# PERSUASION OF A PRIVATELY INFORMED RECEIVER

ANTON KOLOTILIN, TYMOFIY MYLOVANOV, ANDRIY ZAPECHELNYUK
AND MING LI

ABSTRACT. We study persuasion mechanisms in linear environments. A receiver has a private type and chooses between two actions. A sender designs a persuasion mechanism or an experiment to disclose information about a payoff-relevant state. A persuasion mechanism conditions information disclosure on the receiver's report about his type, whereas an experiment discloses information independent of the receiver's type. We establish the equivalence of implementation by persuasion mechanisms and by experiments, and characterize optimal persuasion mechanisms.

*JEL Classification:* D82, D83, L82

*Keywords:* Bayesian persuasion, information disclosure, information design, mechanism design without transfers, experiments, persuasion mechanisms

## 1. Introduction

In numerous situations, an uninformed sender (she) wishes to influence an action of a receiver (he) who privately knows his preference type but is uncertain about a payoff-relevant state. The sender has full control over information disclosure about the state but cannot use any other incentive tools such as monetary transfers. If the sender knew the receiver's type, she would tailor information disclosure to the type. However, the sender can only ask the receiver to report his type, which the receiver may misreport in order to affect the information he receives. This begs the question of whether the sender can benefit by designing a complex *persuasion mechanism* that tailors information disclosure to the receiver's report, as compared to an *experiment* that provides the same information about the state, regardless of the receiver's report.

In our model, the receiver must choose one of two actions. The sender's and receiver's utilities depend linearly on the state and receiver's type.[1] The sender and receiver share a common prior about the state; the receiver privately knows his type. The sender commits to a persuasion mechanism that asks the receiver to report his type and returns a stochastic message that depends on the state and report. After receiving the message, the receiver updates his beliefs about the state and takes an action that maximizes his expected utility.

We characterize the set of the receiver's interim utilities implementable by persuasion mechanisms.[2] The upper and lower bounds are achieved by *full disclosure* and *no disclosure* of information about the state. For any mapping from messages to actions under a persuasion mechanism, the receiver's expected utility is linear in his type. Because each type chooses an optimal mapping, the receiver's interim utility is an upper envelope of linear functions and, hence, is convex. To sum up, an implementable interim utility is necessarily a convex function that lies between the upper and lower bounds generated by full and no disclosure.

Our main theorem shows that (1) this necessary condition is also sufficient and that (2) any implementable interim utility can be attained by an experiment. Moreover, in our model the receiver's interim utility uniquely determines the sender's expected utility. Therefore, there is no loss of generality in restricting attention to experiments. In particular, the sender need not consider more complex persuasion mechanisms that condition information disclosure on the receiver's report.

We now outline the argument behind this result. Because the utilities are linear in the state, the only relevant content of a message about the state is the posterior mean. Therefore, an experiment can be fully described by the distribution of the posterior mean state $H$ that it generates. By Blackwell (1953), there exists an experiment that generates $H$ if and only if the prior distribution of the state is a mean-preserving spread of $H$. By linearity, the receiver's interim utility $U$ can be represented as an

---

[1]For our analysis to apply, we only need to assume that the utilities are linear in some transformation of the state and are arbitrary functions of the receiver's type. The state and receiver's type are one-dimensional independent random variables.

[2]The receiver's interim utility is the mapping from the receiver's type to his expected utility given optimal behavior under a persuasion mechanism.

appropriate integral of $H$. Using the integral form of the mean-preserving spread condition as in Blackwell (1951), we show that the prior is a mean-preserving spread of $H$ if and only if $U$ is convex and lies between the upper and lower bounds generated by full and no disclosure. Therefore, any $U$ that satisfies the necessary conditions that we identified above is implementable by an experiment.

This characterization allows us to formulate the sender's problem as the maximization of a linear functional on a bounded set of convex functions. We derive a general structure of optimal persuasion mechanisms and use it to simplify the sender's problem into a finite-variable optimization problem.

**Related Literature.** Our model is a variation of Kamenica and Gentzkow (2011), who show that an optimal experiment corresponds to a point on the concave closure of the sender's value function. This concavification approach does not rely on any structure of the problem such as the linearity of utility functions or the two actions of the receiver. However, as Gentzkow and Kamenica (2016) point out, this approach has limited applicability when the state space is infinite because it requires to calculate a concave closure on the infinite-dimensional space of distributions over the state space.

Rayo and Segal (2010) impose more structure on the receiver's utility and the distribution of the receiver's type than our paper. At the same time, they allow the sender's utility to be nonlinear in the state. Rayo and Segal (2010) partially characterize optimal experiments and represent the sender's problem as the maximization problem of a concave objective function subject to linear constraints. In our setting, their assumptions about the receiver would imply that either full or no disclosure is optimal.

Kolotilin (2016) allows for nonlinear sender's and receiver's utility functions and an arbitrary joint distribution of the state and the receiver's type. The linear-programming approach in Kolotilin (2016) permits to verify whether a candidate experiment is optimal. But this approach has limited applicability because it does not allow to directly characterize the optimal experiments.

The three papers above study experiments.[3] In contrast, we consider persuasion mechanisms, which can tailor information disclosure to the receiver's report.[4] Linear utility functions with two possible actions of the receiver enable us to use the envelope representation of incentive compatibility as in Myerson (1981). However, the characterization of implementable mechanisms in our setting differs from Myerson (1981) because there are no transfers between the sender and receiver, and there are

---

[3]There is a rapidly growing Bayesian persuasion literature that studies optimal experiments. Bayesian persuasion with a privately informed sender is considered in Gill and Sgroi (2008, 2012), Perez-Richet (2014), Alonso and Câmara (2016b), and Hedlund (2017). Bayesian persuasion with multiple senders is analyzed in Ostrovsky and Schwarz (2010), Board and Lu (2016), Gentzkow and Kamenica (2017a, 2017b), and Li and Norman (2017). Dynamic Bayesian persuasion is examined in Au (2015), Ely et al. (2015), and Ely (2017). Information acquisition and the value of information in Bayesian persuasion are explored in Gentzkow and Kamenica (2014, 2017c), Kolotilin (2015), Alonso and Câmara (2016a), and Bergemann and Morris (2016).

[4]Focusing on experiments is *with* loss of generality in settings with multiple receivers in Bergemann and Morris (2013), Alonso and Câmara (2016c), Mathevet et al. (2017), and Taneva (2016).

obedience constraints instead of participation constraints.[5] The equivalence between persuasion mechanisms and experiments relies on the majorization theory of Hardy et al. (1929) and, thus, relates to the equivalence between Bayesian and dominant-strategy implementation in Manelli and Vincent (2010) and Gershkov et al. (2013).

## 2. Model

2.1. **Setup.** There are two players: sender and receiver. The receiver makes a choice $a \in A = \{0, 1\}$ between action ($a = 1$) and inaction ($a = 0$). There are two payoff-relevant random variables: the state of the world $\omega \in \Omega$ and the receiver's private type $r \in R$, where $\Omega$ and $R$ are intervals in the real line. Random variables $\omega$ and $r$ are independent and have distributions $F$ and $G$.

Let the receiver's and sender's utilities be

$$
\begin{aligned}
u(\omega, r, a) &= a \cdot (\omega - r), \\
v(\omega, r, a) &= a + \rho(r) u(\omega, r, a),
\end{aligned}
\tag{1}
$$

where $\rho$ is a bounded measurable function.[6]

The receiver's utility from inaction ($a = 0$) is normalized to zero, whereas his utility from action ($a = 1$) equals the benefit $\omega$ less the private cost $r$. The sender's utility is a weighted sum of the receiver's utility and action. The sender is biased towards the receiver's action but also puts a type-specific weight $\rho(r)$ on the receiver's utility. In particular, if the weight $\rho$ is large, then the sender's and receiver's interests are aligned, whereas if the weight is zero, then the sender cares only about whether the receiver acts or not.

We assume that the set of states is $\Omega = [0, 1]$ and $\mathbb{E}[\omega] \in (0, 1)$. Because the optimal action of the receiver of any type above 1 or below 0 is independent of the state, we assume that the set of types is $R = [0, 1]$. These assumptions allow for elegance of presentation; relaxing them poses no difficulty.

2.2. **Persuasion Mechanisms.** In order to influence the action taken by the receiver, the sender can design a mechanism that asks the receiver to report his private information and sends a message to the receiver conditional on his report and the realized state.

A *persuasion mechanism* $\pi$ asks the receiver to report $\hat{r} \in R$ and then recommends him to take one of two actions: for every $\omega \in \Omega$, it recommends to act ($\hat{a} = 1$) with probability $\pi(\omega, \hat{r})$ and not to act ($\hat{a} = 0$) with complementary probability $1 - \pi(\omega, \hat{r})$, where $\pi : \Omega \times R \to [0, 1]$ is a measurable function.

A persuasion mechanism is *incentive compatible* if the receiver finds it optimal to report his true type and to follow the mechanism's recommendation. Denote the

---

[5]Bayesian persuasion with monetary transfers is investigated in Bergemann and Pesendorfer (2007), Esö and Szentes (2007), Bergemann et al. (2015, 2017a, 2016b), Li and Shi (2016), and Hörner and Skrzypacz (2016).

[6]For our analysis, it is sufficient to assume that $u(\omega, r, a)$ and $v(\omega, r, a)$ are linear in (a transformation of) the state $\omega$. The unidimensionality of the type $r$ and the linearity of the utilities in the type are assumed only for elegance of exposition. We further comment on this in Section 3.3, and refer an interested reader to Appendix C of the Supplemental Material for technical details.

interim utility of a receiver of type $r \in R$ who reports $\hat{r} \in R$ and takes actions $a_0 \in \{0, 1\}$ and $a_1 \in \{0, 1\}$ after recommendations $\hat{a} = 0$ and $\hat{a} = 1$, respectively, by

$$U_\pi(r, \hat{r}, a_0, a_1) = \int_\Omega \left( a_0(1 - \pi(\omega, \hat{r})) + a_1\pi(\omega, \hat{r}) \right) (\omega - r)\mathrm{d}F(\omega).$$

The interim utility of the truthful ($\hat{r} = r$) and obedient ($a_0 = 0$ and $a_1 = 1$) receiver is equal to

$$U_\pi(r) = U_\pi(r, r, 0, 1) = \int_\Omega \pi(\omega, r)(\omega - r)\mathrm{d}F(\omega).$$

We consider mechanisms that satisfy the incentive compatibility constraint

$$U_\pi(r) \geq U_\pi(r, \hat{r}, a_0, a_1) \text{ for all } r, \hat{r} \in R \text{ and } a_0, a_1 \in A. \tag{2}$$

It is convenient to introduce the notation for the interim utility of the *obedient* receiver of type $r$, who reports $\hat{r}$ and then obeys the recommendation of the mechanism:

$$U_\pi(r, \hat{r}) = U_\pi(r, \hat{r}, 0, 1) = p_\pi(\hat{r}) - q_\pi(\hat{r}) r,$$

where $q_\pi(\hat{r})$ is the interim action and $p_\pi(\hat{r})$ is the expected state after action conditional on a report $\hat{r}$:

$$q_\pi(\hat{r}) = \int_\Omega \pi(\omega, \hat{r})\mathrm{d}F(\omega) \quad \text{and} \quad p_\pi(\hat{r}) = \int_\Omega \omega\pi(\omega, \hat{r})\mathrm{d}F(\omega).$$

With these notations, we can draw the parallel to the standard mechanism design problem with transfers, where $r$ is a private value, $\hat{r}$ is a reported value, $q_\pi(\hat{r})$ is the probability of transaction, and $p_\pi(\hat{r})$ is the expected monetary transfer. The classical envelope argument yields the following lemma.

**Lemma 1.** *A mechanism $\pi$ is incentive compatible if and only if*

$$q_\pi \quad \text{is nonincreasing,} \tag{3}$$

$$U_\pi(r) = \int_r^1 q_\pi(s)\mathrm{d}s, \tag{4}$$

$$U_\pi(0) = \mathbb{E}[\omega]. \tag{5}$$

The proof is in the Appendix.

