The Economics of Monotone Function Intervals*

Kai Hao Yang† Alexander K. Zentefis‡

November 9, 2023

Abstract

Monotone function intervals are sets of monotone functions that are bounded pointwise above and below by two monotone functions. We characterize the extreme points of such intervals and apply this result to various economic subjects. Using the extreme points, we characterize the distributions of posterior quantiles, leading to an analog of a classical result on the distributions of posterior means. We apply this analog to political economy, Bayesian persuasion, and the psychology of judgment. Monotone function intervals provide a common structure to security design, and we use their extreme points to unify and generalize seminal results in that literature when either adverse selection or moral hazard pertains.

JEL classification: D72, D82, D83, D86, G23

Keywords: Extreme points, monotone functions, pointwise dominance, posterior quantiles, gerrymandering, Bayesian persuasion, misconfidence, security design

---

*We thank the coeditors and three anonymous reviewers for their generous comments. We are also grateful to Ian Ball, Dirk Bergemann, Alex Bloedel, Simon Board, Ben Brooks, Peter Caradonna, Roberto Corrao, Xiaohong Chen, Joyee Deb, Piotr Dworczak, Mira Frick, Tan Gan, Marina Halac, Ryota Iijima, Emir Kamenica, Nicolas Lambert, Elliot Lipnowski, Alessandro Lizzeri, Jay Lu, Leslie Marx, Moritz Meyer-ter-Vehn, Stephen Morris, Barry Nalebuff, Pietro Ortoleva, Aniko Öry, Wolfgang Pesendorfer, Benjamin Polak, Doron Ravid, Anne-Katrin Roesler, Hamid Sabourian, Fedor Sandomirskiy, Christoph Schlom, Ludvig Sinander, Vasiliki Skreta, Nicholas Stephanopoulos, Philipp Strack, Roland Strausz, Tomasz Strzalecki, Omer Tamuz, Can Urgun, Quitzé Valenzuela-Stookey, Martin Vaeth, Dong Wei, Mark Whitmeyer, Alexander Wolitzky, Wenji Xu, Nathan Yoder, Jidong Zhou, Pavel Zryumov, and the participants of various seminars and conferences for their helpful comments. We also thank Jialun (Neil) He and Nick Wu for their research assistance. All errors are our own.

†Yale School of Management, Email: kaihao.yang@yale.edu

‡Yale School of Management, Email: alexander.zentefis@yale.edu

1 Introduction

Monotone functions play a crucial role in many economic settings. In standard equilibrium analyses, demand curves and supply curves are monotone. In moral hazard problems, many contracts are monotone. In information economics, distributions of a one-dimensional unknown state can be summarized by monotone cumulative distribution functions (CDFs). Among all orderings, the pointwise dominance order is one of the most natural ways to compare monotone functions: Outward/inward shifts of supply and demand, limited liability in contract theory, and the first-order stochastic dominance order of CDFs are all expressed in terms of pointwise dominance of monotone functions.

In this paper, we provide a systematic way to study an arbitrary set of monotone functions that are bounded pointwise from above and below by two monotone functions. Without loss, we focus on sets of nondecreasing and right-continuous functions bounded by two nondecreasing functions, such as the blue and red curves in Figure I. We refer to these sets as *monotone function intervals* and show that many economic problems are connected to such intervals. Our main result (Theorem 1) characterizes the *extreme points* of monotone function intervals. We show that a nondecreasing, right-continuous function is an extreme point of a monotone function interval if and only if the function either coincides with one of the two bounds or is constant on an interval in its domain. Wherever the function is constant on an interval, it must coincide with one of the two bounds at one of the endpoints of the interval, as illustrated by the black curve in Figure I.

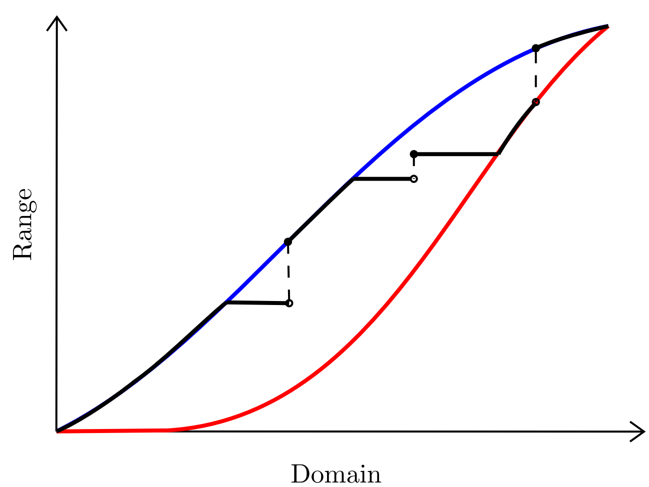


Figure I
An Extreme Point of a Monotone Function Interval

Since monotone function intervals are convex, characterizing their extreme points is useful by virtue of two known properties of extreme points. The first property, formally known as

Choquet's theorem, is that any element of a compact and convex set can be represented as a mixture of the extreme points. This allows one to focus only on extreme points when trying to establish properties that are preserved under mixtures. The second property is that for any convex optimization problem that admits a solution, one of the solutions must be an extreme point of the feasible set. We demonstrate how the extreme point characterization can be applied to economics through two classes of applications, each one exploiting one of the two aforementioned properties of extreme points.

In the first class of applications, we use Theorem 1 and Choqet's theorem to characterize the distributions of posterior quantiles. Consider a random variable and a signal for it. Each signal realization induces a posterior belief. For every posterior belief, one can compute the posterior mean. Strassen's theorem (Strassen 1965) implies that the distribution of these posterior means is a mean-preserving contraction of the prior. Conversely, every mean-preserving contraction of the prior is the distribution of posterior means under some signal. Instead of posterior means, one can derive many other statistics of a posterior. The characterization of the extreme points of monotone function intervals leads to an analog of Strassen's theorem, which characterizes the distributions of posterior *quantiles* (Theorem 2 and Theorem 3). The set of feasible distributions of posterior quantiles coincides with an interval of CDFs bounded by a natural upper and lower truncation of the prior.

We apply Theorem 2 and Theorem 3 to three settings: gerrymandering, quantile-based persuasion, and apparent over/underconfidence (misconfidence). These settings all share concerns over ordinal rather than cardinal outcomes. First, gerrymandering is connected to the distributions of posterior quantiles since voters' political ideologies are only ordinal. When the distribution of voters' political ideologies in an election district is interpreted as a posterior, the median voter theorem implies that the ideological position of the elected representative in that district is a posterior median. Since an electoral map corresponds to a distribution of posteriors under this interpretation, Theorem 2 and Theorem 3 characterize the compositions of the legislative body that a gerrymandered map can create. Second, in Bayesian persuasion, Theorem 2 and Theorem 3 bring tractability to persuasion problems where the sender's indirect payoff is a function of posterior *quantiles*: an ordinal analog of the widely studied environment where the sender's indirect payoff is a function of posterior *means*. Ordinal outcomes matter if the receiver is not an expected utility maximizer, but a *quantile maximizer* (Manski 1988; Rostek 2010), or if the sender's payoff is state independent and the receiver chooses an action to minimize the expected *absolute*—as opposed to *quadratic*—distance to the state. Third, the literature on the psychology of judgment documents that individuals appear to be over or under confident when evaluating themselves relative to a population. Theorem 3 implies the seminal result of Benoît and Dubra (2011), who provide a necessary and sufficient condition for *apparent* overconfidence (e.g., more than 50% of individuals ranking themselves above the population median) to imply *true* overconfidence (i.e., individuals are not Bayesian).

In the second class of applications, we use Theorem 1, together with the optimality of extreme points in convex problems, to study security design with limited liability. Consider the canonical security design problem where the security issuer designs a security that specifies payments contingent on the realized return of an asset. Two assumptions are commonly adopted in the security design literature. The first assumption is that any security must be nondecreasing in the asset's return.[1] The second is limited liability, which places natural upper and lower bounds on the security's payoff given each realized return. Under these two assumptions, the set of securities coincides with a monotone function interval bounded by the identity function and the constant function 0.

Two seminal papers adopt these assumptions in their analysis of the security design problem. Innes (1990) studies the problem under moral hazard, whereas DeMarzo and Duffie (1999) study it under adverse selection. Both papers show that a standard debt contract is optimal, which promises either a constant payment or the asset's realized return, whichever is smaller. Many papers in security design that followed built upon the Innes (1990) or DeMarzo and Duffie (1999) environment. (See, for example, Schmidt 1997; Casamatta 2003 and Eisfeldt 2004; Biais and Mariotti 2005.)

The optimality of standard debt in Innes (1990) and DeMarzo and Duffie (1999) relies on a crucial assumption: The distribution of the asset return satisfies the monotone likelihood ratio property (MLRP). Therefore, the structure of optimal securities without MLRP remains relatively under-explored. Nonetheless, since security design in these settings is a convex optimization problem, and since the set of securities is a monotone function interval, there must be an extreme point of the feasible set that is optimal. Using the characterization in Theorem 1, we show that these extreme points correspond to *contingent* debt contracts, where the payment due depends on the realized return of the asset. Part of the nature of standard debt contracts—which grants the entrepreneur only residual rights and never has the entrepreneur share equity with investors—is still preserved even without assuming MLRP, but now the face value of the debt may depend on the realized asset return. In essence, this result separates the effects of limited liability from those of MLRP on the optimal security.

Overall, this paper uncovers the common underlying role of monotone function intervals in many topics in economics, and it offers a unifying approach to answering canonical economic questions that have been previously answered by separate, case-specific approaches.

[1] Nondecreasing securities are desirable, as a security holder would not have an incentive to sabotage the asset if the payment they receive is increasing in the asset's return.

Related Literature. This paper relates to several areas. The main result connects to characterizations of extreme points of convex sets. Pioneering in this area, Hardy, Littlewood and Pólya (1929) characterize the extreme points of a set of vectors x majorized by another vector x_0 in $\mathbb{R}^n$.[2] Ryff (1967) extends this result to infinite dimensional spaces. Kleiner, Moldovanu and Strack (2021) characterize the extreme points with an additional monotonicity assumption, which is equivalent to focusing on the set of probability distributions that are either a mean-preserving spread or mean-preserving contraction of a probability distribution on $\mathbb{R}$.[3]

The results of Kleiner, Moldovanu and Strack (2021) can be regarded as characterizations of extreme points of *increasing convex* functions, defined on a bounded interval and agreeing at the endpoints, that dominate another increasing convex function *or* are dominated by another increasing convex function. By comparison, this paper characterizes the extreme points of *increasing* functions that dominate an increasing function *and* are dominated by another increasing function. Moreover, the applications of Kleiner, Moldovanu and Strack (2021) pertain to the dispersion of expected values (i.e., problems related to *second*-order stochastic dominance), such as allocation problems with quasi-linear preferences, two-sided matching, delegation problems, and mean-based persuasion. The applications in this paper pertain to levels and orderings of variables of interest (i.e., problems related to *first*-order stochastic dominance), such as voting, quantile-based persuasion, self-ranking, and monotone securities with limited liability.

The first application of the extreme point characterization to the distributions of posterior quantiles is related to belief-based characterizations of signals, which date back to the seminal contributions of Blackwell (1953) and Harsanyi (1967-68). The characterization of distributions of posterior means can be derived from Strassen (1965). This paper's characterization of posterior quantiles is a complement.

The application to gerrymandering relates to the literature on redistricting, particularly to Owen and Grofman (1988), Friedman and Holden (2008), Gul and Pesendorfer (2010), and Kolotilin and Wolitzky (2023), who also adopt the belief-based approach and model a district map as a way to split the population distribution of voters. Existing work mainly focuses on a political party's optimal gerrymandering when maximizing either its expected number of seats or its probability of winning a majority. In contrast, this paper characterizes the *feasible compositions* of a legislative body that a district map can induce. The application to Bayesian

[2] A vector $x \in \mathbb{R}^n$ majorizes $y \in \mathbb{R}^n$ if $\sum_{i=1}^k x_{(i)} \geq \sum_{i=1}^k y_{(i)}$ for all $k \in \{1, \dots, n\}$, with equality at $k = n$, where $x_{(j)}$ and $y_{(j)}$ are the j-th smallest component of x and y, respectively.

[3] Several recent papers exploit properties of extreme points to derive economic implications. See, for instance, Bergemann, Brooks and Morris (2015); Lipnowski and Mathevet (2018); and Arieli, Babichenko, Smorodinsky and Yamashita (2023).

persuasion relates to that large literature (see Kamenica 2019 for a comprehensive survey), in particular to communication problems where only posterior means are payoff-relevant (e.g., Gentzkow and Kamenica 2016; Roesler and Szentes 2017; Dworczak and Martini 2019; Ali, Haghpanah, Lin and Siegel 2022). This paper complements that literature by providing a foundation for solving communication problems where only the posterior *quantiles* are payoff-relevant.

Finally, the application to security design connects this paper to that large literature. Allen and Barbalau (2022) provide a recent survey. In this application, we base our economic environments on Innes (1990), which involves moral hazard, and DeMarzo and Duffie (1999), which involves adverse selection. This paper generalizes and unifies results in those seminal works under a common structure.

Outline. The rest of the paper proceeds as follows. Section 2 presents the paper's central theorem: the characterization of the extreme points of monotone function intervals (Theorem 1). Section 3 applies Theorem 1 to characterize the distributions of posterior quantiles. Economic applications related to the quantile characterization (gerrymandering, quantile-based persuasion, and apparent misconfidence) follow in Section 3.2. Section 4 applies Theorem 1 to security design with limited liability. Section 5 concludes.

2 Extreme Points of Monotone Function Intervals

2.1 Notation

Let $\mathcal{F}$ be the set of nondecreasing and right-continuous functions on $\mathbb{R}$.[4] For any $\underline{F}, \overline{F} \in \mathcal{F}$ such that $\underline{F}(x) \leq \overline{F}(x)$ for all $x \in \mathbb{R}$ ($\underline{F} \leq \overline{F}$ henceforth), let

$$\mathcal{I}(\underline{F}, \overline{F}) := \{H \in \mathcal{F} \,|\, \underline{F}(x) \leq H(x) \leq \overline{F}(x), \, \forall x \in \mathbb{R}\}.$$

Namely, $\mathcal{I}(\underline{F}, \overline{F})$ is the collection of nondecreasing, right-continuous functions that dominate $\underline{F}$ and simultaneously are dominated by $\overline{F}$ pointwise. We refer to $\mathcal{I}(\underline{F}, \overline{F})$ as the *interval* of monotone functions bounded by $\underline{F}$ and $\overline{F}$. For any $F \in \mathcal{F}$ and for any $x \in \mathbb{R}$, let $F(x^-) := \lim_{y \uparrow x} F(y)$ denote the left-limit of F at x.

[4]Whenever needed, $\mathcal{F}$ is endowed with the topology defined by weak convergence (i.e., $\{F_n\} \to F$ if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x at which F is continuous), as well as the Borel σ-algebra induced by this topology.

2.2 Extreme Points of Monotone Function Intervals

For any $\underline{F}, \overline{F} \in \mathcal{F}$ with $\underline{F} \leq \overline{F}$, the interval $\mathcal{I}(\underline{F}, \overline{F})$ is a convex set. Recall that $H \in \mathcal{I}(\underline{F}, \overline{F})$ is an *extreme point* of the convex set $\mathcal{I}(\underline{F}, \overline{F})$ if H cannot be written as a convex combination of two distinct elements of $\mathcal{I}(\underline{F}, \overline{F})$. Our main result, Theorem 1, characterizes the extreme points of $\mathcal{I}(\underline{F}, \overline{F})$.

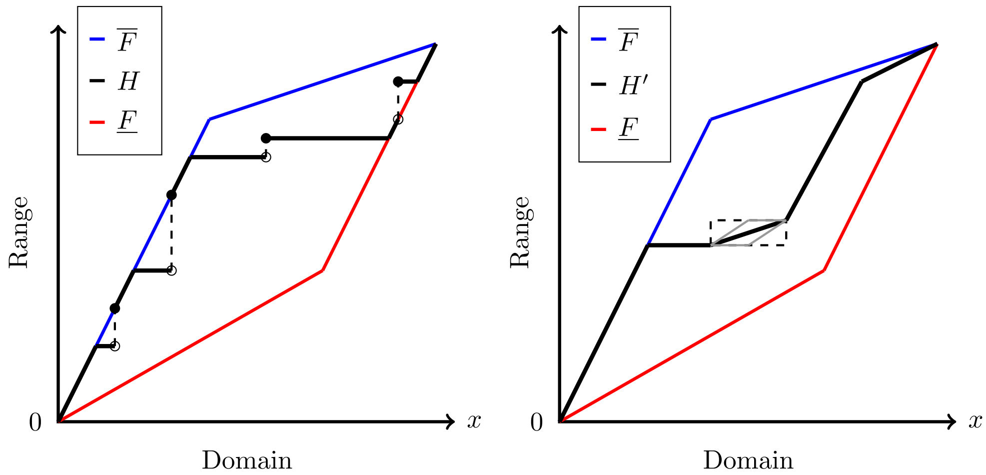
Theorem 1 (Extreme Points of $\mathcal{I}(\underline{F}, \overline{F})$)**.** *For any $\underline{F}, \overline{F} \in \mathcal{F}$ such that $\underline{F} \leq \overline{F}$, H is an extreme point of $\mathcal{I}(\underline{F}, \overline{F})$ if and only if there exists a countable collection of intervals $\{[\underline{x}_n, \overline{x}_n)\}_{n=1}^{\infty}$ such that:*

1. *$H(x) \in \{\underline{F}(x), \overline{F}(x)\}$ for all $x \notin \cup_{n=1}^{\infty}[\underline{x}_n, \overline{x}_n)$.*
2. *For all $n \in \mathbb{N}$, H is constant on $[\underline{x}_n, \overline{x}_n)$ and either $H(\overline{x}_n^-) = \underline{F}(\overline{x}_n^-)$ or $H(\underline{x}_n) = \overline{F}(\underline{x}_n)$.*

Figure IIA depicts an extreme point H of a monotone function interval $\mathcal{I}(\underline{F}, \overline{F})$, where the blue curve is the upper bound $\overline{F}$, and the red curve is the lower bound $\underline{F}$. According to Theorem 1, any extreme point H of $\mathcal{I}(\underline{F}, \overline{F})$ must either coincide with one of the bounds, or be constant on an interval in its domain, where at least one end of the interval reaches one of the bounds.

Appendix A.1 contains the proof of Theorem 1. We briefly summarize the argument here. For the sufficiency part, consider any H that satisfies conditions 1 and 2 of Theorem 1. Suppose that H can be expressed as a convex combination of two distinct H_1 and H_2 in $\mathcal{I}(\underline{F}, \overline{F})$. Then, for any $x \notin \cup_{n=1}^{\infty}[\underline{x}_n, \overline{x}_n)$, it must be that $H_1(x) = H_2(x) = H(x)$, since otherwise at least one of $H_1(x)$ and $H_2(x)$ would be either above $\overline{F}(x)$ or below $\underline{F}(x)$. Thus, since $H_1 \neq H_2$, there exists $n \in \mathbb{N}$ such that $H_1(x) \neq H_2(x)$ and $\lambda H_1(x) + (1-\lambda)H_2(x) = H(x)$ for all $x \in [\underline{x}_n, \overline{x}_n)$, for some $\lambda \in (0,1)$. Since H is constant on $[\underline{x}_n, \overline{x}_n)$, and since H_1 and H_2 are nondecreasing, both H_1 and H_2 must also be constant on $[\underline{x}_n, \overline{x}_n)$. Suppose that, without loss, $H_1(x) < H(x) < H_2(x)$ for all $x \in [\underline{x}_n, \overline{x}_n)$. If $H(\underline{x}_n) = \overline{F}(\underline{x}_n)$, then $\overline{F}(\underline{x}_n) = H(\underline{x}_n) < H_2(\underline{x}_n)$; whereas if $H(\overline{x}_n^-) = \underline{F}(\overline{x}_n^-)$, then $H_1(\overline{x}_n^-) > H(\overline{x}_n^-) = \underline{F}(\overline{x}_n^-)$. In either case, one of H_1 and H_2 must not be an element of $\mathcal{I}(\underline{F}, \overline{F})$, a contradiction.

For the necessity part, consider any H' that does not satisfy conditions 1 and 2 of Theorem 1. In this case, as depicted in Figure IIB, there exists a rectangle that lies between the graphs of $\underline{F}$ and $\overline{F}$, so that when restricted to this rectangle, the graph of H' is not a step function with only one jump. Then, since extreme points of uniformly bounded, nondecreasing functions are step functions with only one jump (see, for example, Skreta 2006; Börgers 2015), H' can be written as a convex combination of two distinct nondecreasing functions when restricted to this rectangle. Since the rectangle lies in between the graphs of $\underline{F}$ and



(A) An Extreme Point, H **(B) Not an Extreme Point, H'**

Figure II
EXTREME POINTS OF $\mathcal{I}(\underline{F}, \overline{F})$

$\overline{F}$, this, in turn, implies that H' can be written as a convex combination of two distinct distributions in $\mathcal{I}(\underline{F}, \overline{F})$.

Remark 1. Several assumptions in the setup are for ease of exposition and can be relaxed. First, the domain of $F \in \mathcal{F}$ does not need to be $\mathbb{R}$. Theorem 1 holds for any monotone function intervals defined on a totally ordered topological space. Second, right-continuity of $F \in \mathcal{F}$ serves as a convention that dictates how a function behaves whenever the function is discontinuous, and is consistent with the natural topology of weak convergence. Lastly, Theorem 1 can be extended even if the bounds $\underline{F}$ and $\overline{F}$ are nonmonotonic. Indeed, for arbitrary functions $\underline{F}, \overline{F}$ such that $\underline{F} \leq \overline{F}$, and for any nondecreasing function H, $\underline{F} \leq H \leq \overline{F}$ if and only if $\text{mon}_+(\underline{F}) \leq H \leq \text{mon}_-(\overline{F})$, where $\text{mon}_+(\underline{F})$ is the smallest nondecreasing function above $\underline{F}$ and $\text{mon}_-(\overline{F})$ is the largest monotone function below $\overline{F}$.

It is also noteworthy that Theorem 1 extends to the case where one of the two bounds $\overline{F}$, $\underline{F}$ equals $\pm\infty$, respectively. Consider when $\overline{F}$ is the bound that takes a finite value for all x and $\underline{F} = -\infty$. (The other case follows symmetrically.) Then, H is an extreme point of $\mathcal{I}(\underline{F}, \overline{F})$ if and only if there exists a countable collection of intervals $\{[\underline{x}_n, \overline{x}_n)\}_{n=1}^{\infty}$ such that $H(x) = \overline{F}(x)$ for all $x \notin \cup_{n=1}^{\infty}[\underline{x}_n, \overline{x}_n)$ and $H(x) = \overline{F}(\underline{x}_n)$ for all $x \in [\underline{x}_n, \overline{x}_n)$ and for all n.

Remainder of the Paper. In the ensuing sections, we demonstrate how the characterization of the extreme points of monotone function intervals can be applied to various economic settings. These applications rely on two crucial properties of extreme points. The first property—formally known as Choquet's theorem—allows one to express any element H of $\mathcal{I}(\underline{F}, \overline{F})$ as a mixture of its extreme points if $\mathcal{I}(\underline{F}, \overline{F})$ is compact. As a result, if one wishes to establish some property for every element of $\mathcal{I}(\underline{F}, \overline{F})$, and if this property is preserved under convex combinations, then it suffices to establish the property for all extreme points of $\mathcal{I}(\underline{F}, \overline{F})$, which is a much smaller set. Section 3 uses this first property to characterize the distributions of posterior quantiles. The second property of extreme points is that, for any convex optimization problem, one of the solutions must be an extreme point of the feasible set. This property is useful for economic applications because it immediately provides knowledge about the solutions to the underlying economic problem if that problem is convex and if the feasible set is related to a monotone function interval. Section 4 uses this second property to analyze security design.

3 Distributions of Posterior Quantiles

Theorem 1 alongside Choquet's theorem permits the characterization of the distributions of posterior quantiles. This characterization is an analog of the celebrated characterization of the distributions of posterior means that follows from Strassen's theorem (Strassen 1965). Knowing the distributions of posterior quantiles is important for settings where only the ordinal values or relative rankings of the relevant variables are meaningful, rather than the cardinal values or numeric differences (e.g., voting, grading or rating schemes, job performance rankings, measures of inequality). Moreover, posterior quantiles are also useful for studying distributions without well-defined moments, which arise in finance and insurance for instance.

