A Rothschild-Stiglitz approach to Bayesian persuasion

Matthew Gentzkow and Emir Kamenica*
Stanford University and University of Chicago

December 2015

Abstract

Rothschild and Stiglitz (1970) represent random variables as convex functions (integrals of the cumulative distribution function). Combining this representation with Blackwell's Theorem (1953), we characterize distributions of posterior means that can be induced by a signal. This characterization provides a novel way to analyze a class of Bayesian persuasion problems.

JEL classification: D83
Keywords: mean-preserving spreads; information design

*We thank the Sloan Foundation and the University of Chicago Booth School of Business for financial support.

1 Introduction

Consider a situation where one person, call him Sender, generates information in order to persuade another person, call her Receiver, to change her action. Sender and Receiver share a common prior about the state of the world. Sender can publicly generate any signal about the state and Receiver observes the signal realization before she takes her action.[1]

Kamenica and Gentzkow (2011) analyze a general version of this 'Bayesian persuasion' problem.[2] They draw on an insight from Aumann and Maschler (1995) to develop a geometric approach to Sender's optimization problem. They derive a value function over beliefs and then construct the optimal signal from the concavification of that value function.[3] This approach provides ample intuition about the structure of the optimal signal, but has limited applicability when the state space is large. The dimensionality of the space of beliefs is roughly the same as the cardinality of the state space,[4] so the value function and its concavification can be visualized easily only when there are two or three states of the world. When the state space is infinite, the concavification approach requires working in an infinite-dimensional space.

In this paper we analyze a class of Bayesian persuasion problems where the state space may be large but Sender and Receiver's preferences take a simple form: the state ω is a random variable, Receiver's optimal action (taken from a finite set) depends only on $\mathbb{E}[\omega]$, and Sender's preferences over Receiver's action are independent of the state.

This environment captures a number of economically relevant settings. For example, it might be the case that Sender is a firm, Receiver is a consumer, and ω is the match quality between the attributes of firm's product and the consumer's preferences. The interpretation of the signal in this case is the firm's choice of what information about the product to provide to the consumer. For example, a software company can decide on the features of the trial version of the product.

Kamenica and Gentzkow (2011) also examine this specific environment, but do not characterize the optimal signal. They show that if one considers the value function over the posterior mean,

[1] A *signal*, in our terminology, is a map from the true state of the world to a distribution over some signal realization space. Others terms for a signal include *experiment*, *signal structure*, and *information structure*.

[2] Gentzkow and Kamenica (2014) extend the analysis to the case of costly signals and Gentzkow and Kamenica (2015a, 2015b) to situations with multiple senders.

[3] A *concavification* of a function f is the smallest concave function that is everywhere weakly greater than f.

[4] When the state space Ω is finite, the space of beliefs has $|\Omega| - 1$ dimensions.

the concavification of that value function pins down whether Sender can benefit from generating information but does not determine the optimal signal.

The problem is that it is difficult to characterize the set of feasible distributions of posterior means. Any distribution of posterior beliefs whose expectation is the prior can be induced by some signal; but, it is not possible to induce every distribution of posterior means whose expectation is the prior mean.

In this paper, we combine insights from Blackwell (1953) and Rothschild and Stiglitz (1970) to derive the characterization of all feasible distributions of the posterior mean.[5] We then use this characterization to analyze the aforementioned class of Bayesian persuasion problems.

Kolotilin (2014) and Kolotilin *et al.* (2015) examine closely related environments. They make the same assumptions on preferences but allow for Receiver to have private information. They focus exclusively on the case where Receiver takes a binary action. Kolotilin (2014) shows that neither Sender's nor Receiver's payoff is necessarily monotone in the precision of Receiver's private information. Kolotilin *et al.* (2015) consider "private persuasion" where Receiver reports his private type before Sender generates information. They show that Sender never strictly benefits by allowing for private persuasion.[6] While the focus of these papers is somewhat different, our proof draws on a result in Kolotilin (2014).