Interestingly, the obedience constraints for the intermediate types are implied by the boundary conditions, $U_\pi(1) = 0$ and $U_\pi(0) = \mathbb{E}[\omega]$, and truth telling, $U_\pi(r) \geq U_\pi(r, \hat{r})$. To disobey by ignoring the recommendation, that is, to act (not to act) irrespective of what is recommended, is not better than pretending to be the lowest type $\hat{r} = 0$ (the highest type $\hat{r} = 1$, respectively). Due to the linearity of the receiver's utility, to misreport and disobey by taking the opposite action to the recommended one is always inferior to reporting either the lowest type, $\hat{r} = 0$, or the highest type, $\hat{r} = 1$, and then following the recommendation.

In our model, there are no transfers, and there are obedience constraints instead of individual rationality constraints. These differences between our and the standard environment with transfers translate into the following differences in characterization.

First, there are two boundary conditions, $U_\pi(1) = 0$ and $U_\pi(0) = \mathbb{E}[\omega]$:

(a) We have $\omega - r \leq 0$ for all $\omega \in \Omega = [0, 1]$ and $r = 1$. Hence, type 1's utility is maximized by not acting for any belief about the state, implying $U_\pi(1) = 0$. This is (4) evaluated at $r = 1$.

(b) We have $\omega - r \geq 0$ for all $\omega \in \Omega = [0, 1]$ and $r = 0$. Hence, type 0's utility is maximized by acting for any belief about the state, implying $U_\pi(0) = \mathbb{E}[\omega]$. This is (5).

Second, not every pair $(q, U)$ that satisfies conditions (3)–(5) is feasible, that is, a mechanism $\pi$ that implements such a pair need not exist. For example, if $F$ assigns probability 1 on $\omega = 1/2$, then every nonincreasing $q$ with $\int_0^1 q(r)\mathrm{d}r = 1/2$ satisfies (3)–(5). Among these functions, $q$ is feasible if and only if it satisfies $q(r) = 1$ for $r < 1/2$ and $q(r) = 0$ for $r > 1/2$.

2.3. **Experiments.** An experiment communicates a one-way message to the receiver and does not require the receiver to report his type. Formally, an *experiment* $\sigma$ sends to the receiver a random message $m$ from a message space $M = [0, 1]$ that depends on a realized state $\omega$. Denote by $\sigma(m|\omega)$ the distribution of message $m$ conditional on state $\omega$.

A persuasion mechanism can be interpreted as a menu of experiments from which the receiver can freely choose one experiment. For a given mechanism $\pi$, the corresponding menu of experiments is $\{\sigma_{\hat{r}}\}_{\hat{r} \in R}$, where, conditional on a state $\omega$, an experiment $\sigma_{\hat{r}}$ sends messages 1 and 0 with probabilities $\pi(\omega, \hat{r})$ and $1 - \pi(\omega, \hat{r})$. Conversely, by the revelation principle, any equilibrium outcome induced by any menu of experiments offered to the receiver can be replicated by a persuasion mechanism.

For a given experiment $\sigma$, each message $m$ induces a posterior belief of the receiver about the state. Because the receiver's utility is monotonic in his type, we can identify every message $m$ with the cutoff type $r$ who is indifferent between the two actions conditional on receiving this message. An experiment is *direct* if its messages are equal to the cutoff types, $m = \mathbb{E}[\omega|m]$. Without loss of generality, we focus on direct experiments (as in, e.g., Kamenica and Gentzkow, 2011).

A persuasion mechanism $\pi$ is *equivalent to an experiment* if, conditional on each state, the behavior of the receiver as stipulated by $\pi$ is a best-reply behavior to the posterior beliefs generated by some direct experiment $\sigma$,

$$\pi(\omega, r) \in \left[1 - \sigma(r|\omega), 1 - \sigma(r_-|\omega)\right] \quad \text{for all } \omega \in \Omega \text{ and all } r \in R, \qquad (6)$$

where $\sigma(r_-|\omega)$ denotes the left limit of $\sigma(.|\omega)$ at $r$. To understand (6), note that, upon receiving a direct message $m$, every type $r < m$ strictly prefers to choose $a = 1$, and type $r = m$ is indifferent between $a = 0$ and $a = 1$ and may optimally choose any lottery over the two actions. Consequently, the probability that type $r$ acts conditional on state $\omega$ can take any value in the interval $[1 - \sigma(r|\omega), 1 - \sigma(r_-|\omega)]$.

## 3. Implementation Equivalence

In this section, we characterize the pairs of the interim utility $U$ and action $q$ implementable by persuasion mechanisms, and show that the same pairs of the interim utility and action are implementable by experiments.

This is a step towards solving the sender's optimization problem, because the sender's interim utility $V_\pi(r)$ when the receiver's type is $r$ is a weighted sum of the receiver's interim utility and action,

$$V_\pi(r) = \int_\Omega (1 + \rho(r)(\omega - r))\pi(\omega, r)\mathrm{d}F(\omega) = q_\pi(r) + \rho(r)U_\pi(r). \tag{7}$$

### 3.1. Bounds on Receiver's Interim Utility.

Consider two simple mechanisms. The *full disclosure* mechanism informs the receiver about the state, so the receiver acts ($\pi(\omega, r) = 1$) if $\omega > r$ and does not act ($\pi(\omega, r) = 0$) if $\omega < r$. Therefore, the interim utility is

$$\overline{U}(r) = \int_r^1 (\omega - r)\,\mathrm{d}F(\omega).$$

The *no disclosure* mechanism does not convey any information to the receiver, so the receiver acts if $\mathbb{E}[\omega] > r$ and does not act if $\mathbb{E}[\omega] < r$. Therefore, the interim utility is

$$\underline{U}(r) = \max\{\mathbb{E}[\omega] - r, 0\}.$$

Thus, $\underline{U}(r)$ is the receiver's interim utility based on the prior information about $\omega$ as given by $F$, whereas $\overline{U}(r)$ is the receiver's interim utility if he observes $\omega$.

Note that every mechanism $\pi$ must satisfy

$$\underline{U}(r) \le U_\pi(r) \le \overline{U}(r) \quad \text{for all } r \in R. \tag{8}$$

The left-hand side inequality of (8) is implied by incentive compatibility: the receiver cannot be better off by ignoring the sender's recommendation. The right-hand side inequality of (8) is the *feasibility constraint*: the receiver's utility cannot exceed the utility attained under full disclosure of $\omega$.

### 3.2. Implementable Interim Utility and Action.

The receiver's interim utility $U$ and action $q$ are *implementable* if there exists an incentive-compatible persuasion mechanism $\pi$ such that $U(r) = U_\pi(r)$ and $q(r) = q_\pi(r)$ for all $r \in R$. Moreover, $U$ and $q$ are *implementable by an experiment* if $\pi$ is equivalent to an experiment.

Let $\mathcal{U}$ be the set of all convex functions bounded by $\underline{U}$ and $\overline{U}$ (see Fig. 1).

**Theorem 1.** *The following statements are equivalent:*
*(a) $U$ is a convex function between $\underline{U}$ and $\overline{U}$;*
*(b) $U$ is implementable;*
*(c) $U$ is implementable by an experiment.*

*Proof.* Observe that $(a)$ states the necessary conditions for the implementation of $U$. The incentive-compatibility constraint requires the convexity of $U$ by Lemma 1, and the feasibility constraint (8) requires $\underline{U}(r) \le U(r) \le \overline{U}(r)$. Hence, $(b) \Rightarrow (a)$. Also, the implication $(c) \Rightarrow (b)$ is trivial by definition. It remains to show that $(a) \Rightarrow (c)$.

Let $U \in \mathcal{U}$. Define $H(r) = 1 + U'(r_+)$, where $U'(r_+)$ denotes the right-derivative of $U$ at $r$, so that $H$ is right-continuous. Since $\underline{U}(0) = \overline{U}(0) = \mathbb{E}[\omega]$ and $\underline{U}'(0_+) = -1$, we have $H(0) = 0$; since $\underline{U}(1) = \overline{U}(1) = 0$ and $\underline{U}'(1_+) = \overline{U}'(1_+) = 0$, we have $H(1) = 1$.
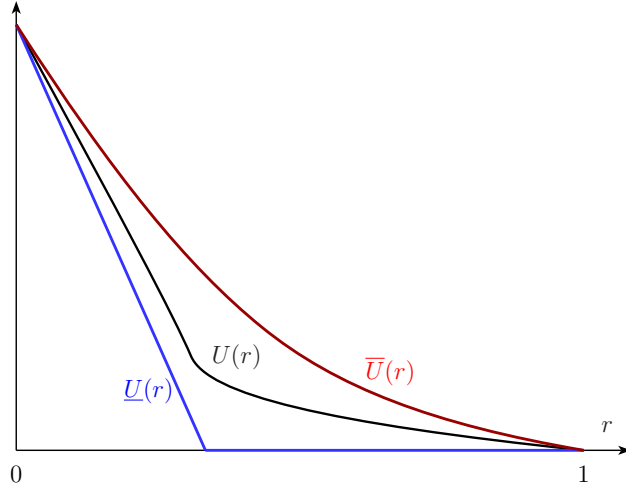
FIGURE 1. Set $\mathcal{U}$ contains every convex function between $\underline{U}$ and $\overline{U}$.

Also, since $U$ is convex, $H$ is nondecreasing. Hence, $H$ is a distribution function. Next, observe that

$$\int_r^1 (1 - H(s))\mathrm{d}s = U(r) \leq \overline{U}(r) = \int_r^1 (1 - F(s))\mathrm{d}s \tag{9}$$

for all $r$, with equality at $r = 0$, because $\underline{U}(0) = \overline{U}(0) = \mathbb{E}[\omega]$. That is, $F$ is a mean-preserving spread of $H$. By Blackwell (1953), there exists a joint distribution function $P(\omega, m)$ such that the marginal distribution of $\omega$ is $F$, the marginal distribution of $m$ is $H$, and $\mathbb{E}_P[\omega|m] = m$ for all $m$.

Now, consider an experiment $\sigma(m|\omega) = P(m|\omega)$. By construction, $\mathbb{E}[\omega|m] = m$ for all $m$ of $\sigma$, and the probability that $\sigma$ generates message $m \leq x$, for any given value $x$, is $H(x)$. Thus, $\sigma$ induces type $r$ to act with probability $q_\sigma(r) \in [1 - H(r), 1 - H(r_-)]$, where indeterminacy arises at each discontinuity point $r$ of $H$, because type $r$ is indifferent between the two actions and can, therefore, optimally choose any lottery over these actions. Finally, (4) implies that $\sigma$ implements $U$. □

Theorem 1 yields the following characterization of implementable interim actions.

**Corollary 1.** *The following statements are equivalent:*
*(a) $q$ is a nonincreasing function that satisfies*

$$\int_r^1 q(\omega)\mathrm{d}\omega \leq \int_r^1 (1 - F(\omega))\mathrm{d}\omega \quad \text{for all } r, \text{ with equality at } r = 0; \tag{10}$$

*(b) $q$ is implementable;*
*(c) $q$ is implementable by an experiment.*

The heart of the characterization result, Theorem 1, is that every implementable pair of $U$ and $q$ is implementable by an experiment. This result relies on the following connection between the Mirrlees (1971) representation of incentive compatibility and

the Blackwell (1953) representation of garbling. Incentive compatibility (4) and feasibility (8) imply that every implementable interim utility $U$ and action $q$ must satisfy the requirements in parts (a) of Theorem 1 and Corollary 1. Part (a) of Corollary 1 states that $F$ is a mean-preserving spread of $1-q$. But any such $q$ can be implemented by an appropriate garbling of the full disclosure experiment.[7] Because each interim utility $U$ is pinned down by some interim action $q$ through (4), we obtain the result.

