. 90–104.

Horner, Johannes and Andrzej Skrzypacz (2012) "Selling Information." Mimeo, Yale University.

Ivanov, Maxim (2010) "Informational Control and Organization Design," *Journal of Economic Theory*, **145** (2), pp. 721–751.

Kamenica, Emir and Matthew Gentzkow (2011) "Bayesian Persuasion," *American Economic Review*, **101** (6), pp. 2590–2615.

Kolmogorov, Andrey N. and Sergey V. Fomin (1975) *Introductory Real Analysis*, New York: Dover Publications.

Kolotilin, Anton (2014) "Experimental Design to Persuade." Mimeo, UNSW.

Lai, Ernest K. (2013) "Expert Advice for Amateurs." Mimeo, Lehigh University.

Milgrom, Paul (1981) "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, **12** (2), pp. 380–391.

Milgrom, Paul and Robert Weber (1985) "Distributional Strategies for Games with Incomplete Information," *Mathematics of Operations Research*, **10** (4), pp. 619–632.

Olszewski, Wojciech (2004) "Informal Communication," *Journal of Economic Theory*, **117** (2), pp. 180–200.

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

Rayo, Luis and Ilya Segal (2010) "Optimal Information Disclosure," *Journal of Political Economy*, **118** (5), pp. 949–987.

Seidmann, Daniel J (1990) "Effective Cheap Talk with Conflicting Interests," *Journal of Economic Theory*, **50** (2), pp. 445–458.

Sher, Itai (2014) "Persuasion and Dynamic Communication," *Theoretical Economics*, **9** (1), pp. 99–136.

Watson, Joel (1996) "Information Transmission when the Informed Party is Confused," *Games and Economic Behavior*, **12** (1), pp. 143–161.