1-18-2023

Extreme Points of First-Order Stochastic Dominance Intervals: Theory and Applications

Kai Hao Yang

Alexander K. Zentefis

EXTREME POINTS OF FIRST-ORDER STOCHASTIC DOMINANCE INTERVALS: THEORY AND APPLICATIONS

By

Kai Hao Yang and Alexander K. Zentefis

January 2023

COWLES FOUNDATION DISCUSSION PAPER NO. 2355

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Extreme Points of First-Order Stochastic Dominance Intervals: Theory and Applications*

Kai Hao Yang† Alexander K. Zentefis‡

January 18, 2023

Abstract

We characterize the extreme points of first-order stochastic dominance (FOSD) intervals and show how these intervals are at the heart of many topics in economics. Using knowledge of these extreme points, we characterize the distributions of posterior quantiles under a given prior, leading to an analogue of a classical result regarding the distribution of posterior means. We apply this analogue to various economic subjects, including the psychology of judgement, political economy, and Bayesian persuasion. In addition, FOSD intervals provide a common structure to security design. We use the 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, first-order stochastic dominance, posterior quantiles, overconfidence, gerrymandering, Bayesian persuasion, security design

*This paper subsumes a previously circulated paper titled "Distributions of Posterior Quantiles and Economic Applications." We are 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, Paul Goldsmith-Pinkham, Marina Halac, Ángel Hernando-Veciana, Ryota Iijima, Emir Kamenica, Nicolas Lambert, Elliot Lipnowski, Jay Lu, Leslie Marx, Moritz Meyer-ter-Vehn, Stephen Morris, Barry Nalebuff, Aniko Öry, Wolfgang Pesendorfer, Benjamin Polak, Doron Ravid, Anne-Katrin Roesler, Hamid Sabourian, Fedor Sandomirskiy, Christoph Schlom, Chris Shannon, Quitzé Valenzuela-Stookey, Vasiliki Skreta, Nicholas Stephanopoulos, Philipp Strack, Roland Strausz, Tomasz Strzalecki, Omer Tamuz, 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. 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

The notion of first-order stochastic dominance has been part of economics since at least the late 1960s. At that time, several authors established its importance for the analysis of choice under risk (Hadar and Russell 1969; Hanoch and Levy 1969; Rothschild and Stiglitz 1970; Whitmore 1970). See Kroll and Levy (1980), Bawa (1982), and Levy (1990) for surveys of the body of work that followed. In this paper, we show that many well-known economic questions can be recast in terms of first-order stochastic dominance. This reframing connects seemingly unrelated subjects in economics—including optimal security design, Bayesian persuasion, the psychology of judgment, and partisan redistricting—revealing that many of these subjects' insights share a common structure.

Our main result characterizes the extreme points of first-order stochastic dominance (FOSD) *intervals*. These intervals describe sets of distributions that dominate a distribution and are simultaneously dominated by another distribution, in the sense of FOSD. The convexity of FOSD intervals means that their *extreme points* are fundamental to understanding their properties. We show that a distribution is an extreme point of an FOSD interval if and only if the distribution coincides with one of the FOSD interval's bounds, or is constant on an interval that has at least one end attached to at least one of the FOSD interval's bounds.

This characterization is useful to economics because various settings studied in different literatures can be reformulated into problems involving FOSD intervals. Many canonical and novel results in the relevant literatures follow from the characterization. We demonstrate this through two broad classes of economic applications.

In the first class of applications, we prove an analogue to a celebrated result in probability theory that has been widely used in economics. Consider a random variable and a signal for it. For each signal realization, a posterior belief is determined by Bayes' rule. 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, and vice versa. Rothschild and Stiglitz (1970) made clear the economic implications of Strassen's theorem, in particular toward the theory of risk. The Bayesian persuasion literature has extensively applied this theorem to obtain explicit solutions to many persuasion problems (see, for example, Gentzkow and Kamenica 2016 and Dworczak and Martini 2019).

Instead of posterior means, one can derive many other statistics of a posterior. Using the characterization of the extreme points of FOSD intervals, we characterize the distributions of posterior *quantiles*, leading to an analogue of Strassen's theorem. The distributions of posterior quantiles coincide with an FOSD interval bounded by an upper and a lower truncation of the prior.

The characterization of the distributions of posterior quantiles further leads to many economic applications. For example, in the psychology of judgement, a seminal result on identifying overconfidence follows immediately. It is well documented that individuals can appear to be over or under confident when evaluating themselves (Alicke, Klotz, Breitenbecher, Yurak and Vredenburg 1995; De Bondt and Thaler 1995; Moore 2007; Kruger, Windschitl, Burrus, Fessel and Chambers 2008). Observing this in the literature, Benoît and Dubra (2011) show that this finding alone does not imply irrationality. They consider a setting where individuals are asked to rank their ability on a certain task (e.g., driving skills) relative to a given population. The main result of Benoît and Dubra (2011) is a characterization of the set of self-ranking data that are rationalizable by a Bayesian model. From this characterization, they 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). As an immediate corollary, our characterization of the distributions of posterior quantiles generalizes this result. This generalization extends the setting beyond self-ranking questions on a relative scale to self-evaluation questions on an absolute scale, such as raw test scores or the probability of employment after graduation, as studied in Weinstein (1980).