We now highlight a connection to the literature on the equivalence of Bayesian and dominant-strategy incentive compatibility in linear environments with transfers (Manelli and Vincent, 2010, and Gershkov et al., 2013). Using Gutmann et al. (1991), Gershkov et al. (2013) show that for a given monotonic expected allocation (Bayesian incentive-compatible mechanism) there exists a monotonic ex-post allocation (dominant-strategy incentive-compatible mechanism) that delivers the same expected allocation. Relatedly, using Blackwell (1953), we show that for a given nonincreasing $q$ that satisfies (10) there exists an incentive compatible $\pi(\omega, r)$ that is nonincreasing in $r$ for each $\omega \in \Omega$ and satisfies

$$\int_\Omega \pi(\omega, r)\mathrm{d}F(\omega) = q(r) \ \text{ for each } r \in R.$$

Since $\pi(r, \omega) \in [0, 1]$ is nonincreasing and, w.l.o.g., right-continuous in $r$, we can interpret the function $\sigma(r|\omega) = 1-\pi(r, \omega)$ as a distribution of messages for every given state $\omega$. That is, $\sigma$ is an experiment, and mechanism $\pi$ is equivalent to experiment $\sigma$. Both Blackwell (1953) and Gutmann et al. (1991) are based on the majorization theory initiated by Hardy et al. (1929).

3.3. **Discussion.** We now discuss extensions and limitations of Theorem 1, when the sender's and receiver's utility functions are nonlinear and the receiver has more than two actions.

**Linearity in Type.** The result in Theorem 1 can be easily extended to the case where the utilities are linear in (a transformation of) the state $\omega$, but are arbitrary functions of the type $r$. Suppose that

$$u(\omega, r, a) = a \cdot (c(r) + b(r)d(\omega)) \quad \text{and} \quad v(\omega, r, a) = a + \rho(r)u(\omega, r, a),$$

where $b$, $c$, $d$, and $\rho$ are bounded measurable functions. In addition, assume that $b(r)$ is strictly positive, and $c(r)/b(r)$ is strictly monotone.[8] All other details of the model remain as in Section 2; in particular, the receiver chooses one of two actions, $a \in \{0, 1\}$. We now demonstrate how a change of variables gives back the problem formulated in Section 2. Define new variables

$$\tilde{\omega} = d(\omega) \quad \text{and} \quad \tilde{r} = -c(r)/b(r),$$

---

[7] Suppose that an experiment induces the distribution of posterior means equal to $H = 1 - q$. It is optimal for a type $r$ to act if the posterior mean is at least $r$, which occurs with probability $1 - H(r) = q(r)$. Therefore, this experiment implements $q$.

[8] These assumptions are unnecessary for our results and are made for elegance of exposition. A more general case of utility functions linear in $\omega$ is analyzed in Appendix C of the Supplemental Material.

and let

$$\tilde{u}(\tilde{\omega}, \tilde{r}, a) = \frac{u(\omega, r, a)}{b(r)} = a \cdot \left( \frac{c(r)}{b(r)} + d(\omega) \right) = a \cdot (\tilde{\omega} - \tilde{r}) \, .$$

The utility of each type $r$ is now multiplied by the factor $1/b(r)$, which has no effect on the optimal choice of action $a \in \{0, 1\}$ for this type. Next, define $\tilde{\rho}(\tilde{r}) = \rho(r)b(r)$ and let

$$\tilde{v}(\tilde{\omega}, \tilde{r}, a) = v(\omega, r, a) = a + \rho(r)b(r)\frac{u(\omega, r, a)}{b(r)} = a + \tilde{\rho}(\tilde{r})\tilde{u}(\tilde{\omega}, \tilde{r}, a).$$

Thus, after this change of variables, the sender's and receiver's utilities have the functional forms identical to (1).

**Linearity in State.** The linearity of the utilities in (a transformation of) the state implies that a posterior distribution of the state can be summarized into a scalar, the posterior mean. For any message of any experiment, the receiver acts if and only if his type is below the cutoff equal to the posterior mean state conditional on that message. An experiment induces a probability distribution over these cutoffs. A persuasion mechanism, which is a menu of experiments, induces a compound distribution over cutoffs, which is itself an experiment.

Without the linearity in the state, a posterior distribution of the state cannot be summarized into a scalar. Even if the receiver's utility is increasing in the state and decreasing in the type, different types of the receiver may have different attitudes to risk. Thus, there may be a value of screening types by offering the receiver a menu of experiments.

For example, consider an experiment that sends one message if $\omega \geq \omega'$ and another message otherwise. For a type $r'$ that is exactly indifferent between $a = 0$ and $a = 1$ upon learning that $\omega \geq \omega'$, this experiment leaves type $r'$ with zero surplus and uniquely maximizes the probability that type $r'$ acts. Under this experiment, a nearby type $r'' < r'$ also acts if and only if $\omega \geq \omega'$. Without linearity, it is possible to design another experiment, preferred by $r''$ but not by $r'$, which induces type $r''$ to act with a different probability. Thus, the persuasion mechanism corresponding to the menu of these two experiments implements the interim actions that cannot be implemented by any experiment.

In Appendix B of the Supplemental Material, we formulate the model in which the utilities are nonlinear in the state, characterize persuasion mechanisms that are equivalent to experiments, and demonstrate that the equivalence of implementation by persuasion mechanisms and by experiments generally fails.

**Binary Action Space.** The assumption that the receiver chooses one of *two* actions is crucial for Theorem 1 to hold. Relaxing this assumption will affect the result in two ways.

First, the number of actions affects the complexity of the obedience constraints. With two actions, there are only two relevant ways to disobey a mechanism, by ignoring the mechanism's message and either always act or never act. However, the outcomes of these deviations can be achieved by reporting the extreme types, $\hat{r} = 0$ and $\hat{r} = 1$, which always act and never act, respectively. Thus, the incentive

to report the type truthfully automatically implies obedience. With more than two actions, there are more ways to disobey than just ignore the messages. The obedience constraints are nontrivial and are not implied by the truthtelling. Consequently, the implication $(a) \Rightarrow (b)$ of Theorem 1 (that every convex function between $\underline{U}$ and $\overline{U}$ is implementable) is false, because the obedience constraints impose additional necessary conditions for implementability.

Second, with more than two actions, there exist outcomes implementable by persuasion mechanisms but not by experiments. For example, suppose that there are three actions available to the receiver, $A = \{0, 1, 2\}$, and consider a persuasion mechanism described by a menu of experiments that send binary messages, "high" or "low". Under this mechanism, each type of the receiver will choose a preferred experiment from the menu, and then choose one of two best-response actions to the messages. Generally, this mechanism is not implementable by an experiment. If an experiment sends a binary message (the same for all types), then, generally, there will be types who would have chosen a different experiment from the menu and would have acted differently under the original persuasion mechanism. If, instead, an experiment sends multiple messages, then, generally, there will be types that have more than two best-response actions to those messages, thus behaving differently than under the original mechanism. Consequently, the implication $(b) \Rightarrow (c)$ of Theorem 1 (that every implementable $U$ is implementable by an experiment) does not hold.

In Appendix A of the Supplemental Material, we formulate the model with multiple actions of the receiver, characterize the implementable receiver's interim utilities, and show that the sender can generally implement a strictly larger set of the receiver's interim utilities and actions by persuasion mechanisms than by experiments.

## 4. Optimal Mechanisms

In this section, we use Theorem 1 to characterize the persuasion mechanisms that are optimal for the sender under additional smoothness assumptions. We assume that the weight $\rho$ in the sender's utility is continuous in the receiver's type and that the distribution of types $G$ admits a continuously differentiable density $g$.

4.1. **Sender's Problem.** The sender seeks an incentive-compatible persuasion mechanism $\pi$ that maximizes

$$\int_R V_\pi(r)\mathrm{d}G(r),$$

where $V_\pi(r)$ is the sender's interim utility when the receiver's type is $r$, as defined by (7). The following lemma is a useful tool for finding optimal persuasion mechanisms. It expresses the sender's expected utility as a function of the receiver's interim utility.

**Lemma 2.** *For every incentive-compatible mechanism $\pi$,*

$$\int_R V_\pi(r)\mathrm{d}G(r) = g(0)\mathbb{E}[\omega] + \int_R U_\pi(r)I(r)\mathrm{d}r,$$

*where*
$$I(r) = g'(r) + \rho(r)g(r) \quad \textit{for all } r \in R.$$

This formulation of the sender's objective is similar to the Myerson (1981) formulation of the seller's objective in terms of virtual valuations. Both formulations account for the designer's biased preferences and information rents left to the privately informed parties. Lemma 2 relies on the assumption that the sender's utility is a weighted sum of the receiver's utility and action. By incentive compatibility (Lemma 1), the receiver's interim action is the derivative of his interim utility. The lemma then follows by integration by parts.

*Proof.* Observe that, by (4) and (7), we have
$$\int_R V_\pi(r)\mathrm{d}G(r) = \int_R \left(q_\pi(r) + \rho(r)U_\pi(r)\right)g(r)\mathrm{d}r = \int_R \left(-U'_\pi(r_+) + \rho(r)U_\pi(r)\right)g(r)\mathrm{d}r.$$
Using integration by parts, we obtain
$$-\int_R U'_\pi(r_+)g(r)\mathrm{d}r = -U_\pi(r)g(r)\big|_0^1 + \int_R U_\pi(r)g'(r)\mathrm{d}r.$$
Since $-U_\pi(r)g(r)\big|_0^1 = \mathbb{E}[\omega]g(0)$ by (5), the lemma follows.  □

By Theorem 1, the receiver's interim utility is implementable by some persuasion mechanism if and only if it is in $\mathcal{U}$. Hence, the sender's problem can be expressed as
$$\max_{U \in \mathcal{U}} \int_R U(r)I(r)\mathrm{d}r. \tag{11}$$
We say that $U$ is *optimal* if it solves the above problem. For a given optimal $U$, the corresponding optimal experiment is defined by the distribution $H$ of the posterior mean state,
$$H(r) = 1 + U'(r_+) \quad \text{for all } r \in [0,1].$$

4.2. **Structure of Optimal Mechanisms.** We characterize the structure of optimal mechanisms under the assumption that function $I$ is nonzero almost everywhere and changes sign $n \geq 0$ times.[9] Let $\{r_1, r_2, \ldots, r_n\} \subset (0,1)$ be the set of types at which $I$ changes its sign, and let $r_0 = 0$ and $r_{n+1} = 1$.

Clearly, as follows from (11), on any interval $(r_i, r_{i+1})$ where $I(r)$ is positive, for any given values of $U(r_i)$ and $U(r_{i+1})$, the optimality requires that $U(r)$ is pointwise maximized subject to feasibility ($U \leq \overline{U}$) and the convexity of $U$ on $[r_i, r_{i+1}]$. That is, the interim utility $U$ on $[r_i, r_{i+1}]$ is a straight line that passes through the endpoints $U(r_i)$ and $U(r_{i+1})$, unless $U(r) = \overline{U}(r)$ for some $r \in [r_i, r_{i+1}]$, as shown in Fig. 2. Formally,

(P$_1$) On every interval $(r_i, r_{i+1})$ where $I(r)$ is positive, $U$ is the greatest convex function that passes through the endpoints $U(r_i)$ and $U(r_{i+1})$ and does not exceed $\overline{U}$.