2 The model

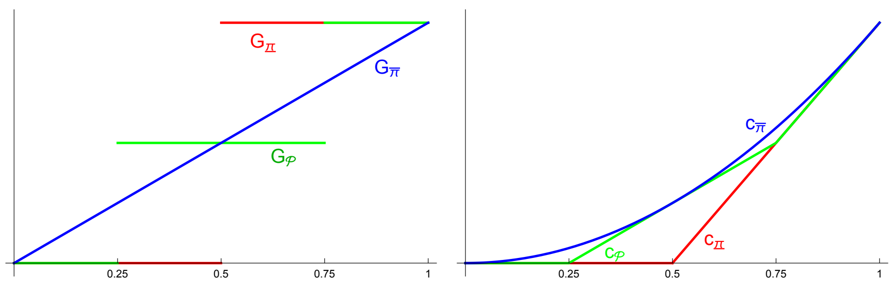
The state of nature is a random variable ω on $[0,1]$. Sender and Receiver share a common prior F_0. Throughout the paper we denote any distribution over real numbers by its cumulative distribution function (CDF); hence, under the prior $Pr(\omega \leq x) = F_0(x)$. Let m_0 denote the mean of F_0. A *signal* π consists of a *signal realization space* S and a family of distributions $\{\pi_\omega\}$ over S. Sender chooses a signal. Receiver observes the choice of the signal π and the signal realization s. Receiver

[5] We take the characterization of second-order stochastic dominance in terms of the integral of the cumulative distribution function from Rothschild and Stiglitz (1970). As Rothschild and Stiglitz (1972) acknowledge, this result was previously established by Blackwell and Girshick (1954). However, because most economists associate second-order stochastic dominance with Rothschild and Stiglitz (1970), we emphasize that reference even though it is not the earliest formulation of the relevant theorem.

[6] Gupta (2014) and Wang (2015) also contrast public and private persuasion but they have a different notion of private persuasion. Specifically, they consider a case with multiple receivers and contrast the case where all of them observe the same signal realization with the case where they observe independent draws of the signal.

Figure 1: Signals as convex functions

(a) CDFs of posterior means (b) Functions induced by signals



then chooses an action from a finite set. Her optimal action depends on her expectation of the state, $\mathbb{E}[\omega]$. Without loss of generality we label the actions so that action a_i is optimal if $\gamma_i \leq \mathbb{E}[\omega] \leq \gamma_{i+1}$ given some set of cutoffs $\gamma_0 \leq \gamma_1 \leq ... \leq \gamma_n \in [0,1]$. Sender has some state-independent utility function over Receiver's action.

3 Signals as convex functions

Given a signal π, a signal realization s induces a posterior F_s. Let m_s denote the mean of F_s. A signal induces a distribution of posteriors and hence a distribution of posterior means. Let G_π denote the distribution of posterior means induced by signal π. Then, for each signal π, let c_π denote the integral of G_π, i.e., $c_\pi(x) = \int_0^x G_\pi(t)\,dt$. If c_π is thus obtained from π we say that π *induces* c_π.

We illustrate this definition with some examples. Suppose that F_0 is uniform. Consider a totally uninformative signal $\underline{\pi}$. This signal induces a degenerate distribution of posterior means always equal to $m_0 = \frac{1}{2}$. Hence, $G_{\underline{\pi}}$ is a step function equal to 0 below $\frac{1}{2}$ and equal to 1 above $\frac{1}{2}$. The function $c_{\underline{\pi}}$ induced in turn is thus flat on $\left[0,\frac{1}{2}\right]$ and then linearly increasing from $\frac{1}{2}$ to 1 with a slope of 1. At the other extreme, consider a fully informative signal $\overline{\pi}$ that generates a distinct signal realization in each state. With this signal, each posterior has a degenerate distribution with all the mass on the true state and thus $G_{\overline{\pi}} = F_0$. Since F_0 is uniform, $G_{\overline{\pi}}$ is linear, and thus $c_{\overline{\pi}}$ is quadratic: $c_{\overline{\pi}}(x) = \frac{1}{2}x^2$. Finally, consider a "partitional" signal $\mathscr{P}$ that gives a distinct signal realization depending on whether the state is in $\left[0,\frac{1}{2}\right]$, or $\left(\frac{1}{2},1\right]$. Then, $G_{\mathscr{P}}$ is a step function and $c_{\mathscr{P}}$ is piecewise-linear. Figure 1 depicts these CDFs and functions.

If we consider an arbitrary signal π, what can we say about c_π? Since G_π is a CDF and thus increasing, c_π as its integral must be convex. Moreover, since any signal π is a garbling of $\overline{\pi}$, we must have that $G_{\overline{\pi}}$ is a mean-preserving spread of G_π (Blackwell 1953); hence, $c_{\overline{\pi}} \geq c_\pi$ by Rothschild and Stiglitz (1970). Similarly, since $\underline{\pi}$ is a garbling of π, G_π is a mean-preserving spread of $G_{\overline{\pi}}$ and thus $c_\pi \geq c_{\underline{\pi}}$. Putting these observations together, we obtain:

Remark 1. Given any signal π, the induced function c_π is convex. Moreover, $c_{\overline{\pi}} \geq c_\pi \geq c_{\underline{\pi}}$.