3.1 Characterization of the Distributions of Posterior Quantiles

Let $\mathcal{F}_0 \subseteq \mathcal{F}$ be the collection of cumulative distribution functions (CDFs) in $\mathcal{F}$.[5] Consider a one-dimensional variable $x \in \mathbb{R}$ that is drawn from a prior F. A *signal* for x is defined as a probability measure $\mu \in \Delta(\mathcal{F}_0)$ such that

$$\int_{\mathcal{F}_0} G(x)\mu(\mathrm{d}G) = F(x), \tag{1}$$

for all $x \in \mathbb{R}$. Let $\mathcal{M}$ denote the collection of all signals.[6]

For any CDF $G \in \mathcal{F}_0$ and for any $\tau \in (0,1)$, denote the set of τ-quantiles of G by $[G^{-1}(\tau), G^{-1}(\tau^+)]$, where $G^{-1}(\tau) := \inf\{x \in \mathbb{R} | G(x) \geq \tau\}$ is the *quantile function* of G and

[5] That is, $G \in \mathcal{F}_0$ if and only if $G \in \mathcal{F}$ and $\lim_{x\to\infty} G(x) = 1$ and $\lim_{x\to-\infty} G(x) = 0$.

[6] Here, we take the belief-based approach and model signals as distributions of posteriors that average to the prior. From Blackwell's theorem (Blackwell 1953; Strassen 1965), given any $\mu \in \mathcal{M}$, each $G \in \mathrm{supp}(\mu)$ can be interpreted as a *posterior* for x (i.e., a version of the regular conditional distribution of x conditional on a signal realization). The marginal distribution of this signal is summarized by μ.

$G^{-1}(\tau^+) := \lim_{q\downarrow\tau} G^{-1}(q)$ denotes the right-limit of G^{-1} at τ.[7] Since the τ-quantile for an arbitrary CDF may not be unique, we further introduce a notation for selecting a quantile. We say that a transition probability $r : \mathcal{F}_0 \to \Delta(\mathbb{R})$ is a *τ-quantile selection rule* if, for all $G \in \mathcal{F}_0$, $r(\cdot|G)$ assigns probability 1 to $[G^{-1}(\tau), G^{-1}(\tau^+)]$. In other words, a quantile selection rule r selects (possibly through randomization) a τ-quantile of G, for every CDF G, whenever it is not unique. Let $\mathcal{R}_\tau$ be the collection of all τ-quantile selection rules.

For any $\tau \in (0,1)$, for any signal $\mu \in \mathcal{M}$, and for any selection rule $r \in \mathcal{R}_\tau$, let $H^\tau(\cdot|\mu, r)$ denote the distribution of the τ-quantile induced by μ and r. For any $\tau \in (0,1)$, let $\mathcal{H}_\tau$ denote the set of distributions of posterior τ-quantiles that can be induced by some signal $\mu \in \mathcal{M}$ and selection rule $r \in \mathcal{R}_\tau$.

Using Theorem 1, we provide a complete characterization of the distributions of posterior quantiles induced by arbitrary signals and selection rules. To this end, define two distributions F_L^τ and F_R^τ as follows:

$$F_L^\tau(x) := \min\left\{\frac{1}{\tau}F(x), 1\right\}, \quad F_R^\tau(x) := \max\left\{\frac{F(x)-\tau}{1-\tau}, 0\right\}.$$

Note that $F_R^\tau \leq F_L^\tau$ for all $\tau \in (0,1)$. In essence, F_L^τ is the *left-truncation* of the prior F: the conditional distribution of F in the event that x is smaller than a τ-quantile of F; whereas F_R^τ is the *right-truncation* of F: the conditional distribution of F in the event that x is larger than the same τ-quantile. Theorem 2 below characterizes the distributions of posterior quantiles $\mathcal{H}_\tau$.

Theorem 2 (Distributions of Posterior Quantiles)**.** *For any $\tau \in (0,1)$,*

$$\mathcal{H}_\tau = \mathcal{I}(F_R^\tau, F_L^\tau).$$

Theorem 2 characterizes the distributions of posterior τ-quantiles by the monotone function interval $\mathcal{I}(F_R^\tau, F_L^\tau)$. Notice that, because F_R^τ and F_L^τ are CDFs, their pointwise dominance relation means that F_R^τ first-order stochastically dominates F_L^τ. Figure III illustrates Theorem 2 for the case when $\tau = 1/2$. The distribution $F_L^{1/2}$ is colored blue, whereas the distribution $F_R^{1/2}$ is colored red. The green dotted curve represents the prior, F. According to Theorem 2, any distribution H bounded by $F_L^{1/2}$ and $F_R^{1/2}$ (for instance, the black curve in the figure) can be induced by a signal $\mu \in \mathcal{M}$ and a selection rule $r \in \mathcal{R}_{1/2}$. Conversely, for any signal and for any selection rule, the induced graph of the distribution of posterior τ-quantiles must fall in the area bounded by the blue and red curves. For example, under the signal that

[7] Note that G^{-1} is nondecreasing and left-continuous for all $G \in \mathcal{F}_0$. Moreover, for any $\tau \in (0,1)$ and for any $x \in \mathbb{R}$, $G^{-1}(\tau) \leq x$ if and only if $G(x) \geq \tau$.

reveals all the information, the distribution of posterior $1/2$-quantiles coincides with the prior, whereas under the signal that does not reveal any information, the distribution of posterior $1/2$-quantiles coincides with the step function that has a jump (of size 1) at $F^{-1}(1/2)$.

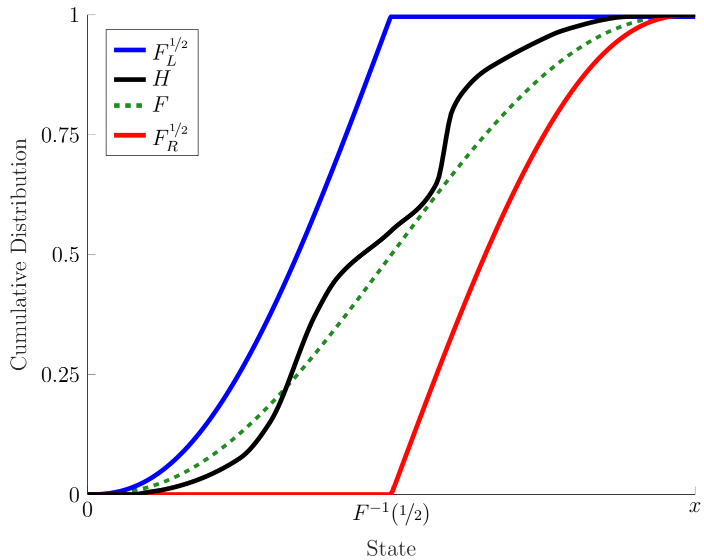
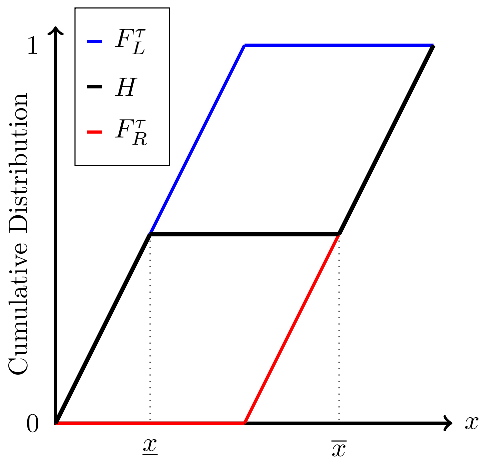


Figure III
Distributions of Posterior Medians

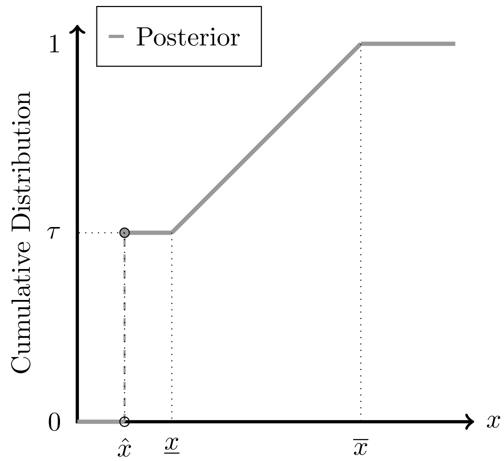
Theorem 2 can be regarded as a natural analog of the well-known characterization of the distributions of posterior *means* that follows from Strassen (1965). Strassen's theorem implies that a CDF $H \in \mathcal{F}_0$ is a distribution of posterior means if and only if H is a mean-preserving contraction of the prior F. Instead of posterior means, Theorem 2 pertains to posterior quantiles. According to Theorem 2, H is a distribution of posterior τ-quantiles if and only if H first-order stochastically dominates the left-truncation F_L^τ and is dominated by the right-truncation F_R^τ.

The fact that $\mathcal{H}_\tau \subseteq \mathcal{I}(F_R^\tau, F_L^\tau)$ follows from the martingale property of posterior beliefs. Proving the other direction ($\mathcal{I}(F_R^\tau, F_L^\tau) \subseteq \mathcal{H}_\tau$) is more challenging. To prove this, one would in principle need to construct a signal that generates the desired distribution of posterior quantiles for *every* distribution $H \in \mathcal{I}(F_R^\tau, F_L^\tau)$. Although it might be easier to construct a signal that induces some specific distribution of posterior quantiles, constructing a signal for any arbitrary distribution $H \in \mathcal{I}(F_R^\tau, F_L^\tau)$ does not seem tractable.[8] Nonetheless, Theorem 1

[8]The bounds F_R^τ and F_L^τ can be attained using a modified version of the "matching extreme" signal introduced by Friedman and Holden (2008). However, matching extremes cannot induce any distribution $H \in \mathcal{I}(F_R^\tau, F_L^\tau)$ that assigns probability zero to some interval containing $[F^{-1}(\tau), F^{-1}(\tau^+)]$. The reason is that matching extreme signals would inevitably assign positive probability to posteriors whose quantiles are nearby the prior quantile.



(A) An Extreme Point H



(B) Posterior Containing $\hat{x}$

Figure IV
CONSTRUCTING A SIGNAL THAT INDUCES H

bypasses this challenge and puts focus on distributions that satisfy its conditions 1 and 2. Indeed, since the mapping $(\mu, r) \mapsto H^\tau(\cdot|\mu, r)$ is affine, it suffices to construct signals that induce the extreme points of $\mathcal{I}(F_R^\tau, F_L^\tau)$ as posterior quantile distributions. The proof of Theorem 2 in Appendix A.2 explicitly constructs a signal (and a selection rule) for each extreme point of $\mathcal{I}(F_R^\tau, F_L^\tau)$. To see the intuition, consider an extreme point H of $\mathcal{I}(F_R^\tau, F_L^\tau)$ that takes the following form:

$$H(x) = \begin{cases} F_L^\tau(x), & \text{if } x < \underline{x} \\ F_L^\tau(\underline{x}), & \text{if } x \in [\underline{x}, \overline{x}) \\ F_R^\tau(x), & \text{if } x \geq \overline{x} \end{cases},$$

for some $\underline{x}, \overline{x}$ such that $F_L^\tau(\underline{x}) = F_R^\tau(\overline{x}^-)$, as depicted in Figure IVA. To construct a signal that has H as its distribution of posterior quantiles, separate all the states $x \notin [\underline{x}, \overline{x}]$. Then, take α fraction of the states in $[\underline{x}, \overline{x}]$ and pool them uniformly with each separated state below $\underline{x}$, while pooling the remaining $1 - \alpha$ fraction uniformly with the separated states above $\overline{x}$. Since $F_L^\tau(\underline{x}) = F_R^\tau(\overline{x}^-)$, when α is chosen correctly,[9] each $x < \underline{x}$, after being pooled with states in $[\underline{x}, \overline{x}]$, would become a τ-quantile of the posterior it belongs to, as illustrated in Figure IVB. Similarly, each $x > \overline{x}$ would become a τ-quantile of the posterior it belongs to. Together, by properly selecting the posterior quantiles, the induced distribution of posterior quantiles under this signal would indeed be H.

Although the characterization of Theorem 2 may seem to rely on selection rules $r \in \mathcal{R}_\tau$,

[9]Specifically, $\alpha = \frac{1-\tau}{\tau} F(\underline{x}) / (\frac{\tau}{1-\tau}(1 - F(\overline{x}^-)) + \frac{1-\tau}{\tau} F(\underline{x}))$.

the result remains (essentially) the same even when restricted to signals that always induce a unique posterior τ-quantile, provided that the prior F has full support on an interval. Theorem 3 below formalizes this statement. To this end, Let $\widetilde{\mathcal{H}}_\tau \subseteq \mathcal{H}_\tau$ be the collection of distributions of posterior τ-quantiles that can be induced by some signal where (almost) all posteriors have a unique τ-quantile. The characterization of $\widetilde{\mathcal{H}}_\tau$ relates to a family of perturbations of the set $\mathcal{I}(F_R^\tau, F_L^\tau)$, denoted by $\{\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})\}_{\varepsilon>0}$, where

$$F_L^{\tau,\varepsilon}(x) := \begin{cases} \frac{1}{\tau+\varepsilon}F(x), & \text{if } x < F^{-1}(\tau) \\ 1, & \text{if } x \geq F^{-1}(\tau) \end{cases} \text{ ; and } F_R^{\tau,\varepsilon}(x) := \begin{cases} 0, & \text{if } x < F^{-1}(\tau) \\ \frac{F(x)-(\tau-\varepsilon)}{1-(\tau-\varepsilon)}, & \text{if } x \geq F^{-1}(\tau) \end{cases},$$

for all $\varepsilon \geq 0$ and for all $x \in \mathbb{R}$. Note that $\mathcal{I}(F_R^{\tau,0}, F_L^{\tau,0}) = \mathcal{I}(F_R^\tau, F_L^\tau)$, and $\{\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})\}_{\varepsilon>0}$ is decreasing in ε under the set-inclusion order.[10]

Theorem 3 (Distributions of Unique Posterior Quantiles)**.** *For any $\tau \in (0,1)$ and for any $F \in \mathcal{F}_0$ that has a full support on an interval,*

$$\bigcup_{\varepsilon>0} \mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon}) \subseteq \widetilde{\mathcal{H}}_\tau \subseteq \mathcal{I}(F_R^\tau, F_L^\tau).$$

According to Theorem 3, for any $\varepsilon > 0$ and for any $H \in \mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})$, there exists a signal μ such that H is the distribution of *unique* posterior τ-quantiles. In other words, the distributions of *unique* posterior quantiles are given by the "interior" of $\mathcal{I}(F_R^\tau, F_L^\tau)$, and only the "boundaries" of $\mathcal{I}(F_R^\tau, F_L^\tau)$ (such as F_R^τ and F_L^τ themselves) are lost by requiring uniqueness.

As an immediate corollary of Theorem 2 and Theorem 3, an analog of the law of iterated expectations emerges, which we refer to as the *law of iterated quantiles*.

Corollary 1 (Law of Iterated Quantiles)**.** *Consider any $\tau, q \in (0,1)$.*

1. *For any closed interval $Q \subseteq \mathbb{R}$, $Q = [H^{-1}(\tau), H^{-1}(\tau^+)]$ for some $H \in \mathcal{H}_q$ if and only if $Q \subseteq [(F_R^q)^{-1}(\tau), (F_L^q)^{-1}(\tau^+)]$.*

2. *Suppose that the prior F is continuous and has full support on an interval. Then for any $\hat{x} \in \mathbb{R}$, $\hat{x} \in [H^{-1}(\tau), H^{-1}(\tau^+)]$ for some $H \in \widetilde{H}_q$ if and only if $\hat{x} \in [(F_R^q)^{-1}(\tau), (F_L^q)^{-1}(\tau)]$.*

The intuition of Corollary 1 is summarized in Figure V. For any $q, \tau \in (0,1)$, Figure V plots the interval $\mathcal{I}(F_R^q, F_L^q)$, which, according to Theorem 2 (and Theorem 3), equals all possible distributions of posterior q-quantiles. Therefore, the τ-quantiles of posterior q-quantiles must coincide with the interval $[(F_L^q)^{-1}(\tau), (F_R^q)^{-1}(\tau^+)]$. According to Corollary 1,

[10]As a convention, let $\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon}) := \emptyset$ when $\varepsilon \geq \max\{\tau, 1-\tau\}$.

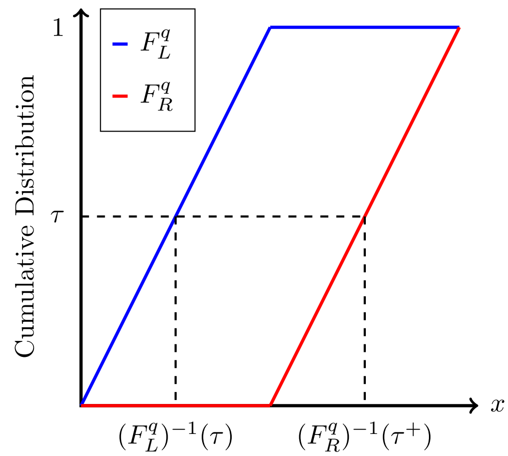


Figure V
LAW OF ITERATED QUANTILES

while the expectation of posterior means under any signal is always the expectation under the prior, the possible τ-quantiles of posterior q-quantiles are exactly $[(F_L^q)^{-1}(\tau), (F_R^q)^{-1}(\tau^+)]$. For example, the collection of all possible *medians* of posterior *medians* is the *interquartiles* $[F^{-1}(1/4), F^{-1}(3/4^+)]$ of the prior.

3.2 Economic Applications

In what follows, we illustrate economic applications of Theorem 2 and Theorem 3 through three examples. These examples are connected by their concerns over ordinal rankings, instead of cardinal values, of relevant outcomes. The first application is to gerrymandering; here, citizens rank candidates' positions relative to their own ideal positions, and the median voter theorem determines who is elected. The second application is to Bayesian persuasion when payoffs depend only on posterior quantiles. The third application is to apparent misconfidence, which explains why people rank themselves better or worse than others.

Limits of Gerrymandering

We first apply Theorem 2 and Theorem 3 to gerrymandering. Existing economic theory on gerrymandering has primarily focused on optimal redistricting or fair redistricting mechanisms (e.g., Owen and Grofman 1988; Friedman and Holden 2008; Gul and Pesendorfer 2010; Pegden, Procaccia and Yu 2017; Ely 2019; Friedman and Holden 2020; Kolotilin and Wolitzky 2023). Another fundamental question is the scope of gerrymandering's impact on a legislature. If *any* electoral map can be drawn, what kinds of legislatures can be created? In other words, what are the "limits of gerrymandering"?

Theorem 2 and Theorem 3 can be used to answer this question. Consider an environment in which a continuum of citizens vote, and each citizen has single-peaked preferences over positions on political issues. Citizens have different ideal positions $x \in \mathbb{R}$, and these positions are distributed according to some $F \in \mathcal{F}_0$. In this setting, a signal $\mu \in \mathcal{M}$ can be thought of as an electoral *map*, which segments citizens into electoral *districts*, such that a district $G \in \text{supp}(\mu)$ is described by the conditional distribution of the ideal positions of citizens who belong to it. Each district elects a *representative*, and election results at the district-level follow the median voter theorem. That is, given any map $\mu \in \mathcal{M}$, the elected representative of each district $G \in \text{supp}(\mu)$ must have an ideal position that is a median of G. When there are multiple medians in a district, the representative's ideal position is determined by a selection rule $r \in \mathcal{R}_{1/2}$, which is either flexible or stipulated by election laws.[11]

Given any $\mu \in \mathcal{M}$ and any selection rule $r \in \mathcal{R}_{1/2}$, the induced distribution of posterior medians $H^{1/2}(\cdot|\mu, r)$ can be interpreted as the distribution of the ideal positions of the elected representatives. Meanwhile, the bounds $F_L^{1/2}$ and $F_R^{1/2}$ can be interpreted as distributions of representatives that only reflect one side of voters' political positions relative to the median of the population. Specifically, $F_L^{1/2}$ describes an "all-left" legislature, which only reflects citizens' ideal positions that are left of the population median. Likewise, $F_R^{1/2}$ represents an "all-right" legislature, which only reflects citizens' ideal positions that are right of the population median. As an immediate implication of Theorem 2 and Theorem 3, Proposition 1 below characterizes the set of possible compositions of the legislature across all election maps.

Proposition 1 (Limits of Gerrymandering). *For any $H \in \mathcal{F}_0$, the following are equivalent:*

1. *$H \in \mathcal{I}(F_R^{1/2}, F_L^{1/2})$.*
2. *H is a distribution of the representatives' ideal positions under some map $\mu \in \mathcal{M}$ and some selection rule $r \in \mathcal{R}_{1/2}$.*

Furthermore, for any fixed selection rule $\hat{r} \in \mathcal{R}_{1/2}$, every $H \in \cup_{\varepsilon>0}\mathcal{I}(F_R^{1/2,\varepsilon}, F_L^{1/2,\varepsilon})$ is a distribution of the representatives' ideal positions under some map $\mu \in \mathcal{M}$ and selection rule $\hat{r}$.

Proposition 1 characterizes the *compositions* of the legislature that gerrymandering can induce. According to Proposition 1, *any* composition of the legislative body ranging from the "all-left" to the "all-right," and anything in between those two extremes, can be created

[11] Any voting method that meets the Condorcet criterion (e.g., majority voting with two office-seeking candidates) satisfies the median voter property in this setting (Downs 1957; Black 1958).

by some gerrymandered map. Meanwhile, *any* composition that is more extreme than the "all-left" or the "all-right" bodies is not possible, regardless of how districts are drawn.[12]

If we further specify the model for the legislature to enact legislation, we may explore the set of possible *legislative outcomes* that can be enacted. One natural assumption for the outcomes, regardless of the details of the legislative model, is that the enacted legislation must be a median of the representatives (i.e., the median voter property holds at the legislative level).[13] Under this assumption, an immediate implication of Corollary 1 is that the set of achievable legislative outcomes coincides with the interquartile range of the citizenry's ideal positions, as summarized by Corollary 2 below.

Corollary 2 (Limits of Legislative Outcomes)**.** *Suppose that the median voter property holds both at the district and legislative level. Then an outcome $x \in \mathbb{R}$ can be enacted as legislation under some map if and only if $x \in [F^{-1}(1/4), F^{-1}(3/4^+)]$.*

According to Corollary 2, while the only Condorcet winners in this setting are the population medians, gerrymandering expands the set of possible legislation to the entire interquartile range of the population's views. Conversely, Corollary 2 also suggests it is impossible to enact any legislative outcome *beyond* the interquartile range, regardless of how the districts are drawn. Studying these downstream effects of gerrymandering on enacted legislation is less common in the political economy literature, which tends to stop at the solution of an optimal map. Work that has examined possible legislation under gerrymandering typically focuses on"policy bias," which is the gap between majority rule (i.e., the ideal point of the population's median voter) and the ultimate policy that could come out of the legislature under some gerrymandered map (Shotts 2002; Buchler 2005; Gilligan and Matsusaka 2006). Corollary 2 unifies existing results on bounding the potential magnitude of policy bias.

Furthermore, as the population becomes more polarized, so that the interquartile range becomes wider, more extreme legislation can pass. For instance, consider two population distributions F and $\widetilde{F}$ with the same unique median x^*, and suppose that $\widetilde{F}$ is more dispersed than F under the rotation order around the common median. That is, $F(x) \geq \widetilde{F}(x)$ for all $x > x^*$ and $F(x) \leq \widetilde{F}(x)$ for all $x < x^*$. Then it must be that $\widetilde{F}^{-1}(1/4) \leq F^{-1}(1/4) \leq F^{-1}(3/4^+) \leq \widetilde{F}^{-1}(3/4^+)$. By Corollary 2, it then follows that the range of legislation that can be enacted becomes wider as the population distribution becomes more dispersed.

[12] Gomberg, Pancs and Sharma (2023) also study how gerrymandering affects the composition of the legislature. However, the authors assume that each district elects a *mean* candidate as opposed to the median.

[13] See McCarty, Poole and Rosenthal 2001; Bradbury and Crain 2005; and Krehbiel 2010 for evidence that the median legislator is decisive. See also Cho and Duggan (2009) for a microfoundation.