---

[9] If there are intervals where $I(r) = 0$, then on those intervals the sender is indifferent about the choice of $U$. Hence, multiple solutions emerge in this case. Characterization of these solutions is a straightforward but tedious extension of the result in this section.
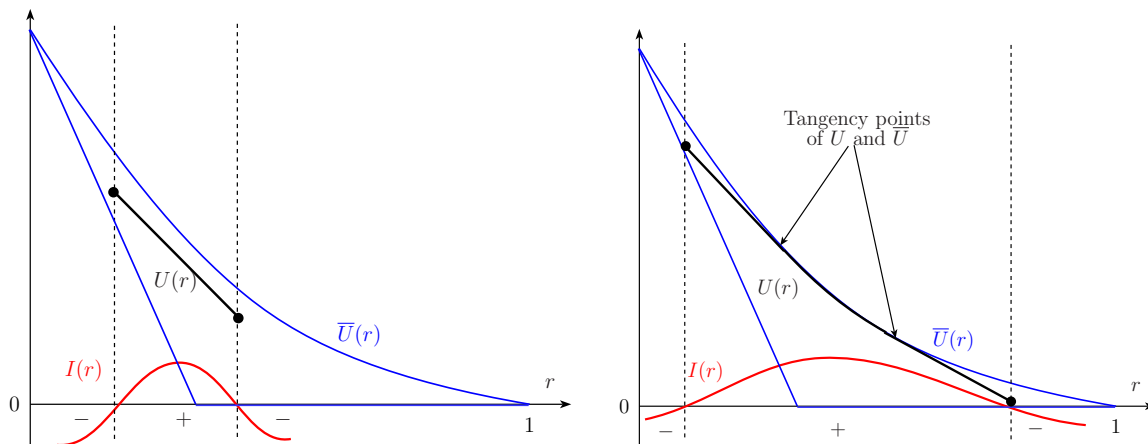
FIGURE 2. The optimal interim utility on the interval where $I(r)$ is positive.
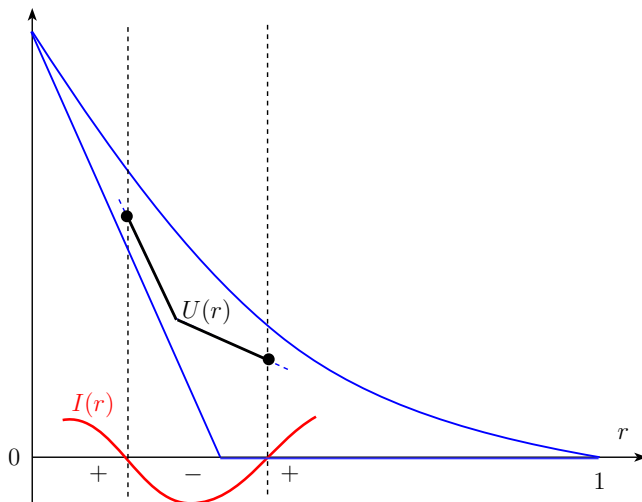


FIGURE 3. The optimal interim utility on the interval where $I(r)$ is negative.

Similarly, on any interval $(r_i, r_{i+1})$ where $I(r)$ is negative, for any given pairs of $(U(r_i), U'(r_i))$ and $(U(r_{i+1}), U'(r_{i+1}))$, the optimality requires that $U(r)$ is pointwise minimized subject to $U \geq \underline{U}$ and the convexity of $U$ on $[r_i, r_{i+1}]$. That is, the interim utility $U$ on $[r_i, r_{i+1}]$ is an upper envelope of two straight lines that pass through the endpoints $U(r_i)$ and $U(r_{i+1})$ and have slopes $U'(r_i)$ and $U'(r_{i+1})$, as shown in Fig. 3. Formally,

(P$_2$) On every interval $(r_i, r_{i+1})$ where $I(r)$ is negative, $U$ is piecewise linear with at most one kink and satisfies

$$U(r) = \max \left\{ U(r_i) + U'(r_i)(r - r_i), U(r_{i+1}) + U'(r_{i+1})(r - r_{i+1}) \right\}.$$

Therefore, we can reduce the sender's problem (11) of optimization on the function space $\mathcal{U}$ to an $n$-variable optimization problem. The optimal $U$ is pinned down by properties (P$_1$) and (P$_2$) within each interval $(r_i, r_{i+1})$ where the sign of $I$ is constant.

Thus, the optimal $U$ is fully defined by the utility values at the points $\{r_1, ..., r_n\}$ where $I$ changes its sign.

For every vector $\bar{y} = (y_1, ..., y_n) \in \mathbb{R}^n$, consider the set of all $U \in \mathcal{U}$ that satisfy $U(r_i) = y_i$ for all $i = 1, ..., n$. If this set is nonempty, the properties (P$_1$) and (P$_2$) uniquely determine the optimal $U$ on this set, denoted by $U_{\bar{y}}^*$. For completeness, define $U_{\bar{y}}^* = \underline{U}$ if this set is empty. Thus, we obtain:

**Theorem 2.** *The sender's problem* (11) *is an $n$-variable optimization problem,*

$$\max_{\bar{y} \in \mathbb{R}^n} \int_R U_{\bar{y}}^*(r) I(r) \mathrm{d}r.$$

Properties (P$_1$) and (P$_2$) imply that the optimal $U$ is piecewise linear except for the intervals where $U(r) = \overline{U}(r)$, as shown in Figs. 2–3. Recall that, for a given optimal $U$, the corresponding optimal experiment induces the distribution of the mean state given by $H(r) = 1 + U'(r_+)$ for all $r \in [0, 1]$. Therefore, the intervals on which $U$ is linear correspond to constant $H$ (i.e., zero density). The kinks of $U$ correspond to the mass points of the distribution equal to the difference between the right and left derivatives of $U$ at those points. The intervals on which $U(r) = \overline{U}(r)$ correspond to the full disclosure intervals, $H(r) = F(r)$.

**Example 1.** Suppose that $I(r) = g'(r) + \rho(r)g(r)$ crosses the horizontal axis at most once and from above. This is the case, for instance, if the density $g(r)$ is single-peaked and $\rho(r) = 0$. Then, Theorem 2 implies that there exists a threshold $\omega^*$ such that the optimal experiment reveals the state if it is below $\omega^*$ and sends the same pooling message for all states above $\omega^*$.

## 5. Conclusion

In this paper, we study the model of persuasion of a privately informed receiver in linear environments, and show the equivalence of persuasion by experiments and by persuasion mechanisms. A privately informed receiver can be interpreted as a random draw from a heterogeneous population of receivers. Within this interpretation, a persuasion mechanism establishes a private communication channel with each individual in the population; it asks each individual to report his type, and returns a private message about the state. In contrast, an experiment broadcasts the same message to the entire population. Our equivalence result can be viewed as the equivalence between persuasion via private communication to receivers and by broadcasting information to the whole population. In particular, so long as the receivers' types are unobserved by the sender, there is no loss of optimality in publicity of messages and free information sharing among the receivers.

Within the above interpretation, it would be natural to assume that the receivers' payoffs depend not only on their own actions, but also on collective actions in the population, thus being affected by network externalities. This extension is beyond the scope of this paper and it is left open for future research.

## APPENDIX. PROOF OF LEMMA 1

*Necessity.* Incentive compatibility (2) requires that for each $\hat{r} > r$, both $r$ and $\hat{r}$ prefer truth telling

$$
\begin{aligned}
U_\pi(r) &\geq U_\pi(r, \hat{r}) = U_\pi(\hat{r}) + q_\pi(\hat{r})(\hat{r} - r), \\
U_\pi(\hat{r}) &\geq U_\pi(\hat{r}, r) = U_\pi(r) + q_\pi(r)(r - \hat{r}).
\end{aligned}
$$

Therefore,

$$
-q_\pi(r)(\hat{r} - r) \leq U_\pi(\hat{r}) - U_\pi(r) \leq -q_\pi(\hat{r})(\hat{r} - r),
$$

which implies (3). Letting $\hat{r} \to r$ and then integrating from $r$ to $1$ gives

$$
U_\pi(1) - U_\pi(r) = -\int_r^1 q_\pi(s)\mathrm{d}s.
$$

Also, observe that type $r = 1$ can secure his maximal attainable utility of $0$ by always acting (irrespective of a recommendation); so $U_\pi(1) = 0$ and (4) follows. Finally, the maximal attainable utility of type $r = 0$ is $\mathbb{E}[\omega]$, which can be secured by never acting; so (5) follows.

*Sufficiency.* It remains to show that (3)–(5) imply (2). If either $\tilde{r} \geq \hat{r} \geq r$ or $\tilde{r} \leq \hat{r} \leq r$, then (3) and (4) imply that

$$
\begin{aligned}
U_\pi(r, \hat{r}) &= U_\pi(r) + q_\pi(\hat{r})(\hat{r} - r) = \int_{\hat{r}}^1 q_\pi(s)\mathrm{d}s + q_\pi(\hat{r})(\hat{r} - r) \\
&\geq \int_{\hat{r}}^{\tilde{r}} q_\pi(\tilde{r})\mathrm{d}s + \int_{\tilde{r}}^1 q_\pi(s)\mathrm{d}s + q_\pi(\tilde{r})(\hat{r} - r) \\
&= \int_{\tilde{r}}^1 q_\pi(s)\mathrm{d}s + q_\pi(\tilde{r})(\tilde{r} - r) = U_\pi(r, \tilde{r}),
\end{aligned}
$$

meaning that $U_\pi(r, \hat{r})$ is single-peaked in $\hat{r}$, with the peak located at $\hat{r} = r$. Therefore,

$$
U_\pi(r) \geq U_\pi(r, \hat{r}) = U_\pi(r, \hat{r}, 0, 1) \text{ for all } r, \hat{r} \in R.
$$

Moreover, letting $\hat{r} = 1$ and $\hat{r} = 0$ gives

$$
\begin{aligned}
U_\pi(r) &\geq U_\pi(r, 1) = U_\pi(1) + q_\pi(1)(1 - r) \geq 0 = U_\pi(r, \hat{r}, 0, 0) \text{ for all } r, \hat{r} \in R, \\
U_\pi(r) &\geq U_\pi(r, 0) = U_\pi(0) - q_\pi(0)r \geq \mathbb{E}[\omega] - r = U_\pi(r, \hat{r}, 1, 1) \text{ for all } r, \hat{r} \in R.
\end{aligned}
$$

Thus, we are left to show that $U_\pi(r) \geq U_\pi(r, \hat{r}, 1, 0)$ for all $r, \hat{r} \in R$. Notice that

$$
U_\pi(r, \hat{r}, 1, 0) = \int_\Omega (1 - \pi(\omega, r))(\omega - r)\mathrm{d}F(\omega) = \mathbb{E}[\omega] - r - U_\pi(r, \hat{r}).
$$

Since $U_\pi(r, \hat{r})$ is single-peaked, we have

$$
\begin{aligned}
U_\pi(r, \hat{r}, 1, 0) &\leq \mathbb{E}[\omega] - r - \min\{U_\pi(r, 0), U_\pi(r, 1)\} = \max\{0, \mathbb{E}[\omega] - r\} \\
&= \max\{U_\pi(r, 1), U_\pi(r, 0)\} \leq U_\pi(r) \text{ for all } r, \hat{r} \in R.
\end{aligned}
$$

## References

ALONSO, R., AND O. CÂMARA (2016a): "Bayesian Persuasion with Heterogeneous Priors," *Journal of Economic Theory*, 165, 672–706.

——— (2016b): "On the Value of Persuasion by Experts," mimeo.

——— (2016c): "Persuading Voters," *American Economic Review*, 106, 3590–3605.

AU, P. H. (2015): "Dynamic Information Disclosure," *RAND Journal of Economics*, 46, 791–823.

BERGEMANN, D., A. BONATTI, AND A. SMOLIN (2017): "The Design and Price of Information," mimeo.

BERGEMANN, D., B. BROOKS, AND S. MORRIS (2015): "The Limits of Price Discrimination," *American Economic Review*, 105, 921–957.

BERGEMANN, D., B. A. BROOKS, AND S. MORRIS (2016): "First Price Auctions with General Information Structures: Implications for Bidding and Revenue," *Econometrica*, 85, 107–143.

BERGEMANN, D., AND S. MORRIS (2013): "Robust Predictions in Games With Incomplete Information," *Econometrica*, 81, 1251–1308.

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

BERGEMANN, D., AND M. PESENDORFER (2007): "Information Structures in Optimal Auctions," *Journal of Economic Theory*, 137, 580–609.

BLACKWELL, D. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 93–102.

——— (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24, 265–272.

BOARD, S., AND J. LU (2016): "Competitive Information Disclosure in Search Markets," mimeo.

ELY, J., A. FRANKEL, AND E. KAMENICA (2015): "Suspense and Surprise," *Journal of Political Economy*, 103, 215–260.

ELY, J. C. (2017): "Beeps," *American Economic Review*, 107(1), 31–53.

ESÖ, P., AND B. SZENTES (2007): "Optimal Information Disclosure in Auctions and the Handicap Auction," *Review of Economic Studies*, 74, 705–731.

GENTZKOW, M., AND E. KAMENICA (2014): "Costly Persuasion," *American Economic Review, Papers & Proceedings*, 104, 457–462.

——— (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review, Papers & Proceedings*, 106, 597–601.

——— (2017a): "Bayesian Persuasion with Multiple Senders and Rich Signal Spaces," *Games and Economic Behavior*, 104, 411–429.

——— (2017b): "Competition in Persuasion," *Review of Economic Studies*, 84, 300–322.