Note that this result applies for any prior, not just for the uniform example depicted in Figure 1. In general, functions $c_{\underline{\pi}}$ and $c_{\overline{\pi}}$ are determined by the prior with $c_{\underline{\pi}}$ flat to the left of m_0 and then increasing with a slope of 1, and $c_{\overline{\pi}}$ equal to the integral of F_0.

Finally, note that these arguments rely on the fact that we are interested in the distribution of posterior mean rather than in the distribution of some other moment. If a distribution of posteriors τ is a mean-preserving spread of τ', then the distribution of posterior mean under τ is a mean-preserving spread of the distribution of posterior mean under τ'. One cannot make the same claim about, say the distribution of posterior variance.[7]

4 Convex functions as signals

Now suppose that we are given some function that satisfies the properties from Remark 1. Is it always the case that there is some signal that induces this function? The answer to this question turns out to be affirmative:

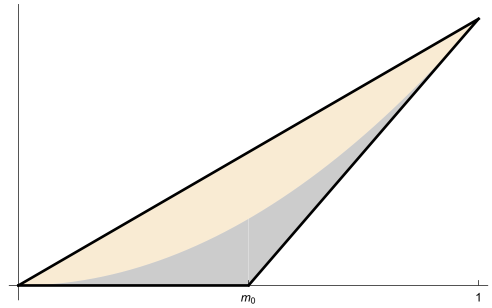
Proposition 1. *Given any convex function $c : [0,1] \to \mathbb{R}$ such that $c_{\overline{\pi}} \geq c \geq c_{\underline{\pi}}$, there exists a signal that induces it.*

Proof. Consider some function c satisfying the given properties. Define a function

$$G(x) = \begin{cases} 0 & : \quad x < 0 \\ c'(x^+) & : \quad 0 \leq x < 1 \\ 1 & : \quad x \geq 1 \end{cases}$$

[7] Our results do apply to any setting where Receiver's action depends on $\mathbb{E}[f(\omega)]$ for some monotone f. Allowing for an f other than the identity function is equivalent to simply "rescaling" the units of ω.

Figure 2: Feasible distributions of posterior means vs. all random variables with $\mathbb{E}[\tilde{m}] = m_0$



where $c'(x^+)$ denotes the right derivative of c at x. Since c is convex, its right derivative must exist. Moreover, since c is convex and $0 \leq c'(x^+) \leq 1$ for all $x \in [0, 1)$ (cf: Lemma 1 in the Appendix), G is weakly increasing and continuous. We also have that $lim_{x \to -\infty} G(x) = 0$ and $lim_{x \to \infty} G(x) = 1$. Hence, G is a CDF. Now, since $\int_0^x F_0(t)\, dt = c_{\overline{\pi}}(x) \geq c(x) = \int_0^x G(t)\, dt$, we know that F_0 is a mean-preserving spread of G. By Proposition 1 in Kolotilin (2014), this in turn implies that there must exist a signal that induces G as the distribution of posterior means. □

Proposition 1 thus provides us with a simple characterization of the distributions of posterior means that can be induced by a signal. Figure 2 contrasts the space of functions induced by all random variables whose expectation is the prior mean (any convex function in the lightly shaded area) with the subset of those that represent feasible distributions of the posterior means (any convex function in the darker area in the bottom right).

It is also easy to extend the analysis to situations with public, exogenous information. Let π^e denote an exogenous signal that Sender and Receiver observe prior to Sender generating additional information. Sender then effectively chooses any convex function between c_{π^e} and $c_{\overline{\pi}}$.

5 Optimal signals

In the previous section we transformed Sender's problem from choosing a signal to choosing a convex function. Our next step is to analyze how to determine Sender's payoff for a given function in this new budget set.

The key observation is that – under the preference structure we have assumed – Sender's payoff is entirely determined by the local behavior of the induced function at the action cutoffs, i.e., γ_i's.

Specifically, the left and/or the right derivatives (depending on how Receiver breaks her indifferences) of c_π at γ_i's determine how often Receiver takes each action. Hence, once we know these derivatives, we can back out Sender's payoff. We illustrate this idea in the next two subsections.