Remark 2 (Districts on a Geographic Map). In practice, election districts are drawn on a geographic map. Drawing districts in this manner can be regarded as partitioning a two-dimensional space that is spanned by latitude and longitude. More specifically, let a convex and compact set $\Theta \subseteq [0,1]^2$ denote a geographic map. Suppose that every citizen who resides at the same location $\theta \in \Theta$ shares the same ideal position $\boldsymbol{x}(\theta)$, where $\boldsymbol{x} : \Theta \to \mathbb{R}$ is a measurable function. Furthermore, suppose that citizens are distributed on Θ according to a density function $\phi > 0$. Under this setting, Theorem 1 of Yang (2020) ensures that for any $\mu \in \mathcal{M}$ with a countable support, there exists a countable partition of Θ, such that the distributions of citizens' ideal positions within each element coincide with the distributions in the support of μ. If we further assume that $\boldsymbol{x}$ is non-degenerate, in the sense that each of its indifference curves $\{\theta \in \Theta | \boldsymbol{x}(\theta) = x\}_{x \in \mathbb{R}}$ is isomorphic to the unit interval, then Theorem 2 of Yang (2020) ensures that for any $\mu \in \mathcal{M}$, there exists a partition on Θ that generates the same distributions in each district. Therefore, the splitting of the distribution of citizens' ideal positions has an exact analog to the splitting of geographic areas on a physical map.

Not only can Proposition 1 characterize the set of feasible maps based on the citizenry's distribution of ideal positions, but it can also help identify that distribution itself. Suppose that H is the observed distribution of representatives' ideal positions. Proposition 1 implies that the population distribution F must have H be dominated by $F_R^{1/2}$ and dominate $F_L^{1/2}$ at the same time. This leads to Corollary 3 below.

Corollary 3 (Identification Set of F)**.** *Suppose that $H \in \mathcal{F}_0$ is the distribution of ideal positions of a legislature. Then the distribution of citizens' ideal positions F must satisfy*

$$\frac{1}{2}H(x) \leq F(x) \leq \frac{1+H(x)}{2}, \tag{2}$$

for all $x \in \mathbb{R}$. Conversely, for any $F \in \mathcal{F}_0$ satisfying (2), there exists a map $\mu \in \mathcal{M}$ and a selection rule $r \in \mathcal{R}_{1/2}$, such that H is the distribution of ideal positions of the legislature.

According to Corollary 3, the distribution of citizens' ideal positions can be identified by (2), even when only the distribution of the representatives' ideal positions can be observed.[14]

Quantile-Based Persuasion

Naturally, Theorem 2 and Theorem 3 can also be applied to a Bayesian persuasion setting where the receiver's payoff depends only on posterior quantiles. Consider the Bayesian persuasion problem formulated by Kamenica and Gentzkow (2011): A state $x \in \mathbb{R}$ is distributed

[14]In Yang and Zentefis (2022), we apply the same logic and use Theorem 2 and Theorem 3 to characterize the identification set of a nonparametric quantile regression function.

according to a common prior F. A sender chooses a signal $\mu \in \mathcal{M}$ to inform a receiver, who then picks an action $a \in A$ after seeing the signal's realization. The ex-post payoffs of the sender and receiver are $u_S(x, a)$ and $u_R(x, a)$, respectively. Kamenica and Gentzkow (2011) show that the sender's optimal signal and the value of persuasion can be characterized by the concave closure of the function $\hat{v} : \mathcal{F}_0 \to \mathbb{R}$, where $\hat{v}(G) := \mathbb{E}_F[u_S(x, a^*(G))]$ is the indirect utility of the sender, and $a^*(G) \in A$ is the optimal action of the receiver under posterior $G \in \mathcal{F}_0$.[15]

When $|\text{supp}(F)| > 2$, this "concavafication" method requires finding the concave closure of a multi-variate function, which is known to be computationally challenging, especially when $|\text{supp}(F)| = \infty$. For tractability, many papers have restricted attention to preferences where the only payoff-relevant statistic of a posterior is its mean (i.e., $\hat{v}(G)$ is measurable with respect to $\mathbb{E}_G[x]$). See, for example, Gentzkow and Kamenica (2016); Kolotilin, Li, Mylovanov and Zapechelnyuk (2017); Dworczak and Martini (2019); Kolotilin, Mylovanov and Zapechelnyuk (2022b); and Arieli, Babichenko, Smorodinsky and Yamashita (2023).

A natural analog of this "mean-based" setting is for the payoffs to depend only on the posterior quantiles. Just as mean-based persuasion problems are tractable because distributions of posterior means are mean-preserving contractions of the prior, Theorem 2 and Theorem 3 provide a tractable formulation of any "quantile-based" persuasion problem, as described in Proposition 2 below.

Proposition 2 (Quantile-Based Persuasion)**.** *Suppose that the prior F has full support on some interval, and suppose that there exists $\tau \in (0, 1)$, a selection rule $r \in \mathcal{R}_\tau$, and a measurable function $v_S : \mathbb{R} \to \mathbb{R}$ such that $\hat{v}(G) = \int_{\mathbb{R}} v_S(x) r(\mathrm{d}x|G)$, for all $G \in \mathcal{F}_0$. Then*

$$\text{cav}(\hat{v})[F] = \sup_{H \in \mathcal{I}(F_R^\tau, F_L^\tau)} \int_{\mathbb{R}} v_S(x) H(\mathrm{d}x). \tag{3}$$

Proof. Let $\bar{v}(G) := \sup_{x \in [G^{-1}(\tau), G^{-1}(\tau^+)]} v_S(x)$ for all $G \in \mathcal{F}_0$. Then, by Theorem 2,

$$\text{cav}(\hat{v})[F] \leq \text{cav}(\bar{v})[F] = \sup_{H \in \mathcal{I}(F_R^\tau, F_L^\tau)} \int_{\mathbb{R}} v_S(x) H(\mathrm{d}x).$$

Meanwhile, by Theorem 3,

$$\sup_{H \in \cup_{\varepsilon > 0} \mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})} \int_{\mathbb{R}} v_S(x) H(\mathrm{d}x) \leq \text{cav}(\hat{v})[F].$$

[15]When there are multiple optimal actions, subgame-perfection would always select the one that the sender prefers most.

Together, since $\mathrm{cl}(\{\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})\}) = \mathcal{I}(F_R^{\tau}, F_L^{\tau})$, (3) then follows. ■

By Proposition 2, any τ-quantile-based persuasion problem can be solved by simply choosing a distribution in $\mathcal{I}(F_R^{\tau}, F_L^{\tau})$ to maximize the expected value of $v_S(x)$, rather than concavafying the infinite-dimensional functional $\hat{v}$. Furthermore, since the objective function of (3) is affine, Theorem 1 further reduces the search for the solution to only distributions that satisfy its conditions 1 and 2.[16]

Proposition 2 can be used to analyze a persuasion problem where the receiver is not an expected utility maximizer but makes decisions according to ordinal models of utility (i.e., quantile maximizers), a class of preferences studied in Manski (1988), Chambers (2007), Rostek (2010), and de Castro and Galvao (2021). When selecting among lotteries, a τ-quantile-maximizer chooses the one that gives the highest τ-quantile of the utility distribution.[17]

In addition, Proposition 2 provides further insights into the structure of optimal signals in a class of canonical persuasion problems. Consider the standard setting where a receiver chooses an action to match the state and a sender has a state-independent payoff (i.e., $u_S(x, a) = v_S(a)$). The typical assumption is that the receiver minimizes a quadratic loss function (i.e., $u_R(x, a) := -(x - a)^2$).[18] Under this assumption, the receiver's optimal action $a^*(G)$, given a posterior G, equals the posterior expected value $\mathbb{E}_G[x]$, and hence, the sender's problem is mean-measurable. Parameterizing the receiver's loss function in this way makes the sender's persuasion problem tractable since the distributions of the receiver's actions are equivalent to mean-preserving contractions of the prior.[19] However, the shape of the loss function imposes a specific cardinal structure on the receiver's preferences, and it remains unclear how different parameterizations of the receiver's loss could affect the structure of the optimal signal.

With Proposition 2, one may now examine the sender's problem when the receiver has

[16]A recent elegant contribution by Kolotilin, Corrao and Wolitzky (2022a) provides a tractable method that simplifies persuasion problems in certain environments. One of these environments is when the receiver's payoff is supermodular and the sender's payoff is state-independent and increasing in the receiver's action. One of their examples in this environment is for the receiver's optimal action for each posterior to be quantile-measurable. When one further assumes that the sender's payoff is increasing, the conditions of Proposition 2 lead to the same example. Since we allow for arbitrary (state-independent) sender payoffs, Proposition 2 generalizes this example in an orthogonal direction and complements their method.

[17]For a quantile-maximizer, the parameter τ provides a complete ranking of risk attitudes, from extreme downside risk aversion ($\tau = 0$) to extreme downside risk tolerance ($\tau = 1$). In Yang and Zentefis (2022), we apply Proposition 2 to a setting where a monopolist can choose to disclose product information to a quantile-maximizing buyer, and we characterize the optimal signal.

[18]To fix ideas, we can let the sender be a financial advisor and the receiver be a client. The financial advisor wishes to persuade the client to allocate a fraction $a \in [0, 1]$ of wealth in stocks and the remaining $1 - a$ fraction in bonds. The client would prefer different portfolio allocations under different states $x \in [0, 1]$ of the economy.

[19]See Dworczak and Martini (2019) for a characterization of the solutions and an interpretation of the Lagrange multipliers.

a different loss function. When the receiver has an *absolute* loss function (i.e., $u_R(x,a) = -|x-a|$), the optimal action under any posterior must be a posterior median. More generally, when the receiver has a "pinball" loss function (i.e., $u_R(x,a) = -\rho_\tau(x-a)$, with $\rho_\tau(y) := y(\tau - \mathbf{1}\{y<0\})$), the optimal action under any posterior must be a posterior τ-quantile. Since the sender's payoff is state-independent, Proposition 2 applies, and the sender's problem can be rewritten via (3).[20] The pinball loss function imposes a different cardinality structure on the receiver's payoff, where the marginal loss remains constant instead of being linear as the action moves further away from the state.

With Proposition 2 and (3), one can solve the sender's problem when the receiver has a pinball loss function for some specific sender payoffs. Specifically, for any continuous prior F that has full support on some interval and for any $a \in \mathbb{R}$, let

$$H_a^L(x) := \begin{cases} 0, & \text{if } x < a \\ F_L^\tau(x), & \text{if } x \geq a \end{cases}; \quad \text{and } H_a^R(x) := \begin{cases} F_R^\tau(x), & \text{if } x < a \\ 1, & \text{if } x \geq a \end{cases},$$

for all $x \in \mathbb{R}$. Also, for any $\underline{a}, \overline{a} \in \mathbb{R}$ such that $F_L^\tau(\underline{a}) = F_R^\tau(\overline{a}) =: \eta$, let

$$H_{\underline{a},\overline{a}}^C(x) := \begin{cases} F_L^\tau(x), & \text{if } x < \underline{a} \\ \eta, & \text{if } x \in [\underline{a}, \overline{a}) \\ F_R^\tau(x), & \text{if } x \geq \overline{a} \end{cases}.$$

Corollary 4 summarizes the sender's optimal signal under various sender payoffs v_S.

Corollary 4. *Suppose that F is continuous and has full support on a compact interval. Suppose that $v_S : \mathbb{R} \to \mathbb{R}$ is upper-semicontinuous. Then*

(i) If v_S is quasi-concave and attains its maximum at $a \leq F^{-1}(\tau)$, then H_a^L solves (3).

(ii) If v_S is quasi-concave and attains its maximum at $a > F^{-1}(\tau)$, then H_a^R solves (3).

(iii) If v_S is strictly quasi-convex, then $H_{\underline{a},\overline{a}}^C$ solves (3) for some $\underline{a}, \overline{a}$ such that $F_L^\tau(\underline{a}) = \overline{F}^\tau(\overline{a})$.

(iv) F is never the unique solution of (3).

The distribution H_a^L (H_a^R) can be induced by separating all states below (above) $F^{-1}(\tau)$ and pooling all states above (below) $F^{-1}(\tau)$ with each of these separated states, and then pooling all the posteriors with states below (above) a together. This signal is optimal for the sender if the sender's payoff is quasi-concave and is maximized at $a \leq F^{-1}(\tau)$ ($a > F^{-1}(\tau)$). In particular, for any strictly concave v_S that is maximized at some $a \in \mathbb{R}$, it is optimal for the sender to reveal no information at all if the receiver's loss function is quadratic, but not optimal if the receiver's loss function is an absolute value. Moreover, the nature of the

[20] When applying Proposition 2 to this problem, one may take the selection rule r to be the one that always selects the sender-preferred τ-quantile.

receiver's loss function affects how the optimal signal changes when monotone transformations are applied to v_S. Since any monotone transformation of v_S remains quasi-concave and a remains its maximizer, the sender's optimal signal remains unchanged when the receiver's loss function is an absolute value; however, the optimal signal can be very different if the receiver has quadratic loss since the curvature of v_S may be different.[21]

Likewise, the distribution $H^C_{\underline{a},\overline{a}}$ can be induced by separating all states above $\overline{a}$ and below $\underline{a}$, while pooling all the states in $[\underline{a},\overline{a}]$ with each of these separated states. In particular, for any strictly convex v_S, it is optimal for the sender to reveal all the information if the receiver's loss function is quadratic, but not optimal if the receiver's loss function is an absolute value. In fact, since F is the distribution of posterior τ-quantiles under the fully revealing signal, it is never the unique optimal signal if the receiver's loss function is an absolute value.[22]

Apparent Misconfidence

Another application of the characterization of distributions of posterior quantiles relates to the literature on over/underconfidence (i.e., misconfidence) in the psychology of judgment. The experimental literature has documented that, when individuals are asked to predict their own abilities, a prediction dataset can be very different from the population distribution. Instead of attributing this observation to individuals being irrationally overconfident or underconfident, Benoît and Dubra (2011) proposed an alternative explanation: This difference can be caused by noise in each individual's signal. Individuals can still be fully Bayesian even if the prediction dataset is different from the population distribution. That is, individuals can be *apparently* misconfident due to dispersion of posterior beliefs. Here, we show how Benoît and Dubra (2011)'s insight follows immediately from Theorem 3.

Consider the following setting due to Benoît and Dubra (2011). There is a unit mass of individuals, and each one of them is attached to a "type" $x \in [0,1]$ that is distributed according to a CDF $F \in \mathcal{F}_0$. Common interpretations of x in the literature include skill levels, scores on a standardized test, the probability of being successful at a task, or simply an individual's ranking in the population in percentage terms. Individuals are asked to predict their own type x. Given a finite partition $0 = z_0 < z_1 < \ldots < z_K = 1$ of $[0,1]$, a *prediction dataset* is a vector $(\theta_k)_{k=1}^K \in [0,1]^K$ with $\sum_{k=1}^K \theta_k = 1$, where θ_k denotes the share

[21] As an example of a quasi-concave but not concave sender payoff, consider the case of the financial advisor and the client. The advisor's commission might be tied to cross-selling some of the firm's newer mutual funds over others. If one of those newer funds is a blended portfolio of stocks and bonds, the advisor's payoff might be quasi-concave, but not necessarily concave, in the client's chosen portfolio weight, with a peak at some $a \in (0,1)$ that has the client put some wealth in stocks and the remainder in bonds.

[22] As seen in the literature, optimal gerrymandering problems can be studied via a belief-based approach (e.g., Friedman and Holden 2008; Gul and Pesendorfer 2010; Kolotilin and Wolitzky 2023). As a result, quantile-based persuasion problems are also connected to gerrymandering when finding optimal or equilibrium election maps with only aggregate uncertainty.

of individuals who predict their own type is in $[z_{k-1}, z_k)$.

A prediction dataset $(\theta_k)_{k\in K}$ is said to be *median rationalizable (τ-quantile rationalizable)*, if there exists a signal for types x such that the induced posterior has a unique median (τ-quantile) with probability 1, and that for all $k \in \{1, \ldots, K\}$, the probability of the posterior median (τ-quantile) being in the interval $[z_{k-1}, z_k)$ is θ_k.[23] In other words, a prediction dataset $(\theta_k)_{k=1}^K$ is median (τ-quantile) rationalizable if there exists a Bayesian framework under which the share of individuals who predict $[z_{k-1}, z_k)$ equals θ_k when individuals are asked to predict their types based on the median (τ-quantiles) of their beliefs.[24] Theorem 1 (Theorem 4) of Benoît and Dubra (2011) characterizes the median (τ-quantile) rationalizable datasets. As we show below, this characterization can be derived immediately from Theorem 3.

Corollary 5 (Rationalizable Apparent Misconfidence)**.** *For any $\tau \in (0,1)$, for any $F \in \mathcal{F}_0$ with full support on $[0,1]$, and for any partition $0 = z_0 < z_1 < \ldots < z_K = 1$ of $[0,1]$, a prediction dataset $(\theta_k)_{k=1}^K$ is τ-quantile rationalizable if and only if for all $k \in \{1, \ldots, K\}$,*

$$\sum_{i=1}^{k} \theta_i < \frac{1}{\tau} F(z_k) \tag{4}$$

and

$$\sum_{i=k}^{K} \theta_i < \frac{1 - F(z_{k-1}^-)}{1-\tau} \tag{5}$$

Proof. For necessity, consider any $H \in \widetilde{\mathcal{H}}_\tau$ such that $H(z_k^-) - H(z_{k-1}^-) = \theta_k$ for all $k \in \{1, \ldots, K\}$. Then for any $k \in \{1, \ldots, K\}$, $\sum_{i=1}^k \theta_i = H(z_k^-)$. Since $H \in \widetilde{\mathcal{H}}_\tau$, there exists a signal $\mu \in \mathcal{M}$ for which μ-almost all posteriors have a unique τ-quantile and $H(z_k^-) = \mu(\{G \in \mathcal{F}_0 | G^{-1}(\tau) < z_k\}) = \mu(\{G \in \mathcal{F}_0 | \tau < G(z_k)\})$. Since $\mu \in \mathcal{M}$, $G(z_k)$ is a mean-preserving spread of $F(z_k)$ when $G \sim \mu$. Thus, $\mu(\{G \in \mathcal{F}_0 | \tau < G(z_k)\}) < F(z_k)/\tau$, and hence (4) holds. Analogous arguments can be applied to show that (5) holds as well.

For sufficiency, consider any prediction dataset $(\theta_k)_{k=1}^K$ such that (4) and (5) hold. Let H be the distribution that assigns probability θ_k at $(z_k + z_{k-1})/2$. Then, there exists $\varepsilon > 0$ such that $H \in \mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})$. By Theorem 3, there exists a signal μ with $\mu(\{G \in \mathcal{F}_0 | G^{-1}(\tau) <$

[23] Namely, $(\theta_k)_{k=1}^K$ is τ-quantile rationalizable if there exists $H \in \widetilde{\mathcal{H}}_\tau$ such that $H(z_k^-) - H(z_{k-1}^-) = \theta_k$. Technically speaking, Benoît and Dubra (2011) use a less stringent requirement regarding multiple quantiles. However, as shown below, Theorem 3 generalizes their conclusion even with this stringent requirement.

[24] Among experiments with clear instructions on how to make a prediction, the most common ones ask individuals to make predictions based on their posterior means or medians. When subjects use the posterior mean to predict their types, the set of rationalizable data would be given by mean-preserving contractions of the prior, which follows immediately from Strassen's theorem, as noted by Benoît and Dubra (2011). Therefore, the interesting characterization of rationalizable datasets would be when the individuals use other statistics to predict, such as posterior *medians* or *quantiles*.

$G^{-1}(\tau^+)\}) = 0$ such that $H(x) = H^\tau(x|\mu)$ for all $x \in \mathbb{R}$, which in turn implies that μ τ-quantile-rationalizes $(\theta_k)_{k=1}^K$, as desired. ∎

Remark 3. Benoît and Dubra (2011) further assume that $F(z_k) = {}^k\!/\!_K$ for all k (i.e., individuals are asked to place themselves into one of the equally populated K-ciles of the population). With this assumption, Corollary 5 specializes to Theorem 4 of Benoît and Dubra (2011), whose proof relies on projection and perturbation arguments and is not constructive. In addition to having a more straightforward proof and yielding a more general result, another benefit of Theorem 3 is that the signals rationalizing a feasible prediction dataset are semi-constructive: The extreme points of $\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})$ are attained by explicitly constructed signals, as shown in the proof of Theorem 3.[25]

4 Security Design with Limited Liability

In this second class of applications, we show how monotone function intervals pertain to security design with limited liability. Security design searches for optimal ways to divide the cash flows of assets across financial claims as a way to mitigate informational frictions. Monotone function intervals embed two widely adopted economic assumptions in the security design literature. The first is limited liability, which places natural upper and lower bounds on the security's payoff—namely, the asset's cash flow and zero, respectively. The second is that the security's payoff is monotone in the asset's cash flow. These two assumptions imply that all feasible securities can be described by monotone function intervals. Recognizing this, we use the second crucial property of extreme points—namely, for any convex optimization problem, one of the solutions must be an extreme point of the feasible set—to generalize and unify several results in security design under a common framework. To do so, we revisit the environments of two seminal papers in the literature: Innes (1990), which has moral hazard, and DeMarzo and Duffie (1999), which has adverse selection.

4.1 Security Design with Moral Hazard

A risk-neutral entrepreneur issues a security to an investor to fund a project. The project needs an investment $I > 0$. If the project is funded, the entrepreneur then exerts costly effort to develop the project. If the effort level is $e \in [0, \bar{e}]$, the project's profit is distributed

[25] It is also noteworthy that, although Theorem 4 of Benoît and Dubra (2011) can be used to prove Theorem 2 indirectly when F admits a density (by taking $K \to \infty$ and establishing proper continuity properties), the same argument cannot be used to prove Theorem 3, which is crucial for the proof of Corollary 5. This is because of the failure of upper-hemicontinuity when signals that induce multiple quantiles are excluded.

according to $\Phi(\cdot|e) \in \mathcal{F}_0$, and the (additively separable) effort cost to the entrepreneur is $C(e) \geq 0$.[26]

A security specifies the return to the investor for every realized profit $x \geq 0$ of the project. Both the entrepreneur and the investor have limited liability, and therefore, any security must be a (measurable) function $H : \mathbb{R}_+ \to \mathbb{R}$ such that $0 \leq H(x) \leq x$ for all $x \geq 0$. Moreover, a security is required to be monotone in the project's profit.[27] Given a security H, the entrepreneur chooses an effort level to solve

$$\sup_{e \in [0,\bar{e}]} \int_0^\infty (x - H(x))\Phi(\mathrm{d}x|e) - C(e). \tag{6}$$

For simplicity, we make the following technical assumptions: 1) The supports of the profit distributions $\{\Phi(\cdot|e)\}_{e\in[0,\bar{e}]}$ are all contained in a compact interval, which is normalized to $[0,1]$. 2) $\Phi(\cdot|e)$ admits a density $\phi(\cdot|e)$ for all $e \in [0,\bar{e}]$. 3) $\phi(x|e) > 0$ and is differentiable with respect to e for all $x \in [0,1]$ and for all $e \geq 0$, with its derivative, $\phi_e(x|e)$, dominated by an integrable function in absolute value. 4) $\{\Phi(\cdot|e)\}_{e\in[0,\bar{e}]}$ and C are such that (6) admits a solution, and every solution to (6) can be characterized by the first-order condition.[28]

The entrepreneur's goal is to design a security to acquire funding from the investor while maximizing the entrepreneur's expected payoff. Specifically, let $\overline{F}(x) := x$ and let $\underline{F}(x) := 0$ for all $x \in [0,1]$. The set of securities can be written as $\mathcal{I}(\underline{F}, \overline{F})$. The entrepreneur solves

$$\begin{aligned} &\sup_{H\in\mathcal{I}(\underline{F},\overline{F}),\, e\in[0,\bar{e}]} \left[\int_0^1 [x - H(x)]\phi(x|e)\,\mathrm{d}x - C(e)\right] \\ &\quad \text{s.t.} \quad \int_0^1 [x - H(x)]\phi_e(x|e)\,\mathrm{d}x = C'(e) \\ &\qquad\qquad \int_0^1 H(x)\phi(x|e)\,\mathrm{d}x \geq (1+r)I, \end{aligned} \tag{7}$$

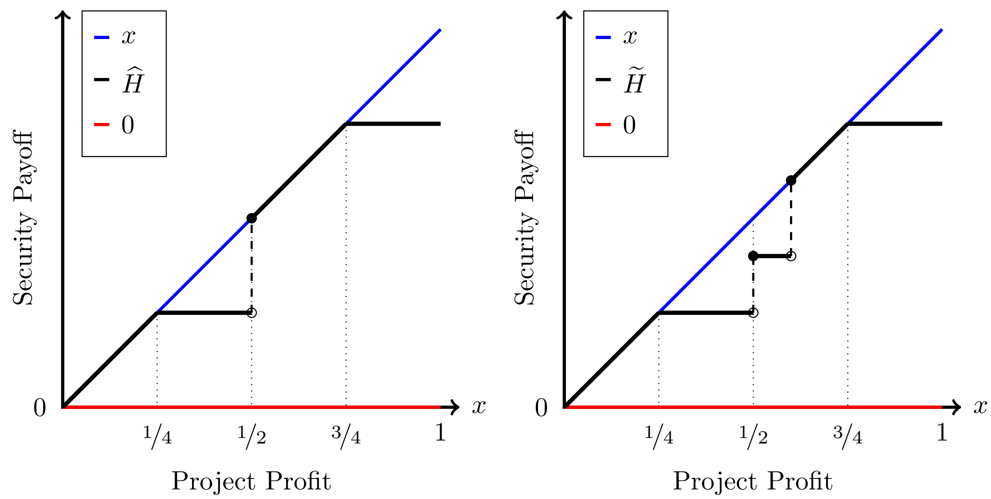
where $r > 0$ is the rate of return on a risk-free asset.