——— (2017c): "Disclosure of Endogenous Information," *Economic Theory Bulletin*, 5, 47–56.

GERSHKOV, A., J. K. GOEREE, A. KUSHNIR, B. MOLDOVANU, AND X. SHI (2013): "On the Equivalence of Bayesian and Dominant Strategy Implementation," *Econometrica*, 81, 197–220.

GILL, D., AND D. SGROI (2008): "Sequential Decisions with Tests," *Games and Economic Behavior*, 63, 663–678.

——— (2012): "The Optimal Choice of Pre-launch Reviewer," *Journal of Economic Theory*, 147, 1247–1260.

GUTMANN, S., J. H. B. KEMPERMAN, J. A. REEDS, AND L. A. SHEPP (1991): "Existence of Probability Measures with Given Marginals," *Annals of Probability*, 19, 1781–1797.

HARDY, G. H., J. E. LITTLEWOOD, AND G. PÓLYA (1929): "Some Simple Inequalities Satisfied by Convex Functions," *Messenger Math*, 58, 145–152.

HEDLUND, J. (2017): "Bayesian Persuasion by a Privately Informed Sender," *Journal of Economic Theory*, 167, 229–268.

HÖRNER, J., AND A. SKRZYPACZ (2016): "Selling Information," *Journal of Political Economy*, 124(6), 1515–1562.

KAMENICA, E., AND M. GENTZKOW (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

KOLOTILIN, A. (2015): "Experimental Design to Persuade," *Games and Economic Behavior*, 90, 215–226.

——— (2016): "Optimal Information Disclosure: A Linear Programming Approach," mimeo.

LI, F., AND P. NORMAN (2017): "On Persuasion with Multiple Senders," mimeo.

LI, H., AND X. SHI (2016): "Discriminatory Information Disclosure," *American Economic Review*, forthcoming.

MANELLI, A., AND D. VINCENT (2010): "Bayesian and Dominant-Strategy Implementation in the Independent Private-Values Model," *Econometrica*, 78, 1905–1938.

MATHEVET, L., J. PEREGO, AND I. TANEVA (2017): "On Information Design in Games," mimeo.

MILGROM, P., AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208.

MYERSON, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

OSTROVSKY, M., AND M. SCHWARZ (2010): "Information Disclosure and Unraveling in Matching Markets," *American Economic Journal: Microeconomics*, 2, 34–63.

PEREZ-RICHET, E. (2014): "Interim Bayesian Persuasion: First Steps," *American Economic Review, Papers & Proceedings*, 104, 469–474.

RAYO, L., AND I. SEGAL (2010): "Optimal Information Disclosure," *Journal of Political Economy*, 118, 949 – 987.

ROTHSCHILD, M., AND J. STIGLITZ (1970): "Increasing Risk: I. A definition," *Journal of Economic Theory*, 2, 225–243.

TANEVA, I. (2016): "Information Design," mimeo.

*Supplementary Material*

SUPPLEMENT TO "PERSUASION OF A PRIVATELY INFORMED RECEIVER"

APPENDIX A. MULTIPLE ACTIONS

In this Appendix, we allow the receiver to make a choice among multiple actions. We characterize the implementable receiver's interim utilities and show that the sender can generally implement a strictly larger set of the receiver's interim utilities by persuasion mechanisms than by experiments. We also formulate the sender's optimization problem and show that the sender can achieve a strictly higher expected utility by persuasion mechanisms than by experiments.

A.1. **Preferences.** Let $A = \{0, 1, ..., n\}$ be a finite set of actions available to the receiver. The state $\omega \in \Omega = [0, 1]$ and the receiver's type $r \in R = [0, 1]$ are independent and have distributions $F$ and $G$. We continue to assume that the receiver's utility is linear in the state for every type and every action.

It is convenient to define the receiver's and sender's utilities, $u(\omega, r, a)$ and $v(\omega, r, a)$, recursively by the utility difference between each two consecutive actions. For each $a \in \{1, \ldots, n\}$,

$$u(\omega, r, a) - u(\omega, r, a - 1) = b_a(r)(\omega - x_a(r)),$$
$$v(\omega, r, a) - v(\omega, r, a - 1) = z_a(r) + \rho(r)(u(\omega, r, a) - u(\omega, r, a - 1)),$$

and the utilities from action $a = 0$ are normalized to zero, $u(\omega, r, 0) = v(\omega, r, 0) = 0$ for all $\omega$ and all $r$.

For each $a \in \{1, \ldots, n\}$, the receiver's and sender's utilities can be expressed as

$$u(\omega, r, a) = \left(\sum_{i=1}^{a} b_i(r)\right) \omega - \left(\sum_{i=1}^{a} b_i(r) x_i(r)\right)$$
$$v(\omega, r, a) = \left(\sum_{i=1}^{a} z_i(r)\right) + \rho(r) u(\omega, r, a).$$

We assume that $b_a(r) > 0$ for all $r$ and all $a \in \{1, \ldots, n\}$. This assumption means that every type $r$ prefers higher actions in higher states. Note that $x_a(r)$ is the cutoff state at which the receiver of type $r$ is indifferent between two consecutive actions $a - 1$ and $a$. Define $x_0(r) = -\infty$ and $x_{n+1}(r) = \infty$.

Denote by $\bar{x}_a(r)$ the cutoff truncated to the unit interval,

$$\bar{x}_a(r) = \max\{0, \min\{1, x_a(r)\}\}.$$

We assume that the cutoffs are ordered on $[0, 1]$ such that

$$\bar{x}_1(r) \leq \bar{x}_2(r) \leq \ldots \leq \bar{x}_n(r) \text{ for all } r \in R.$$

Thus, type $r$ optimally chooses action $a$ on the interval of states $(\bar{x}_a(r), \bar{x}_{a+1}(r))$.[10]

---

[10]This assumption ensures that the actions that can be optimal for type $r$ are consecutive. If actions $a - 1$ and $a + 1$ are optimal for type $r$ under states $\omega'$ and $\omega''$, then there must be a state between $\omega'$ and $\omega''$ where action $a$ is optimal. This assumption simplifies the exposition. Relaxing this assumption poses no difficulty: it will only require us for each type $r$ to omit from the analysis the actions that are never optimal for this type.

A.2. **Experiments.** Because the receiver's utility is linear in the state for every type and every action, every experiment $\sigma$ can be equivalently described by the probability that the posterior mean state is below a given value $x \in \Omega$,

$$H_\sigma(x) = \int_\Omega \sigma(x|\omega)\mathrm{d}F(\omega).$$

In fact, as in Blackwell (1951), Rothschild and Stiglitz (1970), and Gentzkow and Kamenica (2016), it is convenient to describe an experiment by a convex function $C_\sigma : \mathbb{R} \to \mathbb{R}$ defined as

$$C_\sigma(x) = \int_x^\infty (1 - H_\sigma(m))\mathrm{d}m.$$

Observe that by (9), for every experiment $\sigma$, we have $C_\sigma(r) = U_\sigma(r)$ for all $r$, where $U_\sigma(r)$ is the receiver's interim utility under $\sigma$ in the problem of Section 2, with two actions and $u(\omega, r, a) = a(\omega - r)$. Hence, by Theorem 1, the set of all $C_\sigma$ is equal to

$$\mathcal{C} = \{C : \underline{C} \le C \le \overline{C} \text{ and } C \text{ is convex}\},$$

where $\overline{C}$ and $\underline{C}$ correspond to the full and no disclosure experiments,

$$\overline{C}(x) = \int_x^\infty (1 - F(m))\mathrm{d}m,$$

$$\underline{C}(x) = \max\{\mathbb{E}[\omega] - x, 0\}.$$

A.3. **Implementable Interim Utilities.** The expected utility of type $r$ under experiment $\sigma$ is equal to

$$U_\sigma(r) = \int_\Omega \left( \max_{a \in A} u(m, r, a) \right) \mathrm{d}H_\sigma(m) \text{ for all } r \in R.$$

**Proposition 1.** *$U$ is implementable by an experiment if and only if there exists $C \in \mathcal{C}$ such that*

$$U(r) = \sum_{a=1}^n b_a(r)C(x_a(r)) \text{ for all } r \in R. \tag{12}$$

*Proof.* The receiver's interim utility under experiment $\sigma$ is

$$U_\sigma(r) = \sum_{a=0}^n \int_{x_a(r)}^{x_{a+1}(r)} u(m, r, a)\mathrm{d}H_\sigma(m) = \sum_{a=1}^n \int_{x_a(r)}^\infty b_a(r)(m - x_a(r))\mathrm{d}H_\sigma(m),$$

where we used $u(m, r, 0) = 0$. For each $a \in \{1, \ldots, n\}$, integration by parts yields

$$\int_{x_a(r)}^\infty b_a(r)(m - x_a(r))\mathrm{d}H_\sigma(m) = -b_a(r)(m - x_a(r))(1 - H_\sigma(m))\Big|_{x_a(r)}^\infty$$

$$+ b_a(r) \int_{x_a(r)}^\infty (1 - H_\sigma(m))\mathrm{d}m = b_a(r)C_\sigma(x_a(r)).$$

Summing up the above over $a \in \{1, \ldots, n\}$ yields (12).

It follows from Section A.2 that the set of the receiver's interim utilities implementable by experiments is equal to the set of functions that satisfy (12) for every $C \in \mathcal{C}$. $\square$

A persuasion mechanism can be described by a (possibly, infinite) menu of experiments, $\Sigma$. The receiver of type $r$ chooses one experiment from the menu and then observes messages only from this experiment. Obviously, the receiver chooses the experiment that maximizes his expected utility,

$$U_\Sigma(r) = \max_{\sigma \in \Sigma} U_\sigma(r) \text{ for all } r \in R.$$

By Proposition 1, it is immediate that the receiver's interim utility $U$ is implementable if and only if there exists a menu $\mathcal{C}_\Sigma \subset \mathcal{C}$ such that

$$U(r) = \max_{C \in \mathcal{C}_\Sigma} \left\{ \sum_{a=1}^{n} b_a(r) C(x_a(r)) \right\} \quad \text{for all } r \in R.$$

Theorem 1 shows that the sender can implement the same set of receiver's interim utilities by experiments as by persuasion mechanisms. With more than two actions, however, the sender can generally implement a strictly larger set of interim utilities by persuasion mechanisms than by experiments, as shown in Example 2.
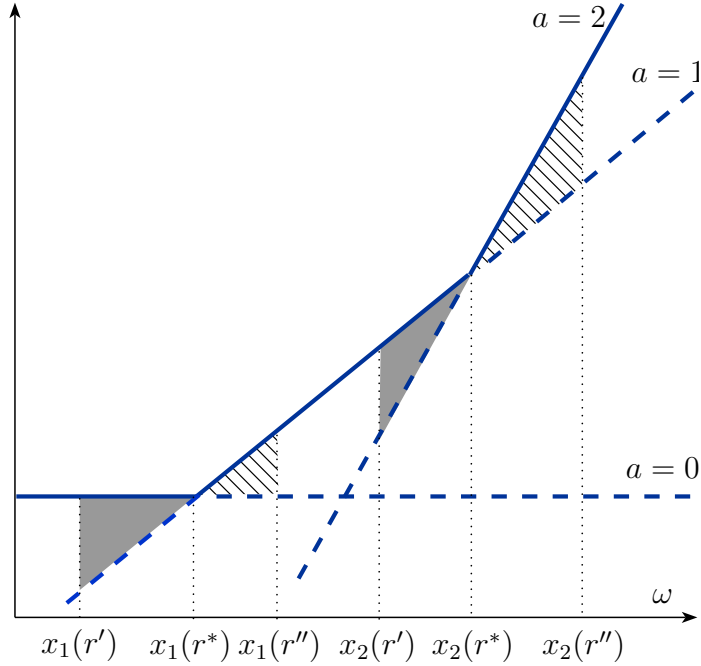


FIGURE 4. The utility of type $r^*$ in Example 1.

