Conjugate Persuasion

Ian Jewitt and Daniel Quigley*

First draft: November 2020 This version: August 2023

Abstract

We consider a class of persuasion games in which the sender has rank-dependent (Yaari (1987)) preferences. Like much of the recent Bayesian persuasion literature, we allow the sender to choose from a rich set of information structures and assume the receiver's action depends only on her posterior expectation of a scalar state variable. Conjugate to the standard problem, our sender's utility is linear in posterior the mean, but may be nonlinear in probabilities. We geometrically characterize the sender's optimal commitment payoff and identify the corresponding optimal information structure. When the state is continuously distributed, communication takes a monotone partitional form. Our characterization admits a simple analysis of comparative statics—for instance, we find that "grading on a curve" is a feature of optimal design. Finally, we apply our analysis to several problems of economic interest including information design in auctions and elections, as well as the design of equilibrium insurance contracts in the face of the 'favorite-longshot' bias. (JEL D02, D30, D81, D82, D83)

1 Introduction

If competition authorities wish to promote competition and protect consumers, what advertising standards should they adopt? By contrast, if a monopoly platform wishes to raise sellers' profits what should it allow them to advertise? Can a lobbyist use information to shift public opinion and thereby distort an election campaign? How should we expect competitive insurance markets to look when agents hold the well-known 'favorite-longshot' bias?

*Jewitt: Department of Economics and Nuffield College, University of Oxford (email: ian.jewitt@nuffield.ox.ac.uk). Quigley: Department of Economics and Exeter College, University of Oxford (email: daniel.quigley@economics.ox.ac.uk). We thank Daghan Carlos Akkar, Meg Meyer, Paula Onuchic, Vasiliki Skreta, Balázs Szentes, and Tom Wiseman for their very helpful feedback, as well as seminar participants at the University of Texas at Austin, London School of Economics, University of Calgary, Simon Fraser University, University of British Columbia, and the Transatlantic Theory Workshop.

These are basic questions of distributional design. Yet, they introduce two kinds of novelty: first, admitting the favorite-longshot bias involves a failure of expected utility. Second, in many settings information design at a population level appears to be conducted in an approximately i.i.d. way. Advertising rules are often applied to firms independently, while news articles can induce different reactions across independent readers. Moreover—perhaps due to a need to avoid conflicts of interest, regulatory scrutiny or other costs—sellers frequently enjoy equal treatment under advertising rules; and to a degree, anonymity grants the majority of readers with symmetric access to mass media outlets.[1] In this paper we study a class of problems that allows us to tractably address these kind of questions.

We study a model similar to—but distinct from—Bayesian persuasion. Developing the persuasion analogy, consider the following setting: a sender (he) commits to provide a signal to a receiver (she), (perhaps stochastically) mapping some state of the world into a message. Following much of the current literature, we assume that the state is drawn from an interval on the real line, and that the sender's ex post payoff (i.e. after the message is observed) depends only on the first moment of the receiver's posterior distribution. We depart from most of the literature by dropping the assumption that the sender is an expected utility maximizer. In standard Bayesian persuasion problems, his payoff may be nonlinear in the posterior mean, but is necessarily linear in the probabilities with which those posteriors are induced. By contrast, we assume his payoff is linear in posteriors, but may be nonlinear in probabilities; this nonlinearity is captured by a (probability) weighting function, $\nu : [0, 1] \to \mathbb{R}$, defined on the rank (i.e., cumulative probability) of each outcome in the chosen distribution. In this sense, our problem is conjugate to classic Bayesian persuasion.

Theorem 1 solves the sender's problem, assuming the state is continuously distributed.[2] While we make no restrictions on his choice of information structures beyond those implied by Bayes' rule, we find that optimal communication takes a very simple and natural form. In particular, the sender's problem can always be solved by *monotone partitional* communication. In a monotone partitional information structure, the sender commits to partition the state into a countable collection of 'pooling intervals', and one residual set. Each interval is assigned a unique message which identifies it; on the residual set, the sender simply reveals the state. Hence, the receiver either learns the state exactly or an interval in which it lies, and in equilibrium her beliefs are unambiguously increasing in the state.

We show that geometric tools reminiscent of those used in Bayesian persuasion identify optimal communication and payoffs. Central to our analysis is the *convexified* weighting

[1] This is natural in offline contexts. Even online, targeted political ads—and the platforms that support them—face increasing scrutiny. According to a YouGov poll conducted in October 2020, 68% (63%) of Democrat (Republican)-leaning respondents support a ban on such targeting.

[2] In Appendix B, we extend our arguments to allow for discontinuities in the distribution of the state.

function, ω—the largest convex function nowhere greater than ν. We show that every optimal pooling interval is an interquantile range defined by a corresponding interval in the rank space for which $\omega(p) < \nu(p)$. Moreover, the sender's optimal payoff is equal to that he would earn if his weighting function were ω and he chose to reveal full information. To establish this, we show the sender's payoff from any feasible information structure is bounded by the full information payoff under weighting ω, via stochastic dominance arguments. By appealing to a linearity property of ω, we then verify that the optimal monotone partition attains the bound.

Owing to the simplicity of optimal information structures, the model readily admits comparative statics results. For example, we show that as the sender becomes more 'risk loving' over posterior means, optimal pooling intervals contract (in the sense of set inclusion). Moreover, as the geometries of ω and ν are independent of the distribution of the state, it is easy to characterize how optimal communication varies with the latter. Most notably, the optimal information structure "grades on a curve", in that optimal pooling intervals shift and scale in direct correspondence with rescaling of the state.

In section 4, we show how our model can be applied to study a natural class of *i.i.d. Bayesian persuasion* problems. Here, the sender is an expected utility maximizer who faces N independent state variables, about which he can provide information in an i.i.d. way. Payoffs from an outcome may depend on that outcome's rank among the draws. This is a reasonably flexible class: for instance, it incorporates a sender who cares only about maximizing the value of the highest draw, the second highest draw, the difference between the two, or the median, among others. Moreover, it incorporates a range of economically interesting applications, including advertising in oligopoly markets and media influence in elections. To draw the connection to our model, we show that every such problem can be cast as one of conjugate persuasion; on the other hand, every conjugate persuasion problem is also the the (uniform) limit of a sequence of i.i.d. Bayesian persuasion games with rank-dependent marginal utility. To illustrate the value of this connection, we show how Theorem 1 can be used to solve for optimal information in oligopolistic markets and competitive elections and develop policy consequences for both large and small markets/electorates. In both settings, we find that censorship of extreme outcomes is critical to optimal design.

In section 5, we develop a behavioral interpretation for the conjugate persuasion problem. In particular, we show that conjugate persuasion applies to agents with a class of preferences studied within cumulative prospect theory (Quiggin (1982); Yaari (1987); Tversky & Kahneman (1992)): several empirical studies have found behavioral patterns well-captured by 'rank-dependent' models with a weighting function which is first concave, and then convex. For instance, it predicts Tversky & Kahneman (1992)'s "four-fold pattern": subjects appear

appear to systematically 'overweight' low probability events (appearing risk-loving/averse for gambles on/against a "longshot"), and 'underweight' frequent events. We consider the implications of these preferences for equilibrium insurance contracts in competitive markets. While this is not an information design problem, we nevertheless show how it can be studied in our framework. We assume insurers are risk-neutral and find they offer only partial insurance—pooling all risks below some threshold of the consumer's wealth distribution, but allowing her to retain the upside in sufficiently good times. We also describe how this framework can be used to assess optimal income redistribution when a policymaker is concerned about inequality, as measured using Lorenz curves. Finally, in section 6 we briefly describe the implications of our analysis for Bayesian persuasion with heterogeneous priors.

The rest of the paper runs as follows. We briefly discuss the literature, before introducing our model in 2. Section 3 provides our main theorem, and some comparative statics. Sections 4 through 6 show how the model applies to a variety of classic economic problems. Section 7 concludes. Most proofs are relegated to appendices.

Related Literature

Our work relates to the growing literature on persuasion and information design. In different contexts, Aumann & Maschler (1995) and Kamenica & Gentzkow (2011) showed that, when a sender can commit to an information structure his optimal payoff can be characterized by his concavified utility function. Their finding applies to the sender's utility defined on the space of posterior distributions, where geometric analysis can become difficult beyond settings with a limited number of states. Subsequently, the literature has focused on the case where the sender cares only about the receiver's posterior mean (see for example, Gentzkow & Kamenica (2016), Kolotilin *et al.* (2017)). The methods of Kamenica & Gentzkow (2011) do not generally apply in the space of posterior means; taking a dual approach Dworczak & Martini (2019) show how to verify the optimality of candidate information structures. We show conjugate persuasion is amenable to geometric analysis, even as the state space grows large.

In a recent paper, Kleiner *et al.* (2021) characterize the extreme points of the (convex) sets of distributions which constitute *(i)* a mean preserving contraction, or *(ii)* a mean preserving spread, of some fixed, exogenous distribution. Most closely related to our paper is concurrent work, Bergemann *et al.* (2021). In contrast to our direct argument, Bergemann *et al.* (2021) show how the second characterization of Kleiner *et al.* (2021) can be 'inverted' to identify a result similar to our Theorem 1, in the context of a second-price auction. Their results relate closely to the application we present in section 4.1. We identify a novel stochastic dominance argument for Theorem 1, and show how it extends to atoms in the distribution of the state (see Appendix B). Moreover, we show how the conjugate persuasion problem applies to a wider range of economically relevant design problems.

There has been much recent interest in partitional information structures—perhaps due to their simplicity, tractability and practical relevance. Kolotilin *et al.* (2017), Dworczak & Martini (2019), Mensch (2021), Kolotilin & Li (2021) and Best & Quigley (2021) identify sufficient conditions under which (monotone) partitional information structures are optimal. In a problem involving heterogeneous priors, Onuchic & Ray (2020) study optimal design of monotone partitional signals. They show the best such partition can be characterized by convexifying a function which represents their prior disagreement. We identify a new argument for the optimality of monotone partitional communication, from which we also find that "grading on a curve" is a feature of optimal communication.

In section 4, we apply our results to study information in oligopolistic markets and in elections. In classic auction settings, Milgrom & Weber (1982) show a seller benefits by revealing information to buyers, when her information and buyers' values are affiliated. Several papers consider the effects of information in oligopolistic markets, including Anderson & Renault (2006); Johnson & Myatt (2006); Ivanov (2013); Boleslavsky *et al.* (2019). In most of these papers, the choice of information is limited to a parametric class. Boleslavsky *et al.* (2019) allow sellers with independently valued goods to advertise about their own product in a rich way. When prices are set before information is revealed, they characterize firms' (symmetric) equilibrium advertising. We study a different, policy-oriented question, characterizing the set of distributions of consumer and producer surplus attainable with i.i.d. information.

Several papers have considered the effects of information on voting (for instance, Gershkov & Szentes (2009); Duggan & Martinelli (2011); Anderson & McLaren (2012); Alonso & Câmara (2016); Bardhi & Guo (2018)).[3] In most of these papers, options under consideration by voters are exogenous. We incorporate endogenous policy choice, in a Downsian election model (Downs (1957)). Strömberg (2004) and Chan & Suen (2008, 2009) also consider models with media coverage and endogenous party policies. In Strömberg (2004), the media acts as a commitment device for making policy offers to special interest groups; in Chan & Suen (2008, 2009), the media is informative, but signals are either of a binary threshold form or cheap talk statements. We allow a media outlet to choose from a rich class of signals, subject to an i.i.d. constraint which we view as a model of mass media.

[3]Gentzkow & Shapiro (2006) develop a related theory in which media bias for a candidate emerges from an outlet's desire to signal it has high quality information.

2 Model

In Bayesian Persuasion problems, a sender (he) may commit to provide information to a receiver (she) about a random variable X, drawn from a distribution F with bounded support $[\underline{x}, \overline{x}] \subset \mathbb{R}$. In the literature F is frequently taken to be continuous, and for the most part we assume the same (unless otherwise stated). In what follows, we sometimes find it useful to refer to the quantile function $G^{-1}(p) := \inf\{x \in [\underline{x}, \overline{x}] : p \leq G(x)\}$ associated with an arbitrary distribution G on $[\underline{x}, \overline{x}]$; of course, $G^{-1}(p)$ is nondecreasing in $p \in [0, 1]$—the *rank* of outcome $G^{-1}(p)$ in distribution G (i.e, the frequency with which $X \leq G^{-1}(p)$).

In these problems the sender commits to send information to a receiver via a choice of experiment, which consists of signal space S and a measurable map $\pi : [\underline{x}, \overline{x}] \to \Delta S$, where ΔS is the set of probability distributions on S. In an increasingly studied class of Bayesian Persuasion problems, the sender is an expected utility maximizer with Bernoulli utility $u : [\underline{x}, \overline{x}] \to \mathbb{R}$, a lower semicontinuous function defined on the receiver's posterior mean $x_s = \mathbb{E}[X \mid s]$, where the expectation is taken with respect to beliefs induced by F and the experiment π.[4] As Gentzkow & Kamenica (2016) and Dworczak & Martini (2019) show, this problem corresponds to choosing a distribution G to solve:[5]

$$\max_{G \preceq_{co} F} \int u(x) dG(x). \tag{1}$$

In problem (1), the sender's objective is linear in probabilities but may be nonlinear in the induced posterior mean. In this paper, we study a different class of problems where the sender's payoff is linear in the receiver's posterior mean, but may be nonlinear in probabilities. Specifically, we consider the problem

$$V(F, \nu) = \max_{G \preceq_{co} F} \int_{\underline{x}}^{\overline{x}} x d\nu(G(x)), \tag{2}$$

where $\nu : [0, 1] \to \mathbb{R}$ is assumed to be a continuous function of bounded variation, and the integral is taken to be Lebesgue-Stieltjes.[6] We refer to the function ν as a *(Yaari) weighting*.

Several economic problems—persuasion and others—fall naturally into this class. For concreteness, we briefly introduce a few of them below. Despite the functional form in (2), in several of these applications the sender is an expected utility maximizer. In sections 4

[4] $x_s = \mathbb{E}[X \mid s]$ is the Radon-Nikodym derivative associated with the sub σ-field generated by s.