Innes (1990) characterizes the optimal security in this setting under an additional crucial

[26] The upper bound $\bar{e}$ can be arbitrarily large. Effort levels are bounded so that the entrepreneur's problem always admits a solution.

[27] Requiring securities to be monotone is a standard assumption in the security design literature (Innes 1990; Nachman and Noe 1994; DeMarzo and Duffie 1999). Monotonicity can be justified without loss of generality if the entrepreneur could contribute additional funds to the project so that only monotone profits would be observed.

[28] For example, we may assume that C is strictly increasing and strictly convex and that $\int_0^1 x \max\{\phi_{ee}(x|e), 0\}\,\mathrm{d}x < C''(e)$ for all e. Another sufficient condition would be $\phi_{ee}(1|e) < 2C''(e)$ and $\phi'_{ee}(x|e) \geq 0$ for all x and for all e. Moreover, if there are only finitely many effort levels available to the entrepreneur, the first-order approach would not be necessary to establish the results below, as suggested by Theorem A.1 in the Appendix.



(A) Face Values: 1/4 and 3/4 **(B) Defaultable Face Values: 1/4 and 3/4**

Figure VI
CONTINGENT DEBT CONTRACTS

assumption: The project profit distributions $\{\phi(\cdot|e)\}_{e\in[0,\bar{e}]}$ satisfy the monotone likelihood ratio property (Milgrom 1981). With this assumption, he shows that every optimal security must be a standard debt contract $H^d(x) := \min\{x, d\}$ for some face value $d > 0$. While the simplicity of a standard debt contract is a desirable feature, the monotone likelihood ratio property is arguably a strong condition (Hart 1995), where higher effort leads to higher probability weights on all higher project profits at any profit level. It remains unclear what the optimal security might be under a more general class of distributions.

Using Theorem 1, we can generalize Innes (1990) and solve the entrepreneur's problem (7) without the monotone likelihood ratio property. As we show in Proposition 3 below, *contingent* debt contracts are now optimal. We say that a security $H \in \mathcal{I}(\underline{F}, \overline{F})$ is a *contingent debt contract* if there exists a countable collection of intervals $\{[\underline{x}_n, \overline{x}_n)\}_{n=1}^N$, with $N \in \mathbb{N} \cup \{\infty\}$, such that H is constant on $[\underline{x}_n, \overline{x}_n)$, $H(x) = x$ for all $x \notin \cup_{n=1}^{\infty}[\underline{x}_n, \overline{x}_n)$, and that $H(\underline{x}_n) \neq H(\underline{x}_m)$ for all $n \neq m$. In other words, a contingent debt contract H has N possible face values, $\{d_n = H(\underline{x}_n)\}_{n=1}^N$, where the entrepreneur pays face value $d_n = H(\underline{x}_n)$ if the project's profit falls in $[\underline{x}_n, \overline{x}_n)$, and defaults otherwise.

Figure VIA illustrates a contingent debt contract $\widehat{H}$ with $N = 2$, $[\underline{x}_1, \overline{x}_1) = [1/4, 1/2)$, $[\underline{x}_2, \overline{x}_2) = [3/4, 1)$, $d_1 = 1/4$, and $d_2 = 3/4$. Under $\widehat{H}$, if the project's profit x is below $1/2$, the entrepreneur owes debt with face value $1/4$; instead, if the profit is above $1/2$, the entrepreneur owes debt with a higher face value $3/4$. The entrepreneur's required debt payment to the investor is contingent on the entrepreneur's capacity to pay, which itself is linked to the realized profit of the project.

Clearly, every standard debt contract with face value d is a contingent debt contract where $N = 1$. Moreover, a contingent debt contract never involves the entrepreneur and investor sharing in the equity of the project (i.e., the derivative of H, whenever defined, must be either 0 or 1). To see how the cash flow is split between parties, consider the contingent debt contract $\widehat{H}$ depicted by Figure VIA. Suppose the project earned $x \in (1/2, 3/4)$. The entrepreneur would default on the high face-value debt contract ($d_2 = 3/4$), and the investor would take claim of all project profits x. If, rather, the project earned $x \in (1/4, 1/2)$, the investor would receive the low face-value amount ($d_1 = 1/4$), and the entrepreneur would retain the amount $x - 1/4$. In general, under any contingent debt contract, either the entrepreneur defaults and the investor absorbs all rights to the project's worth, or the entrepreneur pays a certain face value and retains the residual profit.

Contingent debt contracts are similar in spirit to many fixed-income securities observed in practice. The first are state-contingent debt instruments (SCDIs) from the sovereign debt literature, which link a country's principal or interest payments to its nominal GDP (Lessard and Williamson 1987; Shiller 1994; Borensztein and Mauro 2004). The second are versions of contingent convertible bonds (CoCos) issued by corporations, which write down the bond's face value after a triggering event like financial distress (Albul, Jaffee and Tchistyi 2015; Oster 2020). The third are commodity-linked bonds common to mineral companies and resource-rich developing countries, which tie the amount paid at maturity to the market value of a reference commodity like silver (Lessard 1977; Schwartz 1982).

Using Theorem 1, we show that a natural class of contingent debt contracts is optimal. Given a contingent debt contract H with face values $\{d_n = H(\underline{x}_n)\}_{n=1}^{\infty}$, a face value d_n is said to be *non-defaultable* if $d_n < x$, for all $x \in [\underline{x}_n, \overline{x}_n)$.[29] That is, a face value d_n is non-defaultable if, conditional on d_n being in effect, the project's profit is always higher than that face value and the entrepreneur always retains some residual surplus after paying off d_n. Figure VIB illustrates a contingent debt contract $\widetilde{H}$ with three possible face values, $1/4$, $1/2$, and $3/4$. Here, the face values $1/4$ and $3/4$ are defaultable, whereas the face value $1/2$ is non-defaultable.

Proposition 3. *There is a contingent debt contract with at most two non-defaultable face values that solves the entrepreneur's problem* (7).

According to Proposition 3, even without the MLRP assumption, the nature of standard debt contracts, which allocates any additional dollar of the project's profit either fully to the entrepreneur or to the investor, is preserved even without the monotone likelihood ratio

[29]This is equivalent to saying that $d_n = H(\underline{x}_n) < \underline{x}_n$.

assumption. Nonetheless, the entrepreneur may be liable for more when the project earns more.[30]

The proof of Proposition 3 can be found in Appendix A.6. In essence, since the entrepreneur's objective in (7) is affine and the set of feasible contracts is convex, there must exist an extreme point of the feasible set that solves (7). Thus, it suffices to show that any extreme point of the feasible set must correspond to a contingent debt contract with at most two non-defaultable face values. To this end, first note that, by Proposition 2.1 of Winkler (1988), any extreme point H^* of the feasible set of (7) can be written as convex combinations of at most three extreme points of $\mathcal{I}(\underline{F}, \overline{F})$. Then, by Theorem 1, whenever $H^*(x_0) < x_0$ for some $x_0 \in (0,1)$, there must be an interval $[\underline{x}_0, \overline{x}_0)$ containing x_0 such that $H^*(x) < x$ for all $x \in [\underline{x}_0, \overline{x}_0)$. Moreover, H^* must be constant on any such an interval, since otherwise H^* must be strictly increasing and affine. A similar argument as in the proof of Theorem 1 can then be applied to show that such H^* can be written as a convex combination of two distinct step functions on the interval $[\underline{x}_0, \overline{x}_0]$. Therefore, any extreme point H^* of the feasible set of (7) must be such that, for any $x \in (0,1)$, either $H^*(x) = x$ or H^* is constant on an interval that contains x, which implies H^* is a contingent debt contract. Finally, if the contingent debt contract H^* has more than two non-defaultable face values, then there must be a way to linearly perturb H^* only on these non-defaultable face values while keeping the perturbation feasible. Therefore, for H^* to be an extreme point of the feasible set, it must be a contingent debt contract with at most two non-defaultable face values.[31]

To better understand Proposition 3, recall that the optimality of standard debt contracts in Innes (1990) is due to (i) the risk-neutrality and the limited-liability structure of the problem, and (ii) the monotone likelihood ratio property of the profit distributions. Indeed, for any incentive-compatible and individually-rational contract, risk neutrality allows one to construct a standard debt contract with the same expected payment. Meanwhile, the mono-

[30]Although contingent debt contracts include all step functions in $\mathcal{I}(\underline{F}, \overline{F})$ and thus can approximate any other contract in $\mathcal{I}(\underline{F}, \overline{F})$, such as an equity contract $H(x) = \alpha x$ for some $\alpha \in (0,1)$, according to Proposition 3, there is an *optimal* contract H^* belonging to a family of contingent debts contracts that *does not* approximate any such contracts, since there are at most two intervals on which H^* is constant and is strictly below x.

[31]The observation of Winkler (1988) bounds the number of elements involved in randomization by the number of additional linear constraints when solving a constrained convex problem. In mechanism design, this observation is equivalent to the optimality of randomized posted price mechanisms when side constraints are present (see, e.g., Samuelson 1984; Loertscher and Muir 2022; Kang 2022). In information design, this observation is related to bounds on the cardinality of the support of optimal signals of information design problems with side constraints (Doval and Skreta forthcoming). The structure of monotone function intervals allows us to sharpen this general observation here and bound the number of intervals on which an extreme point H^* is constant but does not reach the upper and lower bounds of the monotone function interval (non-defaultable face values in this setting), instead of bounding the number of extreme points involved in the randomization (see Candogan and Strack 2023 and Nikzad 2023 for a similar discovery regarding second-order stochastic dominance and majorization).

tone likelihood ratio property ensures that this debt contract incentivizes the entrepreneur to exert higher effort, leading to a higher expected project profit. Without the monotone likelihood ratio assumption, simply replicating an individually-rational contract with a standard debt contract may distort incentives and lead to less efficient effort and suboptimal outcomes. In this regard, Proposition 3 shows that contingent debt contracts are enough to replicate the profit level of all other feasible contracts while preserving incentive compatibility and individual rationality. In essence, the proposition separates the effects of risk neutrality and limited liability on security design from the effects of the monotone likelihood ratio property.

Additional assumptions on the project's profit distributions $\{\Phi(\cdot|e)\}_{e\in[0,\bar{e}]}$, can further simplify the structure of the optimal contracts. For any $N \in \mathbb{N}$ and for any $e \in [0, \bar{e}]$, we say that the function $\phi_e(\cdot|e)/\phi(\cdot|e)$ is *N-peaked* if there exists N disjoint intervals $\{I_n\}_{n=1}^N$ in $[0, 1]$ such that $\phi_e(x|e)/\phi(x|e)$ is increasing in x on I_n for all $n \in \{1, \dots, N\}$, and is decreasing in x on $[0, 1]\backslash \cup_{n=1}^N I_n$. Note that if $\phi_e(\cdot|e)/\phi(\cdot|e)$ is increasing on $[0, 1]$, then it is N-peaked with $N = 1$. In particular, profit distributions that satisfy MLRP are N-peaked with $N = 1$.

Furthermore, assume that the risk-free rate of return r and the required investment I are such that $(1 + r)I$ is in the interior of the set

$$\left\{ \int_0^1 H(x)\phi(x|e)\,\mathrm{d}x \,\middle|\, H \in \mathcal{I}(\underline{F}, \overline{F}),\ \int_0^1 (x - H(x))\phi_e(x|e)\,\mathrm{d}x = C'(e) \right\}, \tag{8}$$

for all $e \in [0, \bar{e}]$.[32]

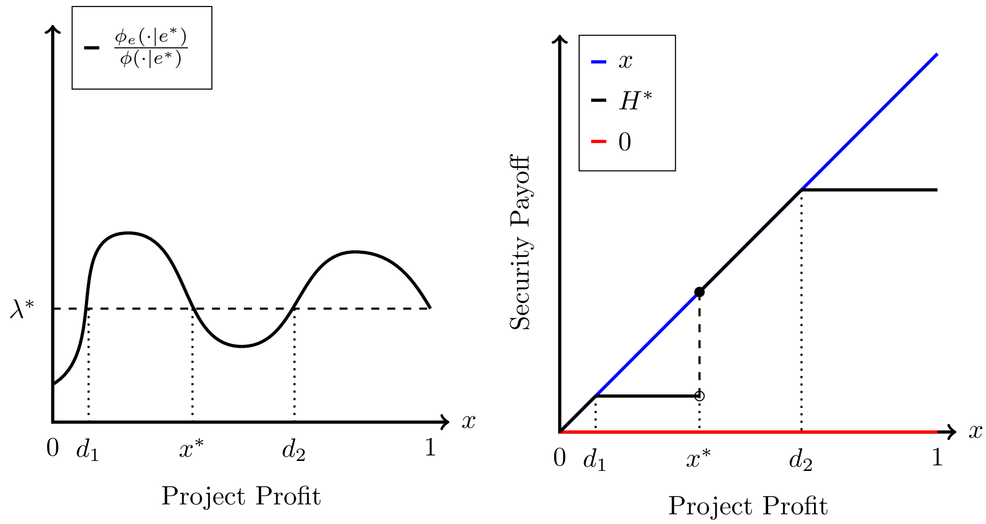
By establishing strong duality of the entrepreneur's problem (7), together with Theorem 1, Proposition 4 below identifies a sufficient condition for there to be an optimal contingent debt contract with finitely many face values.

Proposition 4. *Suppose that there exists $N \in \mathbb{N}$ such that for any $e \in [0, \bar{e}]$, the function $\phi_e(\cdot|e)/\phi(\cdot|e)$ is at most N-peaked. Then there is a contingent debt contract with at most $N+1$ face values (with at most two of them being non-defaultable) that solves the entrepreneur's problem* (7)*.*

The proof of Proposition 4 can be found in Appendix A.7. The essence of the proof is the following observation: Under (8), strong duality holds for the entrepreneur's problem (7). Thus, for the optimal effort level $e^* \in [0, \bar{e}]$, an optimal security H^* must also solve

$$\sup_{H\in\mathcal{I}(\underline{F},\overline{F})} \left[\int_0^1 H(x)[(1 + \lambda_2^*)\phi(x|e^*) - \lambda_1^*\phi_e(x|e^*)]\,\mathrm{d}x \right], \tag{9}$$

[32] This a technical assumption to ensure strong duality and serves a similar role as Slater's condition in finite-dimensional problems.



(A) **2-peaked** $\frac{\phi_e(\cdot|e)}{\phi(\cdot|e)}$ **(B)** **Optimal Security** H^*

Figure VII
OPTIMAL CONTINGENT DEBT WITH 2 FACE VALUES

where $\lambda_1^* \neq 0, \lambda_2^* \geq 0$ are the Lagrange multipliers for the incentive compatibility and individually rationality constraints, respectively. Since

$$(1+\lambda_2^*)\phi(x|e^*) - \lambda_1^*\phi_e(x|e^*) \geq 0 \iff \frac{\phi_e(x|e^*)}{\phi(x|e^*)} \leq \frac{1+\lambda_2^*}{\lambda_1^*} =: \lambda^*,$$

and since $\phi_e(\cdot|e^*)/\phi(\cdot|e^*)$ is at most N-peaked, the set of profits x under which $\phi_e(x|e^*)/\phi(x|e^*)$ is greater than or smaller than λ^* must form an interval partition with at most $2N$ elements, as depicted in Figure VIIA. It can then be shown that, for a contingent debt contract H^* to be optimal, H^* cannot take two distinct values on any element where $\phi_e(x|e^*)/\phi(x|e^*) > \lambda^*$, and that $H^*(x) = x$ whenever $\phi_e(x|e^*)/\phi(x|e^*) < \lambda^*$, as depicted in Figure VIIB. Thus, there must be at most $N+1$ partition elements on which H^* is constant, and hence H^* must be a contingent debt contract with at most $N+1$ face values.[33]

According to Proposition 4, if the project's profit distributions $\{\Phi(\cdot|e)\}_{e\in[0,\bar{e}]}$ satisfy the regularity condition so that $\phi_e(\cdot|e)/\phi(\cdot|e)$ is at most N-peaked for all e, then not only would a contingent debt contract be optimal for the entrepreneur, that contract would be simple, in that it would have at most finitely many face values.

[33]Using this argument, the optimality of standard debt contracts under MLRP also follows immediately, as MLRP would imply that $\phi_e(x|e)/\phi(x|e) < \lambda^*$ if and only if $x < d$ for some d.

4.2 Security Design with Adverse Selection

There is a risk-neutral security issuer with discount rate $\delta \in (0,1)$ and a unit mass of risk-neutral investors. The issuer has an asset that generates a random cash flow $x \geq 0$ in period $t = 1$. The cash flow is distributed according to $\Phi_0 \in \mathcal{F}_0$, which is supported on a compact interval normalized to $[0,1]$. Because $\delta < 1$, the issuer has demand for liquidity in period $t = 0$ and therefore has an incentive to sell a limited-liability security backed by the asset to raise cash. A security is a nondecreasing, right-continuous function $H : [0,1] \to \mathbb{R}_+$ such that $0 \leq H(x) \leq x$ for all x. Let $\overline{F}(x) := x$ and $\underline{F}(x) := 0$ for all $x \in [0,1]$. The set of securities can again be written as $\mathcal{I}(\underline{F}, \overline{F})$.

Given any security $H \in \mathcal{I}(\underline{F}, \overline{F})$, the issuer first observes a signal $s \in S$ for the asset's cash flow. Then, taking as given an inverse demand schedule $P : [0,1] \to \mathbb{R}_+$, she chooses a fraction $q \in [0,1]$ of the security to sell. If a fraction q of the security is sold and the signal realization is s, the issuer's expected return is

$$\underbrace{qP(q)}_{\text{revenue raised in } t=0} + \delta \cdot \underbrace{\mathbb{E}[x - qH(x)|s]}_{\text{residual return in } t=1} = q(P(q) - \delta\mathbb{E}[H(x)|s]) + \delta\mathbb{E}[x|s].$$

Investors observe the quantity q, update their beliefs about x, and decide whether to purchase.

DeMarzo and Duffie (1999) show that, in the unique equilibrium that survives the D1 criterion,[34] the issuer's profit under a security H, when the posterior expected value of the security is $\mathbb{E}[H(x)|s] = z$, is given by

$$\Pi(z|H) := (1-\delta) z_0^{\frac{1}{1-\delta}} z^{-\frac{\delta}{1-\delta}},$$

where z_0 is the lower bound of the support of $\mathbb{E}[H(x)|s]$. Therefore, let $\Phi(\cdot|s)$ be the conditional distribution of the cash flow x given signal s, and let $\Psi : S \to [0,1]$ be the marginal distribution of the signal s. The expected value of a security H is then

$$\Pi(H) := (1-\delta)\left(\inf_{s \in S} \int_0^1 H(x)\Phi(\mathrm{d}x|s)\right)^{\frac{1}{1-\delta}} \int_S \left(\int_0^1 H(x)\Phi(\mathrm{d}x|s)\right)^{-\frac{\delta}{1-\delta}} \Psi(\mathrm{d}s).$$

As a result, the issuer's security design problem can be written as

$$\sup_{H \in \mathcal{I}(\underline{F}, \overline{F})} \Pi(H).$$

[34] An equilibrium in this market is a pair (P, Q) of measurable functions such that $Q(\mathbb{E}[H(x)|s])(P \circ Q(\mathbb{E}[H(x)|s]) - \delta\mathbb{E}[H(x)|s]) \geq q(P(q) - \delta\mathbb{E}[H(x)|s])$ for all $q \in [0,1]$ with probability 1, and $P \circ Q(\mathbb{E}[H(x)|s]) = \mathbb{E}[H(x)|Q(\mathbb{E}[H(x)|s])]$ with probability 1.

Using a variational approach, DeMarzo and Duffie (1999) characterize several general properties of the optimal securities without solving for them explicitly. They then specialize the model by assuming that the signal structure $\{\Phi(\cdot|s)\}_{s\in S}$ has a *uniform worst case*, a condition slightly weaker than the monotone likelihood ratio property that requires the cash flow distribution to be smallest in the sense of fist-order stochastic dominance (FOSD) under some s_0, conditional on every interval I of $[0,1]$.[35] With this assumption, DeMarzo and Duffie (1999) show that a standard debt contract $H^d(x) := \min\{x, d\}$ is optimal.

With Theorem 1, we are able to generalize this result and solve for an optimal security while relaxing the uniform-worst-case assumption. Instead of a uniform worst case, we only assume that there is a worst signal s_0 such that $\Phi(\cdot|s)$ dominates $\Phi(\cdot|s_0)$ in the sense of FOSD for all $s \in S$. With this assumption, the issuer's security design problem can be written as

$$\begin{aligned} \sup_{H\in\mathcal{I}(\underline{F},\overline{F}),\underline{z}\in[0,\mathbb{E}[x|s_0]]} & \left[(1-\delta)\underline{z}^{\frac{1}{1-\delta}}\int_S\left(\int_0^1 H(x)\Phi(\mathrm{d}x|s)\right)^{-\frac{\delta}{1-\delta}}\Psi(\mathrm{d}s)\right] \\ \text{s.t.} & \int_0^1 H(x)\Phi(\mathrm{d}x|s_0) = \underline{z}. \end{aligned} \tag{10}$$

As shown by Proposition 5 below, a particular class of contingent debt contracts is always sufficient for the issuer to consider.

Proposition 5. *There is a contingent debt contract with at most one non-defaultable face value that solves the issuer's problem* (10)*. Furthermore, if $\Phi(\cdot|s)$ has full support on $[0,1]$ for all $s \in S$, this solution is unique.*

Overall, this section showcases the unifying role of the extreme points of monotone function intervals in security design. The security design literature has rationalized the existence of different financial securities observed in practice under a variety of economic environments and assumptions. Doing so has strengthened the robustness of these securities as optimal contracts. But that variety also makes it hard to sort the essential modeling ingredients from the inessential ones. And the core features that connect these environments are not readily apparent.

An advantage of recasting the set of feasible securities as a monotone function interval is that it strips the problem down to its basic elements. Whether the setting has moral hazard or adverse selection, and whether the asset's cash flow distributions exhibit MLRP, are not

[35]Specifically, they assume that there exists some $s_0 \in S$ such that, for any $s \in S$ and for any interval $I \subset [0,1]$, (i) $\Phi(I|s_0) = 0$ implies $\Phi(I|s) = 0$, and (ii) the conditional distribution of the asset's cash flow given signal realization s and given that the cash flow falls in an interval I, which is denoted $\Phi|_I(\cdot|s)/\Phi(I|s)$, dominates the conditional distribution given signal realization s_0, denoted $\Phi|_I(\cdot|s_0)/\Phi(I|s_0)$, in the sense of FOSD.

defining. Limited liability, monotone contracts, and risk neutrality are the core elements that deliver debt as an optimal security. The terms of the debt contract somewhat differ from those of a standard one, as the face value is now contingent on the asset's cash flow, but the nature of debt contracts, which never has the issuer and investor share in the asset's equity and grants the issuer only residual rights, still prevails.

Without knowledge of the extreme points of monotone function intervals, solving the security design problem without the MLRP assumption would have been substantially harder. Thus, just as in the other economic applications of this paper, Theorem 1 offers a unified approach to answering classic economic questions that have been previously answered by case-specific approaches.