**Example 2.** Let $A = \{0, 1, 2\}$, $F$ admit a strictly positive density, and $u(\omega, r, a)$ be continuous in $r$ for all $\omega$ and all $a$. Furthermore, suppose that there exist two types $r' < r''$, such that for all $r \in (r', r'')$,

$$0 < x_1(r') < x_1(r) < x_1(r'') < x_2(r') < x_2(r) < x_2(r'') < 1.$$

Consider a persuasion mechanism consisting of the menu of two experiments represented by partitions $\{P', P''\}$, where $P'$ and $P''$ are the first-best partitions for types $r'$ and $r''$,

$$P' = \{[0, x_1(r')), [x_1(r'), x_2(r')), [x_2(r'), 1]\},$$
$$P'' = \{[0, x_1(r'')), [x_1(r''), x_2(r'')), [x_2(r''), 1]\}.$$

Types $r'$ and $r''$ choose, respectively, $P'$ and $P''$ and get their maximum possible utilities $\overline{U}(r')$ and $\overline{U}(r'')$. By the continuity of $u(\omega, r, a)$ in $r$, there exists a type $r^* \in (r', r'')$ who is indifferent between choosing $P'$ and $P''$. By this indifference,

$$L_1(P') + L_2(P') = L_1(P'') + L_2(P''),$$

where, for each $a \in \{1, 2\}$,

$$L_a(P') = \int_{x_a(r^*)}^{x_a(r')} \Big( u(\omega, r^*, a) - u(\omega, r^*, a-1) \Big) dF(\omega)$$

denotes the utility loss of type $r^*$ from using cutoff $x_a(r')$ rather than his first-best cutoff $x_a(r^*)$ to decide between actions $a-1$ and $a$. Analogously, for each $a \in \{1, 2\}$, we define $L_a(P'')$.

Fig. 4 illustrates this example. The three blue lines depict the utility of the receiver of type $r^*$ from taking action $a$, $u(\omega, r^*, a)$, for each $a = 0, 1, 2$. The kinked solid blue line is the utility of type $r^*$ from taking the optimal action, $\max_{a \in \{0,1,2\}} u(\omega, r^*, a)$. In Fig. 4, the loss of type $r^*$ from experiment $P'$ relative to the first best, $L_1(P') + L_2(P')$, is the total area of the two shaded triangles (assuming that $\omega$ is uniformly distributed). Similarly, the loss of type $r^*$ from experiment $P''$ relative to the first best, $L_1(P'') + L_2(P'')$, is the total area of the two hatched triangles. For type $r^*$, these shaded and hatched areas are equal, so type $r^*$ is indifferent between the two experiments.

An experiment that gives the maximum possible utilities $\overline{U}(r')$ and $\overline{U}(r'')$ to types $r'$ and $r''$ must at least communicate the common refinement of partitions $P'$ and $P''$. Therefore, the utility of type $r^*$ under such an experiment is at least

$$\overline{U}(r^*) - \min\{L_1(P'), L_1(P'')\} - \min\{L_2(P'), L_2(P'')\},$$

which is strictly larger than his utility under the persuasion mechanism,

$$\overline{U}(r^*) - L_1(P') - L_2(P'),$$

unless $L_1(P') = L_1(P'')$ and $L_2(P') = L_2(P'')$.

In Fig. 4, for type $r^*$, the loss $L_1(P'')$ (left hatched triangle) is smaller than the loss $L_1(P')$ (left shaded triangle). Similarly, the loss $L_2(P')$ (right shaded triangle) is smaller than the loss $L_2(P'')$ (right hatched triangle). Hence, the total loss is smaller under the experiment that is the coarsest common refinement of $P'$ and $P''$ (the area of the smaller shaded and hatched triangles) than under either experiment $P'$ or experiment $P''$.

A.4. **Sender's Problem.** In this section, we impose the following additional assumptions. For each $a \in \{0, 1, \ldots, n\}$, function $x_a(r)$ is strictly increasing, and its range contains $\Omega$. Let function $r_a(x)$ be the inverse of function $x_a(r)$. Moreover, for each $a \in \{0, 1, \ldots, n\}$, function $z_a(r)$ is differentiable, and function $G(r_a(x))$ is twice differentiable.

For a given experiment $\sigma$, the sender's expected utility conditional on the receiver's type being $r$ is[11]

$$V_\sigma(r) = \rho(r)U(r) + \sum_{a=1}^{n} \Big( \sum_{i=1}^{a} z_i(r) \Big) \big( H_\sigma(x_{a+1}(r)) - H_\sigma(x_a(r)) \big).$$

We now express the sender's expected utility as a function of $C_\sigma$, and all the model parameters are summarized in function $I$, the same way as in Lemma 2 in Section 4.

**Lemma 3.** *For every experiment $\sigma$,*

$$\int_R V_\sigma(r) \mathrm{d}G(r) = K + \int_\Omega C_\sigma(x) I(x) \mathrm{d}x,$$

*where $K$ is a constant independent of $\sigma$ and*

$$I(x) = \sum_{a=1}^{n} \Big( \tfrac{\mathrm{d}}{\mathrm{d}x} \big( z_a(r_a(x)) \tfrac{\mathrm{d}}{\mathrm{d}x} G(r_a(x)) \big) + \rho(r_a(x)) b_a(r_a(x)) \tfrac{\mathrm{d}}{\mathrm{d}x} G(r_a(x)) \Big).$$

*Proof.* We have

$$\sum_{a=1}^{n} \Big( \sum_{i=1}^{a} z_i(r) \Big) \big( H_\sigma(x_{a+1}(r)) - H_\sigma(x_a(r)) \big) = \sum_{a=1}^{n} z_a(r)(1 - H_\sigma(x_a(r)))$$

$$= -\sum_{a=1}^{n} z_a(r) C'_\sigma(x_a(r)),$$

where we used $x_{n+1}(r) = \infty$ (hence, $H_\sigma(x_{n+1}(r)) = 1$) and $1 - H_\sigma(x) = -C'_\sigma(x)$. By Proposition 1, we, thus, obtain

$$V(r) = \rho(r)U(r) + \sum_{a=1}^{n} \Big( \sum_{i=1}^{a} z_i(r) \Big) \big( H_\sigma(x_a(r)) - H_\sigma(x_{a+1}(r)) \big)$$

$$= \rho(r)U(r) - \sum_{a=1}^{n} z_a(r) C'_\sigma(x_a(r)) = \sum_{a=1}^{n} \Big( \rho(r) b_a(r) C_\sigma(x_a(r)) - z_a(r) C'_\sigma(x_a(r)) \Big).$$

---

[11]For each $r$ where $H_\sigma(x_{a+1}(r))$ is discontinuous, this formula assumes that type $r$ breaks the indifference in favor of action $a$ if the posterior mean state is $x_{a+1}(r)$. This assumption is innocuous because $G$ admits a density, and there are at most countably many discontinuities of $H_\sigma$.

Fix $a \in \{1, \dots, n\}$ and define the variable $x = x_a(r)$. Hence, $r = r_a(x)$. Using this variable change, we have

$$\int_R \Big( \rho(r) b_a(r) C_\sigma(x_a(r)) - z_a(r) C'_\sigma(x_a(r)) \Big) \mathrm{d}G(r)$$
$$= \hat{K}_a + \int_\Omega \Big( \rho(r_a(x)) b_a(r_a(x)) C_\sigma(x) - z_a(r_a(x)) C'_\sigma(x) \Big) \mathrm{d}G(r_a(x)),$$

where $\hat{K}_a$ is a constant independent of $\sigma$ because $\Omega = [0, 1] \subset [x_a(0), x_a(1)]$, and, for all $\sigma$ and all $x \notin \Omega$, we have $C_\sigma(x) = \max\{0, \mathbb{E}[\omega] - x\}$. Now we integrate by parts

$$\int_\Omega z_a(r_a(x)) C'_\sigma(x) \mathrm{d}G(r_a(x)) = \tilde{K}_a - \int_\Omega C_\sigma(x) \tfrac{\mathrm{d}}{\mathrm{d}x} \left( z_a(r_a(x)) \tfrac{\mathrm{d}}{\mathrm{d}x} G(r_a(x)) \right) \mathrm{d}x$$
$$= \tilde{K}_a - \int_\Omega C(x) \tfrac{\mathrm{d}}{\mathrm{d}x} \left( z_a(r_a(x)) \tfrac{\mathrm{d}}{\mathrm{d}x} G(r_a(x)) \right) \mathrm{d}x,$$

where $\tilde{K}_a$ is a constant independent of $\sigma$ because, for all $\sigma$, we have $C_\sigma(0) = \mathbb{E}[\omega]$ and $C_\sigma(1) = 0$. Thus, we obtain

$$\int_R \Big( \rho(r) b_a(r) C_\sigma(x_a(r)) - z_a(r) C'_\sigma(x_a(r)) \Big) \mathrm{d}G(r)$$
$$= K_a + \int_\Omega \Big( \rho(r_a(x)) b_a(r_a(x)) \frac{\mathrm{d}G(r_a(x))}{\mathrm{d}x} + \tfrac{\mathrm{d}}{\mathrm{d}x} \left( z_a(r_a(x)) \tfrac{\mathrm{d}}{\mathrm{d}x} G(r_a(x)) \right) \Big) C_\sigma(x) \mathrm{d}x,$$

where $K_a = \hat{K}_a + \tilde{K}_a$. Summing the above over $a \in \{1, \dots, n\}$, we obtain $K + \int_\Omega C_\sigma(x) I(x) \mathrm{d}x$, where $K = \sum_a K_a$, and $I$ is defined in Lemma 3. $\qquad \square$

The sender's optimal experiment is described by a function $C \in \mathcal{C}$ that solves

$$\max_{C \in \mathcal{C}} \int_\Omega C(x) I(x) \mathrm{d}x.$$

The solutions to this problem are characterized by Theorem 2.

As shown in Section A.3, when the receiver has more than two actions, the sender can implement a strictly larger set of receiver's interim utilities by persuasion mechanisms than by experiments. We now show that the set of implementable interim actions is also strictly larger under persuasion mechanisms. Therefore, even if the sender cares only about the receiver's action, and not his utility, $\rho(r) = 0$ for all $r$, the sender can achieve a strictly larger expected utility under persuasion mechanisms.

**Example 2** (Continued). In addition, let there exist $x_2^* \in (x_1(r''), x_2(r'))$ such that $\mathbb{E}[\omega | \omega \geq x_2^*] = x_2(r'')$ and $\mathbb{E}[\omega | \omega < x_2^*] < x_1(r')$.

An experiment that maximizes the probability of action $a = 2$ for type $r''$ must send message $x_2(r'')$ if and only if $\omega \in [x_2^*, 1]$. Under any such experiment, type $r'$ takes action $a = 2$ if and only if $\omega \in [x_2^*, 1]$, because for $\omega < x_2^*$, this experiment must generate messages distinct from $x_2(r'')$ and, thus, below $x_2^*$, which is in turn below $x_2(r')$.

Consider now a persuasion mechanism consisting of the menu of two experiments represented by the following partitions:

$$P' = \{[0, x_2^* - \varepsilon) \setminus [x_1(r'), x_1(r'')), [x_1(r'), x_1(r'')), [x_2^* - \varepsilon, 1]\},$$
$$P'' = \{[0, x_2^* - \varepsilon), [x_2^* - \varepsilon, x_2^*), [x_2^*, 1]\}$$

where $\varepsilon > 0$ is sufficiently small. Type $r'$ strictly prefers $P'$ (to $P''$) because, for a sufficiently small $\varepsilon$, the benefit of taking action $a = 1$ (rather than $a = 0$) on $[x_1(r'), x_1(r''))$ exceeds the cost of taking action $a = 2$ (rather than $a = 1$) on $[x_2^* - \varepsilon, x_2^*)$. Type $r''$ is indifferent between $P''$ and $P'$, because under both partitions he weakly prefers to take action $a = 0$ on $[0, x_2^* - \varepsilon)$ and action $a = 1$ on $[x_2^* - \varepsilon, 1]$. Therefore, under this mechanism, type $r''$ takes action $a = 2$ if and only if $\omega \in [x_2^*, 1]$, but type $r'$ takes action $a = 2$ if and only if $\omega \in [x_2^* - \varepsilon, 1]$. As shown above, these probabilities of $a = 2$ for types $r'$ and $r''$ cannot be achieved by any experiment.