[5] We write $\preceq_{co}$ to represent the *convex order*: $G \preceq_{co} F$ if any only if $\int u(x)dG(x) \leq \int u(x)dF(x)$ for all convex functions u.

[6] As continuity naturally arises in all the applications we develop in sections 4 and 5, we maintain that assumption throughout.

and 5, we show how each example maps to conjugate persuasion, solve for the corresponding optimal designs and discuss their economic implications.

Example 1. (*Order statistics and i.i.d. information design*)

Each of N receivers has a value X_i, distributed i.i.d. on $[\underline{x}, \overline{x}]$ according to F. Suppose that the sender can provide each receiver with information about her own value, and that information structures must be identical across receivers. Then his choice corresponds to selecting an i.i.d. distribution $G \preceq_{co} F$.

The sender is an expected utility maximizer, whose Bernoulli utility $u : [\underline{x}, \overline{x}]^N \to \mathbb{R}$ depends on the order statistics associated with N i.i.d. draws from G. Letting $x^{(k)}$ be the k^{th} highest order statistic, and $\alpha_k \in \mathbb{R}$ the sender's associated marginal utility for the k^{th} highest outcome, his payoff is:

$$u(x_1, \ldots, x_n) = \sum_{k=1}^{N} \alpha_k x^{(k)}.$$

The sender chooses G to maximize his expected utility. As we elaborate in section 4, this information design problem arises naturally in several economically interesting environments, including auctions and elections.

Example 2. (*A sender with non-expected utility*)

Consider a sender whose preferences violate the standard independence axiom of EU theory. Rather, his preferences satisfy 'dual independence' (Yaari, 1987) (see section 5). Then, (Yaari, 1987) shows there exists a continuous, nondecreasing function $f : [0, 1] \to \mathbb{R}$ such that his preferences can be described over lotteries can be described by the integral

$$\int f(G(x))dx.$$

We refer to such preferences as *Yaari utilities.* With dual independence in place of the standard independence axiom, problem (2) is an otherwise standard persuasion problem.

Example 3. (*Competitive insurance with a Yaari consumer*)

A consumer with Yaari utility faces an uncertain income $X \sim F$. Two competitive, risk-neutral insurance companies may simultaneously offer the consumer an insurance contract, which specifies an upfront premium in exchange for a payment $Y_i(X)$ in state X, where $i = 1, 2$ indexes insurers. Insurer i may only offer payment profiles $Y_i(X)$ that constitute a mean preserving contraction of X. That is, the insurer can only offer random variables which "hedge" the consumer's income risk.[7] The consumer chooses her most preferred contract

[7]While this is a natural feature of competitive insurance contracts, we nonetheless show how our frame-

among those offered. We are interested in equilibrium contracts and the consumer's final utility as a function of her risk preferences.

Example 4. (*Income redistribution with relative inequality aversion*)

Faced with a continuum of citizens with income distribution F, a government considers budget-balanced redistribution via taxation. It is perhaps natural to suppose that a government might measure the degree of inequality by the population Gini coefficient, $\int_0^1 L(p)dp$, where $L(p) = \frac{\int_0^p G^{-1}(p)dp}{\int_0^1 G^{-1}(p)dp}$ is the value of the Lorenz curve at the p^{th} quantile of the income distribution. If the government's objective were to minimize the degree of inequality as measured by the Gini, it would of course choose a completely flat distribution. More permissive preferences for redistribution arise from the generalized Gini coefficient:

$$\int_0^1 \gamma(p)L(p)dp, \tag{3}$$

3 The Main Result

In this section we present our main result, which establishes existence of a solution to problem (2) and characterizes both the sender's optimal payoff and the structure of optimal persuasion.

To aid our discussion, we first briefly observe a connection between problem (2) and Dworczak & Martini (2019)'s solution method to (1). Let $C(u)$ be the set of convex functions on $[\underline{x}, \overline{x}]$ which pointwise dominate u: for any $v \in \overline{C}(u)$, $v(x) \geq u(x)$, for all $x \in [\underline{x}, \overline{x}]$. Then for any $v \in \overline{C}(u)$, (1) is bounded above by $\int vdF$. To see this, suppose $G^\star$ solves (1). Then

$$\int vdF \geq \int vdG^\star \geq \int udG^\star.$$

In other words,

$$\min_{v \in \overline{C}(u)} \int vdF \geq \max_{G \preceq_{co} F} \int udG$$

is the dual problem corresponding to (1). Under some regularity conditions, Dworczak & Martini (2019) show these problems in fact attain the same value. Of course, to characterize the value of the dual, one still needs to identify the function v which minimizes $\int vdF$.

In a similar way, we can identify a natural upper bound on problem (2). To ease the discussion, suppose that ν is an increasing function with $\nu(0) = 0$, $\nu(1) = 1$, so that $\nu \circ G$

work also allows for 'fully relaxed' contract offers (such as opportunities for gambling, which need not have 0 expected return) in section 5.

can be interpreted as a probability measure on $[\underline{x}, \overline{x}]$. Let $\underline{C}(\nu)$ be the set of convex functions w which satisfy $w(p) \leq \nu(p)$, for all $p \in [0,1]$, and with equality at $p \in \{0,1\}$. To extend the analogy, suppose for a moment that each $w \in \underline{C}(v)$ is also increasing. Then, for any distribution G, $w \circ G$ first order stochastically dominates $\nu \circ G$ on $[\underline{x}, \overline{x}]$. Hence,

$$\int xd(w \circ G)(x) \geq \int xd(\nu \circ G)(x).$$

Moreover, a sender with convex weighting function w is risk-loving: $G \preceq_{co} F$ implies $\int xdw \circ F(x) \geq \int xdw \circ G(x)$ (see, for instance, Yaari (1987)). To see some intuition, suppose w were differentiable. Then, by a change-of-variables, the sender's objective is $\int x(p)w'(p)dp$, where $x(p) = G^{-1}(p)$. With a convex weighting function, the 'marginal utility of income' w' is increasing in the quantile and hence, pooling is never desirable—to do so would sacrifice outcomes at the (more valuable) top of the distribution to support those at the bottom.

As we show in the Appendix (see Lemma 4), this basic argument does not rely on ν or w being increasing. Thus, (2) is bounded above by

$$\min_{\omega \in \underline{C}(u)} \int xdw(F). \tag{4}$$

In contrast to regular Bayesian persuasion, the solution to the dual problem is remarkably simple. The set $\underline{C}(\mu)$ has a (pointwise) greatest member, the **convexification** $\omega : [0,1] \to \mathbb{R}$ of ν. Repeating the argument above, the convexification must in fact solve (4).

Our main result solves the conjugate persuasion problem by showing that there exists an information structure which attains the bound (4). We call an information structure **monotone partitional** if there exists a set of disjoint, open intervals $\mathcal{J} = \{J_k\}_{k \in K}$, $J_k = (x_k, x'_k) \subset [\underline{x}, \overline{x}]$, for some countable index set K, and a (monotone) mapping $\pi : [\underline{x}, \overline{x}] \to [\underline{x}, \overline{x}]$ such that (i) if $X \in J_k$ for some $k \in K$ then $\pi(X) = \mathbb{E}_F[X \mid X \in J_k]$, and (ii) $\pi(X) = X$, otherwise. Note that we label signals by their induced posterior means (hence, they interpret as recommended beliefs). Monotone partitional information structures have two properties worth mentioning: they deterministically map the state into signals, and when multiple states are pooled into a single signal, pooling is always between 'adjacent' states. Associated with any monotone partitional information structure is its induced distribution over posterior means, which satisfies:

$$G(x) = \begin{cases} F(x), & \text{if } x \notin \bigcup_{k \in K} J_k \\ F(x_k), & \text{if } x_k < x < \mathbb{E}_F[X \mid X \in J_k] \\ F(x'_k), & \text{if } \mathbb{E}_F[X \mid X \in J_k] \leq x < x'_k. \end{cases}$$

G concentrates all the mass from F at its conditional mean $\mathbb{E}_F[X \mid X \in J_k]$ on any pooling interval J_k, and leaves F unchanged otherwise.

Let $\mathcal{I}$ be the subset of $[0,1]$ on which $\nu(p) > \omega(p)$.[8] As ν and ω are continuous, $\mathcal{I}$ is the countable union of disjoint open intervals $I_k^\star = (p_k, p'_k)$, where $p_k < p'_k$ for k in some index set $K^\star$. Let $J_k^\star = (F^{-1}(p_k), F^{-1}(p'_k))$ be the interquantile range (a subset of $[\underline{x}, \overline{x}]$) corresponding to $I_k^\star$, and $\pi^\star$ the monotone partitional information structure that pairs with this convexification in the obvious way: $\pi^\star$ pools all values $X \in J_k^\star$ into a single signal $\mu_k^\star = \mathbb{E}[X \mid X \in J_k^\star]$, and for $X \notin \cup_{k \in K^\star} J_k^\star$ fully reveals the state.

Theorem 1. *A solution to problem (2) exists. The sender's optimal payoff $V(F,\nu)$ is given by*

$$V(F,\nu) = \int x d\omega(F).$$

Moreover, the monotone partitional information structure $\pi^\star$—with associated distribution $G^\star$—solves (2).

Proof. We need only argue that $G^\star$ attains $\int x d\omega(F)$. Since $G^\star$ is generated by $\pi^\star$, clearly $G^\star \preceq_{co} F$. Note that on $\left(\cup_k \overline{J}_k^\star\right)^c$, $G^\star = F$ and $\nu = \omega$ everywhere. Hence

$$\int\limits_{\left(\cup_k J_k^\star\right)^c} x d\nu(G^\star) = \int\limits_{\left(\cup_k J_k^\star\right)^c} x d\nu(F) = \int\limits_{\left(\cup_k J_k^\star\right)^c} x d\omega(G^\star).$$

Now, consider any interval $J_k^\star = (F^{-1}(p_k), F^{-1}(p'_k))$. As $(G^\star)^{-1}(p) = \mu_k^\star$ on $J_k^\star$ we have $\int_{J_k^\star} x d\nu(G^\star) = \mu_k^\star \cdot \int_{J_k^\star} d\nu(G^\star)$. As is well-known, ω is linear in p over each $I_k^\star \subset [0,1]$. Hence, there exists a constant δ such that

$$\int\limits_{J_k^\star} x d\omega(F) = \delta \int\limits_{J_k^\star} x dF, \quad \text{and} \quad \int\limits_{J_k^\star} d\omega(F) = \delta \int\limits_{J_k^\star} dF,$$

or $\int_{J_k^\star} x d\omega(F) = \frac{\int_{J_k^\star} x dF}{\int_{J_k^\star} dF} \int_{J_k^\star} d\omega(F) = \mu_k^\star \int_{J_k^\star} d\omega(F)$. Now,

$$\int_{J_k^\star} d\nu(G^\star) = \nu(G^\star(F^{-1}(p'_k))) - \nu(G^\star(F^{-1}(p_k))) = \omega(p'_k) - \omega(p_k)$$

where the second equality follows because $\nu = \omega$ and $G^\star = F$ at the endpoints $F^{-1}(p'_k)$, $F^{-1}(p_k)$. Hence, we conclude that $\int x d\nu(G^\star) = \int x d\omega(F)$. □

[8] As is well-known, ω is linear on any interval over which $\nu > \omega$. Hence, since by assumption ν has bounded variation, so too does ω.

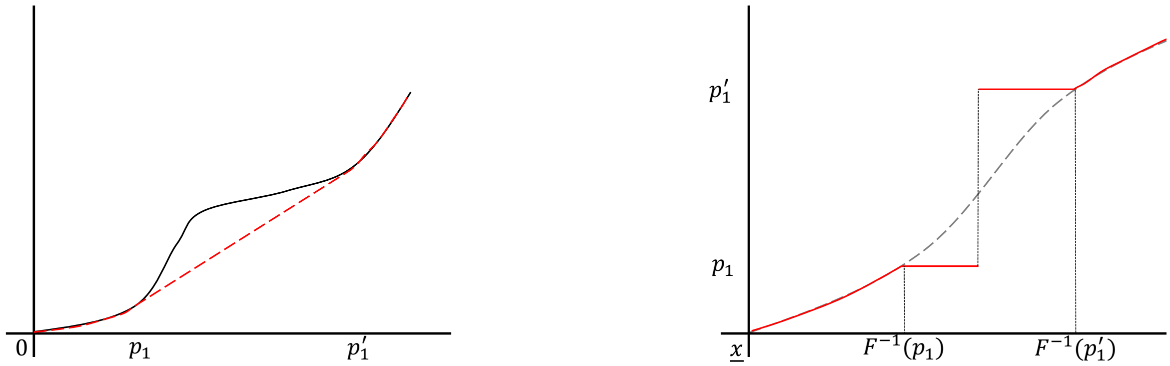

Figure 1: An illustration of Theorem 1. Left panel: weighting function ν and its convexification identify pooling intervals in rank space. Right panel: implied optimal information structure.

Theorem 1 shows that geometric arguments reminiscent of Kamenica & Gentzkow (2011) can be applied to solve conjugate persuasion problems. In particular, the sender's optimal payoff can be found by evaluating payoffs using full information and the *convexified* weighting function.[9] Moreover, despite making no such restrictions at the outset, we find that optimal communication takes a very natural form: (2) is solved by a monotone partitional information structure, whose pooling intervals correspond exactly to the relevant interquantile regions on which ω and ν differ. Of course, these regions can be identified geometrically. For example, when ν is convex, then $\nu = \omega$ everywhere and full information is optimal; when ν is strictly concave, ω is the chord connecting $(0, \nu(0))$ to $(1, \nu(1))$, and hence 'no information' is optimal.

Theorem 1 applies to environments where payoffs depend on posterior means. In many economic problems the expected state is the variable of interest, and hence a large body of research—in persuasion and beyond—has grown to address them.[10] Dworczak & Martini (2019) provide a verification method for establishing optimality in such Bayesian persuasion problems. Theorem 1 gives a simple method for solving the information design problem (2).