5 Conclusion

We characterize the extreme points of monotone function intervals and apply this result to various economic problems. We show that any extreme point of a monotone function interval must either coincide with one of the montone function interval's bounds, or be constant on an interval in its domain, where at least one end of the interval reaches one of the bounds. Using this result, we characterize the distributions of posterior quantiles, which coincide with a monotone function interval. We apply this insight to topics in political economy, Bayesian persuasion, and the psychology of judgment. Furthermore, monotone function intervals provide a common structure to security design. We unify and generalize seminal results in that literature when either adverse selection or moral hazard afflicts the environment.

It is worthwhile acknowledging the paper's limitations. Regarding the distributions of posterior quantiles, the analysis is restricted to a one-dimensional state space. Moreover, while the characterization parallels the well-known characterization of distributions of posterior means, it provides little intuition for how distributions of other statistics (say, the posterior k-th moment) may behave. In particular, while the characterization of distributions of posterior quantiles allows one to compare Bayesian persuasion problems when the receiver has either an absolute loss function or a quadratic loss function, optimal signals under other loss functions remain largely under-explored.

Regarding security design, the clearest limitation is the absence of risk aversion. This is due to the lack of convexity of the objectives and constraints of security design problems with risk averse agents. The majority of the security design literature features risk neutral agents, and this risk neutrality makes the design problem amenable to being analyzed using extreme points of monotone function intervals. Nevertheless, security design with risk averse agents has gotten less attention among researchers and deserves further study. Allen and Gale (1988); Malamud, Rui and Whinston (2010); Gershkov, Moldovanu, Strack and Zhang (2023) study the problem and provide many intriguing results thus far.

Notwithstanding these limitations, other applications involving monotone function intervals undoubtedly exist. For instance, their link to the distributions of posterior quantiles opens many potential research avenues. For instance, inequality is often measured as an upper percentile of the wealth or income distribution, making it eligible for analysis. Likewise, settings in which the feasible set can be represented as a monotone function interval, such as R&D investments and screening problems with stochastic inventories, are yet other directions for future work.

References

Albul, B., D. M. Jaffee, and A. Tchistyi (2015) "Contingent Convertible Bonds and Capital Structure Decisions," Working Paper.

Ali, S. N., N. Haghpanah, X. Lin, and R. Siegel (2022) "How to Sell Hard Information," *Quarterly Journal of Economics*, 137 (1), 619–678.

Allen, F. and A. Barbalau (2022) "Security Design: A Review," Working Paper.

Allen, F. and D. Gale (1988) "Optimal Security Design," *Review of Financial Studies*, 1 (3), 229–263.

Anderson, E. J. and P. Nash (1987) *Linear Programming in Infinite-Dimensional Space*: Wiley.

Arieli, I., Y. Babichenko, R. Smorodinsky, and T. Yamashita (2023) "Optimal Persuasion via Bi-Pooling," *Theoretical Economics*, 18 (1), 15–36.

Benoît, J.-P. and J. Dubra (2011) "Apparent Overconfidence," *Econometrica*, 79 (5), 1591–625.

Bergemann, D., B. Brooks, and S. Morris (2015) "The Limits of Price Discrimination," *American Economic Review*, 105 (3), 921–957.

Biais, B. and T. Mariotti (2005) "Strategic Liquidity Supply and Security Design," *Review of Economic Studies*, 72 (3), 615–649.

Black, D. (1958) *The Theory of Committees and Elections*: Springer.

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

Borensztein, E. and P. Mauro (2004) "The Case for GDP-Indexed Bonds," *Economic Policy*, 19 (38), 166–216.

Börgers, T. (2015) *An Introduction to the Theory of Mechanism Design*: Oxford University Press.

Bradbury, J. C. and W. M. Crain (2005) "Legislative District Configurations and Fiscal Policy in American States," *Public Choice*, 125 (3), 385–407.

Buchler, J. (2005) "Competition, Representation and Redistricting: The Case Against Competitive Congressional Districts," *Journal of Theoretical Politics*, 17 (4), 431–463.

Candogan, O. and P. Strack (2023) "Optimal Disclosure of Information to Privately Informed Agents," *Theoretical Economics*, 18, 1225–1269.

Casamatta, C. (2003) "Financing and Advising: Optimal Financial Contracts with Venture Capitalists," *Journal of Finance*, 58 (5), 2059–2085.

de Castro, L. and A. F. Galvao (2021) "Static and Dynamic Quantile Preferences," *Economic Theory*, 1–33.

Chambers, C. P. (2007) "Ordinal Aggregation and Quantiles," *Journal of Economic Theory*, 137 (1), 416–431.

Cho, S.-j. and J. Duggan (2009) "Bargaining Foundations of the Median Voter Theorem," *Journal of Economic Theory*, 144 (2), 851–868.

Çinlar, E. (2010) *Probability and Stochastics*: Springer.

DeMarzo, P. and D. Duffie (1999) "A Liquidity-Based Model of Security Design," *Econometrica*, 67 (1), 65–99.

Doval, L. and V. Skreta (forthcoming) "Constrained Information Design," *Mathematics of Operations Research*.

Downs, A. (1957) *An Economic Theory of Democracy*: Harper & Row New York.

Dworczak, P. and G. Martini (2019) "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, 127, 1993–2048.

EISFELDT, A. L. (2004) "Endogenous Liquidity in Asset Markets," *Journal of Finance*, 59 (1), 1–30.

ELY, J. C. (2019) "A Cake-Cutting Solution to Gerrymandering," Working Paper.

FARMER, J. D. (1994) "Extreme Points of the Unit Ball of the Space of Lipschitz Functions," *Proceedings of the American Mathematical Society*, 121 (3), 807–813.

FRIEDMAN, J. N. AND R. HOLDEN (2020) "Optimal Gerrymandering in a Competitive Environment," *Economic Theory Bulletin*, 8 (2), 347–367.

FRIEDMAN, J. N. AND R. T. HOLDEN (2008) "Optimal Gerrymandering: Sometimes Pack, but Never Crack," *American Economic Review*, 98 (1), 113–44.

GENTZKOW, M. AND E. KAMENICA (2016) "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review: Papers and Proceedings*, 106 (5), 597–601.

GERSHKOV, A., B. MOLDOVANU, P. STRACK, AND M. ZHANG (2023) "Optimal Security Design for Risk-Averse Investors," Working Paper.

GILLIGAN, T. W. AND J. G. MATSUSAKA (2006) "Public Choice Principles of Redistricting," *Public Choice*, 129 (3), 381–398.

GOMBERG, A., R. PANCS, AND T. SHARMA (2023) "Electoral Maldistricting," *International Economic Review*, 64 (3).

GUL, F. AND W. PESENDORFER (2010) "Strategic Redistricting," *American Economic Review*, 100 (4), 1616–41.

HARDY, G., J. E. LITTLEWOOD, AND G. PÓLYA (1929) "Some Simple Inequalities Satisfied by Convex Functions," *Messenger Math*, 58 (4), 145–152.

HARSANYI, J. C. (1967-68) "Games of Incomplete Information Played by Bayesian Players, Part I, II, III," *Management Science*, 14, 159–182, 320–334, 486–502.

HART, O. (1995) *Firms, Contracts, and Financial Structure*: Clarendon Press.

INNES, R. D. (1990) "Limited Liability and Incentive Contracting with Ex-ante Action Choices," *Journal of Economic Theory*, 52 (1), 45–67.

KAMENICA, E. (2019) "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11, 249–272.

Kamenica, E. and M. Gentzkow (2011) "Bayesian Persuasion," *American Economic Review*, 101 (6), 2560–2615.

Kang, Z. Y. (2022) "The Public Option and Optimal Redistribution," Working Paper.

Kleiner, A., B. Moldovanu, and P. Strack (2021) "Extreme Points and Majorization: Economic Applications," *Econometrica*, 89 (4), 1557–1593.

Kolotilin, A., R. Corrao, and A. Wolitzky (2022a) "Persuasion with Nonlinear Preferences," Working Paper.

Kolotilin, A., M. Li, T. Mylovanov, and A. Zapechelnyuk (2017) "Persuasion of a Privately Informed Receiver," *Econometrica*, 85 (6), 1949–1964.

Kolotilin, A., T. Mylovanov, and A. Zapechelnyuk (2022b) "Censorship as Optimal Persuasion," *Theoretical Economics*, 17 (2), 561–585.

Kolotilin, A. and A. Wolitzky (2023) "The Economics of Partisan Gerrymandering," Working Paper.

Krehbiel, K. (2010) *Pivotal Politics*: University of Chicago Press.

Lessard, D. (1977) "Commodity-Linked Bonds from Less-Developed Countries: An Investment Opportunity?" *Working Paper*.

Lessard, D. R. and J. Williamson (1987) *Capital Flight and the Third World Debt*: Institute for International Economics.

Lipnowski, E. and L. Mathevet (2018) "Disclosure to a Psychological Audience," *American Economic Journal: Microeconomics*, 10 (4), 67–93.

Loertscher, S. and E. Muir (2022) "Monopoly Pricing, Optimal Randomization, and Resale," *Journal of Political Economy*, 130 (3), 566–635.

Malamud, S., H. Rui, and A. B. Whinston (2010) "Optimal Securitization with Heterogeneous Investors," Working Paper.

Manski, C. F. (1988) "Ordinal Utility Models of Decision Making under Uncertainty," *Theory and Decision*, 25 (1), 79–104.

McCarty, N., K. T. Poole, and H. Rosenthal (2001) "The Hunt for Party Discipline in Congress," *American Political Science Review*, 95 (3), 673–687.

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

Nachman, D. C. and T. H. Noe (1994) “Optimal Design of Securities under Asymmetric Information,” *Review of Financial Studies*, 7 (1), 1–44.

Nikzad, A. (2023) “Constrained Majorization: Applications in Mechanism Design,” Working Paper.

Oster, P. (2020) “Contingent Convertible Bond Literature Review: Making Everything and Nothing Possible?” *Journal of Banking Regulation*, 21, 343–381.

Owen, G. and B. Grofman (1988) “Optimal Partisan Gerrymandering,” *Political Geography Quarterly*, 7 (1), 5–22.

Pegden, W., A. D. Procaccia, and D. Yu (2017) “A Partisan Districting Protocol with Provably Nonpartisan Outcomes,” Working Paper.

Roesler, A.-K. and B. Szentes (2017) “Buyer-Optimal Learning and Monopoly Pricing,” *American Economic Review*, 107 (7), 2072–2080.

Rolewicz, S. (1984) “On Optimal Observability of Lipschitz Systems.,” in Hammer, G. and D. Pallaschke eds. *Selected Topics in Operations Research and Mathematical Economics: Proceedings of the 8th Symposium on Operations Research*, 152–158, Berlin: Springer.

——— (1986) “On Extremal Points of the Unit Ball in the Banach Space of Lipschitz Continuous Functions,” *Journal of the Australian Mathematical Society*, 41 (1), 95–98.

Rostek, M. (2010) “Quantile Maximization in Decision Theory,” *Review of Economic Studies*, 77 (1), 339–371.

Ryff, J. V. (1967) “Extreme Points of Some Convex Subsets of $L^1(0,1)$,” *Proceedings of the American Mathematical Society*, 18 (6), 1026–1034.

Samuelson, W. (1984) “Bargaining under Asymmetric Information,” *Econometrica*, 52 (4), 995–1005.

Schmidt, K. M. (1997) “Managerial Incentives and Product Market Competition,” *Review of Economic Studies*, 64 (2), 191–213.

Schwartz, E. S. (1982) “The pricing of commodity-linked bonds,” *The journal of Finance*, 37 (2), 525–539.

SHILLER, R. J. (1994) *Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks*: OUP Oxford.

SHOTTS, K. W. (2002) "Gerrymandering, Legislative Composition, and National Policy Outcomes," *American Journal of Political Science*, 398–414.

SKRETA, V. (2006) "Sequentially Optimal Mechanisms," *Review of Economic Studies*, 73 (4), 1085–1111.

SMARZEWSKI, R. (1997) "Extreme Points of Unit Balls in Lipschitz Function Spaces," *Proceedings of the American Mathematical Society*, 125 (5), 1391–1397.

STRASSEN, V. (1965) "The Existence of Probability Measures with Given Marginals," *Annuals of Mathematical Statistics*, 36, 423–439.

WINKLER, G. (1988) "Extreme Points of Moment Sets," *Mathematics of Operations Research*, 13 (4), 581–587.

YANG, K. H. (2020) "A Note on Generating Arbitrary Joint Distributions Using Partitions," Working Paper.

YANG, K. H. AND A. K. ZENTEFIS (2022) "Distributions of Posterior Quantiles and Economic Applications," Working Paper.

Appendix

A.1 Proof of Theorem 1

Consider any $\overline{F}, \underline{F}, H \in \mathcal{F}$ such that $\underline{F}(x) \leq H(x) \leq \overline{F}(x)$ for all $x \in \mathbb{R}$. We first show that if H satisfies 1 and 2 for a countable collection of intervals $\{[\underline{x}_n, \overline{x}_n)\}_{n=1}^{\infty}$, then H must be an extreme point of $\mathcal{I}(\underline{F}, \overline{F})$. To this end, first note that $\mathcal{I}(\underline{F}, \overline{F}) \subseteq \mathcal{F}$ is a convex subset of the collection of Borel-measurable functions on $\mathbb{R}$. Since the collection of Borel-measurable functions on $\mathbb{R}$ is a real vector space, it suffices to show that for any Borel-measurable $\widehat{H}$ with $\widehat{H} \neq 0$, either $H + \widehat{H} \notin \mathcal{I}(\underline{F}, \overline{F})$ or $H - \widehat{H} \notin \mathcal{I}(\underline{F}, \overline{F})$. Clearly, if either $H + \widehat{H} \notin \mathcal{F}$ or $H - \widehat{H} \notin \mathcal{F}$, then it must be that either $H + \widehat{H} \notin \mathcal{I}(\underline{F}, \overline{F})$ or $H - \widehat{H} \notin \mathcal{I}(\underline{F}, \overline{F})$. Thus, we may suppose that both $H + \widehat{H}$ and $H - \widehat{H}$ are in $\mathcal{F}$. Now notice that since $\widehat{H} \neq 0$, there exists $x_0 \in \mathbb{R}$ such that $\widehat{H}(x_0) \neq 0$. If $x_0 \notin \cup_{n=1}^{\infty} [\underline{x}_n, \overline{x}_n)$, then $H(x_0) \in \{\underline{F}(x_0), \overline{F}(x_0)\}$ and hence either $H(x_0) + |\widehat{H}(x_0)| > \overline{F}(x_0)$ or $H(x_0) - |\widehat{H}(x_0)| < \underline{F}(x_0)$. Thus, it must be that either $H + \widehat{H} \notin \mathcal{I}(\underline{F}, \overline{F})$ or $H - \widehat{H} \notin \mathcal{I}(\underline{F}, \overline{F})$. Meanwhile, if $x_0 \in [\underline{x}_n, \overline{x}_n)$ for some $n \in \mathbb{N}$, then $\widehat{H}$ must be constant on $[\underline{x}_n, \overline{x}_n)$ as H is constant on $[\underline{x}_n, \overline{x}_n)$ and both $H + \widehat{H}$ and $H - \widehat{H}$ are nondecreasing. Thus, either $H(\underline{x}_n) + |\widehat{H}(\underline{x}_n)| = \overline{F}(\underline{x}_n) + |\widehat{H}(x_0)| > \overline{F}(\underline{x}_n)$, or $H(\overline{x}_n^-) - |\widehat{H}(\overline{x}_n^-)| = \underline{F}(\overline{x}_n^-) - |\widehat{H}(x_0)| < \underline{F}(\overline{x}_n^-)$, and hence either $H + \widehat{H} \notin \mathcal{I}(\underline{F}, \overline{F})$ or $H - \widehat{H} \notin \mathcal{I}(\underline{F}, \overline{F})$, as desired.

Conversely, suppose that H is an extreme point of $\mathcal{I}(\underline{F}, \overline{F})$. To show that H must satisfy 1 and 2 for some countable collection of intervals $\{[\underline{x}_n, \overline{x}_n)\}_{n=1}^{\infty}$, we first claim that if $\underline{F}(x_0^-) < H(x_0) := \eta < \overline{F}(x_0)$ for some $x_0 \in \mathbb{R}$, then it must be that either $H(x) = H(x_0)$ for all $x \in [\overline{F}^{-1}(\eta^+), x_0]$ or $H(x) = H(x_0)$ for all $x \in [x_0, \underline{F}^{-1}(\eta))$. Indeed, suppose the contrary, so that there exists $\underline{x} \in [\overline{F}^{-1}(\eta^+), x_0)$ and $\overline{x} \in (x_0, \underline{F}^{-1}(\eta))$ such that $H(\underline{x}) < H(x_0) < H(\overline{x}^-)$. Then, since H is right-continuous, and since $H(\underline{x}) < H(x_0) < H(\overline{x}^-)$, it must be that $H^{-1}(\eta) > \overline{F}^{-1}(\eta^+)$ and $H^{-1}(\eta^+) < \underline{F}^{-1}(\eta)$. Moreover, since $x \mapsto F(x^-)$ is left-continuous, $H^{-1}(\eta) > \underline{x} \geq \overline{F}^{-1}(\eta^+)$ implies $\overline{F}(H^{-1}(\eta)^-) > \eta$. Likewise, $H^{-1}(\eta^+) < \overline{x} < \underline{F}^{-1}(\eta)$ implies that $\underline{F}(H^{-1}(\eta^+)) < \eta$. Now define a function $\Phi : [0,1]^2 \to \mathbb{R}^2$ as

$$\Phi(\varepsilon_1, \varepsilon_2) := \begin{pmatrix} \eta - \varepsilon_2 - \underline{F}(H^{-1}((\eta + \varepsilon_1)^+)) \\ \overline{F}(H^{-1}(\eta - \varepsilon_2)^-) - \eta - \varepsilon_1 \end{pmatrix},$$

for all $(\varepsilon_1, \varepsilon_2) \in [0,1]^2$. Then Φ is continuous at $(0,0)$ and $\Phi(0,0) \in \mathbb{R}^2_{++}$. Therefore, there exists $(\hat{\varepsilon}_1, \hat{\varepsilon}_2) \in [0,1]^2 \backslash \{(0,0)\}$ such that $\Phi(\hat{\varepsilon_1}, \hat{\varepsilon}_2) \in \mathbb{R}^2_{++}$. Let $\underline{\eta} := \eta - \hat{\varepsilon}_2$ and $\overline{\eta} := \eta + \hat{\varepsilon}_1$, it then follows that

$$\underline{F}(H^{-1}(\underline{\eta}^+)^-) \leq \underline{F}(H^{-1}(\overline{\eta}^+)) < \underline{\eta} < \eta < \overline{\eta} < \overline{F}(H^{-1}(\underline{\eta})^-) \leq \overline{F}(H^{-1}(\underline{\eta})). \tag{A.11}$$

Now consider the function $h : [H^{-1}(\underline{\eta}), H^{-1}(\overline{\eta}^+)] \to [\underline{\eta}, \overline{\eta}]$, defined as $h(x) := H(x)$, for all $x \in [H^{-1}(\underline{\eta}), H^{-1}(\overline{\eta}^+)]$. Clearly, h is nondecreasing. As a result, since the extreme points of the collection of uniformly bounded monotone functions are step functions with at most one jump (see, for instances, Skreta 2006 and Börgers 2015), $\underline{\eta} < h(x_0) = H(x_0) = \eta < \overline{\eta}$ implies that there exists distinct nondecreasing, right-continuous functions h_1, h_2 that map from $[H^{-1}(\underline{\eta}), H^{-1}(\overline{\eta}^+)]$ to $[\underline{\eta}, \overline{\eta}]$, as well as a constant $\lambda \in (0,1)$ such that

$h(x)=\lambda h_1(x)+(1-\lambda)h_2(x)$, for all $x\in[H^{-1}(\underline{\eta}),H^{-1}(\overline{\eta}^+)]$. Now define $\widehat{H}_1,\widehat{H}_2$ as

$$\widehat{H}_1(x):=\begin{cases} H(x), & \text{if } x\notin[H^{-1}(\underline{\eta}),H^{-1}(\overline{\eta}^+)] \\ h_1(x), & \text{if } x\in[H^{-1}(\underline{\eta}),H^{-1}(\overline{\eta}^+)]\end{cases};$$

and

$$\widehat{H}_2(x):=\begin{cases} H(x), & \text{if } x\notin[H^{-1}(\underline{\eta}),H^{-1}(\overline{\eta}^+)] \\ h_2(x), & \text{if } x\in[H^{-1}(\underline{\eta}),H^{-1}(\overline{\eta}^+)]\end{cases}$$

Clearly, $\lambda\widehat{H}_1+(1-\lambda)\widehat{H}_2=H$.

It now remains to show that $\widehat{H}_1,\widehat{H}_2\in\mathcal{I}(\underline{F},\overline{F})$. Indeed, for any $i\in\{1,2\}$ and for any $x,y\in\mathbb{R}$ with $x<y$, if $x,y\notin[H^{-1}(\underline{\eta}),H^{-1}(\overline{\eta}^+)]$, then $\widehat{H}_i(x)=H(x)\le H(y)=\widehat{H}_i(y)$, since H is nondecreasing. Meanwhile, if $x,y\in[H^{-1}(\underline{\eta}),H^{-1}(\overline{\eta}^+)]$, then $\widehat{H}_i(x)=h_i(x)\le h_i(y)=\widehat{H}_i(y)$. If $x<H^{-1}(\underline{\eta})$ and $y\in[H^{-1}(\underline{\eta}),H^{-1}(\overline{\eta}^+)]$, then $\widehat{H}_i(x)=H(x)\le\underline{\eta}\le h_i(y)=\widehat{H}_i(y)$. Likewise, if $y>H^{-1}(\overline{\eta}^+)$ and $x\in[H^{-1}(\underline{\eta}),H^{-1}(\overline{\eta}^+)]$, then $\widehat{H}_i(x)=h_i(x)\le\overline{\eta}\le H(y)=\widehat{H}_i(y)$. Together, $\widehat{H}_i$ must be nondecreasing, and hence $\widehat{H}_i\in\mathcal{F}$ for all $i\in\{1,2\}$. Moreover, for any $i\in\{1,2\}$ and for all $x\in[H^{-1}(\underline{\eta}),H^{-1}(\overline{\eta}^+)]$, from (A.11), we have

$$\underline{F}(x)\le\underline{F}(H^{-1}(\overline{\eta}^+))<\underline{\eta}\le h_i(x)\le\overline{\eta}<\overline{F}(H^{-1}(\eta)^-)\le\overline{F}(x).$$

Together with $H\in\mathcal{I}(\underline{F},\overline{F})$, it then follows that $\underline{F}(x)\le\widehat{H}_i(x)\le\overline{F}(x)$ for all $x\in\mathbb{R}$, and hence $\widehat{H}_i\in\mathcal{I}(\underline{F},\overline{F})$ for all $i\in\{1,2\}$. Consequently, there exists distinct $\widehat{H}_1,\widehat{H}_2\in\mathcal{I}(\underline{F},\overline{F})$ and $\lambda\in(0,1)$ such that $H=\lambda\widehat{H}_1+(1-\lambda)\widehat{H}_2$. Thus H is not an extreme point of $\mathcal{I}(\underline{F},\overline{F})$, as desired.

As a result, for any extreme point H of $\mathcal{I}(\underline{F},\overline{F})$, the set $\{x\in\mathbb{R}|\underline{F}(x)<H(x)<\overline{F}(x)\}$ can be partitioned into three classes of open intervals: $I^{\overline{F}}$, $I^{\underline{F}}$, and $I^{\overline{F},\underline{F}}$ such that for any open interval $(\underline{x},\overline{x})\in I^{\overline{F}}$, H is a constant on $[\underline{x},\overline{x})$ and $H(\underline{x})=\overline{F}(\underline{x})$; for any open interval $(\underline{x},\overline{x})\in I^{\underline{F}}$, H is a constant on $[\underline{x},\overline{x})$ and $H(\overline{x}^-)=\underline{F}(\overline{x}^-)$; and for any open interval $(\underline{x},\overline{x})\in I^{\overline{F},\underline{F}}$, H is a constant on $[\underline{x},\overline{x})$ and $\overline{F}(\underline{x})=H(\underline{x})=H(\overline{x}^-)=\underline{F}(\overline{x}^-)$. Note that since $\overline{F},\underline{F},H$ are nondecreasing and since $H\in\mathcal{I}(\underline{F},\overline{F})$, every interval in $I^{\overline{F}}$ and $I^{\underline{F}}$ must have at least one of its end points being a discontinuity point of H. Since H has at most countably many discontinuity points, $I^{\overline{F}}$ and $I^{\underline{F}}$ must be countable. Meanwhile, any distinct intervals $(\underline{x}_1,\overline{x}_1),(\underline{x}_2,\overline{x}_2)\in I^{\overline{F},\underline{F}}$ must be disjoint. Moreover, for any pair of these intervals with $\overline{x}_1<\underline{x}_2$, there must exist some $x_0\in(\overline{x}_1,\underline{x}_2)$ at which H is discontinuous. Therefore, since H has at most countably many discontinuity points, $I^{\overline{F},\underline{F}}$ must be countable as well.