Finally, an optimal persuasion mechanism need not be an experiment. Suppose that the sender cares only about action $a = 2$, i.e., $\rho(r) = z_0(r) = z_1(r) = 0$ and $z_2(r) = 1$ for all $r$. Also suppose that the support of $G$ contains only $r'$ and $r''$ with $r''$ being likely enough, so that the sender's optimal experiment maximizes the probability of action $a = 2$ for type $r''$. The persuasion mechanism constructed above gives a strictly larger expected utility to the sender than any experiment.

## Appendix B. Nonlinear Utilities

In this appendix, we allow the sender's and receiver's utilities to be nonlinear in the state. We characterize conditions under which persuasion mechanisms are equivalent to experiments, and show, in particular, that cutoff mechanisms are equivalent to experiments. We also show that the equivalence of implementation by persuasion mechanisms and by experiments generally fails.

B.1. **Preferences.** As in Section 2, the receiver has two actions, $A = \{0, 1\}$, the set of states is $\Omega = [0, 1]$, and the set of receiver's types is $R = [0, 1]$. The receiver's utility, however, is

$$u(\omega, r, a) = au(\omega, r),$$

where $u(\omega, r)$ is differentiable, strictly increasing in $\omega$, and strictly decreasing in $r$. We also normalize the utility such that, for each $\omega \in \Omega$,

$$u(\omega, \omega) = 0.$$

The sender's utility is

$$v(\omega, r, a) = av(\omega, r),$$

where $v(\omega, r)$ is differentiable. State $\omega$ and type $r$ are independent and have distributions $F$ and $G$.[12]

---

[12]Note that if $\omega$ and $r$ are correlated, the analysis below applies if we impose strict monotonicity on function $\tilde{u}(\omega, r) = u(\omega, r) g(r|\omega)/g(r)$ rather than on $u$, where $g(r)$ and $g(r|\omega)$ denote, respectively, the marginal density of $r$ and the conditional density of $r$ for a given $\omega$. This is because the receiver's interim utility under a mechanism $\pi$ can be written as $U(r) = \int_\Omega \tilde{u}(\omega, r)\pi(\omega, r)\mathrm{d}F(\omega)$.

B.2. **Characterization of Experiments.** We start with the characterization of persuasion mechanisms that are equivalent to experiments.

**Proposition 2.** *An incentive-compatible persuasion mechanism $\pi$ is equivalent to an experiment if and only if $\pi(\omega, r)$ is nonincreasing in $r$ for every $\omega \in \Omega$.*

Intuitively, because for each experiment $\sigma$, the distribution $\sigma(r|\omega)$ of $r$ conditional on each state $\omega$ is nondecreasing in $r$, each $\pi(\omega, r) \in \left[1 - \sigma(r|\omega), 1 - \sigma(r_-|\omega)\right]$ (see (6)) is nonincreasing in $r$.

*Proof.* Consider a mechanism $\pi$ that is equivalent to an experiment $\sigma$. Since $\sigma(r|\omega)$ is a distribution function of $r$ conditional on $\omega$, it is nondecreasing in $r$ for each $\omega$. Then, by (6), $\pi(\omega, r)$ is nonincreasing in $r$ for each $\omega$.

Conversely, let $\pi(r, \omega)$ be nonincreasing in $r$ for all $\omega$. For every $\omega$ and $r$ define $\sigma(r|\omega) = 1 - \pi(r_+, \omega)$, where $\pi(r_+, \omega)$ denotes the right limit of $\pi(., \omega)$ at $r$. Since $\pi(r_+, \omega) \in [0, 1]$ is nonincreasing and right-continuous in $r$, the function $\sigma(r|\omega)$ is a distribution, which describes the distribution of messages for every given state $\omega$. Thus, $\sigma$ is an experiment. It remains to verify that the constructed experiment is direct and induces the same action by the receiver as mechanism $\pi$, i.e., when the experiment sends a message $r$, then type $r$ is indifferent between the two actions. For all $r$,

$$
U_\pi(r) = \int_\Omega u(\omega, r)\pi(\omega, r)\mathrm{d}F(\omega) = \int_\Omega u(\omega, r_+)\pi(\omega, r_+)\mathrm{d}F(\omega)
$$
$$
= \int_\Omega u(\omega, r)\pi(\omega, r_+)\mathrm{d}F(\omega) = \int_\Omega u(\omega, r)(1 - \sigma(r|\omega))\mathrm{d}F(\omega) = U_\sigma(r),
$$

where the first equality holds by the definition of $U_\pi$, the second by the absolute continuity of $U_\pi$ (Theorem 1 of Milgrom and Segal, 2002), the third by the continuity of $u$ in $r$, the fourth by the definition of $\sigma$, and the last by the definition of $U_\sigma$ for direct experiments. There exist left and right derivatives of $U_\pi$ for all $r$ (Theorem 3 of Milgrom and Segal, 2002) that satisfy:

$$
U'_\pi(r_+) = \int_\Omega \frac{\partial u(\omega, r)}{\partial r}\pi(r_+, \omega)\mathrm{d}F(\omega),
$$
$$
U'_\pi(r_-) = \int_\Omega \frac{\partial u(\omega, r)}{\partial r}\pi(r_-, \omega)\mathrm{d}F(\omega).
$$

Since $U_\pi(r) = U_\sigma(r)$ and $\sigma(r|\omega) = 1 - \pi(r_+, \omega)$ for all $r$, we have

$$
U'_\sigma(r_+) = \int_\Omega \frac{\partial u(\omega, r)}{\partial r}(1 - \sigma(r|\omega))\mathrm{d}F(\omega),
$$
$$
U'_\sigma(r_-) = \int_\Omega \frac{\partial u(\omega, r)}{\partial r}(1 - \sigma(r_-|\omega))\mathrm{d}F(\omega),
$$

showing that type $r$ is indifferent between the two actions upon receiving message $r$.
$\square$

B.3. **Binary State.** Here we apply Proposition 2 to show that if there are only two states in the support of the prior $F$, then every incentive-compatible mechanism is equivalent to an experiment.

**Corollary 2.** *Let the support of $F$ consist of two states. Then every incentive-compatible mechanism $\pi$ is equivalent to an experiment.*

*Proof.* Consider $F$ whose support consist of two states, without loss of generality, $\{0, 1\}$, and let $\pi$ be an incentive-compatible persuasion mechanism. By Proposition 2, it is sufficient to show that $\pi$ is nonincreasing in $r$ for all $r \in (0, 1)$. Incentive compatibility implies that for all $r, \hat{r} \in (0, 1)$,

$$\sum_{\omega = 0,1} u(\omega, r) \left( \pi(\omega, r) - \pi(\omega, \hat{r}) \right) \Pr(\omega = 1) \geq 0. \tag{13}$$

Rewriting (13) twice, with $(r, \hat{r}) = (r_2, r_1)$ and $(r, \hat{r}) = (r_1, r_2)$, yields the inequalities

$$- \frac{u(0, r_2)}{u(1, r_2)} \delta(r_2, r_1, 0) \leq \delta(r_2, r_1, 1) \leq - \frac{u(0, r_1)}{u(1, r_1)} \delta(r_2, r_1, 0) \tag{14}$$

where $\delta(r_2, r_1, \omega) = (\pi(\omega, r_2) - \pi(\omega, r_1)) \Pr(\omega = 1)$. Because $u(0, r) < 0$ and $u(1, r) > 0$ for $r = r_1, r_2$, the monotonicity of $u$ in $r$ implies that

$$0 < - \frac{u(0, r_2)}{u(1, r_2)} \leq - \frac{u(0, r_1)}{u(1, r_1)} \text{ for } r_2 \leq r_1. \tag{15}$$

Combining (14) and (15) gives $\pi(\omega, r_2) \geq \pi(\omega, r_1)$ if $r_2 \leq r_1$ for each $\omega = 0, 1$. □

Note that if $F$ has a two-point support, then the receiver's utility is linear in the state without loss of generality, and, hence, Theorem 1 applies. However, Corollary 2 makes a stronger statement, because it asserts that every incentive-compatible mechanism is equivalent to an experiment, not just implements the same receiver's interim utility.

B.4. **Beyond Binary State.** Suppose now that the support of the prior $F$ consists of three states $\omega_1 < \omega_2 < \omega_3$ and let $f_i = \Pr(\omega_i) > 0$ for $i = 1, 2, 3$.

When there are at least three states and the utility of the receiver is nonlinear in (any transformation of) the state, then the posterior distribution of the state induced by an experiment can no longer be parametrized by a one-dimensional variable—such as the posterior mean state in the case of linear utilities, the posterior probability of one of the states in the case of binary-valued state, and the cutoff value in the case of cutoff mechanisms.

As a consequence, the interim action $q(r)$ and, hence, the sender's interim utility $V(r)$ are no longer pinned down by the receiver's interim utility $U(r)$.

**Proposition 3.** *Let $\pi_1$ and $\pi_2$ be two mechanisms that are distinct for each $r \in (\omega_1, \omega_3)$ but implement the same differentiable receiver's interim utility $U$. Then, the interim action $q$ is the same for $\pi_1$ and $\pi_2$ if and only if there exist functions $b$, $c$, and $d$ such that $u(\omega, r) = c(r) + b(r) d(\omega)$ for each $(\omega, r) \in \{\omega_1, \omega_2, \omega_3\} \times (\omega_1, \omega_3)$.*

*Proof.* For all $r \in (\omega_1, \omega_3)$ and $j = 1, 2$, we have

$$U(r) = \sum_{i=1}^{3} u(\omega_i, r) \pi_j(\omega_i, r) f_i,$$

$$U'(r) = \sum_{i=1}^{3} \frac{\partial u(\omega_i, r)}{\partial r} \pi_j(\omega_i, r) f_i,$$

where the first line holds by the definition of $U$, and the second line by the incentive compatibility of $\pi$.

The expected action, $q_{\pi_j}(r) = \sum_{i=1}^{3} \pi_j(\omega_i, r) f_i$ is the same across $j = 1, 2$ for each $r$ if and only if the vectors $u(\omega, r)$, $\frac{\partial u(\omega, r)}{\partial r}$, and $\mathbf{1}$ are linearly dependent for each $r$. That is, for each $(\omega, r) \in \{\omega_1, \omega_2, \omega_3\} \times (\omega_1, \omega_3)$, there exist functions $\gamma(r)$ and $\mu(r)$ such that

$$\frac{\partial u(\omega, r)}{\partial r} + \mu(r) u(\omega, r) = \gamma(r). \tag{16}$$

The solution of differential equation (16) is given by

$$u(\omega, r) = e^{-\int_{\omega_1}^{r} \mu(x)\mathrm{d}x} \left( \eta(\omega) + \int_{\omega_1}^{r} \gamma(x) e^{\int_{\omega_1}^{x} \mu(y)\mathrm{d}y} \mathrm{d}x \right),$$

where function $\eta(\omega)$ satisfies the (initial) normalization condition $u(\omega, \omega) = 0$. This completes the proof with $b(r)$, $c(r)$, and $d(\omega)$ given by

$$(b(r), c(r), d(\omega)) = \left( e^{-\int_{\omega_1}^{r} \mu(x)\mathrm{d}x} \int_{\omega_1}^{r} \gamma(x) e^{\int_{\omega_1}^{x} \mu(y)\mathrm{d}y} \mathrm{d}x, e^{-\int_{\omega_1}^{r} \mu(x)\mathrm{d}x}, \eta(\omega) \right).$$

$\square$

When the receiver's utility is nonlinear, the sender can implement a strictly larger set of the receiver's interim actions by persuasion mechanisms than by experiments. Therefore, the sender can achieve a strictly higher expected utility by persuasion mechanisms, even if her utility $v$ is state-independent.

**Example 3.** Let there be three states, $\omega_1 < \omega_2 < \omega_3$, and two types of the receiver, $r' > r''$. Denote $u_i' = u(\omega_i, r')$ and $u_i'' = u(\omega_i, r'')$ for $i \in \{1, 2, 3\}$. Assume $u_2'' < 0 < u_3'$. Moreover, assume $u_1'/u_1'' > u_2'/u_2''$, which means that point $(\omega_1, u_1'')$ lies above the dashed line in Fig. 5. Finally, assume that the probability masses $(f_1, f_2, f_3)$ on the states satisfy $f_1 u_1' + f_3 u_3' < 0$ and $f_2 u_2' + f_3 u_3' < 0$.