Theorem 1 applies when F is continuous. However, the same methods can be used to identify optimal information structures even when F admits discontinuities. In Appendix B, we identify the optimal information structures in this more general setting. In this case, some mixing may be required to accommodate the possibility of atoms at the endpoints of pooling intervals. Hence, optimal information structures need not be partitional in this

[9]In Kamenica & Gentzkow (2011), optimal payoffs can be found by applying 'no information' to a concavified Bernoulli utility.

[10]For a few examples, see Crawford & Sobel (1982); Perloff & Salop (1985); Chakraborty & Harbaugh (2010); Gentzkow & Kamenica (2016); Kolotilin *et al.* (2017); Dworczak & Martini (2019).

case.[11] Yet, as we discuss in the appendix, they continue to be monotone in a strong sense.[12]

Comparative statics

Given the characterization of optimality identified by Theorem 1, comparative statics results on optimal communication are relatively easy to describe. First, note that the optimal pooling intervals in $\pi^\star$ depend only on the interquantile ranges $I_i^\star$ over which $\omega < \nu$. These sets are independent of F. Hence, the optimal pooling intervals always vary to reflect the appropriate quantiles $F^{-1}(J_i^\star)$ of the full information distribution. An immediate consequence is that in conjugate persuasion, grading 'on a curve' is always optimal:

Corollary 1. *Suppose $\tilde{F}(x) = F(\frac{x-b}{a})$, $a > 0$. Then the sender's optimal pooling regions under distribution $\tilde{F}$ satisfy:*

$$J_i^\star(\tilde{F}) = a(J_i^\star(F) + b).$$

It is natural to ask how optimal communication depends on the sender's preferences. As Yaari (1987) shows, the sender becomes more risk-loving as ν becomes more convex. The next proposition verifies that as this happens, he shares more information:

Proposition 1. *Let $\rho : \mathbb{R} \to \mathbb{R}$ be an increasing convex function. Then the optimal information structure with objective $\rho \circ \nu$ is less (Blackwell) informative than that under ν.*

Consider two such weighting functions, ν and $\tilde{\nu} = \rho \circ \nu$. Since monotone partitions are always optimal, Proposition 1 says that optimal pooling intervals, as a function of preferences, are ordered by set inclusion: for each $J_i^\star(\tilde{\nu})$ there is some $J_k^\star(\nu)$ such that $J_i^\star(\tilde{\nu}) \subset J_k^\star(\nu)$. Indeed, this follows immediately from the proof, which shows that the regions on which $\tilde{\omega} := \text{vex}(\tilde{\nu}) < \tilde{\nu}$ are contained in the intervals on which $\omega < \nu$.

Extreme Censorship

Define the class of *convex-concave* functions $\nu : [0, 1] \to \mathbb{R}$ as those for which there exists a $\overline{p} \in (0, 1)$ such that ν is convex on $[0, \overline{p}]$ and concave on $[\overline{p}, 1]$. We define the class of *concave-convex* functions analogously. As will become apparent from our applications, functions from these classes arise naturally in a variety of economically interesting applications. The convexification of such functions is particularly simple to find and describe:

[11] See section 4 for a simple example of this, in the context of insurance markets.

[12] For instance, an implication of the optimal information structure is that the conditional distributions over posterior means (given X) are ordered by first-order stochastic dominance.

Proposition 2. *Suppose* ν *is convex on* $[0,\overline{p}]$ *and concave on* $[\overline{p},1]$ *for some* $\overline{p}\in(0,1)$*. Then there exists a unique* $p^c\in[0,\overline{p}]$ *such that*

$$\omega(p)=\begin{cases}\nu(p) & ,\ 0\le p\le p^c\\ \left(\frac{1-p}{1-p^c}\right)\nu(p^c)+\left(\frac{p-p^c}{1-p^c}\right)\nu(1) & ,\ p^c\le p\le 1.\end{cases}$$

If ν *is differentiable, then* p^c *satisfies*

$$\nu'(p^c)\ge\frac{\nu(1)-\nu(p^c)}{1-p^c}, \tag{5}$$

holding with equality iff $p^c>0$*. If* p^c *is interior, it is the unique solution to* $\nu'(p)=\frac{\nu(1)-\nu(p)}{1-p}$*.*

Proposition 2 is stated for convex-concave functions, but extends to concave-convex functions in the obvious way. We define two classes of 'extreme censorship' policies as follows: right censorship is an information structure which is characterized by a threshold $x_r\in[\underline{x},\overline{x})$ such that (i) if $X_i<x_r$, then X_i is fully revealed and (ii) if $X_i\in[x_r,\overline{x}]$ then no signal is sent. Similarly, left censorship is characterized by a threshold $x_l\in(\underline{x},\overline{x}]$ such that (i) if $X_i>x_l$, then X_i is fully revealed and (ii) if $X_i\in[\underline{x},x_l]$ then no signal is sent. We call a right censorship (left-censorship) policy informative if $x_r<\overline{x}$ ($x_l>\underline{x}$).[13] With ν convex-concave or concave-convex, Theorem 1 implies censoring extreme information is always optimal:

Corollary 2. *If* ν *is convex-concave then a right censorship policy is optimal. If* ν *is concave-convex, then left censorship is optimal.*

In the next three sections, we show how our results apply in three kinds of economic problem. Section 4 shows how we can reframe a class of i.i.d. Bayesian persuasion problems—similar to Example 1—in terms of conjugate persuasion. In section 5, we connect problem (2) to a behavioral model of rank-dependent preferences, and assess the consequences of such preferences for equilibrium insurance and optimal income redistribution. Section 6 briefly discusses the implications of our analysis for Bayesian persuasion with heterogeneous priors.

4 Order Statistics and Persuasion of Multiple Receivers

In Example 1, the sender's task was to solve the following i.i.d. Bayesian Persuasion problem

$$\max_{G\preceq_{co}F}\mathbb{E}[u(x_1,\dots,x_N)]=\sum_{k=1}^{N}\alpha_k\mathbb{E}_G[x^{(k)}], \tag{6}$$

[13]Note the definition of censorship does never includes full information when F is continuous.

where $\mathbb{E}_G[x^{(k)}]$ is the expected value of the k^{th} highest order statistic from N i.i.d. draws from a distribution G. In problem (6) the sender is an expected utility maximizer, whose preferences over x_i may depend on their relative position among the N draws. Still, it can be expressed in the form (2). To see this, recall that the k^{th} highest order statistic from N i.i.d. draws of distribution G has c.d.f.

$$\Pr(x^{(k)} \leq y) = \beta_{N-k+1,k}(G(y)) = N \begin{pmatrix} N-1 \\ k-1 \end{pmatrix} \int_0^{G(y)} G^{N-k}(1-G)^{k-1} dG,$$

where $\beta_{N-k+1,k}$ is a beta distribution with parameters $N-k+1$ and k. Hence, $\mathbb{E}_G[x^{(k)}] = \int x d\beta_{N-k+1,k}(G)$ and the sender's problem can be written

$$\max_{G \preceq_{co} F} \int x d\tilde{\beta}_N(G), \tag{7}$$

a conjugate persuasion problem with $\tilde{\beta}_N = \sum \alpha_k \beta_{N-k+1,k}$. Since each $\beta_{N-k+1,k}$ is obviously continuous and increasing, the conditions of Theorem 1 apply to $\tilde{\beta}_N$. Let the solution to problem (7)—as described by Theorem 1—be $G^\star_N$.

Any i.i.d. Bayesian Persuasion problem in which the sender's marginal utility over outcomes is rank-dependent is therefore expressible as a conjugate persuasion problem. On the other hand, any conjugate persuasion problem is the uniform limit of some such sequence of i.i.d. Bayesian Persuasion problems:

Theorem 2. *Suppose ν is continuous. Then there exists a sequence of functions $\tilde{\beta}_N : [0,1] \to \mathbb{R}$, $N = 1, 2, \ldots$, such that $\tilde{\beta}_N$ converges uniformly to ν, and*

$$\int x d\tilde{\beta}_N(\tilde{G}_N) \to \int x d\nu(G^\star).$$

Moreover, if $G^\star$ uniquely solves (2) (up to a set of measure 0), then $\tilde{G}_N$ converges weakly to $G^\star$.

To prove the result, for each n in the sequence we assign weights $\alpha^n_k = \nu(1-\frac{k-1}{n}) - \nu(1-\frac{k}{n})$. As a result, each $\tilde{\beta}_n$ can be expressed as a linear combination of Bernstein basis polynomials, from which uniform convergence to ν follows naturally. As the proof shows, uniform convergence guarantees the $\tilde{G}_N$ have an upper hemicontinuity property (in particular, there is always a subsequence of $(\tilde{G}_N)$ which converges weakly to a solution to (2)), and when $G^\star$ is unique $\tilde{G}_N$ converges weakly to it.[14]

[14]When $G^\star$ is not unique but F has full support, any weakly convergent subsequence of $(\tilde{G}_N)$ limits to a

Theorem 2 shows a close connection between conjugate persuasion and i.i.d. Bayesian persuasion with rank-based utility. On one hand, this connection allows us to solve i.i.d. persuasion problems with the methods of Theorem 1. On the other, it illustrates that problem (6) captures a rich set of incentives: *any* pattern of monotone partitional disclosures is approximately optimal in some such problem.[15]

Moreover, it provides a set of sufficient conditions under which a sequence of i.i.d. persuasion problems has nice convergence properties, as the number of draws grows large. Indeed, the uniform convergence of $\tilde{\beta}_N$ is important for the usefulness of the limit as a guide to optimal information in large societies; pointwise convergence is insufficient.[16] By establishing natural conditions under which a sequence of i.i.d. Bayesian persuasion problems converge, Theorem 2 allows us to identify robust, approximately optimal policies for large i.i.d. persuasion problems. This can be useful: as our examples below illustrate, it can offer a simple guide to optimal policy in large societies—even when solving for the precise optimum in fixed societies can be tricky.

In the next two subsections, we apply our results to study information design in two classic settings: oligopoly markets and elections.

4.1 Oligopoly, advertising and information design

We adapt the classic Perloff & Salop (1985) model of differentiated products oligopoly to study the implications of information design for the distribution of consumer and producer surplus. A buyer with quasilinear utility has unit demand for a good. There are $N \geq 3$ sellers in the market, who each produce a heterogenous variety at a marginal cost normalized to 0.[17] The buyer's willingness to pay X_n for each item $n = 1, \ldots, N$ is i.i.d. across sellers, distributed according to a continuous distribution function F on bounded support $[0, \overline{x}]$, where $\overline{x} > 0$. The buyer is initially uninformed of her own value for each good; hence, there is a role for advertising.

We assume seller n can costlessly provide information about his own good, modeled as a choice of information structure $\pi_n : X_n \to \Delta S_n$, where S_n is his signal space. As before, this

monotone partitional $\hat{G}$, whose pooling intervals include *(i)* all intervals $\{J_i^\star\}_{i \in \mathcal{I}}$, *(ii)* possibly others *over which ω is linear*; any such $\hat{G}$ differs from $G^\star$ in a trivial way only. Details available on request.

[15] Pooling at arbitrary intervals $I_i = [p_i, p_i'] \in (0,1)$, $i \in \mathcal{I} \subset \mathbb{N}^+$, is optimal for the following choice of ν. Let $\nu_0(p) = p^2$, and $\nu_i(q) = \min\{\nu_0(p_i) + 2p_i'(q - p_i), \nu_0(p_i') + 2p_i(q - p_i')\}$. Then put $\nu(p) = \nu_0(p)$, $p \notin \bigcup_{\mathcal{I}} I_i$, and $\nu(p) = \nu_i(p)$ for $p \in I_i$.

[16] One can find sequences of weighting functions which converge pointwise to some limit, but whose corresponding optimal information structures do not converge to a solution of the limiting weighting function. See section 4.1.

[17] We omit the case $N = 2$—which is trivial but requires slightly different arguments—for the sake of brevity.

corresponds to a choice of distribution H_n over the posterior expectation of X_n satisfying $H_n \preceq_{co} F$. We refer to H_n as the seller's advertising strategy. Notice that we assume sellers can only provide information about their own good. Hence, the buyer's equilibrium information inherits the independence of $\{X_n\}_{n=1}^N$.

We are interested in a policy question: how can information affect the surplus and distribution of gains in oligopolistic markets? To address this, we introduce a policymaker who can impose constraints on seller's advertising strategies. In particular, he can choose an upper limit $G \preceq_{co} F$ on the information sellers can provide, which we assume to be symmetric across sellers.[18,19] Subject to the constraint G, each seller's set of available advertising strategies is $\{H_n : H_n \preceq_{co} G\}$.

The timing of the game is as follows: first, the policymaker chooses G. Then sellers simultaneously commit to advertising strategies, which are observed by the buyer. Each H_n induces a draw x_n of the buyer's expected value for item n, also observed by the buyer and by seller n; rival sellers do not learn x_n.[20] Sellers then compete in prices, which we model as a descending auction.

Let CS be the buyer's ex ante expected utility, and correspondingly PS sellers' joint profits. To study the payoffs feasible with i.i.d. information, we endow the policymaker with payoffs $\alpha_B CS + \alpha_S PS$, where α_B, α_S are real constants. Of course, this accommodates a policymaker who wishes to maximize (or minimize) consumer surplus, producer surplus and aggregate surplus ($\alpha_B = \alpha_S = 1$), among others. Note that this flexibility allows for multiple interpretations of the policymaker.[21]

We now argue that the policymaker's problem can be written as a special case of problem (2). If seller n offers an item worth x_n at price p_n, then the buyer's net utility from item n is $s_n = x_n - p_n$. Clearly, the buyer chooses item n if and only if $s_n > \max_{m \neq n} s_m$, and the winning seller receives $p_n = x_n - s_n$.[22] Hence, we can consider the price-setting subgame as a second-price auction in offers of s_n, treating the each seller 'as if' he holds value $x_n \sim H_n$ for winning. Using this analogy, it is easy to characterize sellers' equilibrium advertising strategies:

[18] One might imagine other forms of constraint would be natural; for instance, the platform could impose a minimum bound on information. As we will see, such a constraint would not be binding.

[19] The symmetry can be motivated by fair treatment considerations or a desire to avoid any perception of favoritism on the part of the policymaker.

[20] In Perloff & Salop (1985), sellers have no information on how much the buyer values their product. We have in mind settings where the seller can observe its customers' interactions with its advertising and thereby learn the 'match quality' between customer and good.