Together, for any extreme point H of $\mathcal{I}(\underline{F},\overline{F})$, there exists countably many intervals $\{[\underline{x}_n,\overline{x}_n)\}_{n=1}^{\infty}:=I^{\overline{F}}\cup I^{\underline{F}}\cup I^{\overline{F},\underline{F}}$ such that H satisfies 1 and 2. This completes the proof. ■

A.2 Proof of Theorem 2

To show that $\mathcal{H}_\tau\subseteq\mathcal{I}(F_R^\tau,F_L^\tau)$, consider any $H\in\mathcal{H}_\tau$. Let $\mu\in\mathcal{M}$ and any $r\in\mathcal{R}_\tau$ be a signal and a selection rule, respectively, such that $H^\tau(\cdot|\mu,r)=H$. By the definition of $H^\tau(\cdot|\mu,r)$, it must be that, for all $x\in\mathbb{R}$,

$$H^\tau(x|\mu,r)\le\mu(\{G\in\mathcal{F}_0|G^{-1}(\tau)\le x\})=\mu(\{G\in\mathcal{F}_0|G(x)\ge\tau\}).$$

Now consider any $x \in \mathbb{R}$. Clearly, $\mu(\{G \in \mathcal{F}_0 | G(x) \geq \tau\}) \leq 1$, since μ is a probability measure. Moreover, let $M_x^+(q) := \mu(\{G \in \mathcal{F}_0 | G(x) \geq q\})$ for all $q \in [0,1]$. From (1), it follows that the left-limit of $1 - M_x^+$ is a CDF and a mean-preserving spread of a Dirac measure at $F(x)$. Therefore, whenever $\tau \geq F(x)$, then $M_x^+(\tau)$ can be at most $F(x)/\tau$ to have a mean of $F(x)$.[36] Together, this implies that $\mu(\{G \in \mathcal{F}_0 | G(x) \geq \tau\}) \leq F_L^\tau(x)$ for all $x \in \mathbb{R}$.

At the same time, by the definition of $H^\tau(\cdot|\mu, r)$, it must be that, for all $x \in \mathbb{R}$,

$$H^\tau(x^-|\mu, r) \geq \mu(\{G \in \mathcal{F}_0 | G^{-1}(\tau^+) < x\}) = \mu(\{G \in \mathcal{F}_0 | G(x) > \tau\}).$$

Consider any $x \in \mathbb{R}$. Since μ is a probability measure, it must be that $\mu(\{G \in \mathcal{F}_0 | G(x) > \tau\}) \geq 0$. Furthermore, let $M_x^-(q) := \mu(\{G \in \mathcal{F}_0 | G(x) > q\})$ for all $q \in [0,1]$. From (1), it follows that $1 - M_x^-$ is a CDF and a mean-preserving spread of a Dirac measure at $F(x)$. Therefore, whenever $\tau \leq F(x)$, then $M_x^-(\tau)$ must be at least $(F(x) - \tau)/(1 - \tau)$ to have a mean of $F(x)$.[37] Together, this implies that $\mu(\{G \in \mathcal{F}_0 | G(x) > \tau\}) \geq F_R^\tau$ for all $x \in \mathbb{R}$, which, in turn, implies that $F_R^\tau(x) \leq H^\tau(x^-|\mu, r) \leq H^\tau(x|\mu, r) \leq F_L^\tau(x)$ for all $x \in \mathbb{R}$, as desired.

To prove that $\mathcal{I}(F_R^\tau, F_L^\tau) \subseteq \mathcal{H}_\tau$, we first show that for any extreme point H of $\mathcal{I}(F_R^\tau, F_L^\tau)$, there exists a signal $\mu \in \mathcal{M}$ and a selection rule $r \in \mathcal{R}_\tau$ such that $H(x) = H^\tau(x|\mu, r)$ for all $x \in \mathbb{R}$. Consider any extreme point H of $\mathcal{I}(F_R^\tau, F_L^\tau)$. By Theorem 1, there exists a countable collection of intervals $\{(\underline{x}_n, \overline{x}_n)\}_{n=1}^\infty$ such that H satisfies 1 and 2. Since $(1 - F_L^\tau(x))F_R^\tau(x) = 0$ for all $x \notin [F^{-1}(\tau), F^{-1}(\tau^+)]$, there exists at most one $n \in \mathbb{N}$ such that $0 < H(\underline{x}_n) = F_L^\tau(\underline{x}_n) = F_R^\tau(\overline{x}_n^-) = H(\overline{x}_n^-) < 1$. Therefore, for $\underline{x}$ and $\overline{x}$ defined as

$$\underline{x} := \sup\{\underline{x}_n | n \in \mathbb{N},\, H(\underline{x}_n) = F_L^\tau(\underline{x}_n)\},$$

and

$$\overline{x} := \inf\{\overline{x}_n | n \in \mathbb{N},\, H(\overline{x}_n^-) = F_R^\tau(\overline{x}_n^-)\},$$

respectively, it must be that $\overline{x} \geq \underline{x}$, and that for all $n \in \mathbb{N}$, either $\overline{x}_n \leq \underline{x}$ and $H(\underline{x}_n) = F_L^\tau(\underline{x}_n)$; or $\underline{x}_n \geq \overline{x}$ and $H(\overline{x}_n^-) = F_R^\tau(\overline{x}_n^-)$. Henceforth, let $\mathbb{N}_1$ be the collection of $n \in \mathbb{N}$ such that $\overline{x}_n \leq \overline{x}$ and $H(\underline{x}_n) = F_L^\tau(\underline{x}_n)$, and let $\mathbb{N}_2$ be the collection of $n \in \mathbb{N}$ such that $\underline{x}_n \geq \underline{x}$ and $H(\overline{x}_n^-) = F_R^\tau(\overline{x}_n^-)$. Note that $\mathbb{N}_1 \cup \mathbb{N}_2 = \mathbb{N}$ and that $|\mathbb{N}_1 \cap \mathbb{N}_2| \leq 1$, with $\underline{x}_n = \underline{x}$ and $\overline{x}_n = \overline{x}$ whenever $n \in \mathbb{N}_1 \cap \mathbb{N}_2$.

We now construct a signal $\mu \in \mathcal{M}$ and a selection rule $r \in \mathcal{R}_\tau$ such that $H^\tau(\cdot|\mu, r) = H$. To this end, let $\eta := H(\overline{x}^-) - H(\underline{x})$ and let $\hat{x} := \inf\{x \in [\underline{x}, \overline{x}] | H(x) = H(\overline{x}^-)\}$. Note that by the definition of $\underline{x}$ and $\overline{x}$, if $\eta > 0$, then $\hat{x} \in (\underline{x}, \overline{x})$ and $H(x) = H(\underline{x})$ for all $x \in [\underline{x}, \hat{x})$, while $H(x) = H(\overline{x}^-)$ for all $x \in [\hat{x}, \overline{x})$. In particular, $F_L^\tau(\hat{x}) \geq H(\hat{x}) = F_L^\tau(\underline{x}) + \eta$, and hence $F(\hat{x}) - \tau\eta \geq F(\underline{x})$. Likewise, $F(\hat{x}) + (1 - \tau)\eta \leq F(\overline{x}^-)$. Let

$$\underline{y} := F^{-1}([F(\hat{x}) - \tau\eta]^+), \quad \text{and} \quad \overline{y} := F^{-1}(F(\hat{x}) + (1 - \tau)\eta)).$$

It then follows that $\underline{x} \leq \underline{y} \leq \hat{x} \leq \overline{y} \leq \overline{x}$, with at least one inequality being strict if $\eta > 0$. Next, define $\widehat{F}$ as

[36] More specifically, to maximize the probability at τ, a mean-preserving spread of $F(x)$ must assign probability $F(x)/\tau$ at τ, and probability $1 - F(x)/\tau$ at 0.

[37] More specifically, to minimize the probability at τ, a mean-preserving spread of $F_0(x)$ must assign probability $(F(x) - \tau)/(1 - \tau)$ at 1, and probability $1 - (F(x) - \tau)/(1 - \tau)$ at 0.

follows: $\widehat{F} \equiv 0$ if $\eta = 0$; and

$$\widehat{F}(x) := \begin{cases} 0, & \text{if } x < \underline{y} \\ \frac{F(x) - (F(\hat{x}) - \tau\eta)}{\eta}, & \text{if } x \in [\underline{y}, \overline{y}) \\ 1, & \text{if } x \geq \overline{y} \end{cases},$$

if $\eta > 0$. Clearly $\widehat{F} \in \mathcal{F}_0$ if $\eta > 0$, and $\hat{x} \in [\widehat{F}^{-1}(\tau), \widehat{F}^{-1}(\tau^+)]$. Moreover, for all $x \in \mathbb{R}$, let

$$\widetilde{F}(x) := \frac{F(x) - \eta\widehat{F}(x)}{1 - \eta}.$$

By construction, $\eta\widehat{F} + (1-\eta)\widetilde{F} = F$. From the definition of $\underline{y}$ and $\overline{y}$, it can be shown that $\widetilde{F} \in \mathcal{F}_0$ as well. Furthermore,

$$\widetilde{F}(\overline{x}^-) - \widetilde{F}(\underline{x}) = \frac{F(\overline{x}^-) - F(\underline{x}) - \eta}{1 - \eta} = \frac{1}{1-\eta}\left[\frac{\tau}{1-\tau}(1 - F(\overline{x}^-)) + \frac{1-\tau}{\tau}F(\underline{x})\right].$$

Next, define $\widetilde{F}_1$ and $\widetilde{F}_2$ as follows:

$$\widetilde{F}_1(x) := \begin{cases} \frac{F(x)}{F(\underline{x}) + \alpha(F(\overline{x}^-) - F(\underline{x}) - \eta)}, & \text{if } x < \underline{x} \\ \frac{F(\underline{x})\alpha(F(x) - F(\underline{x}) - \eta)}{F(\underline{x}) + \alpha(F(\overline{x}^-) - F(\underline{x}) - \eta)}, & \text{if } x \in [\underline{x}, \overline{x}) \\ 1, & \text{if } x \geq \overline{x} \end{cases};$$

and

$$\widetilde{F}_2(x) := \begin{cases} 0, & \text{if } x < \underline{x} \\ \frac{(1-\alpha)(F(x) - F(\underline{x}) - \eta)}{1 - F(\overline{x}^-) + (1-\alpha)(F(\overline{x}^-) - F(\underline{x}) - \eta)}, & \text{if } x \in [\underline{x}, \overline{x}) \\ \frac{F(x) - F(\underline{x}) + (1-\alpha)(F(\overline{x}^-) - F(\underline{x}) - \eta)}{1 - F(\overline{x}^-) + (1-\alpha)(F(\overline{x}^-) - F(\underline{x}) - \eta)}, & \text{if } x \geq \overline{x} \end{cases},$$

where

$$\alpha := \frac{\frac{1-\tau}{\tau}F(\underline{x})}{\frac{\tau}{1-\tau}(1 - F(\overline{x}^-)) + \frac{1-\tau}{\tau}F(\underline{x})}.$$

By construction, $\widetilde{\alpha}\widetilde{F}_1 + (1-\widetilde{\alpha})\widetilde{F}_2 = \widetilde{F}$, where $\widetilde{\alpha} \in (0,1)$ is given by $\widetilde{\alpha} := [F(\underline{x}) + \alpha(F(\overline{x}^-) - F(\underline{x}) - \eta)]/(1-\eta)$. Moreover, $\widetilde{F}_1(\underline{x}) \geq \tau$, and $\widetilde{F}_2(\overline{x}^-) \leq \tau$.

Now define two classes of distributions, $\{\widetilde{F}_1^x\}_{x \leq \underline{x}}$ and $\{\widetilde{F}_2^x\}_{x \geq \overline{x}}$, as follows:

$$\widetilde{F}_1^x(z) := \begin{cases} 0, & \text{if } z < x \\ \widetilde{F}(\underline{x}), & \text{if } z \in [x, \underline{x}) \\ \widetilde{F}(z), & \text{if } z \geq \underline{x} \end{cases}; \text{ and } \widetilde{F}_2^x(z) := \begin{cases} \widetilde{F}(z), & \text{if } z < \overline{x} \\ \widetilde{F}(\overline{x}^-), & \text{if } z \in [\overline{x}, x) \\ 1, & \text{if } z \geq x \end{cases}.$$

Note that, since $\widetilde{F}_1(\underline{x}) \geq \tau$ and $\widetilde{F}_2(\overline{x}^-) \leq \tau$, $x \in [(\widetilde{F}_1^x)^{-1}(\tau), (\widetilde{F}_1^x)^{-1}(\tau^+)]$ for all $x \leq \underline{x}$ and $x \in [(\widetilde{F}_2^x)^{-1}(\tau), (\widetilde{F}_2^x)^{-1}(\tau^+)]$ for all $x \geq \overline{x}$. Moreover, for any $n \in \mathbb{N}_1$ and for any $m \in \mathbb{N}_2$, let

$$\widetilde{F}_1^n(z) := \frac{1}{\widetilde{F}(\overline{x}_n) - \widetilde{F}(\underline{x}_n)} \int_{\underline{x}_n}^{\overline{x}_n} \widetilde{F}_1^x(z)\widetilde{F}(\mathrm{d}x),$$

and

$$\widetilde{F}_2^m(z) := \frac{1}{\widetilde{F}(\overline{x}_m) - \widetilde{F}(\underline{x}_m)} \int_{\underline{x}_m}^{\overline{x}_m} \widetilde{F}_2^x(z)\, \mathrm{d}\widetilde{F}(\mathrm{d}x),$$

for all $z \in \mathbb{R}$. By construction, $\widetilde{F}_1^n, \widetilde{F}_2^m \in \mathcal{F}_0$ and $\overline{x}_n \in [(\widetilde{F}_1^n)^{-1}(\tau), (\widetilde{F}_1^n)^{-1}(\tau^+)]$, $\underline{x}_m \in [(\widetilde{F}_2^m)^{-1}(\tau), (\widetilde{F}_2^m)^{-1}(\tau^+)]$ for all $n \in \mathbb{N}_1$ and $m \in \mathbb{N}_2$.

Next, for any $x \in \mathbb{R}$, let $\widetilde{G}^x \in \mathcal{F}_0$ be defined as

$$\widetilde{G}^x(z) := \begin{cases} \widetilde{F}_1^x(z), & \text{if } x \in (-\infty, \overline{x}] \setminus \cup_{n\in\mathbb{N}_1} [\underline{x}_n, \overline{x}_n) \\ \widetilde{F}_1^n(z), & \text{if } x \in [\underline{x}_n, \overline{x}_n),\ n \in \mathbb{N}_1 \\ \widetilde{F}_2^x(z), & \text{if } x \in [\overline{x}, \infty) \setminus \cup_{m\in\mathbb{N}_2} [\underline{x}_m, \overline{x}_m) \\ \widetilde{F}_2^m(z), & \text{if } x \in [\underline{x}_m, \overline{x}_m),\ m \in \mathbb{N}_2 \end{cases},$$

for all $z \in \mathbb{R}$. Let

$$\widetilde{H}(x) := \begin{cases} \frac{H(x)}{1-\eta}, & \text{if } x < \underline{x} \\ \frac{H(\underline{x})}{1-\eta}, & \text{if } x \in [\underline{x}, \overline{x}) \\ \frac{H(x)-\eta}{1-\eta}, & \text{if } x \geq \overline{x} \end{cases},$$

and define $\tilde{\mu}$ as

$$\tilde{\mu}(\{\widetilde{G}^x \in \mathcal{F}_0 | x \leq z\}) := \widetilde{H}(z),$$

for all $z \in \mathbb{R}$. Then, by construction, for any $z \in \mathbb{R}$,

$$\int_{\mathcal{F}} F(z)\tilde{\mu}(\mathrm{d}F) = \int_{\mathbb{R}} \widetilde{G}^x(z)\widetilde{H}(\mathrm{d}x) = \widetilde{F}(z). \tag{A.12}$$

Moreover, let $\tilde{r} : \mathcal{F}_0 \to \Delta(\mathbb{R})$ be defined as

$$\tilde{r}(G) := \begin{cases} \delta_{\{G^{-1}(\tau^+)\}}, & \text{if } G = \widetilde{G}^x,\ x \geq \overline{x} \\ \delta_{\{G^{-1}(\tau)\}}, & \text{otherwise} \end{cases},$$

for all $G \in \mathcal{F}_0$. It then follows that $H^\tau(x|\tilde{\mu}, \tilde{r}) = \widetilde{H}(x)$ for all $x \in \mathbb{R}$. Next, let $\mu \in \Delta(\mathcal{F}_0), r \in \mathcal{R}_\tau$ together be defined as

$$\mu := (1-\eta)\tilde{\mu} + \eta\delta_{\{\widehat{F}\}},$$

and

$$r(G) := \begin{cases} \delta_{\{\hat{x}\}}, & \text{if } G = \widehat{F} \\ \tilde{r}(G), & \text{otherwise} \end{cases},$$

for all $G \in \mathcal{F}_0$. Since $F = \eta\widehat{F} + (1-\eta)\widetilde{F}$, together with (A.12), we have $\mu \in \mathcal{M}$. Moreover, since $H^\tau(\cdot|\tilde{\mu}, \tilde{r}) = \widetilde{H}$, we have $H^\tau(x|\mu, r) = H(x)$ for all $x \in \mathbb{R}$.

Lastly, let Γ be a collection of probability measures $\gamma \in \Delta(\mathbb{R} \times \mathcal{F}_0)$ such that $\gamma(\{(x, G) \in \mathbb{R} \times \mathcal{F}_0 | x \in [G^{-1}(\tau), G^{-1}(\tau^+)]\}) = 1$ and

$$\int_{\mathbb{R}\times\mathcal{F}_0} G(x)\gamma(\mathrm{d}x, \mathrm{d}G) = F(x),$$

for all $x \in \mathbb{R}$. Define a linear functional $\Xi : \Gamma \to \mathcal{F}_0$ as

$$\Xi(\gamma)[x] := \gamma((-\infty, x], \mathcal{F}_0),$$

for all $\gamma \in \Gamma$ and for all $x \in \mathbb{R}$. Then, since for any $\widehat{H}$ in the set of extreme points $\text{ext}(\mathcal{I}(F_R^\tau, F_L^\tau))$ of $\mathcal{I}(F_R^\tau, F_L^\tau)$, there exists $\hat{\mu} \in \mathcal{M}$ and $\hat{r} \in \mathcal{R}_\tau$ such that $H^\tau(x|\hat{\mu}, \hat{r}) = \widehat{H}(x)$ for all $x \in \mathbb{R}$, it must be that $\text{ext}(\mathcal{I}(F_R^\tau, F_L^\tau)) \subseteq \Xi(\Gamma)$.

Now consider any $H \in \mathcal{I}(F_R^\tau, F_L^\tau)$. Since $\mathcal{I}(F_R^\tau, F_L^\tau)$ is a compact and convex set of a metrizable, locally convex topological space,[38] Choquet's theorem implies that there exists a probability measure $\Lambda_H \in \Delta(\mathcal{I}(F_R^\tau, F_L^\tau))$ with $\Lambda_H(\text{ext}(\mathcal{I}(F_R^\tau, F_L^\tau))) = 1$ such that

$$\int_{\mathcal{I}(F_R^\tau, F_L^\tau)} \widehat{H}(x)\Lambda_H(\mathrm{d}\widehat{H}) = H(x),$$

for all $x \in \mathbb{R}$. Define a measure $\widetilde{\Lambda}_H$ by

$$\widetilde{\Lambda}_H(A) := \Lambda_H(\{\Xi(\gamma)|\gamma \in A\}),$$

for all measurable $A \subseteq \Gamma$. Since $\Lambda_H(\text{ext}(\mathcal{I}(F_R^\tau, F_L^\tau))) = 1$ and $\text{ext}(\mathcal{I}(F_R^\tau, F_L^\tau)) \subseteq \Xi(\Gamma)$, $\widetilde{\Lambda}_H$ is a probability measure on Γ. For any $x \in \mathbb{R}$ and for any measurable $A \subseteq \mathcal{F}_0$, let

$$\gamma((-\infty, x], A) := \int_\Gamma \tilde{\gamma}((-\infty, x], A)\widetilde{\Lambda}_H(\mathrm{d}\tilde{\gamma}),$$

and let $\mu(A) := \gamma(\mathbb{R}, A)$. By construction, for all $x \in \mathbb{R}$,

$$\int_{\mathcal{F}} G(x)\mu(\mathrm{d}G) = \int_\Gamma \left(\int_{\mathbb{R}\times\mathcal{F}_0} G(x)\tilde{\gamma}(\mathrm{d}\tilde{x}, \mathrm{d}G) \right) \widetilde{\Lambda}_H(\mathrm{d}\tilde{\gamma}) = F(x),$$

and hence $\mu \in \mathcal{M}$. Furthermore, by the disintegration theorem (c.f., Çinlar 2010, theorem 2.18), there exists a transition probability $r : \mathcal{F}_0 \to \Delta(\mathbb{R})$ such that $\gamma(\mathrm{d}x, \mathrm{d}G) = r(\mathrm{d}x|G)\mu(\mathrm{d}G)$. Since $\widetilde{\Lambda}_H(\Gamma) = 1$, and since r is a transition probability, we have $r \in \mathcal{R}_\tau$. Finally, for any $x \in \mathbb{R}$, since Ξ is affine,

$$\begin{aligned} H^\tau(x|\mu, r) = \gamma((-\infty, x], \mathcal{F}_0) =& \Xi(\gamma)[x] \\ =& \int_\Gamma \Xi(\tilde{\gamma})[x]\widetilde{\Lambda}_H(\mathrm{d}\tilde{\gamma}) \\ =& \int_{\text{ext}(\mathcal{I}(F_R^\tau, F_L^\tau))} \widehat{H}(x)\Lambda_H(\mathrm{d}\widehat{H}) \\ =& H(x), \end{aligned}$$

as desired. This completes the proof. ■

[38] To see this, recall that for any sequence $\{H_n\} \subseteq \mathcal{I}(F_R^\tau, F_L^\tau)$, Helly's selection theorem implies that there exists a subsequence $\{H_{n_k}\} \subseteq \{H_n\}$ that converges pointwise (and hence, in weak-*) to some $H \in \mathcal{I}(F_R^\tau, F_L^\tau)$.