5.1 Two actions

Consider the simplest case where Receiver takes one of two actions: a_0 or a_1. To make the problem non-trivial, we assume that Sender prefers a_1, but $m_0 < \gamma_1$.[8] In any equilibrium, Receiver must break her indifference at γ_1 in Sender's favor.[9] Hence, Sender wants to design a signal that maximizes the probability of a signal realization s such that $m_s \geq \gamma_1$. This problem can be solved algebraically (Ivanov 2015),[10] but we nonetheless begin with this simplest example as it illustrates our approach in the most transparent way.

As mentioned above, the probability that Receiver takes Sender's preferred action is determined by a derivative of c_π at γ_1. More specifically, since Receiver breaks her indifference in Sender's favor, Receiver's behavior depends on the *left* derivative of c_π. The left derivative at γ_1 equals the likelihood that the posterior mean is strictly below γ_1 – i.e., the probability that Receiver takes action a_0 – so Sender's payoff is decreasing in $c'_\pi\left(\gamma_1^-\right)$.

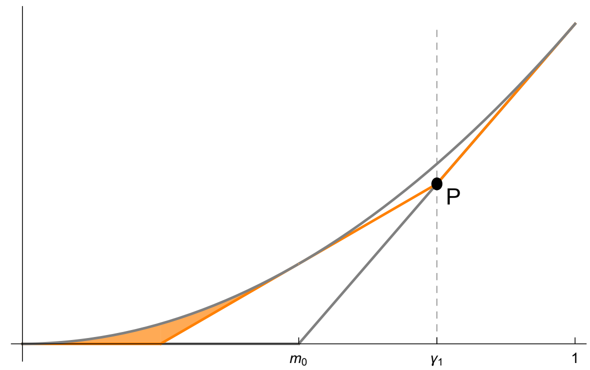
Sender wants to induce a function that minimizes the left derivative at γ_1. As we can see Figure 3, he cannot bring this derivative all the way down to zero. Doing so would violate the restriction that c_π must be both convex and bounded above by $c_{\overline{\pi}}$. In fact, looking at Figure 3, it is easy to see that any optimal c_π – the one that minimizes the left derivative – must satisfy two features. First, it must coincide with $c_{\underline{\pi}}$ at γ_1, as indicated by the "pivot point" labeled P. Second, the "arm" leaving P to the left should be "pivoted up" as much as possible, until it is tangent to $c_{\overline{\pi}}$. This identifies all optimal signals since the behavior of the function to the left of the tangency point is irrelevant. Any convex function within the shaded area of Figure 3 is optimal. These functions correspond to signals that yield a single, deterministic realization s when ω is above the tangency point and generate arbitrary (potentially stochastic) realizations (not equal to s) for other states. The top of the shaded area is induced by a signal that fully reveals all ω below the tangency point

[8] Otherwise, a completely uninformative signal is clearly optimal.

[9] Otherwise, Sender wants to induce a posterior mean "arbitrarily close" to γ_1, which is not a well defined problem and thus cannot be a part of an equilibrium.

[10] An optimal signal is a partition that reveals whether ω belongs to $[x^*, 1]$, with x^* defined by $\int_{x^*}^1 x dF_0(x) = \gamma_1$.

Figure 3: Optimal signals with binary actions



while the bottom of the area is induced by a signal that generates a single realization for all those states.

5.2 More actions

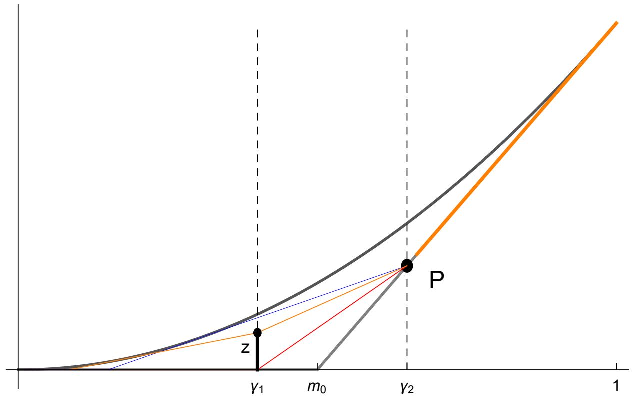
Now suppose Receiver can take one of three actions and Sender's utility is 0 from a_0, 1 from a_1, and $\lambda > 1$ from a_2. As before, we know that Receiver will break indifferences in Sender's favor and take action a_1 at γ_1 and action a_2 at γ_2. Hence, Sender's payoff is determined by $c'_\pi\left(\gamma_1^-\right) = Pr\left(a_0\right)$ and $c'_\pi\left(\gamma_2^-\right) = Pr\left(a_0\right) + Pr\left(a_1\right)$.