[21] Examples include: regulatory/consumer protection agency, a platform through which advertisements can be posted or simply a tool for comparative statics and welfare analysis of exogenous information.

[22] In this setting, the possibility of ties has no consequence for equilibrium strategies and payoffs—and so we ignore them here.

Lemma 1. *Fix $G \preceq_{co} F$. Maximal information, $H_n = G$, is optimal for each seller.*

The intuition behind Lemma 1 is straightforward. The pricing subgame is a second-price auction, and advertising gives the seller more information about his value x_n for winning. Since advertising is costless and not observed by rival sellers, this can never be harmful. Moreover, it strictly improves his expected utility when his bid varies with his type.

By familiar arguments, the consumer's surplus is equal to $CS = \mathbb{E}_G[x^{(N-1)}]$ in equilibrium. Moreover, since the most valuable item is always purchased, aggregate surplus is $\mathbb{E}_G[x^{(N)}]$ and accordingly joint profits are $PS = \mathbb{E}_G[x^{(N)}] - \mathbb{E}_G[x^{(N-1)}]$.

Hence, the policymaker's objective is to maximize the expected weighted sum of the two highest order statistics. From the discussion of the previous section, the policymaker's problem is therefore a conjugate persuasion problem:

Lemma 2. *Given equilibrium behavior of buyer and sellers, the policymaker's problem is*

$$\max_{G \preceq_{co} F} \int x d\tilde{\beta}_N(G), \tag{8}$$

where $\tilde{\beta}_N = (\alpha_B - \alpha_S)\beta_{(N-1,2)} + \alpha_S \beta_{N,1}$.

We can now apply Theorem 1 to characterize the scope for advertising to generate surpluses for both sides of an oligopolistic market. We write

$$\mathcal{E} := \left\{(CS, PS) : \exists G \preceq_{co} F \text{ such that } CS = \mathbb{E}_G[x^{(N-1)}], PS = \mathbb{E}_G[x^{(N)}] - \mathbb{E}_G[x^{(N-1)}]\right\}$$

for the set of equilibrium payoff pairs attainable with some information structure G. Let $r = \frac{\alpha_S}{\alpha_B}$ represent the relative weight the policymaker places on profits (which may be negative).

Proposition 3. *$\mathcal{E}$ is compact and convex. Hence, its extreme points are characterized by the solutions to (8), which are extreme censorship policies for all α_B, α_S. When $\alpha_B > 0$, $G^\star$ is (i) a left-censorship policy, for $r > 1$; (ii) fully informative, $G^\star = F$, for $r \in [\frac{1}{2}, 1]$; (iii) a right-censorship policy, for $r < \frac{1}{2}$. As $r \downarrow 1$ or $r \uparrow \frac{1}{2}$, $G^\star$ becomes (Blackwell) more informative. When $\alpha_B < 0$, $G^\star$ is (i) completely uninformative for $r > \frac{1}{n}$. Otherwise, left-censorship is optimal. As $r \uparrow \frac{1}{n}$, $G^\star$ becomes less (Blackwell) informative.*

Proposition 3 shows that extreme censorship policies play an important role in distributing gains from trade between buyers and sellers in oligopolistic markets.[23] Indeed, any feasible pair (CS, PS) can be attained by a convex combination of such information structures. Moreover, the maximal payoffs in $\mathcal{E}$ are all attainable with some extreme censorship

[23] It is easy to verify that either *(i)* full or *(ii)* no information are always optimal for $N = 2$.

policy. When the policymaker's payoff increases in both CS and PS, there is a range of parameters for which he simply prefers full information. In particular, this occurs whenever he has a mild bias towards the buyer ($r \in [\frac{1}{2}, 1]$). The full information outcome is obviously Pareto efficient; indeed, total surplus is $\mathbb{E}[x^{(1)}] = \int x d\beta_{N,1}(G)$, where $\beta_{N,1}$ is strictly convex. Interestingly, this implies that $\mathcal{E}$ is kinked at the full information payoffs.

However, beyond this point, all other values on the maximal frontier of $\mathcal{E}$—the set $\{(CS, PS) \in \mathcal{E} : \nexists (CS', PS') \in \mathcal{E} \text{ s.t. } (CS, PS) < (CS', PS')\}$—are necessarily inefficient. Indeed, when α_B, $\alpha_S \geq 0$ and $r \notin [\frac{1}{2}, 1]$ the policymaker optimally sacrifices aggregate surplus in order to skew the distribution of the surplus to his preferred side of the market. If the policymaker puts more emphasis on profits ($r > 1$) then left censorship is optimal. Conversely, if he emphasizes consumer surplus ($r < \frac{1}{2}$), then right censorship is best.

To develop some intuition, we briefly discuss two special cases. Let $G^{\star}_B$ be the consumer surplus-maximizing information structure, and analogously define $G^{\star}_S$ for the case of joint profits. Proposition 3 implies the consumer prefers right-censorship:

Corollary 3. *CS is maximized by a right-censorship policy. As N increases, $G^{\star}_B$ becomes (Blackwell) more informative and as $N \to \infty$, the buyer's value for an arbitrary item n is censored with probability approaching* 0.

Right censorship favors the consumer by increasing the expected value of the second order statistic, pooling values likely to be second best with those likely to be best. While this reduces the overall surplus from trade relative to full information the buyer is more than compensated by the lower prices that reduced differentiation between the best sellers brings about. However, more censorship is not always better—if the censorship threshold is set too low, the second-best value is also pooled with those much lower. Here, the costs of inefficient misallocation more than offset the reduction in prices. In the appendix, we apply Proposition 2 to prove that the optimal censorship threshold increases with N. As N grows, the values of the "top two" items are likely to fall further into the right of the distribution. To benefit from optimal pooling, the censorship threshold must also move right.

By contrast, left-censorship benefits sellers:

Corollary 4. *PS is maximized by a left-censorship policy. $G^{\star}_S$ becomes (Blackwell) less informative. As $N \to \infty$, the buyer's value for any item n is censored with probability approaching* 1.

Pooling sufficiently low values weakens the expected competition faced by the seller with the best good. To see why, consider the case where only one item exceeds a censorship threshold. If the seller knew this, he would be in a strong bargaining position: the buyer's outside option would be the mean value among the other items, rather than the highest value among them. On the other hand, if another seller's item also exceeds the threshold this advantage would be diminished. Applying Proposition 2 again, we show that as N increases this trade-off pushes towards greater censorship: with many sellers the policymaker can set a high threshold, set so that on average one seller (with a product close to the maximum value, $\overline{x}$) exceeds it.

Finally, we discuss some efficiency implications of these advertising standards for $N \to \infty$, which are also identified in Bergemann *et al.* (2021):

Corollary 5. *As $N \to \infty$, maximal consumer surplus approaches $\overline{x}$. Maximum producer surplus approaches $((1-2e^{-1}) \cdot z, e^{-1} \cdot z)$, where $z = \overline{x} - \mathbb{E}_F[X]$.*

Of course, when advertising standards are designed to maximize consumer surplus, equilibrium payoffs must converge to an efficient outcome in which the consumer extracts all the surplus, $\overline{x}$. After all, the policymaker could always allow full information. For N large, the probability that at least two sellers have a value close to $\overline{x}$ approaches 1, so that $\mathbb{E}_F[x^{(2)}] \to \overline{x}$.[24]

However, advertising standards which favor sellers are bound to sacrifice surplus. To maximize joint profits, we show in the appendix that profit-maximizing standards censor all realizations of X_n below the $\left(\frac{N-2}{N-1}\right)^{\text{th}}$ quantile of F. As $N \to \infty$, the probability that any seller's product avoids censorship is therefore $\frac{1}{N-1}$ and the overall number of such 'successes' converges to a Poisson distribution with mean 1. Moreover as the censorship threshold approaches $\overline{x}$ in the limit, an item has value $E_F[X]$ if it is censored and $\overline{x}$ otherwise. As a result, the expected gains from trade are $(1-e^{-1})\mathbb{E}_F[X] + e^{-1}\overline{x} = \mathbb{E}_F[X] + (1-e^{-1})z$, where e^{-1} is the probability of no success under the Poisson distribution. Similarly, equilibrium joint profits—which depend on the difference $x^{(N)} - x^{(N-1)}$—are $e^{-1} \cdot (\overline{x} - \mathbb{E}_F[X])$, where here e^{-1} is the probability that exactly one seller's item exceeds the censorship threshold.

Interestingly, joint profit maximization necessitates sacrificing $e^{-1}z$ of surplus — approximately $\frac{1}{3}$ of the difference between the full information and no information outcomes. While censorship is extremely likely for an individual seller, the threshold is calibrated so that in expectation a single item exceeds it. In this event, such a seller enjoys the market power associated with having a good worth more than competing items. However, with the restriction to i.i.d. advertising, these standards must also accept a probability of e^{-1} that all sellers' products fail to meet the threshold and the subsequent inefficiency that arises.

Finally, we briefly on remark the joint profit maximization problem, with reference to Theorem 2. It is easy to verify that, as $N \to \infty$, $\beta_{N,1} - \beta_{N-1,2}$ converges pointwise to 0.

[24]Bergemann *et al.* (2021) consider the case $\overline{x} = \infty$. Here, full information does not approximate the optimal information structure if F has 'fat tails'.

One might be tempted to conclude that this implies that joint profits necessarily converge to 0 irrespective of information. However, as Corollary 5 shows, this is not the case. The reason for the discrepancy is that the convergence of $\beta_{N,1} - \beta_{N-1,2}$ is not uniform. In light of Theorem 2, it is not surprising that the approximation fails in this case.

4.2 Media and Electoral Influence

Consider an election in which two office-motivated politicians compete for the votes of $2N+1$ voters, $N \in \mathbb{N}^{+}$. Before the vote, each politician $j \in \{1, 2\}$ commits to a policy platform, a choice of some level of public good provision, $a_j \in [\underline{x}, \overline{x}]$. Voters have independent values for the public good, each voting for the candidate whose platform she prefers. Specifically, voter n's utility depends on the implemented policy a and her type $x_n \sim F$ as follows:

$$u_n(a, x_n) = -(a - x_n)^2.$$

That is, each voter has a quadratic loss function over the policy choice, with her ideal policy being x_n. To ease exposition, we again assume F is continuous on $[\underline{x}, \overline{x}]$.

Voters are initially uninformed about the costs and benefits surrounding this policy issue, and hence do not observe their own x_n (or anyone else's). However, before politicians commit to their respective choices of a, a biased lobbyist (he) can provide voters with information. This information maps each buyer's value into a signal.[25] We assume the lobbyist is not able to individually identify and target particular voters to receive different information structures. In other words, he must choose symmetric information structures.[26]

Let $x^{\star}$ denote the policy implemented in equilibrium. We consider two forms of lobbyist bias. A lobbyist is *right-biased* if he wishes to maximize $\mathbb{E}[x^{\star}]$, regardless of voters' preferences. He is *left-biased* if he wishes to minimize the expected policy.

Proposition 4. *The lobbyist can influence the election; that is, $\mathbb{E}_G[x^{\star}]$ is not independent of G. A right-biased lobbyist optimally influences the election via an informative right-censorship policy. A left-biased lobbyist optimally employs informative left-censorship. For each voter, censorship occurs with a probability of at least $\frac{1}{2}$, regardless of N.*

[25] As voters are assumed to vote their preference, it does not matter whether they can observe each others' signals. However, politicians are able to see the signals. This is appropriate where, for instance, campaigns conduct private polls of voters.

[26] One could interpret symmetry here in terms of a public campaign which is designed to reveal (perhaps coarsely) to each individual their own preferences. Symmetry then corresponds to the assumption that any two agents with the same type respond the same way to the news. Of course, in some settings symmetry is less appropriate. For instance, one might expect such targeting is possible via social media; our results are therefore better viewed in the context of traditional mass media.

In this setting, the median voter theorem applies to policy offers. As a result, a right-biased lobbyist's problem is simply to maximize the expected value of the median, $\mathbb{E}_G[x^{(N+1)}]$. This is a conjugate persuasion problem, with weighting function $\nu = \beta_{N,N}$, which is convex-concave for any N. Moreover, $\beta_{N,N}$ is strictly concave on $[0, \frac{1}{2}]$ and strictly convex on $[\frac{1}{2}, 1]$. From Proposition 2, we can conclude that right-censorship is optimal, with threshold located at some rank p^c of F satisfying $p^c \in (0, \frac{1}{2})$. In particular, the optimal threshold is interior, so that it is optimal for the lobbyist to provide at least some information. Still, information is withheld from voters with probability exceeding $\frac{1}{2}$. In contrast to the oligopoly advertising problem, in elections partial censorship remains an important tool even in the limit of large populations. In the Appendix, we further prove the following result for $N \to \infty$:

Proposition 5. *As the number of voters increases, the lobbyist provides more information. As $N \to \infty$, each voter's true value is censored with probability $\frac{1}{2}$.*

Does media coverage influence society, or simply reflect it? Our results are consistent with both perspectives. Taking F as given, Proposition 4 shows that the media can influence elections. However, Corollary 1 shows that as society's preferences move to the right (left)—for example, if F is translated—then the lobbyist's censorship threshold moves in sympathy.

5 Persuasion and non-Expected Utility

Conjugate persuasion problems also arise in settings where agents' risk preferences violate the independence axiom (IIA) of EU theory. For instance, the objective function in (2) arises in models of rank-dependent utility (Quiggin (1982), Yaari (1987)), as well as in prospect theory (Kahneman & Tversky (1979), Tversky & Kahneman (1992)). These models have been able to rationalize several empirical puzzles, including the common ratio effect (Allais (1953), Kahneman & Tversky (1979)) and the favorite-longshot bias (Griffith (1949)), as well as to separate risk preferences from assumptions on how utility varies with income (Yaari (1987)). Moreover, they provide a foundation for inequality preferences that can be represented by Gini coefficients (Dorfman (1979))—which cannot be achieved in the EU framework (Newbery (1970)).