A.3 Proof of Theorem 3

By Theorem 2,

$$\widetilde{\mathcal{H}}_\tau \subseteq \mathcal{H}_\tau = \mathcal{I}(F_R^\tau, F_L^\tau).$$

It remains to show that

$$\bigcup_{\varepsilon>0} \mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon}) \subseteq \widetilde{\mathcal{H}}_\tau.$$

To this end, let $\widetilde{\mathcal{M}}_\tau$ be the collection of $\mu \in \mathcal{M}$ such that $\mu(\{G \in \mathcal{F}_0 | G^{-1}(\tau) < G^{-1}(\tau^+)\}) = 0$. Consider any $\varepsilon > 0$ and any extreme point H of $\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})$. By Theorem 1, there exists a countable collection of intervals $\{(\underline{x}_n, \overline{x}_n)\}_{n=1}^\infty$ such that H satisfies 1 and 2. Since $(1 - F_R^{\tau,\varepsilon}(x))F_L^{\tau,\varepsilon}(x) = 0$ for all $x \neq F_0^{-1}(\tau)$, there exists at most one $n \in \mathbb{N}$ such that $0 < H(\underline{x}_n) = F_R^{\tau,\varepsilon}(\underline{x}_n) = F_L^{\tau,\varepsilon}(\overline{x}_n^-) = H(\overline{x}_n^-) < 1$. Therefore, for $\underline{x}$ and $\overline{x}$ defined as

$$\underline{x} := \sup\{\underline{x}_n | n \in \mathbb{N},\, H(\underline{x}_n) = F_R^{\tau,\varepsilon}(\underline{x}_n)\} \quad \text{and} \quad \overline{x} := \inf\{\overline{x}_n | n \in \mathbb{N},\, H(\overline{x}_n^-) = F_L^{\tau,\varepsilon}(\overline{x}_n^-)\},$$

respectively, it must be that $\overline{x} \geq \underline{x}$, and that, for all $n \in \mathbb{N}$, either $\overline{x}_n \leq \underline{x}$ and $H(\underline{x}_n) = F_L^{\tau,\varepsilon}(\underline{x}_n)$, or $\underline{x}_n \geq \overline{x}$ and $H(\overline{x}_n^-) = F_R^{\tau,\varepsilon}(\overline{x}_n^-)$. Henceforth, let $\mathbb{N}_1$ be the collection of $n \in \mathbb{N}$ such that $\overline{x}_n \leq \overline{x}$ and $H(\underline{x}_n) = F_L^{\tau,\varepsilon}(\underline{x}_n)$, and let $\mathbb{N}_2$ be the collection of $n \in \mathbb{N}$ such that $\underline{x}_n \geq \underline{x}$ and $H(\overline{x}_n^-) = F_R^{\tau,\varepsilon}(\overline{x}_n^-)$. Note that $\mathbb{N}_1 \cup \mathbb{N}_2 = \mathbb{N}$ and that $|\mathbb{N}_1 \cap \mathbb{N}_2| \leq 1$, with $\underline{x}_n = \underline{x}$ and $\overline{x}_n = \overline{x}$ whenever $n \in \mathbb{N}_1 \cap \mathbb{N}_2$.

We now construct a signal $\mu \in \widetilde{\mathcal{M}}_\tau$ such that $H^\tau(\cdot|\mu) = H$. First, let $\eta := H(\overline{x}^-) - H(\underline{x})$ and let $\hat{x} := \inf\{x \in [\underline{x}, \overline{x}] | H(x) = H(\overline{x}^-)\}$. Note that, by the definition of $\underline{x}$ and $\overline{x}$, if $\eta > 0$, then $\hat{x} \in (\underline{x}, \overline{x})$ and $H(x) = H(\underline{x})$ for all $x \in [\underline{x}, \hat{x})$, while $H(x) = H(\overline{x}^-)$ for all $x \in [\hat{x}, \overline{x})$. In particular, $F_L^{\tau,\varepsilon}(\hat{x}) \geq H(\hat{x}) = F_L^{\tau,\varepsilon}(\underline{x}) + \eta$, and hence $F(\hat{x}) - (\tau + \varepsilon)\eta \geq F(\underline{x})$. Likewise, $F(\hat{x}) + (1 - \tau + \varepsilon)\eta \leq F(\overline{x}^-)$. Now let

$$\underline{y} := F^{-1}(F(\hat{x}) - (\tau + \varepsilon)\eta), \quad \text{and} \quad \overline{y} := F^{-1}(F(\hat{x}) + (1 - \tau + \varepsilon)\eta).$$

It then follows that $\underline{x} \leq \underline{y} \leq \hat{x} \leq \overline{y} \leq \overline{x}$, with at least one inequality being strict if $\eta > 0$. Next, define $\widehat{F}$ as follows: $\widehat{F} \equiv 0$ if $\eta = 0$; and

$$\widehat{F}(x) := \begin{cases} 0, & \text{if } x < \underline{y} \\ \frac{F(x) - (F(\hat{x}) - (\tau+\varepsilon)\eta)}{\eta}, & \text{if } x \in [\underline{y}, \overline{y}) \\ 1, & \text{if } x \geq \overline{y} \end{cases},$$

if $\eta > 0$. Clearly $\widehat{F} \in \mathcal{F}_0$ if $\eta > 0$, and $\hat{x} = \widehat{F}^{-1}(\tau)$. Moreover, for all $x \in \mathbb{R}$, let

$$\widetilde{F}(x) := \frac{F(x) - \eta\widehat{F}(x)}{1 - \eta}.$$

By construction, $\eta\widehat{F} + (1 - \eta)\widetilde{F} = F$. From the definition of $\underline{y}$ and $\overline{y}$, it can be shown that $\widetilde{F} \in \mathcal{F}_0$ as well. Furthermore,

$$\widetilde{F}(\overline{x}^-) - \widetilde{F}(\underline{x}) = \frac{F(\overline{x}^-) - F(\underline{x}) - \eta}{1 - \eta} = \frac{1}{1 - \eta}\left[\frac{\tau - \varepsilon}{1 - (\tau - \varepsilon)}(1 - F(\overline{x}^-)) + \frac{1 - (\tau + \varepsilon)}{\tau + \varepsilon}F(\underline{x})\right].$$

Next, define $\widetilde{F}_1$ and $\widetilde{F}_2$ as follows:

$$\widetilde{F}_1(x) := \begin{cases} \frac{F(x)}{F(\underline{x})+\alpha(F(\overline{x}^-)-F(\underline{x})-\eta)}, & \text{if } x < \underline{x} \\ \frac{F(\underline{x})\alpha(F(x)-F(\underline{x})-\eta)}{F(\underline{x})+\alpha(F(\overline{x}^-)-F(\underline{x})-\eta)}, & \text{if } x \in [\underline{x}, \overline{x}) \\ 1, & \text{if } x \geq \overline{x} \end{cases};$$

and

$$\widetilde{F}_2(x) := \begin{cases} 0, & \text{if } x < \underline{x} \\ \frac{(1-\alpha)(F(x)-F(\underline{x})-\eta)}{1-F(\overline{x}^-)+(1-\alpha)(F(\overline{x}^-)-F(\underline{x})-\eta)}, & \text{if } x \in [\underline{x}, \overline{x}) \\ \frac{F(x)-F(\underline{x})+(1-\alpha)(F(\overline{x}^-)-F(\underline{x})-\eta)}{1-F(\overline{x}^-)+(1-\alpha)(F(\overline{x}^-)-F(\underline{x})-\eta)}, & \text{if } x \geq \overline{x} \end{cases},$$

where

$$\alpha := \frac{\frac{1-(\tau+\varepsilon)}{\tau+\varepsilon}F(\underline{x})}{\frac{\tau-\varepsilon}{1-(\tau-\varepsilon)}(1-F(\overline{x}^-)) + \frac{1-(\tau+\varepsilon)}{\tau+\varepsilon}F(\underline{x})}.$$

By construction, $\widetilde{\alpha}\widetilde{F}_1+(1-\widetilde{\alpha})\widetilde{F}_2 = \widetilde{F}$, where $\widetilde{\alpha} \in (0,1)$ is given by $\widetilde{\alpha} := [F(\underline{x})+\alpha(F(\overline{x}^-)-F(\underline{x})-\eta)]/(1-\eta)$. Moreover, $\widetilde{F}_1(\underline{x}) = \tau+\varepsilon > \tau$, and $\widetilde{F}_2(\overline{x}^-) = \tau-\varepsilon < \tau$.

Now define two classes of distributions, $\{\widetilde{F}_1^x\}_{x\leq\underline{x}}$ and $\{\widetilde{F}_2^x\}_{x\geq\overline{x}}$, as follows:

$$\widetilde{F}_1^x(z) := \begin{cases} 0, & \text{if } z < x \\ \widetilde{F}(\underline{x}), & \text{if } z \in [x, \underline{x}) \\ \widetilde{F}(z), & \text{if } z \geq \underline{x} \end{cases}; \text{ and } \widetilde{F}_2^x(z) := \begin{cases} \widetilde{F}(z), & \text{if } z < \overline{x} \\ \widetilde{F}(\overline{x}^-), & \text{if } z \in [\overline{x}, x) \\ 1, & \text{if } z \geq x \end{cases}.$$

Note that, since $\widetilde{F}_1(\underline{x}) > \tau$ and $\widetilde{F}_2(\overline{x}^-) < \tau$, $x = (\widetilde{F}_1^x)^{-1}(\tau) = (\widetilde{F}_1^x)^{-1}(\tau^+)$ for all $x \leq \underline{x}$ and $x = (\widetilde{F}_2^x)^{-1}(\tau) = (\widetilde{F}_2^x)^{-1}(\tau^+)$ for all $x \geq \overline{x}$. Moreover, for any $n \in \mathbb{N}_1$ and for any $m \in \mathbb{N}_2$, let

$$\widetilde{F}_1^n(z) := \frac{1}{\widetilde{F}(\overline{x}_n) - \widetilde{F}(\underline{x}_n)} \int_{\underline{x}_n}^{\overline{x}_n} \widetilde{F}_1^x(z)\widetilde{F}(\mathrm{d}x),$$

and

$$\widetilde{F}_2^m(z) := \frac{1}{\widetilde{F}(\overline{x}_m) - \widetilde{F}(\underline{x}_m)} \int_{\underline{x}_m}^{\overline{x}_m} \widetilde{F}_2^x(z)\widetilde{F}(\mathrm{d}x),$$

for all $z \in \mathbb{R}$. By construction, $\widetilde{F}_1^n, \widetilde{F}_2^m \in \mathcal{F}_0$ and $\overline{x}_n = (\widetilde{F}_1^n)^{-1}(\tau) = (\widetilde{F}_1^n)^{-1}(\tau^+)$, $\underline{x}_m = (\widetilde{F}_2^m)^{-1}(\tau) = (\widetilde{F}_2^m)^{-1}(\tau^+)$ for all $n \in \mathbb{N}_1$ and $m \in \mathbb{N}_2$. Next, for any $x \in \mathbb{R}$, let $\widetilde{G}^x \in \mathcal{F}_0$ be defined as

$$\widetilde{G}^x(z) := \begin{cases} \widetilde{F}_1^x(z), & \text{if } x \in (-\infty, \overline{x}] \setminus \cup_{n\in\mathbb{N}_1} [\underline{x}_n, \overline{x}_n) \\ \widetilde{F}_1^n(z), & \text{if } x \in [\underline{x}_n, \overline{x}_n),\ n \in \mathbb{N}_1 \\ \widetilde{F}_2^x(z), & \text{if } x \in [\overline{x}, \infty) \setminus \cup_{m\in\mathbb{N}_2} [\underline{x}_m, \overline{x}_m) \\ \widetilde{F}_2^m(z), & \text{if } x \in [\underline{x}_m, \overline{x}_m),\ m \in \mathbb{N}_2 \end{cases},$$

for all $z \in \mathbb{R}$. Let

$$\widetilde{H}(x) := \begin{cases} \frac{H(x)}{1-\eta}, & \text{if } x < \underline{x} \\ \frac{H(\underline{x})}{1-\eta}, & \text{if } x \in [\underline{x}, \overline{x}) \\ \frac{H(x)-\eta}{1-\eta}, & \text{if } x \geq \overline{x} \end{cases},$$

and define $\tilde{\mu}$ as

$$\tilde{\mu}(\{\widetilde{G}^x \in \mathcal{F}_0 | x \leq z\}) := \widetilde{H}(z),$$

for all $z \in \mathbb{R}$. Then, by construction, for any $z \in \mathbb{R}$,

$$\int_{\mathcal{F}_0} G(z)\tilde{\mu}(\mathrm{d}G) = \int_{\mathbb{R}} \widetilde{G}^x(z)\widetilde{H}(\mathrm{d}x) = \widetilde{F}(z). \tag{A.13}$$

Furthermore, $H^\tau(x|\tilde{\mu}) = \widetilde{H}(x)$ for all $x \in \mathbb{R}$. As a result, from (A.13), for $\mu \in \Delta(\mathcal{F}_0)$ defined as

$$\mu := (1-\eta)\tilde{\mu} + \eta\delta_{\{\widehat{F}\}},$$

since $F = \eta\widehat{F} + (1-\eta)\widetilde{F}$, it must be that $\mu \in \widetilde{\mathcal{M}}_\tau$. Moreover, since $H^\tau(\cdot|\tilde{\mu}) = \widetilde{H}$, we have $H^\tau(x|\mu) = H(x)$ for all $x \in \mathbb{R}$.

Lastly, consider any $H \in \mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})$. Since $\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})$ is a convex and compact set in a metrizable space, Choquet's theorem implies that there exists a probability measure $\Lambda_H \in \Delta(\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon}))$ that assigns probability 1 to $\mathrm{ext}(\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon}))$ such that

$$H(x) = \int_{\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})} \widetilde{H}(x)\Lambda_H(\mathrm{d}\widetilde{H}).$$

Meanwhile, define the linear functional $\Xi : \widetilde{\mathcal{M}}_\tau \to \mathcal{F}_0$ as

$$\Xi(\tilde{\mu})[x] := \tilde{\mu}(\{G \in \mathcal{F}_0 | G^{-1}(\tau) \leq x\}),$$

for all $\tilde{\mu} \in \widetilde{\mathcal{M}}_\tau$ and for all $x \in \mathbb{R}$. Now define a probability measure $\widetilde{\Lambda}$ on $\widetilde{\mathcal{M}}_\tau$ by

$$\widetilde{\Lambda}_H(A) := \Lambda_H(\{\Xi(\tilde{\mu}) | \tilde{\mu} \in A\}),$$

for all $A \subseteq \widetilde{\mathcal{M}}_\tau$. Then, since $\Lambda_H(\mathrm{ext}(\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon}))) = 1$ and since, for any $\widetilde{H} \in \mathrm{ext}(\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon}))$, there exists $\tilde{\mu} \in \widetilde{\mathcal{M}}_\tau$ such that $H(x) = H^\tau(x|\tilde{\mu})$, it must be that $\widetilde{\Lambda}_H(\widetilde{\mathcal{M}}_\tau) = 1$, and hence $\widetilde{\Lambda}_H$ is a probability measure on $\widetilde{\mathcal{M}}_\tau$. Let $\tilde{\mu} \in \widetilde{\mathcal{M}}_\tau$ be defined as

$$\tilde{\mu}(A) := \int_{\widetilde{\mathcal{M}}_\tau} \mu(A)\widetilde{\Lambda}_H(\mathrm{d}\mu),$$

for all measurable $A \subseteq \mathcal{F}_0$. Then, since Ξ is linear, it follows that

$$\begin{aligned} H(x) = \int_{\mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})} \widetilde{H}(x)\Lambda_H(\mathrm{d}\widetilde{H}) = & \int_{\widetilde{\mathcal{M}}_\tau} \Xi(\mu)[x]\widetilde{\Lambda}_H(\mathrm{d}\mu) \\ =& \Xi(\tilde{\mu})[x] \\ =& H^\tau(x|\tilde{\mu}), \end{aligned}$$

and therefore, $H \in \widetilde{\mathcal{H}}_\tau$. Together, for any $\varepsilon > 0$, any $H \in \mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon})$ must be in $\widetilde{\mathcal{H}}_\tau$. In other words,

$$\bigcup_{\varepsilon>0} \mathcal{I}(F_R^{\tau,\varepsilon}, F_L^{\tau,\varepsilon}) \subseteq \widetilde{\mathcal{H}}_\tau.$$

This completes the proof. ■

A.4 Proof of Corollary 1

For 1, consider any $H \in \mathcal{H}_q$. By Theorem 2, $H \in \mathcal{I}(F_R^q, F_L^q)$. Thus, $(F_L^q)^{-1}(\tau) \leq H^{-1}(\tau) \leq H^{-1}(\tau^+) \leq (F_R^q)^{-1}(\tau^+)$, and therefore $[H^{-1}(\tau), H^{-1}(\tau^+)] \subseteq [(F_L^q)^{-1}(\tau), (F_R^q)^{-1}(\tau^+)]$. Conversely, consider any interval $Q = [\underline{x}, \overline{x}] \subseteq [(F_L^q)^{-1}(\tau), (F_R^q)^{-1}(\tau^+)]$. Then, let $\widehat{H}$ be defined as

$$\widehat{H}(x) := \begin{cases} 0, & \text{if } x < \underline{x} \\ \tau, & \text{if } x \in [\underline{x}, \overline{x}) \\ 1, & \text{if } x \geq \overline{x} \end{cases},$$

for all $x \in \mathbb{R}$. Then $\widehat{H} \in \mathcal{I}(F_L^q, F_R^q)$ and $Q = [H^{-1}(\tau), H^{-1}(\tau^+)]$. Moreover, by Theorem 2, $\widehat{H} \in \mathcal{H}_q$, as desired.

For 2, consider any $H \in \widetilde{\mathcal{H}}_q$. By Theorem 2, $H \in \mathcal{I}(F_R^q, F_L^q)$. Thus, it must be that $[H^{-1}(\tau), H^{-1}(\tau^+)] \subseteq [(F_L^q)^{-1}(\tau), (F_R^q)^{-1}(\tau)]$. Conversely, for any $\hat{x} \in ((F_L^q)^{-1}(\tau), (F_R^q)^{-1}(\tau^+))$, note that since $\hat{x} > (F_L^q)^{-1}(\tau)$ and since F is continuous, we have $F(\hat{x})/\tau > q$. Similarly, we also have $(F(\hat{x}) - \tau)/(1-\tau) < q$. Let $\varepsilon := \min\{F(\hat{x})/\tau - q, q - (F(\hat{x}) - \tau)/(1-\tau)\}$. Then, either $\hat{x} = (F_L^{q,\varepsilon})^{-1}(\tau)$ or $\hat{x} = (F_R^{q,\varepsilon})^{-1}(\tau)$. Since both $F_L^{q,\varepsilon}$ and $F_R^{q,\varepsilon}$ are in $\mathcal{I}(F_R^{q,\varepsilon}, F_L^{q,\varepsilon})$, Theorem 3 implies that $\hat{x} = H^{-1}(\tau)$ for some $H \in \widetilde{H}_q$. Lastly, note that under a signal $\mu \in \mathcal{M}$ such that μ assigns probability τ to F_L^τ and probability $1-\tau$ to F_R^τ, we have $\mu \in \widetilde{M_q}$ and $H^q(x|\mu) = \tau$ for all $x \in [(F_L^q)^{-1}(\tau), (F_R^q)^{-1}(\tau)]$. Hence, $[(F_L^q)^{-1}(\tau), (F_R^q)^{-1}(\tau)] \subseteq [H^{-1}(\tau), H^{-1}(\tau^+)]$ for some $H \in \widetilde{H}_q$, as desired. ■

A.5 Proof of Corollary 4

(i) and (ii) follow immediately from the fact that any $H \in \mathcal{I}(F_R^\tau, F_L^\tau)$ is dominated by F_R^τ and dominates F_L^τ, and that $v_S(x)$ is increasing in x for all $x \leq a$ and is decreasing in x for all $x > a$.

For (iii), suppose that for any $\underline{a} \leq \overline{a}$, $H_{\underline{a},\overline{a}}^C$ is not optimal. Then, since at least one extreme point of $\mathcal{I}(F_R^\tau, F_L^\tau)$ must be the solution of (3), consider any such extreme point and denote it by H. By Theorem 1, there exists a countable collection of intervals $\{[\underline{x}_n, \overline{x}_n)\}_{n=1}^\infty$ such that conditions 1 and 2 of Theorem 1 hold. Since $H_{\underline{a},\overline{a}}^C$ is not optimal for any $\underline{a} \leq \overline{a}$, $H \neq H_{\underline{a},\overline{a}}^C$ for all $\underline{a} \leq \overline{a}$. In particular, there must exist $n \in \mathbb{N}$ such that $\underline{x}_n < \overline{x}_n$ and either $H(\overline{x_n}^-) > F_R^\tau(\overline{x}_n)$ or $H(\underline{x}_n) < F_L^\tau(\underline{x}_n)$. Let a be the minimizer of v_S and suppose that $a \leq F^{-1}(\tau)$. Suppose that $H(\overline{x}_n^-) > F_R^\tau(\overline{x}_n)$. Then it must be that $H(\underline{x}_n) = F_L^\tau(\underline{x}_n)$. Moreover, since $H(\overline{x}_n^-) > F_R^\tau(\overline{x}_n)$, $H(\overline{x}_n) > F_R^\tau(\overline{x}_n)$ as well. If $\overline{x}_n \leq a$, then by replacing $H(x)$ with $\min\{F_L^\tau(x), H(\overline{x}_n)\}$ for all $x \in [\underline{x}_n, \overline{x}_n)$ and otherwise leaving H unchanged, the resulting distribution $\widehat{H}$ must still be in $\mathcal{I}(F_R^\tau, F_L^\tau)$. Since v_S is strictly decreasing on $[\underline{x}_n, \overline{x}_n)$, $\widehat{H}$ must give a higher value, a contradiction. If, on the other hand, $\overline{x}_n > a$, then since $H(\overline{x}_n) > F_R^\tau(\overline{x}_n)$ and since F is continuous, there exists $y > \overline{x}_n$ such that $H(\overline{x}_n^-) > F_R^\tau(y)$. Moreover, since H satisfies conditions 1 and 2, $H(x) > H(\overline{x}_n^-)$ for all $x \in [\overline{x}_n, y)$. Therefore, by replacing

$H(x)$ with $H(\overline{x}_n^-)$ for all $x \in [\overline{x}, y)$ and leaving H unchanged otherwise, the resulting $\widehat{H}$ must still be in $\mathcal{I}(F_R^\tau, F_L^\tau)$. Since v_S is strictly increasing on $[\overline{x}_n, y)$, $\widehat{H}$ must give a higher value, a contradiction. Analogous arguments also lead to a contradiction for the case of $H(\underline{x}_n) < F_L^\tau(\underline{x}_n)$, as well as $a > F^{-1}(\tau)$. Therefore, $H^C_{\underline{a}.\overline{a}}$ must be optimal for some $\underline{a} \leq \overline{a}$.