Suppose $m_0 \in \left(\gamma_1, \gamma_2\right)$. Looking at Figure 4, we first note that the optimal function must go through point P. Otherwise, it would be possible to decrease $c'_\pi\left(\gamma_2^-\right)$ while keeping $c'_\pi\left(\gamma_1^-\right)$ unchanged, which would lead to a higher payoff. The only question that remains, therefore, is where the function should cross γ_1; this point determines the tradeoff between how frequently a_1 and a_2 are taken.

At one extreme, we have the blue function that maximizes the probability of a_2. This occurs at the expense of a_1 never happening. At the other extreme is the red function that ensures that a_0 never happens, but consequently leads to a_2 being less frequent than it could be. Finally, the orange function shows a "compromise" solution where all three actions are taken with positive probability. As Figure 4 shows, we can index all potentially optimal functions with a single-dimensional parameter z that denotes the height at which the function crosses γ_1.

How does Sender's payoff vary with z? Probability of a_2 is $1 - c'\left(\gamma_2^-\right)$, which is linearly increasing

Figure 4: Optimal signal with three actions



in z.[11] The probability of a_1, on the other hand, is decreasing in z. This relationship is generally not linear. As can be seen from Figure 4, it depends on $c_{\overline{\pi}}$ which in turn depends on F_0. In the Appendix, we explicitly compute the relationship between z and the probability of a_1 in the case of a uniform prior. It takes the form of $A - \sqrt{B - 2z} - Cz$ where A, B, and C are constants that depend on γ_1 and γ_2. Because the relationship is not linear, we do not necessarily end up at a corner solution with either the blue line (zero probability of a_1) or the red line (zero probability of a_0) being optimal.[12] For example, if the prior is uniform, $\gamma_1 = \frac{1}{3}$, $\gamma_2 = \frac{2}{3}$, and $\lambda = 3$, the optimal z is $\frac{1}{24}$. This function cannot be induced through an interval partition. One signal that that achieves the optimum is a non-monotone partition that reveals whether the state is in $\left[0, \frac{8}{48}\right]$ inducing a_0, in $\left[\frac{11}{48}, \frac{21}{48}\right]$ inducing a_1, or in $\left[\frac{8}{48}, \frac{11}{48}\right] \cup \left[\frac{21}{48}, 1\right]$ inducing a_2.

6 Conclusion

Previous work on Bayesian persuasion built on the observation that a distribution of posterior beliefs is feasible, i.e., can be induced by a signal, if and only if its expectation is the prior. In this paper, we characterize the set of feasible distributions of posterior means. This provides us with a novel way to solve an important class of Bayesian persuasion problems.

[11]Specifically, the probability of a_2 is $1 - \frac{\gamma_2 - m_0 - z}{\gamma_2 - \gamma_1}$.

[12]Of course if λ is particularly high or particularly low, a corner solution will be optimal.

7 Appendix

7.1 Additional proofs

Lemma 1. *Fix any prior. Consider any convex function* $c : [0,1] \to \mathbb{R}$ *such that* $c_{\overline{\pi}}(x) \geq c(x) \geq c_{\underline{\pi}}(x)\ \forall x \in [0,1]$. *Let* $c'(x^+)$ *denote the right derivative of* c *at* x. *Then,* $0 \leq c'(x^+) \leq 1\ \forall x \in [0,1)$.

Proof. Note that since c is convex, it must be continuous.

We first establish the result for the degenerate case where $m_0 = 1$. In that case, $c_{\overline{\pi}}(x) = c_{\underline{\pi}}(x) = 0$ for all $x \in [0,1]$, so we must have $c(x) = 0$ for all $x \in [0,1]$ and hence $c'(x^+) = 0$ for all $x \in [0,1)$. From here on, we assume $m_0 < 1$.

Suppose that $c'(x_*^+) < 0$ for some $x_* \in [0,1)$. Since c is convex, its right derivative must be increasing so $c'(0^+) < 0$. Since $c_{\overline{\pi}} \geq c$ and $c_{\overline{\pi}}(0) = 0$, we have $c(0) \leq 0$. Thus, since $c'(0^+) < 0$, for a small enough x, we have $c(x) < 0$. But this cannot be since for all x, $c(x) \geq c_{\underline{\pi}}(x) \geq 0$.