To fix ideas, we briefly connect our problem to the rank-dependent approach of Yaari (1987). Consider a sender whose preferences over lotteries are determined entirely by the distribution over outcomes in $[\underline{x}, \overline{x}]$. Assume his preferences are complete, transitive, reflexive and continuous. In place of IIA, Yaari (1987) adopts the alternative axiom of *dual independence*: where independence is taken with respect to mixtures over probabilities of each outcome, dual independence is taken with respect to mixing of the quantile functions.

Yaari's axiom can be described as follows. Consider two strictly increasing, continuous distributions F, G on $[\underline{x}, \overline{x}]$, and suppose the sender prefers F over G.[27] For a third, irrelevant distribution H and some weight $\alpha \in (0,1)$, the mixture $\alpha F \boxplus (1-\alpha)H$ generates outcome $\alpha F^{-1}(p) + H^{-1}(p)$ as its p^{th} quantile:

$$\alpha F \boxplus (1-\alpha)H := \left(\alpha F^{-1}(p) + H^{-1}(p)\right)^{-1}. \tag{9}$$

In this notation, dual independence can be stated as:

Axiom 1. *(Dual independence: Yaari (1987)) If $F \prec G$ then $\alpha F \boxplus (1-\alpha)H \prec \alpha G \boxplus (1-\alpha)H$ for all $\alpha \in (0,1)$ and distributions H.*

Under Axiom 1, preferences can be represented by Yaari utilities (Yaari (1987)). A slight rewriting shows that (2) can be considered a persuasion problem in which the sender's risk preferences obey Axiom 1:

Theorem 3. *Suppose the sender's preferences $\succeq$ over distributions on $[\underline{x}, \overline{x}]$ are complete, transitive, reflexive, continuous and obey Axiom 1. Then there exists a continuous, non-decreasing function $\nu[0,1] :\to \mathbb{R}$ such that*

$$\int x d\nu(G) \tag{10}$$

represents $\succeq$.

Proof. From Yaari (1987), there exists a continuous, non-increasing function $\phi : [0,1] \to \mathbb{R}$ such that $F \succeq G$ if

$$\int \phi(F(x))dx \geq \int \phi(G(x))dx. \tag{11}$$

Defining $\nu = -\phi$, and integrating by parts, establishes the result. □

Under Axiom 1, the sender's preferences over lotteries are rank-based: the value he places on the outcome depends on its rank p in the distribution. To see this, suppose ν is differentiable. Then, for an arbitrary distribution G the sender's utility can be written $\int G^{-1}(p)\nu'(p)dp$: a weighted sum, where the marginal utility associated with the outcome $x = G^{-1}(p)$ at rank p is $\nu'(p)dp$.

Empirical studies of decision-making under uncertainty find a consistent, "four-fold pattern" (Tversky & Kahneman (1992)): people are risk-loving (averse) when it comes to gambles on "longshot", low probability gains (losses). However, their risk preferences reverse as

[27] For ease of exposition, we describe the dual independence here for increasing, continuous distributions, where each distribution and quantile function have a well-defined inverse. The property itself applies to the class of all distributions on $[\underline{x}, \overline{x}]$ – see Yaari (1987).

the gain (loss) becomes more likely.[28] Tversky & Kahneman (1992) have shown that these preferences can be represented by (10), if ν is increasing and concave-convex. Similarly, Snowberg & Wolfers (2010) find evidence favoring (10)—with ν increasing, concave-convex—over expected utility as an explanation for the favorite-longshot bias.

Concave-convex weighting functions corresponds to preferences which place 'too much weight' on extreme outcomes. To see this, we define the concept of extreme bias. In what follows, we take ν to be increasing and normalize $\nu(0) = 0$, $\nu(1) = 1$; hence, we can consider it a distribution function. With a small abuse of notation, we also write ν for the corresponding measure. Let λ denote Lebesgue measure.

Definition 1. An agent with preferences represented by (10) is extreme-biased if there exists $p_m \in [0, 1]$ such that $\nu(p_m) = p_m$ and for any Borel sets B, C such that $\inf_{x \in B} \mid x - p_m \mid \geq \sup_{x \in C} \mid x - p_m \mid$:

$$\frac{\nu(B)}{\nu(C)} \geq \frac{\lambda(B)}{\lambda(C)}, \tag{12}$$

with strict inequality for some B, C.

Extreme bias refers to the way in which the agent weights outcomes in at the extreme ranks of a distribution. Recalling that the rank is uniform on $[0, 1]$, an agent suffers extreme bias if ν places relatively more weight on *extreme* outcomes in any distribution (relative to q) than is justified by the true (i.e., Lebesgue) distribution. For instance, when $p_m = 1/2$ Definition 1 requires that the agent put 'too much weight' on outcomes far away from the median. The following lemma shows that extreme-biased preferences are equivalent to a concave-convex weighting function:[29]

Lemma 3. *An agent with utility (10) is extreme biased if and only if ν is concave-convex.*

Motivated by their empirical relevance, we now use our framework to study the implications of agents with extreme bias in applications to insurance and redistribution.

5.1 Competitive insurance under extreme bias

We use the following applications to illustrate how the tools of section 3 can address other (non-informational) design problems. Moreover, they provide contexts in which the constraint in (2) need not be continuous, in spite of the underlying primitives.

Recall Example 3. A consumer (she) faces a risky income $X \sim F$, a continuous distribution on $[\underline{x}, \overline{x}]$. Her preferences over risk are described by (10), and she suffers from extreme

[28] See, for example, Tversky & Kahneman (1992); Tversky & Fox (1995); Wu & Gonzalez (1996)

[29] Alternatively, Prelec (1998) provides a collection of choice axioms under which ν is concave-convex.

bias. To hedge her risks, she may purchase insurance in a perfectly competitive market. Suppose for now that each seller s, $s = 1, \dots,$ can offer an *insurance contract*, which specifies *(i)* an upfront premium $\rho_s \geq 0$, and *(ii)* a contingent-consumption $Y_s : [\underline{x}, \overline{x}] \to [\underline{x}, \overline{x}]$, where the random variable $Y_s(X)$ represents a mean preserving contraction of X. Throughout this section, we assume Y_s is measurable and that the consumer has no private information about X.

Of course, in equilibrium firms must earn 0 profits so that $\rho_s = 0$ for all s and by the usual logic firms must choose a profile Y_s that maximizes the consumer's utility. Hence, equilibrium contracts follow as an immediate corollary of Theorem 1:

Corollary 6. *When sellers may only offer insurance contracts, the consumer's equilibrium payoff is* $\int x d\omega(F)$*: all sellers set* $\rho_s = 0$ *and*

$$Y_s(X) = \begin{cases} \mathbb{E}_F[X \mid X \leq F^{-1}(p^c)] & \text{if } X \leq F^{-1}(p^c) \\ X, & \text{otherwise.} \end{cases}$$

where p^c *is identified by Proposition 2 for* ν *concave-convex.*

Under extreme bias, equilibrium insurance offers partial coverage. Each insurer offers to protect the consumer from downside risks to her income. But they do not fully hedge her risk: above a threshold, the contract does not interfere with her income. This is a common feature of insurance contracts in private markets.[30] In the standard theory, partial insurance often arises as an inefficient response to asymmetric information. By contrast, in this setting a realistic form of partial insurance arises as an efficient response to empirically plausible preferences. Moreover, by Corollary 1 consumers continue to use insurance to partially hedge their risks, irrespective of (mean) income—not a general feature under expected utility.

Of course, the consumer is not risk averse: if she were, full insurance would have been an equilibrium. Hence, we might wonder if a seller could do better by offering something other than an insurance contract. To that end, suppose each seller can offer a pair (ρ_x, Y_s), where now the only restriction on Y_s is that $\underline{x} \leq Y_s(X) \leq \overline{x}$ for all X.[31] Define $B : [\underline{x}, \overline{x}] \to \mathbb{R}$ by

$$B(x) = \begin{cases} \frac{\overline{x} - \mathbb{E}_F[x]}{\overline{x} - \underline{x}}, & \underline{x} \leq x < \overline{x}, \\ 1, & x = \overline{x}. \end{cases}$$

[30] And even outside of private markets — for example, unemployment insurance.

[31] For our result, we only require payments be bounded from above by some constant. In particular, a simple application of Theorem 4 (Appendix) shows that we could allow unboundedness from below without affecting the results. In the space of contracts with bounded payments to the consumer, assuming the decomposition (ρ_x, Y_s) is obviously without loss of generality.

B is the distribution function associated with a binary lottery on $\{\underline{x}, \overline{x}\}$ with mean $\mathbb{E}_F[x]$. As such, it obviously satisfies $F \preceq_{co} B$. Exactly as above, in equilibrium firms set $\rho_s = 0$ and choose Y_s to maximize the consumer's payoff, subject to earning zero profits. Hence, equilibrium insurance contracts correspond to a mean preserving contraction of B.

As we needed no continuity assumptions to derive the bound (1), it is immediate that the consumer's equilibrium payoff is bounded above by $\int x d\omega(B)$. However, as B is not continuous, Theorem 1 does not directly apply. Nonetheless, in Appendix B we show how to deal with discontinuities. As a result, equilibrium still admits a simple characterization:

Corollary 7. *An insurance equilibrium exists. The consumer's equilibrium payoff is* $\int x d\omega(B)$. *If* $\frac{\overline{x}-\mathbb{E}_F[x]}{\overline{x}-\underline{x}} \leq p^c$, *then consumption is distributed according to* B. *Otherwise, consumption is distributed according to*

$$G^\star(x) = \begin{cases} \overline{x}, & \textit{with probability } (1-p^c) \\ \frac{\mathbb{E}_F[X]-(1-p^c)\overline{x}}{p^c}, & \textit{with probability } p^c \end{cases}$$

Similarly to Theorem 1, Theorem 4 shows that the optimal distribution $G^\star \preceq_{co} B$ pools all values of $B^{-1}(p)$ for $p \in [0, p^c]$ (to take advantage of the concavity of ν) and otherwise separates (taking advantage of convexity). When $\frac{\overline{x}-\mathbb{E}_F[x]}{\overline{x}-\underline{x}} \leq p^c$, the optimal distribution nonetheless corresponds to full separation of $\overline{x}$ and $\underline{x}$. However, when $\frac{\overline{x}-\mathbb{E}_F[x]}{\overline{x}-\underline{x}} > p^c$, there is an optimal role for insurance. Intuitively, the optimal distribution $G^\star$ achieves the desired pooling probabilistically: with some probability $\overline{x}$ is pooled with $\underline{x}$ for a mean income $\frac{\mathbb{E}_F[X]-(1-q^\star)\overline{x}}{q^\star}$, and is separated otherwise. In this way, competitive insurance markets retain the key features of Corollary 6 even when they can offer gambles to customers. Specifically, they continue to offer partial insurance of downside risks—so long as that downside constitutes a rare loss.

5.2 Redistribution

We end with a brief description of how our results relate to income redistribution using budget-balanced transfers. Consider an economy made up of a continuum of agents with income distribution F. As we mentioned in Example 4, the (generalized) Gini coefficient is a natural measure of inequality. As Newbery (1970) shows, there is no Bernoulli utility whose expected value ranks distributions in the same order as their Gini coefficient. On the other hand, the Gini coefficient orders inequality (for any fixed mean) in the same way as (10) for $\nu(p) = -p^2$ (Dorfman (1979)). Similarly, the generalized Gini coefficient (3) can be expressed in terms of (10). Integrating by parts, the generalized Gini represents the same preferences ordering as:

$$\frac{\int x d\nu(G)}{\mu},$$

where $\mu = \mathbb{E}_F[x]$ is (fixed) mean income and $\nu'(p) = -\int_0^p \gamma(q)dq$. Hence, we can interpret $-\gamma(p)$ as a rate of change in the welfare weights the policymaker assigns to agents at each rank of the distribution:[32] an increasing γ hence corresponds to inequality aversion (concave ν), while single-peaked γ implies a concave-convex weighting function. Note, there is no need for γ to be positive. Indeed, so long as γ is bounded, then we can consider $\nu' \geq 0$ without loss— addition of a linear function to ν leaves the solution to (2) unchanged.

If a government is only willing to consider inequality-reducing transfers, then Theorem 1 applies. If it is willing to engage in any budget-balanced transfers, then (as in section 5.1) Theorem 4 applies, with $F = B$. While the specifics of optimal redistribution depend on the form of γ, our previous results suggest that income support for those in the left tail can be an important feature of policy—even if governments are willing to consider inequality increasing transfers. Moreover, as corollary 1 highlights, the policymaker's demand for redistribution in (2) is invariant to affine transformations of the income distribution. This seems natural where arguments for redistribution are often based on relative rather than absolute grounds.

6 Persuasion with heterogeneous priors

In this section, we revisit the model of Onuchic & Ray (2020) and show how their results can be extended using the results of section 3.

We briefly recap the key elements of their model: there is a state of the world X, drawn from support $[\underline{x}, \overline{x}]$. A sender and receiver have different prior beliefs about the state. Let F_S denote the sender's prior and F_R the receiver's. Before the receiver is due to take an action $a \in \mathbb{R}$, the sender can provide her with information about X. Given any information observed, the receiver updates her beliefs and chooses an action equal to her posterior expectation of the state. The sender wishes to maximize the receiver's expected action, taken with respect to the sender's beliefs.

Onuchic & Ray (2020) show that, if attention is restricted to *monotone partitional* signals then the optimal such signal can be identified from the geometry of the function $F_S \circ F_R^{-1}$. We now argue that Theorem 1 also identifies the sender's optimal information structure, under the weaker assumption that the sender's signal be *monotone*. Without loss of generality, let us label signals by their induced 'receiver posterior' mean. That is, letting $\mathbb{E}_i$ denote expectations with respect to prior F_i, $i \in \{S, R\}$, we define our signals y so that they satisfy

[32] Recall the discussion following Theorem 1.

$y = \mathbb{E}_R[X \mid y]$. An experiment is a Markov kernel K, a conditional distribution of Y given X obeying $Y = \mathbb{E}_R[X \mid Y]$. Together, F_i and K induce subjective probability distributions $\Pr_i$, $i = S, R$, over (X, Y). Let G_i be the induced marginal distribution over Y, for $i = S, R$. Of course, since $Y = \mathbb{E}_R[Y \mid X]$, from the receiver's perspective Y is a mean preserving contraction of X: $G_R \preceq F_R$.