Let $\Sigma'$ be the set of all experiments that maximize the probability of action for type $r'$. It is easy to check that any $\sigma' \in \Sigma'$ induces type $r'$ to act with probability 1 if $\omega = \omega_3$, with probability $-f_3 u_3'/(f_2 u_2')$ if $\omega = \omega_2$, and with probability 0 if $\omega = \omega_1$.

Observe that by the monotonicity of $u$ in $r$, each message of $\sigma' \in \Sigma'$ that induces type $r'$ to act also induces type $r'' < r'$ to act. Moreover, by the definition of $\Sigma'$, each message of $\sigma' \in \Sigma'$ that induces type $r'$ not to act can be sent only in states $\omega_1$ or $\omega_2$ where the utility of type $r''$ is negative by assumption, $u_1'' < u_2'' < 0$, so type $r''$ does not act either. Thus, for each experiment $\sigma$ under which type $r'$ acts with probability $f_3(1 - u_3'/u_2')$ (i.e., $\sigma \in \Sigma'$), type $r''$ acts with the same probability as type $r'$.
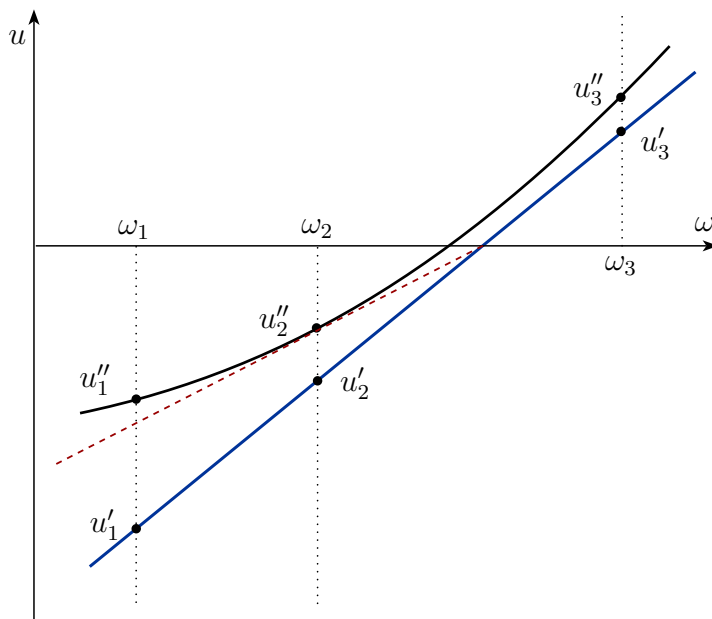
FIGURE 5. The utilities of types $r'$ (blue) and $r''$ (black) in Example 2.

We now construct a persuasion mechanism that also maximizes the probability that type $r'$ acts, but induces type $r''$ to act with a different probability. Let $\Sigma''$ be the set of all experiments $\sigma''$ that induce type $r'$ to act with probability 1 if $\omega = \omega_3$, with probability 0 if $\omega = \omega_2$, and with probability $-f_3 u'_3/(f_1 u'_1)$ if $\omega = \omega_1$. Consider a persuasion mechanism that consists of a menu of two experiments $\{\sigma', \sigma''\}$ with $\sigma' \in \Sigma'$ and $\sigma'' \in \Sigma''$. Notice that type $r'$ is indifferent between $\sigma'$ and $\sigma''$ as he obtains zero expected utility in either case. However, type $r''$ strictly prefers $\sigma''$ to $\sigma'$, because $u'_1/u''_1 > u'_2/u''_2$ by assumption. Therefore, under this persuasion mechanism, type $r'$ acts with probability $f_3 \left(1 - u'_3/u'_2\right)$, but type $r''$ acts with different probability $f_3(1 - u'_3/u'_1)$.

Finally, when the receiver's utility is nonlinear, the set of receiver's interim utilities implementable by persuasion mechanisms (as compared to experiments) can be strictly larger.

**Example 3** (Continued). In addition, let $r^* \in (r', r'')$ be such that $u'_1/u^*_1 < u'_2/u^*_2$, where $u^*_i = u(\omega_i, r^*)$ for $i \in \{1, 2, 3\}$.

It is easy to check that $\Sigma'$ and $\Sigma''$ are the sets of experiments $\sigma'$ and $\sigma''$ that maximize the utility of types $r^*$ and $r''$, respectively, subject to the constraint that type $r'$ gets utility $\underline{U}(r') = 0$. Because $\Sigma'$ and $\Sigma''$ do not intersect, no experiment can achieve the interim utility induced by a persuasion mechanism that consists of the menu of two experiments $\sigma' \in \Sigma'$ and $\sigma'' \in \Sigma''$.

## Appendix C. Linear Utilities

This appendix extends our main results to the class of utility functions that are linear in the state. Specifically, we consider the model defined in Section 2, with the modification that the utilities are linear in the state and are arbitrary functions of the receiver's type.

Let the receiver's and sender's utilities be normalized to zero if the receiver does not act, $a = 0$, and be linear in the state if the receiver acts, $a = 1$,

$$u(\omega, r, a) = a \cdot b(\omega - t),$$
$$v(\omega, r, a) = a \cdot (c(\omega - t) + d),$$

where $r = (b, c, d, t) \in \mathbb{R}^4$ denotes the receiver's type. The type has distribution $G$ that admits a differentiable density $g$, which is strictly positive on a compact set in $\mathbb{R}^4$ and zero everywhere else. The state $\omega \in \Omega = [0, 1]$ is independent of $r$ and has distribution $F$.

Let $H_\sigma$ be the distribution of the posterior mean induced by an experiment $\sigma$. As in Section A.2, it is convenient to describe $\sigma$ by

$$C_\sigma(t) = \int_t^\infty (1 - H_\sigma(m)) \mathrm{d}m.$$

**Proposition 4.** *For each experiment $\sigma$, the receiver's interim utility is*

$$U_\sigma(r) = |b| C_\sigma(t) + \min\{0, b\}(\mathbb{E}[\omega] - t). \tag{17}$$

*There exist $K \in \mathbb{R}$ and $I : \mathbb{R} \to \mathbb{R}$ such that for each $\sigma$ the sender's expected utility is*

$$V_\sigma = K + \int_{t \in \mathbb{R}} C_\sigma(t) I(t) \mathrm{d}t. \tag{18}$$

Proposition 4 allows us to extend Theorems 1 and 2 to this setting. Recall from Section A.2 that the set of all $C_\sigma$ is equal to

$$\mathcal{C} = \{C : \underline{C} \leq C \leq \overline{C} \text{ and } C \text{ is convex}\},$$

where $\overline{C}$ and $\underline{C}$ correspond to the full and no disclosure experiments. Each persuasion mechanism can be described by a (possibly, infinite) menu of experiments, $\Sigma$, which the receiver chooses from. By (17), for a given menu $\Sigma$, the receiver's interim utility is

$$\max_{\sigma \in \Sigma} U_\sigma(r) = |b| \left( \max_{\sigma \in \Sigma} C_\sigma(t) \right) + \min\{0, b\}(\mathbb{E}[\omega] - t).$$

Notice that $\max_{\sigma \in \Sigma} C_\sigma$ is the upper envelope of convex functions $C_\sigma \in \mathcal{C}$ and hence it is in $\mathcal{C}$. Therefore, by Proposition 4, any implementable pair of the sender's and receiver's expected utilities is implementable by an experiment. Moreover, the sender's problem can be expressed as

$$\max_{C \in \mathcal{C}} \int_{\mathbb{R}} C(t) I(t) \mathrm{d}t,$$

and Theorem 2 holds with $U$ replaced by $C$.

*Proof of Proposition 4.* Fix a type $r = (b, c, d, t) \in \mathbb{R}^4$ and evaluate

$$U_\sigma(r) = \int_0^1 \max\{0, b(m - t)\} \mathrm{d}H_\sigma(m).$$

Clearly, if $b = 0$, then $U_\sigma(r) = 0$. We now consider two cases, $b > 0$ and $b < 0$.

*Case 1: $b > 0$.* Given a posterior mean $m$, the receiver acts if and only if $t < m$. By integration by parts, the receiver's interim utility is

$$U_\sigma(r) = \int_t^1 b(m - t) \mathrm{d}H_\sigma(m) = bC_\sigma(t).$$

Again, by integration by parts, the sender's interim utility is

$$V_\sigma(r) = \int_t^1 (c(m - t) + d) \mathrm{d}H_\sigma(m) = cC_\sigma(t) - dC'_\sigma(t).$$

*Case 2: $b < 0$.* Given a posterior mean $m$, the receiver acts if and only if $t \geq m$. By integration by parts, the receiver's interim utility is

$$U_\sigma(r) = \int_0^t b(m - t) \mathrm{d}H(m) = -bC(t) + b(\mathbb{E}[\omega] - t).$$

Again, by integration by parts, the sender's interim utility is

$$V_\sigma(r) = \int_0^t (c(m - t) + d) \mathrm{d}H(m) = -cC(t) + dC'(t) + d + c(\mathbb{E}[\omega] - t).$$

We, thus, obtain (17).

We now show that the sender's expected utility is given by (18). Let $g(b, c, d|t)$ be the density of $(b, c, d)$ conditional on $t$, and let $g_t(t)$ be the marginal density of $t$. Define

$$c_+(t) = \int cg(b, c, d|t) \mathbf{1}_{b>0} \mathrm{d}(b, c, d),$$

and

$$c_-(t) = \int cg(b, c, d|t) \mathbf{1}_{b<0} \mathrm{d}(b, c, d).$$

Similarly, define $d_+(t)$ and $d_-(t)$.

Fix $t$ and take the expectation with respect to $(b, c, d)$ on the set of $b > 0$:

$$\int_{(b,c,d)} V_\sigma(b, c, d, t) \mathbf{1}_{b>0} g(b, c, d|t) \mathrm{d}(b, c, d) = c_+(t)C_\sigma(t) - d_+(t)C'_\sigma(t).$$

Now, integrating with respect to $t$,

$$\int_{(b,c,d,t)} V_\sigma(b, c, d, t) \mathbf{1}_{b>0} \mathrm{d}G(b, c, d, t) = \int_t (c_+(t)C_\sigma(t) - d_+(t)C'_\sigma(t)) g_t(t) \mathrm{d}t$$

$$= \int_t (c_+(t)g_t(t) + \tfrac{\mathrm{d}}{\mathrm{d}t}[d_+(t)g_t(t)]) C_\sigma(t) \mathrm{d}t.$$

Similarly, for $b < 0$,

$$\int_{(b,c,d)} V_\sigma(b,c,d,t)\mathbf{1}_{b<0}g(b,c,d|t)\mathrm{d}(b,c,d) = -c_-(t)C_\sigma(t)+d_-(t)C'_\sigma(t)+d_-(t)+c_-(t)(\mathbb{E}[\omega]-t).$$

Now, integrating with respect to $t$,

$$\int_{(b,c,d,t)} V_\sigma(b,c,d,t)\mathbf{1}_{b<0}\mathrm{d}G(b,c,d,t) = -\int_t (c_-(t)C_\sigma(t) - d_-(t)C'_\sigma(t))g_t(t)\mathrm{d}t + K$$

$$= -\int_t (c_-(t)g_t(t) + \tfrac{\mathrm{d}}{\mathrm{d}t}[d_-(t)g_t(t)])C_\sigma(t)\mathrm{d}t + K,$$

where

$$K = \int_t \Big(d_-(t) + c_-(t)(\mathbb{E}[\omega] - t)\Big)g_t(t)\mathbf{1}_{b<0}\mathrm{d}t$$

is a constant independent of $C_\sigma$.

Since the measure of types with $b = 0$ is zero, we obtain

$$\int_r V_\sigma(r)\mathrm{d}G(r) = \int_r V_\sigma(r)(\mathbf{1}_{b>0} + \mathbf{1}_{b<0})\mathrm{d}G(r) = \int_t I(t)C_\sigma(t)\mathrm{d}t + K,$$

where

$$I(t) = (c_+(t) - c_-(t))g_t(t) + \tfrac{\mathrm{d}}{\mathrm{d}t}[(d_+(t) - d_-(t))g_t(t)].$$

$\square$