Suppose that $c'(x_*^+) > 1$ for some $x_* \in [0,1)$. Since c is convex, we have that $c'(x^+) > 1$ for all $x > x_*$. Since $c_{\overline{\pi}} \geq c \geq c_{\underline{\pi}}$ and $c_{\overline{\pi}}(1) = c_{\underline{\pi}}(1) = 1 - m_0$, we have $c(1) = c_{\underline{\pi}}(1)$. Since $m_0 < 1$, we have that $lim_{x \to 1} c'_{\underline{\pi}}(x^+) = 1$. But then the fact that $c'(x^+) > 1$ for all $x > x_*$, combined with $c(1) = c_{\underline{\pi}}(1)$, implies that there exists an $x < 1$ s.t. $c(x) < c_{\underline{\pi}}(x)$ so we have reached a contradiction. □

7.2 Optimal signal with three actions

Suppose F_0 is uniform on $[0,1]$. Consider a function $f(x) = a + bx$ that is tangent to $c_{\overline{\pi}}$ to the left of γ_1 and crosses through the point (γ_1, z). Since F_0 is uniform, we know $c'_{\overline{\pi}}(x) = x$. Hence, f must be tangent to $c_{\overline{\pi}}$ at b and we have $f(b) = c_{\overline{\pi}}(b)$, which means $f(x) = -\frac{b^2}{2} + bx$. Since $f(\gamma_1) = z$, we have $\frac{-b^2}{2} + b\gamma_1 = z$. By the quadratic equation, this implies $b = \gamma_1 \pm \sqrt{\gamma_1^2 - 2z}$. From Figure 4 we can clearly see that b is increasing in z, so we know that that the smaller solution is the correct one: $b = \gamma_1 - \sqrt{\gamma_1^2 - 2z}$.

Since the probability of a_0 is b and the the probability of a_2 is $1 - \frac{\gamma_2 - \frac{1}{2} - z}{\gamma_2 - \gamma_1}$, the probability of a_1 is $\frac{\gamma_2 - \frac{1}{2} - z}{\gamma_2 - \gamma_1} - \gamma_1 + \sqrt{\gamma_1^2 - 2z}$. If we have $\gamma_1 = \frac{1}{3}$, $\gamma_2 = \frac{2}{3}$, and $\lambda = 3$, the overall payoff is $\left(\frac{\gamma_2 - \frac{1}{2} - z}{\gamma_2 - \gamma_1} - \gamma_1 + \sqrt{\gamma_1^2 - 2z}\right) + \left(1 - \frac{\gamma_2 - \frac{1}{2} - z}{\gamma_2 - \gamma_1}\right) 3$, which is maximized at $z = \frac{1}{24}$.

References

Aumann, Robert J., and Michael B. Maschler. 1995. *Repeated Games with Incomplete Information.* Cambridge, MA: MIT Press

Blackwell, David. 1953. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics.* Vol. 24, No. 2, 265-272.

Blackwell, David and M. A. Girshick. 1954. *Theory of Games and Statistical Decisions.* Wiley, New York.

Gentzkow, Matthew and Emir Kamenica. 2014. Costly persuasion. *American Economic Review: Papers & Proceedings.* Vol. 104, No. 5, 457-462.

Gentzkow, Matthew and Emir Kamenica. 2015a. Competition in persuasion. Working paper.

Gentzkow, Matthew and Emir Kamenica. 2015b. Bayesian persuasion with multiple senders and rich signal spaces. Working paper.

Gupta, Seher. 2014. Information design in collective decision games. Working paper.

Ivanov, Maxim. 2015. Optimal signals in Bayesian persuasion mechanisms with ranking. Working paper.

Kamenica, Emir and Matthew Gentzkow. 2011. Bayesian persuasion. *American Economic Review.* Vol. 101, No. 6, 2590-2615.

Kolotilin, Anton. Optimal information disclosure: Quantity vs. quality. Working paper.

Kolotilin, Anton, Ming Li, Tymofiy Mylovanov, and Andriy Zapechelnyuk. 2015. Persuasion of a privately informed receiver. Working paper.

Rothschild, Michael and Joseph Stiglitz. 1970. Increasing Risk: I. A definition. *Journal of Economic Theory.* Vol 2, 225-243.

Rothschild, Michael and Joseph Stiglitz. 1972. Addendum to "Increasing Risk: I. A definition." *Journal of Economic Theory.* Vol 5, 306.

Wang, Yun. 2015. Bayesian persuasion with multiple receivers. Working paper.