For (iv), note that F is not an extreme point of $\mathcal{I}(F_R^\tau, F_L^\tau)$ according to Theorem 1. Therefore, it is never the unique solution of (3). ■

A.6 Proofs of Proposition 3 and Proposition 5

We prove the following result that leads to Proposition 3 and Proposition 5 immediately.[39]

Theorem A.1. *Let $\overline{F}(x) := x$ and $\underline{F}(x) := 0$ for all $x \in [0,1]$. For any $J \in \mathbb{N}$, for any collection of bounded linear functionals $\{\Gamma_j\}_{j=1}^J$ on $L^1([0,1])$ and for any collection $\{\gamma_j\}_{j=1}^J \subseteq \mathbb{R}$, let $\mathcal{I}^c$ be a convex subset of $\mathcal{I}(\underline{F}, \overline{F})$ defined as*

$$\mathcal{I}^c := \left\{H \in \mathcal{I}(\underline{F}, \overline{F}) | \Gamma_j(H) \geq \gamma_j, \, \forall j \in \{1, \dots, J\}\right\}.$$

Suppose that $H \in \mathcal{I}^c$ is an extreme point of $\mathcal{I}^c$. Then there exists countably many intervals $\{[\underline{x}_n, \overline{x}_n)\}_{n=1}^\infty$ such that:

1. *$H(x) = x$ for all $x \notin \cup_{n=1}^\infty [\underline{x}_n, \overline{x}_n)$.*
2. *For all $n, m \in \mathbb{N}$, with $n \neq m$, H is constant on $[\underline{x}_n, \overline{x}_n)$ and $H(\underline{x}_n) \neq H(\underline{x}_m)$.*
3. *For all but at most J many $n \in \mathbb{N}$, $H(\underline{x}_n) = \underline{x}_n$.*

Proof. Consider any extreme point H of $\mathcal{I}^c$. We first claim that for any $x \in (0,1)$, it must be either $H(x) = x$ or $H(y) = H(x)$ for all $y \in (x, x+\delta)$ for some $\delta > 0$. To see this, note that since $\mathcal{I}^c$ is a subset of $\mathcal{I}(\underline{F}, \overline{F})$ defined by J linear constraints, Proposition 2.1 of Winkler (1988) implies that there exists $\{H_j\}_{j=1}^{J+1} \subseteq \text{ext}(\mathcal{I}(\underline{F}, \overline{F}))$ and $\{\lambda_j\}_{j=1}^{J+1} \subseteq [0,1]$ such that $H(x) = \sum_{j=1}^{J+1} \lambda_j H_j(x)$ for all $x \in [0,1]$ and $\sum_{j=1}^{J+1} \lambda_j = 1$. Now suppose that $H(x) < x$ for some $x \in (0,1)$. Then there must exist a nonempty subset $\mathcal{J} \subseteq \{1, \dots, J+1\}$ such that $H_j(x) < x$ for all $j \in \mathcal{J}$ and that $H_j(x) = x$ for all $j \in \{1, \dots, J+1\} \backslash \mathcal{J}$. Since H_j is an extreme point of $\mathcal{I}(\underline{F}, \overline{F})$ for all $j \in \mathcal{J}$, Theorem 1 implies that for each $j \in \mathcal{J}$, there exists an interval $[\underline{x}^j, \overline{x}^j)$ containing x on which H_j is constant. Let $(\underline{x}, \overline{x})$ be the interior of the intersection of $\{[\underline{x}^j, \overline{x}^j)\}_{j \in \mathcal{J}}$. Then it must be that

$$H(y) = \alpha y + (1-\alpha)\eta$$

for all $y \in (\underline{x}, \overline{x})$, for some $\eta < x$, and $\alpha \in (0,1)$. Now suppose that for any $\delta > 0$, there exists $y \in (x, x+\delta)$ such that $H(x) < H(y)$. Take any $\hat{\delta} \in (0, \min\{(1-\alpha)(x-\eta)/(1+\alpha), x - \underline{x}, \overline{x} - x\})$ and let $x_* := x - \hat{\delta}$

[39] Rolewicz (1984) characterizes the extreme points of bounded Lipschitz functions defined on the unit interval that vanish at zero, and he shows that a function is an extreme point of the unit ball of this set if and only if the absolute value of its derivative equals 1 almost everywhere (see also Rolewicz 1986; Farmer 1994; Smarzewski 1997). The convex set of interest here is different. First, functions in $\mathcal{I}(\underline{F}, \overline{F})$ are subject to an additional monotonicity constraint. Second, these functions are bounded by $\underline{F}$ and $\overline{F}$ under the pointwise dominance order, rather than the Lipschitz (semi) norm. In particular, functions in $\mathcal{I}(\underline{F}, \overline{F})$ may have unbounded derivatives, whenever well-defined. Lastly, Theorem A.1 below characterizes the extreme points of this set subject to finitely many other linear constraints, which are not present in the characterization of Rolewicz (1984).

and $x^* := x + \hat{\delta}$. Then it must be that $H(y) < x$ for any $y \in [x_*, x^*]$ and that $H(x^*) < x_*$. Moreover, the function $h : [x_*, x^*] \to [H(x_*), H(x^*)]$ defined as $h(y) := H(y)$ for all $y \in [x_*, x^*]$ must not be a step function, since otherwise, as h is right-continuous on (x_*, x^*), there must be some $\delta > 0$ such that $H(y) = h(y) = h(x) = H(x)$ for all $y \in [x, x + \delta)$, a contradiction. Meanwhile, since each functional $\Gamma_j : L^1([0,1]) \to \mathbb{R}$ is bounded, Riesz's representation implies that there must exist $\Phi_j \in L^\infty([0,1])$ such that

$$\Gamma_j(\widetilde{H}) = \int_0^1 \widetilde{H}(x)\Phi_j(x)\,\mathrm{d}x,$$

for all $\widetilde{H} \in \mathcal{I}(\underline{F}, \overline{F})$. Therefore, since any extreme point of the collection of nondecreasing, right-continuous functions $\tilde{h}$ from $[x_*, x^*]$ to $[H(x_*), H(x^*)]$ such that

$$\int_{x_*}^{x^*} \tilde{h}(x)\Phi_j(x)\,\mathrm{d}x \geq \gamma_j$$

for all $j \in \{1, \dots, J\}$ is a step function with at most $J + 1$ steps, as implied by Proposition 2.1 of Winkler (1988), the function h is not an extreme point of this collection. Thus, there exists two distinct functions $h_1, h_2 : [x_*, x^*] \to [H(x_*), H(x^*)]$ and $\lambda \in (0, 1)$ such that $h(y) = \lambda h_1(y) + (1 - \lambda)h_2(y)$ for all $y \in [x_*, x^*]$ and that

$$\int_{x_*}^{x^*} h_l(x)\Phi_j(x)\,\mathrm{d}x = \int_{x_*}^{x^*} H(x)\Phi_j(x)\,\mathrm{d}x, \tag{A.14}$$

for all $j \in \{1, \dots, J\}$ and for all $l \in \{1, 2\}$. Now let H_1, H_2 be defined as

$$H_1(y) := \begin{cases} H(y), & \text{if } y \notin [x_*, x^*] \\ h_1(y), & \text{if } y \in [x_*, x^*] \end{cases} ; \quad H_2(y) := \begin{cases} H(y), & \text{if } y \notin [x_*, x^*] \\ h_2(y), & \text{if } y \in [x_*, x^*] \end{cases} .$$

Then, $H = \lambda H_1 + (1 - \lambda)H_2$ and $H_1 \neq H_2$. Moreover, since $h_1(y), h_2(y) \leq H(x^*) < x_*$ for all $y \in [x_*, x^*]$, and since $H \in \mathcal{I}(\underline{F}, \overline{F})$, it must be that both H_1 and H_2 are in $\mathcal{I}(\underline{F}, \overline{F})$. Furthermore, by (A.14), it must be that

$$\begin{aligned}
\Gamma_j(H_l) = \int_0^1 H_l(x)\Phi_j(x)\,\mathrm{d}x &= \int_{[0,1]\setminus[x_*,x^*]} H(x)\Phi_j(x)\,\mathrm{d}x + \int_{x_*}^{x^*} h_l(x)\Phi_j(x)\,\mathrm{d}x \\
&= \int_{[0,1]\setminus[x_*,x^*]} H(x)\Phi_j(x)\,\mathrm{d}x + \int_{x_*}^{x^*} H(x)\Phi_j(x)\,\mathrm{d}x \\
&= \int_0^1 H(x)\Phi_j(x)\,\mathrm{d}x \\
&\geq \gamma_j,
\end{aligned}$$

for all $j \in \{1, \dots, J\}$ and for all $l \in \{1, 2\}$. Thus, $H_1, H_2 \in \mathcal{I}^c$, a contradiction. Together, for any $x \in (0, 1)$, it must be either $H(x) = x$ or $H(y) = H(x)$ for all $y \in (x, x + \delta)$ for some $\delta > 0$.

Let $X \subseteq [0, 1]$ be the collection of $x \in [0, 1]$ such that $H(x) = x$. For any $x \notin X$, let $\overline{\delta}_x := \sup\{y \in [0, 1] | H(y) = H(x)\}$ and $\underline{\delta}_x := \inf\{y \in [0, 1] | H(y) = H(x)\}$. Then it must be $\underline{\delta}_x < \overline{\delta}_x$ for all $x \notin X$. Moreover, for any $x, y \in [0, 1]\setminus X$ with $x < y$, $H(x) < H(y)$ if and only if $\overline{\delta}_x < \underline{\delta}_y$. Therefore, $[0, 1]\setminus X$ can be expressed as a union of a collection I of disjoint intervals. Since I is a collection of disjoint intervals on

$[0,1]$, each element of I must uniquely contain at least one rational number. Thus, there exists an injective map from the collection I to a subset of rational numbers in $[0,1]$, and therefore the collection I must be countable.

Enumerate I as $\{[\underline{x}_n, \overline{x}_n)\}_{n=1}^{\infty}$ and suppose that there is a subset $\mathcal{N}$ of these intervals, with $|\mathcal{N}| > J$, such that $H(\underline{x}_n) < \underline{x}_n$. For each $n \in \mathcal{N}$, since $H(\underline{x}_n) < \underline{x}_n$ and since $H(x) = x$ for all $x \notin \cup_{n=1}^{\infty}[\underline{x}_n, \overline{x}_n)$, H must be discontinuous at $\underline{x}_n$. Let $\eta_n := H(\underline{x}_n) - H(\underline{x}_n^-)$ for all $n \in \mathcal{N}$, and let $\eta := \min\{\eta_n\}_{n\in\mathcal{N}}$. Furthermore, let $\phi_j^n \in \mathbb{R}$ be defined as

$$\phi_j^n := \int_{\underline{x}_n}^{\overline{x}_n} \Phi_j(x)\,\mathrm{d}x,$$

for all $n \in \mathcal{N}$ and for all $j \in \{1,\dots,J\}$. Then the $|\mathcal{N}| \times J$ matrix $\Phi := (\phi_j^n)_{j\in\{1,\dots,J\}}^{n\in\mathcal{N}}$ is a linear map from $\mathbb{R}^{|\mathcal{N}|}$ to $\mathbb{R}^J$. Since $|\mathcal{N}| > J$, $\dim(\text{null}(\Phi)) \geq 1$, and thus there must exists a nonzero vector $\{\hat{h}_n\}_{n\in\mathcal{N}}$ such that

$$\sum_{n\in\mathcal{N}} \phi_j^n \hat{h}_n = 0, \tag{A.15}$$

for all $j \in \{1,\dots,J\}$. Let $\varepsilon := \min\{\eta/4|\hat{h}_n|, (\underline{x}_n - H(\underline{x}_n))/4|\hat{h}_n|\}_{n\in\mathcal{N}}$, and let $\widehat{H}$ be defined as

$$\widehat{H}(x) := \begin{cases} 0, & \text{if } x \notin \cup_{n\in\mathcal{N}}[\underline{x}_n, \overline{x}_n) \\ \varepsilon\hat{h}_n, & \text{if } x \in [\underline{x}_n, \overline{x}_n),\ n \in \mathcal{N} \end{cases}.$$

Then, since $\{\hat{h}_n\}_{n\in\mathcal{N}}$ is a nonzero vector in $\mathbb{R}^{|\mathcal{N}|}$ and since $\varepsilon > 0$, $\hat{H} \neq 0$. Moreover, since $\varepsilon < \eta/4|\hat{h}_n|$ for all $n \in \mathcal{N}$, $H(x) - |\widehat{H}(x)| = H(\underline{x}_n) - \varepsilon|\hat{h}_n| > H(\underline{x}_n) - \eta/2 > H(\underline{x}_n^-) + \eta/4 > H(x) + |\widehat{H}(x)|$ for all $x < \underline{x}_n$ and for all $n \in \mathcal{N}$. Therefore, both $H + \widehat{H}$ and $H - \widehat{H}$ are nondecreasing. Meanwhile, since for any $n \in \mathcal{N}$ and for any $x \in [\underline{x}_n, \overline{x}_n)$, $H(x) + |\widehat{H}(x)| = H(\underline{x}_n) + \varepsilon|\hat{h}_n| < \underline{x}_n$, both $H + \widehat{H}$ and $H - \widehat{H}$ are in $\mathcal{I}(\underline{F}, \overline{F})$. In addition, by (A.15), for any $j \in \{1,\dots,J\}$,

$$\begin{aligned}\int_0^1 (H(x) + \widehat{H}(x))\Phi_j(x)\,\mathrm{d}x &= \int_{[0,1]\setminus\cup_{n\in\mathcal{N}}[\underline{x}_n,\overline{x}_n)} H(x)\Phi_j(x)\,\mathrm{d}x + \int_{\cup_{n\in\mathcal{N}}[\underline{x}_n,\overline{x}_n)} H(x)\Phi_j(x)\,\mathrm{d}x + \varepsilon\sum_{n\in\mathcal{N}}\hat{h}_n\phi_j^n \\ &= \int_0^1 H(x)\Phi_j(x)\,\mathrm{d}x \\ &\geq \gamma_j,\end{aligned}$$

and

$$\begin{aligned}\int_0^1 (H(x) - \widehat{H}(x))\Phi_j(x)\,\mathrm{d}x &= \int_{[0,1]\setminus\cup_{n\in\mathcal{N}}[\underline{x}_n,\overline{x}_n)} H(x)\Phi_j(x)\,\mathrm{d}x + \int_{\cup_{n\in\mathcal{N}}[\underline{x}_n,\overline{x}_n)} H(x)\Phi_j(x)\,\mathrm{d}x - \varepsilon\sum_{n\in\mathcal{N}}\hat{h}_n\phi_j^n \\ &= \int_0^1 H(x)\Phi_j(x)\,\mathrm{d}x \\ &\geq \gamma_j.\end{aligned}$$

Together, both $H + \widehat{H}$ and $H - \widehat{H}$ are in $\mathcal{I}^c$ and hence H is not an extreme point, a contradiction. This completes the proof. ■

Proofs of Proposition 3 and Proposition 5. Note that since $|\phi_e(x|e)|$ is dominated by an integrable function on $[0,1]$, one can apply the dominated convergence theorem to show that the objective function of both (7) and (10) are continuous in (H,e) and $(H,\underline{z})$, respectively. Similarly, the constraint set can be shown to be closed. Therefore, both (7) and (10) admit a solution.

Consequently, since for any fixed e and $\underline{z}$, the objective is continuous in H and the feasible set is compact and convex in (7) and (10), respectively, Proposition 3 and the first part of Proposition 5 follow immediately from Theorem A.1, with $J=2$ and $J=1$, respectively. This is because any H satisfying conditions 1 through 3 corresponds to a contingent debt contract with at most J non-defaultable face values. The uniqueness part of Proposition 5 further follows from the fact that the objective of (10) is strictly convex in H when $\Phi(\cdot|s)$ has full support for all $s\in S$, and hence, every solution must be an extreme point of the feasible set. ■

A.7 Proof of Proposition 4

Let $\Pi^*(e)$ be the value of the entrepreneur's problem (7) for a fixed $e\in[0,\bar{e}]$. We first show that there exists Lagrange multipliers $\lambda_1^*\neq 0$ and $\lambda_2^*\geq 0$ such that

$$\Pi^*(e)=\sup_{H\in\mathcal{I}(\underline{F},\overline{F})}\left[\int_0^1(x-H(x))\phi(x|e)\,\mathrm{d}x+\lambda_1^*\left(\int_0^1(x-H(x))\phi_e(x|e)\,\mathrm{d}x-C'(e)\right)\right.$$
$$\left.+\lambda_2^*\left(\int_0^1 H(x)\phi(x|e)\,\mathrm{d}x-(1+r)I\right)\right]. \tag{A.16}$$

To this end, we adopt a similar argument as Nikzad (2023). For any fixed $e\in[0,\bar{e}]$ and for any $\gamma\in\mathbb{R}$, let $M_e(\gamma)$ be the value of

$$\begin{aligned}\sup_{H\in\mathcal{I}(\underline{F},\overline{F})}&\left[\int_0^1[x-H(x)]\phi(x|e)\,\mathrm{d}x-C(e)\right]\\ \text{s.t.}\ &\int_0^1[x-H(x)]\phi_e(x|e)\,\mathrm{d}x=C'(e)\\ &\int_0^1 H(x)\phi(x|e)\,\mathrm{d}x\geq\gamma.\end{aligned} \tag{A.17}$$

Note that

$$M_e((1+r)I)=\Pi^*(e)=\int_0^1(x-H^*(x))\phi(x|e)\,\mathrm{d}x-C(e), \tag{A.18}$$

where H^* is a solution of (7) with a fixed e. Moreover, M_e is nonincreasing and concave in γ. Indeed, monotonicity follows from the ordered structure of the feasible set as γ increases. For concavity, consider any $\gamma_1,\gamma_2\in\mathbb{R}$ and let $\gamma^\lambda:=\lambda\gamma_1+(1-\lambda)\gamma_2$ for any $\lambda\in(0,1)$. Since (A.17) admits a solution, there exists $H_1,H_2\in\mathcal{I}(\underline{F},\overline{F})$ such that

$$\int_0^1(x-H_1(x))\phi(x|e)\,\mathrm{d}x-C(e)=M(\gamma_1);\quad \int_0^1(x-H_2(x))\phi(x|e)\,\mathrm{d}x-C(e)=M(\gamma_2).$$

Furthermore,

$$\int_0^1 (x - H_i(x))\phi_e(x|e)\,\mathrm{d}x = C'(e)$$
$$\int_0^1 H_i(x)\phi(x|e)\,\mathrm{d}x \geq \gamma_i$$

for $i \in \{1,2\}$. Let $H^\lambda := \lambda H_1 + (1-\lambda)H_2$, we must have $H^\lambda \in \mathcal{I}(\underline{F}, \overline{F})$ and

$$\int_0^1 (x - H^\lambda(x))\phi_e(x|e)\,\mathrm{d}x = C'(e)$$
$$\int_0^1 H^\lambda(x)\phi(x|e)\,\mathrm{d}x \geq \gamma^\lambda.$$

Thus,

$$\begin{aligned}
M_e(\gamma^\lambda) \geq & \int_0^1 (x - H^\lambda(x))\phi(x|e)\,\mathrm{d}x - C(e) \\
= & \lambda \int_0^1 (x - H_1(x))\phi(x|e)\,\mathrm{d}x + (1-\lambda)\int_0^1 (x - H_2(x))\phi(x|e)\,\mathrm{d}x \\
= & \lambda M_e(\gamma_1) + (1-\lambda)M_e(\gamma_2).
\end{aligned}$$

Since M_e is nonincreasing and concave, and since $(1+r)I$ is an interior of the set

$$\left\{ \int_0^1 H(x)\phi(x|e)\,\mathrm{d}x \middle| H \in \mathcal{I}(\underline{F}, \overline{F}),\ \int_0^1 (x - H(x))\phi_e(x|e)\,\mathrm{d}x = C'(e) \right\},$$

there exists $\lambda_2^* \geq 0$ such that

$$M_e(\gamma) \leq M_e((1+r)I) - \lambda_2^*(\gamma - (1+r)I)$$

for all $\gamma \in \mathbb{R}$. Meanwhile, for any $H \in \mathcal{I}(\underline{F}, \overline{F})$ such that

$$\int_0^1 (x - H(x))\phi_e(x|e)\,\mathrm{d}x = C'(e), \tag{A.19}$$

it must be that

$$M_e\left(\int_0^1 H(x)\phi(x|e)\,\mathrm{d}x\right) \geq \int_0^1 (x - H(x))\phi(x|e)\,\mathrm{d}x - C(e),$$

by the definition of M_e. Together with (A.18), we have

$$\begin{aligned}
M_e((1+r)I) = & \int_0^1 (x - H^*(x))\phi(x|e)\,\mathrm{d}x - C(e) \\
\geq & \int_0^1 (x - H(x))\phi(x|e)\,\mathrm{d}x - C(e) + \lambda_2^*\left(\int_0^1 H(x)\phi(x|e)\,\mathrm{d}x - (1+r)I\right),
\end{aligned} \tag{A.20}$$

for all $H \in \mathcal{I}(\underline{F}, \overline{F})$ such that (A.19) holds. Since H^* is feasible for (7) with the fixed e, (A.20) implies

$$\begin{aligned}
&\int_0^1 (x - H^*(x))\phi(x|e)\,\mathrm{d}x + \lambda_2^* \left(\int_0^1 H^*(x)\phi(x|e)\,\mathrm{d}x - (1+r)I \right) \\
\geq &\int_0^1 (x - H(x))\phi(x|e)\,\mathrm{d}x + \lambda_2^* \left(\int_0^1 H(x)\phi(x|e)\,\mathrm{d}x - (1+r)I \right),
\end{aligned} \tag{A.21}$$

for all $H \in \mathcal{I}(\underline{F}, \overline{F})$ satisfying (A.19). Now let

$$\mathcal{L}_e(H;\lambda) := \int_0^1 (x - H(x))\phi(x|e)\,\mathrm{d}x - C(e) + \lambda \left(\int_0^1 H(x)\phi(x|e)\,\mathrm{d}x - (1+r)I \right),$$

and let $\mathcal{L}_e(\lambda)$ be the value of

$$\begin{aligned}
&\sup_{H \in \mathcal{I}(\underline{F}, \overline{F})} \mathcal{L}_e(H;\lambda) \\
&\quad \text{s.t.} \int_0^1 (x - H(x))\phi_e(x|e)\,\mathrm{d}x = C'(e)
\end{aligned} \tag{A.22}$$

Then (A.21) implies that H^* solves (A.22) with $\lambda = \lambda_2^*$ and

$$\mathcal{L}_e(\lambda_2^*) = \int_0^1 (x - H^*(x))\phi(x|e)\,\mathrm{d}x - C(e).$$

Meanwhile, by the definition of $\mathcal{L}_e(\lambda)$,

$$\mathcal{L}_e(\lambda) \geq \int_0^1 (x - H(x))\phi(x|e)\,\mathrm{d}x - C(e)$$

for all feasible H of (7) with fixed e. Finally, since the constraint in (A.22) is an equality, standard results (see, e.g., Theorem 3.12 of Anderson and Nash 1987) implies that there exits $\lambda_1 \neq 0$ such that (A.16) holds.

For any fixed $e \in [0, \bar{e}]$, since the primal problem (7) is convex for any fixed $e \in [0, \bar{e}]$, there exists an extreme point H^* of the feasible set that attains $\Pi^*(e)$. By Theorem A.1, there exists a countable collection of intervals $\{[\underline{x}_n, \overline{x}_n)\}_{n=1}^{\infty}$ such that H^* satisfies conditions 1 through 3 for $J = 2$. Meanwhile, as established above, H^* must also solves the dual problem (A.16) of (7) for this fixed e. Note that the dual problem can be written as

$$\sup_{H \in \mathcal{I}(\underline{F}, \overline{F})} \left[\int_0^1 H(x)[(1+\lambda_2^*)\phi(x|e) - \lambda_1^* \phi_e(x|e)]\,\mathrm{d}x + \kappa \right],$$

with $\kappa \in \mathbb{R}$ being a constant that does not depend on H. Moreover,

$$(1+\lambda_2^*)\phi(x|e) - \lambda_1^* \phi_e(x|e) \geq 0 \iff \frac{\phi_e(x|e)}{\phi(x|e)} \leq \frac{1+\lambda_2^*}{\lambda_1^*}.$$

Since $\phi_e(\cdot|e)/\phi(\cdot|e)$ is at most N-peaked, there must be a finite interval partition $\{I_k\}_{k=1}^K$ of $[0,1]$ with $K \leq 2N$ such that $\phi_e(x|e)/\phi(x|e) - (1+\lambda_2^*)/\lambda_1^*$ takes the same sign for all $x \in I_k$.

Therefore, if there are more than $N+1$ intervals on which H^* is constant, then either there are at least two of them contained in a single interval I_k with $\phi_e(x|e)/\phi(x|e) < (1+\lambda_2^*)/\lambda_1^*$ for all $x \in I_k$, or there is at least one of them contained in an interval I_j with $\phi_e(x|e)/\phi(x|e) > (1+\lambda_2^*)/\lambda_1^*$ for all $x \in I_j$. If there are two intervals $[\underline{x}_n, \overline{x}_n)$, $[\underline{x}_m, \overline{x}_m)$, with $\overline{x}_n \leq \underline{x}_m$, that are contained in some I_k with $\phi_e(x|e)/\phi(x|e) < (1+\lambda_2^*)/\lambda_1^*$ for all $x \in I_k$, then, since by condition 2 of Theorem A.1, $H^*(\underline{x}_n) < H^*(\underline{x}_m)$, for H^{**} defined as

$$H^{**}(x) := \begin{cases} H^*(x), & \text{if } x \notin [\underline{x}_n, \overline{x}_m) \\ H^*(\underline{x}_n), & \text{if } x \in [\underline{x}_n, \overline{x}_m) \end{cases},$$

for all $x \in [0,1]$, $H^{**} \in \mathcal{I}(\underline{F}, \overline{F})$ and yields a higher value to the objective of (A.16) than H^*. Likewise, if there is at least one interval on which H^* is constant that is contained in some I_j such that $\phi_e(x|e)/\phi(x|e) < (1+\lambda_2^*)/\lambda_1^*$ for all $x \in I_j$, then, since $H^*(x) < x$ for all $x \in (\underline{x}_n, \overline{x}_n)$, for H^{**} defined as

$$H^{**}(x) := \begin{cases} H^*(x), & \text{if } x \notin I_j \\ \max\{x, H^*(\overline{x}_n)\}, & \text{if } x \in I_j \end{cases},$$

for all $x \in [0,1]$, $H^{**} \in \mathcal{I}(\underline{F}, \overline{F})$ and yields a higher value to the objective of (A.16) than H^*. Thus, H^* cannot be a solution of the dual problem (A.16) for this fixed e, a contradiction. Consequently, the solution H^* to the primal problem (7) for any fixed $e \in [0, \bar{e}]$ cannot admit more than $N+1$ intervals where H^* is constant. As a result, H^* is a contingent debt contract with at most $N+1$ face values. Since $e \in [0, \bar{e}]$ is arbitrary, this completes the proof. ■