We define a signal to be *monotone* if it induces a conditional distribution over the receiver's action that is nondecreasing with the state. That is, a signal is monotone if

$$\varphi(x) := \int y dK(y \mid x)$$

is increasing in x.[33] Given φ is increasing, we can write the distribution over φ as $H_i(z) := \Pr_i[\varphi(X) \leq z] = F_i(\varphi^{-1}(z))$, $i = S, R$. This implies $H_S(z) = F_S(\varphi^{-1}(z)) = F_S \circ F_R^{-1}(H_R(z))$. Moreover, since φ is a conditional expectation of Y, $H_R \preceq_{co} G_R$ and therefore $H_R \preceq F_R$.

The sender's objective is to maximize

$$\mathbb{E}_S[Y] \quad = \quad \int \int y dK(y \mid x) dF_S(x) \quad = \quad \int \varphi(x) dF_S(x) \quad = \quad \int \varphi H_S(\varphi) \ .$$

Hence, since $H_R \preceq_{co} F_R$ is a necessary condition for any monotone experiment, the problem

$$\max_{H_R \preceq F_R} \quad \int \varphi dF_S \circ F_R^{-1}(H_R(\varphi))$$

provides an upper bound on the value of any monotone experiment. But this is a problem of the form (2), with $\nu = F_S \circ F_R^{-1}$. From Theorem 1, the solution to this problem $H_R^\star$ corresponds to a monotone partition, where the pooling regions are determined by the convexification of $F_S \circ F_R^{-1}$. Since a monotone partition is monotone, it is a feasible. Moreover, since it is deterministic, the implied distribution over Y is simply $G_R^\star = H_R^\star$. The above argument shows:

Proposition 6. *Any monotonic disclosure policy is dominated by a monotone partitional policy. The optimal monotone partition is identified by Theorem1 applied to $\nu = F_S \circ F_R^{-1}$.*

[33] Notice that monotonicity is strictly weaker than monotone partitionality. In particular, while a monotone partitional signal is always implementable with a deterministic map from the state to signals, monotone signals may be stochastic. One immediate implication of this is that while the conditional distributions $K(\cdot \mid x)$, $K(\cdot \mid x')$ for $x < x'$ are necessarily ordered by FOSD for monotone partitional signals, this is not required of a monotone signal.

7 Conclusions

We study a class of information design problems in which the sender has rank-dependent preferences over the receiver's posterior mean. We show that optimal persuasion can always be characterized by a convexification approach, which identifies a simple form of communication—monotone partitions—as optimal. Simply, the sender either reveals the state, or pools states within an interval into a single message. As the sender becomes more risk-loving, communication becomes more informative, as characterized by a shrinking of the intervals on which the state is pooled. Moreover, we find that "grading on a curve" is optimal, in the sense that optimal pooling intervals scale directly with the underlying state.

Our problem applies to several natural economic interactions. For instance, it applies to a broad class of information design problems involving multiple states and i.i.d. constraints, which incorporate the design of advertising standards in oligopoly markets and media influence in elections. It also applies to the study of insurance markets when customers' risk preferences fall into an empirically-relevant behavioral class, and income redistribution when a policymaker has a concern for inequality measured by the (weighted) area under the Lorenz curve. Finally, it provides a new implication for Bayesian persuasion with heterogeneous priors.

References

Allais, M. 1953. Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica*, **21**(4), 503–546.

Alonso, Ricardo, & Câmara, Odilon. 2016. Persuading voters. *The American Economic Review*, **106**(11), 3590–3605.

Anderson, Simon P., & McLaren, John. 2012. Media mergers and media bias with rational consumers. *Journal of the European Economic Association*, **10**(4), 831–859.

Anderson, Simon P., & Renault, Regis. 2006. Advertising Content. *American Economic Review*, **96**(1), 93–113.

Aumann, Robert J, & Maschler, Michael B. 1995. *Repeated games with incomplete information.* MIT press.

Bardhi, Arjada, & Guo, Yingni. 2018. Modes of persuasion toward unanimous consent. *Theoretical Economics*, **13**(3), 1111–1150.

Baron, David P., & Myerson, Roger B. 1982. Regulating a Monopolist with Unknown Costs. *Econometrica*, **50**(4), 911–930.

Bergemann, Dirk, Heumann, Tibor, Morris, Stephen Edward, Sorokin, Constantine, & Winter, Eyal. 2021. *Selling Impressions: Efficiency vs. Competition*.

Best, James, & Quigley, Daniel. 2021. Persuasion for the long run. *Working Paper*.

Billingsley, Patrick. 1995. *Probability and measure*. 3. ed edn. John Wiley & Sons.

Boleslavsky, Raphael, Hwang, Ilwoo, & Kim, Kyungmin. 2019. Competitive advertising and pricing. *Working Paper*.

Chakraborty, Archishman, & Harbaugh, Rick. 2010. Persuasion by cheap talk. *American Economic Review*, **100**(5), 2361–2382.

Chan, Jimmy, & Suen, Wing. 2008. A Spatial Theory of News Consumption and Electoral Competition. *The Review of Economic Studies*, **75**(3), 699–728.

Chan, Jimmy, & Suen, Wing. 2009. Media as watchdogs: The role of news media in electoral competition. *European Economic Review*, **53**(7), 799–814.

Crawford, VP, & Sobel, Joel. 1982. Strategic information transmission. *Econometrica*, **50**(6), 1431–1451.

Dorfman, Robert. 1979. A Formula for the Gini Coefficient. *The Review of Economics and Statistics*, **61**(1), 146–149.

Downs, Anthony. 1957. An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, **65**(2), 135–150.

Duggan, J., & Martinelli, C. 2011. A Spatial Theory of Media Slant and Voter Choice. *The Review of Economic Studies*, **78**(2), 640–666.

Dworczak, Piotr, & Martini, Giorgio. 2019. The Simple Economics of Optimal Persuasion. *Journal of Political Economy*, **127**(5), 1993–2048.

Gentzkow, Matthew, & Kamenica, Emir. 2016. A rothschild-stiglitz approach to Bayesian persuasion. *American Economic Review*, **106**(5), 597–601.

Gentzkow, Matthew, & Shapiro, Jesse M. 2006. Media Bias and Reputation. *Journal of Political Economy*, **114**(2), 280–316.

Gershkov, Alex, & Szentes, Balázs. 2009. Optimal voting schemes with costly information acquisition. *Journal of Economic Theory*, **144**(1), 36–68.

Griffith, R. M. 1949. Odds Adjustments by American Horse-Race Bettors. *The American Journal of Psychology*, **62**(2), 290–294.

Ivanov, Maxim. 2013. Information revelation in competitive markets. *Economic Theory*, **52**(1), 337–365.

Johnson, Justin P., & Myatt, David P. 2006. On the Simple Economics of Advertising, Marketing, and Product Design. *American Economic Review*, **96**(3), 756–784.

Kahneman, Daniel, & Tversky, Amos. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, **47**(2), 263–291.

Kamenica, Emir, & Gentzkow, Matthew. 2011. Bayesian persuasion. *American Economic Review*, **101**(6), 2590–2615.

Kleiner, Andreas, Moldovanu, Benny, & Strack, Philipp. 2021. Extreme Points and Majorization: Economic Applications. *Econometrica*, **89**(4), 1557–1593.

Kolotilin, Anton, & Li, Hongyi. 2021. Relational Communication. *Theoretical Economics*.

Kolotilin, Anton, Mylovanov, Tymofiy, Zapechelnyuk, Andriy, & Li, Ming. 2017. Persuasion of a Privately Informed Receiver. *Econometrica*, **85**(6), 1949–1964.

Mensch, Jeffrey. 2021. Monotone Persuasion. *Games and Economic Behavior*.

Milgrom, Paul R., & Weber, Robert J. 1982. A Theory of Auctions and Competitive Bidding. *Econometrica*, **50**(5), 1089–1122.

Newbery, David. 1970. A theorem on the measurement of inequality. *Journal of Economic Theory*, **2**(3), 264 – 266.

Onuchic, Paula, & Ray, Debraj. 2020. Conveying value via categories. *Working Paper*.

Perloff, Jeffrey M., & Salop, Steven C. 1985. Equilibrium with Product Differentiation. *The Review of Economic Studies*, **52**(1), 107–120.

Prelec, Drazen. 1998. The Probability Weighting Function. *Econometrica*, **66**(3), 497–527.

Quiggin, John. 1982. A theory of anticipated utility. *Journal of Economic Behavior & Organization*, **3**(4), 323–343.

Snowberg, Erik, & Wolfers, Justin. 2010. Explaining the Favorite - Long Shot Bias: Is It Risk-Love or Misperceptions?. *Journal of Political Economy*, **118**(4), 723–746.

Strömberg, David. 2004. Mass Media Competition, Political Competition, and Public Policy. *The Review of Economic Studies*, **71**(1), 265–284.

Tversky, Amos, & Fox, Craig R. 1995. Weighing risk and uncertainty. *Psychological review*, **102**(2), 269.

Tversky, Amos, & Kahneman, Daniel. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**(4), 297–323.

Wu, George, & Gonzalez, Richard. 1996. Curvature of the Probability Weighting Function. *Management Science*, **42**(12), 1676–1690.

Yaari, Menahem E. 1987. The Dual Theory of Choice under Risk. *Econometrica*, **55**(1), 95–115.

Appendix A: Proofs

Lemma 4. *For any fixed G with support on $[\underline{x}, \overline{x}]$,*

$$\int_{\underline{x}}^{\overline{x}} x d\nu(G(x)) \leq \int_{\underline{x}}^{\overline{x}} x d\omega(G(x)). \tag{13}$$

Proof. Let $\eta := \nu - \omega$. As $\nu(p) \geq \omega(p)$ for all $p \in [0,1]$, η is positive. Moreover, it is well-known (see, for instance, Baron & Myerson (1982)) that $\nu(0) = \omega(0)$ and $\nu(1) = \omega(1)$ and hence $\eta(0) = \eta(1) = 0$. As ν and ω have bounded variation, so does η. Hence, $\eta = \eta_1 - \eta_2 \geq 0$, where η_1, η_2 are non-decreasing functions with $\eta_1(0) = \eta_2(0)$ and $\eta_1(1) = \eta_2(1)$.

For any feasible distribution G, notice that $\eta_1 \leq \eta_2$ and $\eta_1(0) = \eta_2(0)$, $\eta_1(1) = \eta_2(1)$ implies

$$\int x d(\eta_1 \circ G)(x) \leq \int x d(\eta_2 \circ G)(x).$$

Hence, by linearity of the integral$\int x d(\eta \circ G)(x) \leq 0$ and by linearity again

$$\int x d(\nu \circ G)(x) \leq \int x d(\omega \circ G)(x).$$

□

Proof of Proposition 1

Let vex$\rho \circ \nu$ be the convexification of $\rho \circ \nu$, and let $I_i^\star(\rho \circ \nu)$, $i \in \mathcal{I}_{\rho\circ\nu}$, denote the associated disjoint intervals over which $\omega < \nu$. Since ω is convex, so too is $\rho \circ \omega$. Moreover as ρ is increasing, $\rho \circ \omega(p) \leq \rho \circ \nu(p)$ for all $p \in [0,1]$. Hence, by definition $\rho \circ \omega(p) \leq \text{vex}\rho \circ \nu(p)$ for all $p \in [0,1]$. In other words, $\{p : \text{vex}\rho \circ \nu(p) = \rho \circ \nu(p)\} \supseteq \{p : \rho \circ \omega(p) = \rho \circ \nu(p)\} = \{p : \omega(p) = \nu(p)\}$. Each $I_i^\star(\rho \circ \nu)$, $i \in \mathcal{I}_{\rho\circ\nu}$, must therefore be a subset of some $I_j^\star(\nu)$, $j \in \mathcal{I}_\nu$. Fix a continuous distribution function F. By Theorem 1, the optimal information structures corresponding to the respective weighting functions satisfy $G^\star(\rho \circ \nu) \preceq_{co} G^\star(\nu)$. □

Proof of Proposition 2

Consider the problem

$$\max_{p\in[0,1]} \ell'(p) := \frac{\nu(1) - \nu(p)}{1 - p}. \tag{14}$$

As ν is concave on $[\underline{p}, 1]$, a simple application of Jensen's inequality shows that any solution to (14) lies in the compact subset $[0, \overline{p}]$. Moreover, as ν is convex on $[0, \overline{p}]$ the objective is continuous. Hence, a solution p^c exists.

Define the line $\ell(p) = \nu(p^c) + \ell'(p^c)(p - q)$, and note that $\nu(1) = \ell(1)$. Moreover, $\nu(p) \geq \ell(p)$ for all $p \in [0,1]$. Indeed, if not, then for some $p' \in [0,1]$ we would have $\frac{\nu(1)-\nu(p')}{1-p'} > \frac{\ell(1)-\ell(p')}{1-p'} = \ell'(p^c)$—a contradiction to the optimality of p^c.

Now, define the function $\hat{\nu}(p) = \nu(p)$ for $0 \leq p \leq p^c$, and $\hat{\nu}(p) = \ell(p)$ for $p^c < p \leq 1$. As ν is concave on $[0, \overline{p}]$, and ℓ is linear, $\hat{\nu}$ is concave if and only if $\lim_{p\uparrow p^c} \frac{\nu(p^c)-\nu(p)}{p^c-p} \geq \ell'(p^c)$: the latter follows immediately from $\ell(p) \leq \nu(p)$ for all $p \leq q$ and $\ell(q) = \nu(q)$.

Hence, $\hat{\nu}$ is a convex function satisfying $\hat{\nu}(p) \leq \nu(p)$, for all p. We argue it is the maximal such function. Clearly this is the case for any $p \leq p^c$. Consider some $p' > p^c$ and any convex h which satisfies $h(p') > \hat{\nu}(p')$. We show that h cannot be pointwise dominated by ν. Suppose $h(p^c) \leq \nu(p^c)$ (otherwise, we are done). Then we would have $\hat{\nu}(p^c) = h(p^c) = \nu(p^c)$. But then

$$\begin{aligned} h(1) &\geq h(p^c) + (1 - p^c)\frac{h(p') - h(p^c)}{p' - p^c} \\ &> \ell(p^c) + \ell'(p^c)(1 - p^c) \\ &= \nu(1). \end{aligned}$$

A contradiction. Therefore, $\hat{\nu} = \omega$, the convexification of ν.

Finally, the first order condition for differentiable ν follows since ℓ is by definition a tangent line of ν at p^c. As ν is convex on $[0, \overline{p}]$ ℓ is a support line of ν on $[0, \overline{p}]$ and hence has a unique tangency point. By comparison with ℓ, it is easy to see that no other line passing through $(1,\nu(1))$ can be a support of the convex function at p^c. Hence, there can be no alternative solution to $\nu'(p)=\frac{\nu(1)-\nu(p)}{1-p}$. □

Proof of Theorem 2

We prove the result assuming ν is increasing; the extension to BV ν follows immediately, noting it can be expressed as the difference of increasing functions. Throughout, fix some continuous F. We first establish the existence of a sequence $\tilde{\beta}_n$ which converges uniformly to ν:

Lemma 5. *Fix a continuous function $\nu:[0,1]\to\mathbb{R}$, and normalize $\nu(0)=0$. Let $\alpha_k^n=\nu(1-\frac{k}{n})-\nu(1-\frac{k+1}{n})$, and $\tilde{\beta}_n=\sum\alpha_k^n\beta_{n-k+1,k}$. Then $\tilde{\beta}_n\to\nu$ uniformly on $[0,1]$.*

Proof. As is well-known, the beta distribution satisfies the following

$$\beta_{n-k+1,k}(p)=\sum_{r=n-k}^{n}\begin{pmatrix} n \\ r \end{pmatrix}p^r(1-p)^{n-r}=b_{n-k,n}(p)+\beta_{n-k,k+1}(p),$$

where $\beta_{0,n+1}:=0$, and each $b_{n-k,n}(p)=\begin{pmatrix} n \\ n-k \end{pmatrix}p^{n-k}(1-p)^k$, $k=0,\ldots,n$ is a Bernstein basis polynomial. Using the recursive formulation, $\tilde{\beta}_n$ can be expressed $\tilde{\beta}_n=\sum_{k=0}^{n}\nu(1-\frac{k}{n})b_{n-k,n}$. By Theorem 6.2 Billingsley (1995), $\tilde{\beta}_n\to\nu$ uniformly. □

Let $G_n^\star\in\arg\max_{G\preceq_{co}F}\int xd\tilde{\beta}_n(G)$ be the solution identified by Theorem 1 corresponding to weighting function $\tilde{\beta}_n$, and let $u^\star=\limsup\int xd\tilde{\beta}_n(G_n^\star)$. Trivially, there exists a subsequence $\{G_{n_k}^\star\}$ for which $\int xd\tilde{\beta}_{n_k}(G_{n_k}^\star)\to u^\star$. As $[\underline{x},\overline{x}]$ is bounded, Helly's theorem implies there exists a further subsequence $\{G_j^\star\}$ and a distribution function G_∞ such that $G_j^\star\Rightarrow G_\infty$, where $\Rightarrow$ denotes weak convergence.[34] Moreover, $G_\infty\preceq_{co}F$ since, by $G_j^\star\preceq_{co}F$ and the Bounded Convergence Theorem

$$\int_s^{\overline{x}}\big(1-G_\infty(x)\big)dx=\lim_{j\to\infty}\int_s^{\overline{x}}\big(1-G_j^\star(x)\big)dx\leq\int_s^{\overline{x}}\big(1-F(x)\big)dx$$

for each $s\in[\underline{x},\overline{x}]$.

Lemma 6. *Along subsequence $\{G_j^\star\}$,* $\lim_{j\to\infty}\int xd\tilde{\beta}_j(G_j^\star)=\int xd\nu(G_\infty)$.

[34]To ease notation, we write $G_j^\star$ in place of $G_{n_{k_j}}^\star$ along the latter subsequence.

Proof. For each n, k, $\beta_{n-k+1,k}(p)$ corresponds to the complementary c.d.f. of a binomial distribution, and hence is trivially increasing in p. As ν is increasing, this implies each $\tilde{\beta}_n$ is also increasing. Hence, $\tilde{\beta}_n \circ G^\star_n$ represents a measure. We argue that $\tilde{\beta}_n \circ G^\star_n \Rightarrow \nu \circ G_\infty$, from which the lemma follows immediately.

We must show that, for any continuity point x of $\nu \circ G_\infty$, $\tilde{\beta}_n \circ G^\star_n(x) \to \nu \circ G_\infty(x)$. There are two ways such a point can arise: (i) x is a continuity point of G_∞, or (ii) $G_\infty(x_-) := \lim_{y \uparrow x} G_\infty(y) < G_\infty(x)$ with ν constant on $[G_\infty(x_-), G_\infty(x)]$. The former case is simpler, and so we prove the latter. To do this, bound $|\, \tilde{\beta}_n \circ G^\star_n(x) - \nu \circ G_\infty(x) \,|$ as follows:

$$\begin{aligned} |\, \tilde{\beta}_n \circ G^\star_n(x) - \nu \circ G_\infty(x) \,| &\leq |\, \tilde{\beta}_n \circ G^\star_n(x) - \tilde{\beta}_n \circ G_\infty(x) \,| + |\, \tilde{\beta}_n \circ G_\infty(x) - \nu \circ G_\infty(x) \,| \\ &\leq |\, \nu \circ G^\star_n(x) - \nu \circ G_\infty(x) \,| + |\, \tilde{\beta}_n \circ G^\star_n(x) - \nu \circ G^\star_n(x) \,| \\ &\quad +2 \,|\, \tilde{\beta}_n \circ G_\infty(x) - \nu \circ G_\infty(x) \,| \,. \end{aligned}$$

Fix some $\varepsilon > 0$. By Lemma 5, there exists an m such that for $n \geq m$, the final two terms are bounded above by $\frac{3}{4}\varepsilon$. Hence, we will be done if we can find m' such that $|\, \nu \circ G^\star_n(x) - \nu \circ G_\infty(x) \,| \leq \frac{\varepsilon}{4}$ for all $n \geq m'$. Since $G^\star_n$ is increasing and the set of continuity points of G_∞ is dense, it is easy to see that $\lim_{y \uparrow x} G^\star(y) \leq \liminf G^\star_n(x) \leq \limsup G^\star_n(x) \leq G^\star(x)$. By the continuity of ν at $p \in \{G_\infty(x_-), G_\infty(x)\}$ and its constancy on $[G_\infty(x_-), G_\infty(x)]$, $\nu(G^\star_n(x)) - \nu(G_\infty(x)) \to 0$, guaranteeing existence of the required m'. □

We can now prove $\int x d\tilde{\beta}_n(G^\star_n) \to \int x d\nu(G^\star)$, where $G^\star$ is characterized by Theorem 1. Since $G^\star \preceq_{co} F$, we have $\int x d\tilde{\beta}_n(G^\star_n) \geq \int x d\tilde{\beta}_n(G^\star)$ for each n. Moreover, for fixed $G^\star$ it is easy to see that uniform convergence $\tilde{\beta}_n \to \nu$ implies weak convergence $\tilde{\beta}_n \circ G^\star \Rightarrow \nu \circ G^\star$, so that $\int x d\tilde{\beta}_n(G^\star) \to \int x d\nu(G^\star)$. Hence, $\liminf \int x d\tilde{\beta}_n(G^\star_n) \geq \int x d\nu(G^\star)$. Moreover, since each $G_\infty \preceq_{co} F$, we have $\limsup \int x d\tilde{\beta}_n(G^\star_n) = \int x d\nu(G_\infty) \leq \int x d\nu(G^\star)$—which proves the claim.

Finally, we argue that if $G^\star$ is unique then $G_n \Rightarrow G^\star$. For sake of a contradiction, suppose $G^\star$ is the unique solution to (2), but $G_n \not\Rightarrow G^\star$. Applying Helly's theorem again, there must some distribution function $G' \neq G^\star$ and a subsequence $G_l \Rightarrow G'$. But, this implies $\lim \int x d\tilde{\beta}(G_l) = \int x d\nu(G') < \int x d\nu(G^\star)$—a contradiction to $\lim \int x d\tilde{\beta}(G_n) = \int x d\nu(G^\star)$. □

Proof of Lemma 1

Let $p_s : [\underline{x}, \overline{x}] \to \mathbb{R}$ be the pricing strategy of seller s, a measurable function of x_s. Let $v_s := x_s - p_s$ be the consumer surplus delivered by s, and define the best rival offer $v^m_{-s} = \max_{t \in \{1,\ldots,N\} \setminus \{s\}} v_t$. v^m_{-s} has some distribution function induced by $H_{-s} = \{H_t\}_{t \neq s}$; call it

Q^m_{-s}. Then–given private observation of x_s—s's optimal pricing problem is

$$\max_{p_s\in\mathbb{R}} p_s Q^m_{-s}(x_s - p_s) = \max_{v_s\in\mathbb{R}}(x_s - v_s)Q^m_{-s}(v_s). \tag{15}$$

Anticipating this, at the advertising stage seller s solves

$$\max_{G\preceq_{co}F} \int \psi(x_s)dG(x_s),$$

where $\psi(x_s) := \max_{v_s\in\mathbb{R}}(x_s - v_s)Q^m_{-s}(v_s)$ is the upper envelope of a family of functions which are all affine in x_s, and hence convex. Hence, $G = F$ is a solution to (15) for all s; maximal disclosure is an equilibrium. □

Proof of Proposition 3

$\mathcal{E}$ is trivially bounded by the triangle $\{CS + \Pi \leq E_F[x], CS, \Pi \geq 0\}$. Suppose $\mathcal{E}$ is not closed. Then there exists a sequence (v_n, π_n) and some (v', π') such that $(v_n, \pi_n) \in \mathcal{E}$, $\forall n$, $(v_n, \pi_n) \to (v', \pi')$ and $(v', \pi') \notin \mathcal{E}$. By the separating hyperplane theorem, there exist constants α_B, α_S such that $\alpha_B v' + \alpha_S \pi' > \alpha_B v + \alpha_S \pi$ for all $(v, \pi) \in \mathcal{E}$. By continuity, $\alpha_B v_n + \alpha_S \pi_n \to \alpha_B v' + \alpha_S \pi'$. Hence, $\alpha_B v' + \alpha_S \pi'$ is a supremum for $\alpha_B v + \alpha_S \pi$, which is unattainable on $\mathcal{E}$. Hence, (8) has no solution—a contradiction to Theorem 1.

To show $\mathcal{E}$ is convex, suppose $(v_1, \pi_1), (v_2, \pi_2) \in \mathcal{E}$. Then there exist $G_1, G_2 \preceq_{co} F$ such that $v_i = \int x d\beta_{(N-1,2)}(G_i)$, $\pi_i = \int x d\Delta\beta_N(G_i)$, $i = 1, 2$, where $\Delta\beta_N = \beta_{(N,1)} - \beta_{(N-1,2)}$. Consider the distribution $G_\alpha = \alpha G_1(1-\alpha)G_2$, defined by (9). By a change of variables, it is easy to see that $v_\alpha = \int x d\beta_{(N-1,2)}(G_\alpha) = \alpha v_1 + (1-\alpha)v_2$ and similarly $\pi_\alpha = \alpha\pi_1 + (1-\alpha)\pi_2$. Convexity of $\mathcal{E}$ will therefore follow if we show $G_\alpha \preceq_{co} F$. We appeal to conjugate duality: integrating by parts and changing variables shows $\int_q^1 G^{-1}(p)dp = \max_s\{s(1-q) + \int_s^{\overline{x}}(1 - G(x))dx\}$ for any distribution G and $q \in (0, 1)$. But as $G_i \preceq_{co} F$, $\int_s^{\overline{x}}(1 - G_i(x))dx \geq \int_s^{\overline{x}}(1 - F(x))dx$, $\forall s \in [\underline{x}, \overline{x}]$. Hence, $\int_q^1 G_i^{-1}(p)dp \geq \int_q^1 F^{-1}(p)dp$, $i = 1, 2$, and by linearity $\int_q^1 G_\alpha^{-1}(p)dp = \alpha\int_q^1 G_1^{-1}(p)dp + (1-\alpha)\int_q^1 G_2^{-1}(p)dp \geq \int_q^1 F^{-1}(p)dp$, $\forall q \in [0, 1]$. Applying similar logic in the other direction gives $\int_s^{\overline{x}}(1 - G_\alpha(x))dx \geq \int_s^{\overline{x}}(1 - F(x))dx$, $\forall s \in [\underline{x}, \overline{x}]$. Hence, $G_\alpha \preceq_{co} F$.

The extreme points of $\mathcal{E}$ solve (8). We prove these points are characterized by extreme censorship, by showing that for any α_B, α_S, $\tilde{\beta}_N = \alpha_B\beta_{(N-1,2)} + \alpha_S\Delta\beta_N$ is either convex-concave or concave-convex. We show the result assuming $\alpha_B > 0$ (the extension to $\alpha_B \leq 0$ is trivial). By linearity it is without loss to replace $\tilde{\beta}_N$ by $\nu_r = \beta_{(N-1,2)} + r\Delta\beta_N$ in the sender's objective. Taking derivatives twice yields, after some algebra:

$$\begin{aligned}\nu_r'(p) &= Np^{N-2}\big((1-r)(N-1)(1-p)+rp\big)\\ \nu_r''(p) &= N(N-1)p^{N-3}\big((1-r)(N-2)+(1+N(r-1))p\big).\end{aligned}$$

Inspection of ν_r'' shows that v_r is *(i)* convex for $r \in [\frac{1}{2},1]$, *(ii)* convex-concave for $r < \frac{1}{2}$, and *(iii)* concave-convex for $r > 1$. Hence, optimal information structures in each case correspond to full information, right-censorship and left-censorship, respectively. Comparative statics in r for $r < \frac{1}{2}$ ($r > 1$) follow from the implicit function theorem applied to (5) and to its counterpart for concave-convex functions. □

Proof of Corollary 3

CS maximization corresponds to problem (8) with $\tilde{\beta}_N = \beta_{(N-1,2)}$, where $\beta_{(N-1,2)}(p) = N(N-1)\int_0^p q^{N-2}(1-q)dq$. The proof of Proposition 3 (for $\alpha_B = 1$, $\alpha_S = 0$) shows right censorship is optimal. By Proposition 2, the optimal rank threshold $p = p^c(N)$ satisfies, on rearrangement:

$$p^{N-2}(1-p) = \frac{\int_p^1 q^{N-2}(1-q)dq}{1-p}.$$

We show $p^c(N)$ is increasing in N. If N increases by 1, the left hand side multiplies by $p<1$, while the right side becomes

$$\frac{\int_p^1 q^{(N+1)-2}(1-q)dq}{1-p} > \frac{\int_p^1 pq^{N-2}(1-q)dq}{1-p} = p\frac{\int_p^1 q^{N-2}(1-q)dq}{1-p},$$

where the inequality follows because the range of integration is over $p \le q \le 1$. Hence, for $m = N+1$, $p = p^c(N)$ satisfies the inequality

$$p^{m-2}(1-p) < \frac{\int_p^1 q^{m-2}(1-q)dq}{1-p}. \tag{16}$$

We establish that equality is restored for some $p^c(N+2) \in (p^c(N),1)$. Notice that, at $p' = \frac{m-1}{m-2}$, the inequality (16) reverses – this follows from the observation that p' maximizes the value of the integrand. By continuity, there exists $p^c(m) \in (p^c(N), \frac{m-1}{m-2})$ for which (16) becomes an equality. By Proposition 2, this $p^c(m)$ is the unique optimal rank threshold. Hence, information obviously increases in the sense of Blackwell.

As $p^c(N)$ is increasing, it has a limit $p^c(\infty)$. Rearranging the first order condition gives

$$1 - p^c(N) = \int\limits_{p^c(N)}^{1} \left(\frac{q}{p^c(N)}\right)^{n-2} \left(\frac{1-q}{1-p^c(N)}\right) dq.$$

By an almost identical argument to that in the proof of Proposition 5, $p^c(\infty) = 1$. □

Proof of Corollary 4

Consumer surplus maximization corresponds to problem 8 with $\tilde{\beta}_N = \beta_{(N-1,2)}$. From the proof of Proposition 3 ($\alpha_B = 0$, $\alpha_S = 1$), left censorship is optimal. In this case, the first order condition for the optimal rank threshold reduces to

$$\Delta\beta'_N(p^c) = \frac{\Delta\beta_N(p^c)}{p^c},$$

where $\Delta\beta_N = \beta_{N,1} - \beta_{N-1,2}$. Directly solving the first order condition, we find $p^c = \frac{N-2}{N-1}$; hence, the rest of the corollary follows. □

Proof of Proposition 4

We prove the result for a right-biased lobbyist. Similar arguments apply for left-bias, applying instead the weighting function $-\nu$. Given any signal realization $X_i = \mathbb{E}[x_i \mid X_i]$, a voter's expected utility from policy a is

$$\mathbb{E}[u_i \mid X_i] = -a^2 + 2aX_i - X_i^2,$$

which is quadratic, and hence single-peaked, in a. Given an odd number of voters, there is a unique equilibrium of the continuation game between politicians in which both politicians set $a_1 = a_2 = X^{(N+1)}$, the median voter's ideal policy. Hence, a right-biased lobbyist wishes to maximize

$$\mathbb{E}[X^{(N+1)}] = \int xd(\beta_{(N,N)} \circ G)(x)$$

subject to $G \preceq_{co} F$, where $\beta_{(N,N)} : [0,1] \to \mathbb{R}_+$ satisfies $\beta_{(N,N)}(p) = \frac{(2N+1)!}{N!N!} \int_0^p q^N(1-q)^N dq$. Taking derivatives shows $\beta'_{(N,N)}(p) = \frac{(2N+1)!}{N!N!} p^N(1-p)^N$ and $\beta''_{(N,N)}(p) = N\beta'_{(N,N)}(p)\left(\frac{1-2p}{p(1-p)}\right)$. $\beta_{(N,N)}$ is everywhere increasing, convex for $0 \leq p \leq \frac{1}{2}$ and concave for $\frac{1}{2} < p \leq 1$. Moreover, $\beta'_{(N,N)}(0) = 0$. By Proposition 2 and Theorem 1, informative right-censorship is optimal with censorship above rank $p^c \in (0, \frac{1}{2})$ of F. □

Proof of Proposition 5

We prove the result for a right-biased lobbyist. Let $p^c(N)$ be the optimal rank threshold as a function of N. As $\beta_{(N,N)}$ is differentiable and $p^c(N)$ interior, Proposition 2 implies $p^n(1-p)^n = \int_p^1 \frac{q^n(1-q)^n}{1-p}dq$ holds at $p = p^c(N)$. Now consider $m > N$. We have $p^m(1-p)^m = \left(\int_p^1 \frac{q^n(1-q)^n}{1-p}dq\right)^{\frac{m}{n}} < \int_p^1 \frac{q^m(1-q)^m}{1-p}dq$, where the inequality is Jensen's, applied to the strictly convex function $f(x) = x^{\frac{m}{N}}$. By contrast, at $\tilde{p} = \frac{1}{2}$, we have $\tilde{p}^m(1-\tilde{p})^m > \int_{\tilde{p}}^1 \frac{q^m(1-q)^m}{1-\tilde{p}}dq$— which follows from the observation that $q(1-q)$ is maximized at $q = \tilde{p}$. Each side of the inequality is continuous, and hence there is a $p^c(m) \in (p^c(N), \frac{1}{2})$ at which they intersect; by Proposition 2 this is unique. Hence, $p^c(N)$ is increasing in N; which trivially corresponds to a Blackwell increase in information.

Finally, as $p^c(N)$ is increasing and bounded by $\frac{1}{2}$, it has a limit $p^c(\infty) \leq \frac{1}{2}$. Rewriting the first order condition for $p^c(N)$:

$$1 - p^c(N) = \int_{p^c(N)}^{1} \left(\frac{q(1-q)}{p^c(N)(1-p^c(N))}\right)^N dq.$$

Suppose $p^c(\infty) < \frac{1}{2}$. Then there exists an interval $[p', \frac{1}{2}]$, such that $p^c(\infty) \leq p' < \frac{1}{2}$. For every N, the right hand side above is therefore at least $\int_{p'}^{\frac{1}{2}} \left(\frac{q(1-q)}{p^c(N)(1-p^c(N))}\right)^N dq$. But since $q(1-q)$ is single peaked at $\frac{1}{2}$, the integrand increases without bound on $[p', \frac{1}{2}]$ as $N \to \infty$. As the left-hand side of this expression is bounded, the first order condition cannot be satisfied for N sufficiently large – a contradiction. Hence, $p^c(\infty) = \frac{1}{2}$. □

Proof of Lemma 3

(If) Divide $[0,1]$ into $X = [0, p_m]$ and $X^c = [p_m, 1]$. Since ν is concave on X, then property (12) holds by definition on X for any intervals $B = [p, p']$, $C = [q, q']$ such that $p' \leq q \leq p_m$. Moreover, the expression holds with strict inequality for some B, C if ν is not linear. The interval subsets form a π-system generating the Borel σ-algebra on X. If we can show the class of sets satisfying (12) form a λ-system (closed under complements and disjoint unions) then Dynkin's theorem implies it extends to the Borel σ-algebra on X.

To see that (12) is closed under disjoint unions, suppose it holds for pair (B_i, C_j) from $B_1, B_{2,\ldots}$ and $C_1, C_2, \ldots$, where all sets are disjoint and $\sup B_i \leq \inf C_j$ for all i, j. Rewriting (12)

$$\nu(B_i)\lambda(C_j) \geq \nu(C_j)\lambda(B_i),$$

holds for all i, j. Summing the inequality, first over i and then j, and applying disjointness shows that

$$\nu(\cup_i B)\lambda(\cup_j C) \geq \nu(\cup_j C)\lambda(\cup_i B).$$

Since $\nu(p_m) = p_m$, similar steps easily establish that (12) is closed under relative complements on X. Hence, the class of sets satisfying (12) is a λ-system. An almost identical argument establishes the result on $X^c = [\overline{p}, 1]$ for ν convex.

(Only if) This case is trivial, and hence omitted. □

Appendix B: Optimal information design with discontinuities

In Theorem 1, we assumed that the distribution of X, F, was continuous. However our methods extend naturally to settings in which F admits discontinuities, such as when the state space has a finite support or when atoms are present. Indeed, we note that bound (4) did not rely on the assumption that F was continuous. Hence, to extend Theorem 1 we need only identify a feasible distribution G' which still attains the bound in this case. Intuitively, G' will deal with potential atoms at the endpoints of a pooling interval by allowing for mixed revelation of these endpoints. Of course, G' will no longer be partitional in such cases, but—as we argue below—is nonetheless still monotone.

Let $F(x-) = \lim_{y \uparrow x} F(y)$ and define $D := \{x : F(x-) < F(x)\}$, the set of discontinuity points of F. As is well-known, D is countable. Additionally, we write $F(\{x\}) := [F(x-), F(x)]$ for the interquantile range spanned by the event $\{X = x\}$ and $\Delta F(x) = F(x) - F(x-)$. Let x_k (x'_k) be the (unique) value in $[\underline{x}, \overline{x}]$ for which $p_k \in F(\{x\})$ ($p'_k \in F(\{x'_k\})$), define $J^\star_k = (x_k, x'_k)$ and let $\overline{J}^\star_k = [p_k, p'_k]$ be the closure of $J^\star_k$.[35]

Writing λ for the Lebesgue measure on $[0,1]$, let

$$r_k(x) = \begin{cases} \frac{\lambda(F(\{x\}) \cap I^\star_k)}{\lambda(F(\{x\}))} & , \text{ if } x \in D, \\ 1 & , \text{ if } x \notin D, \ F(x) \in I^\star_k, \\ 0 & , \text{ otherwise.} \end{cases}$$

be the proportion of $F(\{x\})$ which intersects $I^\star_k$. Notice that (i) $r_k(x)$ can be interior to $[0,1]$ only at $x \in \{x_k, x'_k\}$, and (ii) by monotonicity of F, $r_k(x) = 0$ for $x \notin [x_k, x'_k]$ and $r_k(x) = 1$ for $x \in (x_k, x'_k)$.

[35]Note how this extends the definitions of x_k, x'_k in section 1.

$$\tilde{\mu}_k = \frac{r_k(x_k)\Delta F(x_k)x_k + r_k(x'_k)\Delta F(x'_k)x'_k + \int_{J^\star_k} x dF}{\lambda(I^\star_k)}.$$

Note that $r_k(x_k)\Delta F(x_k) + r_k(x'_k)\Delta F(x'_k)x'_k + \int_{J^\star_k} dF = \lambda(I^\star_k)$, by definition of r_k.

Suppose that with probability $r_k(x)$ the sender sends a signal $s = \tilde{\mu}_k$. Then, $\tilde{\mu}_k$ clearly represents the conditional mean of X given s is observed. Building on this, we construct an information structure π' as follows. For $x \notin D$:

$$\pi'(X) = \begin{cases} \mu'_k, & \text{with probability } r_k(X),\ \text{k} \in K, \\ X, & \text{with probability 1-}\sum_{k \in K} r_k(X). \end{cases}$$

Fixing some $X = x$, the support of $\pi'(x)$ contains at most three signals. This follows from the definition of r_k and the fact that x can equal x_k (x'_j) for at most one $k \in K$ (respectively, $j \in K$). Moreover, notice that by the monotonicity of F and the definition of r_k, π' is *monotone* in the following sense: if $x > y$, then $\min\ \text{supp}\ \pi'(x) \geq \max\ \text{supp}\ \pi'(y)$. This is a strong form of monotonicity—for example, it obviously implies $\pi'(x)$ first-order stochastically dominates $\pi'(y)$.

Of course, π' induces a distribution G' over posterior means. Hence, $G' \preceq_{co} F$ holds trivially. It is easy to verify that, by construction, G' satisfies

$$G'(x) = \begin{cases} F(x), & x \notin \bigcup_{k \in K} \overline{J}^\star_k, \text{ or } x = x_k, k \in K, \text{ and } x \neq x'_j, \forall j \in K \\ p_k, & x_k \leq x < \tilde{\mu}_k, \\ p'_k, & \tilde{\mu}_k \leq x < x'_k, \end{cases}$$

Notice that the specification of G' is rich enough to accommodate cases where two adjacent intervals $\overline{J}^\star_k$ and $\overline{J}^\star_j$ satisfy $x_k = \tilde{x}_j$. In this case, $G'(x) = p_k$ at $x = x_k$. Otherwise, if $x_k \neq x'_j$ for all $j \in K$ (i.e., $\overline{J}^\star_k$ is not adjacent to another such interval on its left), then $G'(x_k) = F(x_k)$. With these preliminaries, we can now extend Theorem 1.

Theorem 4. *G' solves (2).*

Proof. Exactly as before, the same upper bound (4) applies. Hence, we need only argue that G' satisfies (expressing both sides in terms of p):[36]

$$\int G'^{-1}(p) d\nu(p) = \int F^{-1}(p) d\omega(p).$$

[36] Working in the rank space is slightly easier, allowing us to split the range of integration up into intervals—avoiding the need to describe an equivalent procedure for a mixed splitting in $[\underline{x}, \overline{x}]$.

With $\tilde{\mu}_i$ in place of μ_i, almost identical arguments to the proof of Theorem 1 go through. In particular,

$$\int_{(\bigcup_k J_k^\star)^c} G'^{-1}(p)d\nu(p) = \int_{(\bigcup_k J_k^\star)^c} F^{-1}(p)d\omega(p)$$

again holds trivially (since $G^{-1} = F^{-1}$ and $\omega = \nu$ on $(\bigcup_k J_k^\star)^c$. Moreover, on any $J_k^\star$ essentially the same steps show that

$$\int_{J_k^\star} G'^{-1}(p)d\nu(p) = \tilde{\mu}_i(\nu(p_i') - \nu(p_i)) = \tilde{\mu}_i(\omega(p_i') - \omega(p_i)) = \int_{J_k^\star} F^{-1}(p)d\omega(p).$$

□