As another example, our characterization of the distributions of posterior quantiles leads to novel results in political economy, in particular on gerrymandering, or the manipulation of electoral district boundaries. In this setting, citizens identify with an ideal position on political issues along a spectrum. The variety of positions is represented as a distribution, which we can call a prior. An electoral map segments citizens into districts, which splits the prior distribution into different parts. This electoral map can be regarded as a signal, and the distribution of ideal positions within each district of the map can be interpreted as a posterior. If each district elects a representative holding the district's median position (Downs 1957; Black 1958), the composition of the legislative body (i.e., the distribution of ideal positions of elected representatives) can then be represented as a distribution of posterior medians. Our characterization of the distributions of posterior quantiles fully describes the scope of legislatures that unrestrained gerrymandering can achieve. Gerrymandering can induce *any* legislature within the bounds of two extremes: an "all-left" body and an "all-right" body. In the former, the composition of the legislature only reflects citizens' ideal positions that are left of the population median; whereas in the latter, the composition of the legislature only reflects citizens' ideal positions that are right of the population median. At the same time, any compositions *beyond* the "all-left" and the "all-right" bodies (e.g., anything more right-leaning than the distribution of citizens' ideal positions that are to the right of the population median) are not possible through *any* kind of gerrymandering. Thus, the scope of unrestrained gerrymandering is identified by the all-left and all-right bodies, as well as *anything in between*.

Our third application of the distributions of posterior quantiles is to Bayesian persuasion. Kamenica and Gentzkow (2011) provide a framework for studying a sender's communication to a receiver under the commitment assumption. A practical challenge, however, is that the concavification approach used in this literature loses tractability as the number of states increases. An exception is when the state is one-dimensional and only posterior *means* are payoff-relevant to the sender. Our characterization complements this literature, as it brings tractability to settings where only posterior *quantiles* are payoff-relevant to the sender. For example, our characterization leads to explicit solutions to a persuasion problem where the sender's payoff is state-independent, and the receiver chooses an action to match the state and minimizes the *absolute* loss, rather than the quadratic loss. We show how this simple change has substantive implications for the the type of information the sender optimally discloses.

In addition to characterizing the distributions of posterior quantiles, we apply our characterization of the extreme points of FOSD intervals to the security design literature. We show how FOSD intervals present a unifying structure to security design, and we uncover common features of the optimal securities in a wide class of security design problems. A typical setting in security design involves an entrepreneur with an asset but no money, and investors with money but no asset. The entrepreneur considers the type of security to issue to investors in exchange for funding. The entrepreneur typically has more information than the investors about how much the asset actually earned or about the effort the entrepreneur exerted to jump-start the asset.

Two widely adopted assumptions in the literature make the security design problem amenable to FOSD intervals. The first is limited liability. The entrepreneur cannot pay the investors any more than all the asset's cash flow, and the investors cannot receive anything less than zero. Limited liability places natural upper and lower stochastic bounds on the security's payoff. The second assumption is that the security's payoff is monotone in the asset's cash flow. See Innes (1990), Nachman and Noe (1994), and DeMarzo and Duffie (1999) for justifications of this assumption. Monotonicity introduces a natural first-order stochastic dominance between the asset and the security.

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) consider a situation with adverse selection. Both papers derive a standard debt contract as an optimal security, which promises either a constant payment or the asset's realized cash flow, whichever is smaller. Many papers in security design that followed were influenced by the Innes (1990) or DeMarzo and Duffie (1999) environment. (See, for example, Schmidt 1997; Casamatta 2003 and Eisfeldt 2004; Biais and Mariotti 2005.)

But the optimality of standard debt in Innes (1990) and DeMarzo and Duffie (1999) relies on another crucial assumption: The asset's cash flow distribution (or signal about the cash flow's distribution) satisfies the monotone likelihood ratio property (MLRP). The assumption is reasonable, but not without limitations (Hart 1995). By recasting the security design problem using FOSD intervals, our characterization of the extreme points allows us to solve for the optimal security without reliance on MLRP. This reframing also demonstrates that many security design problems, whether afflicted by moral hazard or adverse selection, can be unified under a common framework.

Without assuming MLRP, we show that the optimal security is not necessarily standard debt, but *contingent* debt. For this security, the face value of the entrepreneur's debt to investors is contingent on the realized cash flow of the asset. The nature of standard debt contracts—which grants the entrepreneur only residual rights and never has the entrepreneur share partial equity with the investor—is preserved even without assuming MLRP. The only difference is that the entrepreneur may be liable for more when the asset earns more.[1]

Overall, this paper uncovers the common underlying role of FOSD 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. Not only do several classical results follow from our main characterization, but we also use that characterization to develop new findings that otherwise would have been challenging to obtain without it.

Related Literature. This paper relates to several areas. The main result connects to characterizations of extreme points of convex sets. 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 $\mathbb{R}^n$, which is often referred to as majorization orbits.[2] They show that the extreme points of this set coincide with the permutations of x_0. Ryff (1967) extends this result to infinite dimensional spaces. Kleiner, Moldovanu and Strack (2021) characterize the extreme points of a subset of orbits under an additional monotonicity assumption, which is equivalent to the set of probability distributions being either a mean-preserving spread or mean-preserving contraction of a probability distribution on $\mathbb{R}$. Independently, Arieli, Babichenko, Smorodinsky and Yamashita (forthcoming) also characterize the extreme points

[1] Contingent debt contracts share some similarity with state-contingent debt instruments (SCDIs) from the sovereign debt literature, which tie a country's principal or interest payments to its nominal GDP (Lessard and Williamson 1987; Shiller 1994; Borensztein and Mauro 2004).

[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.

of mean-preserving contractions of a probability distribution on $\mathbb{R}$ and show that they coincide to a class of signals they refer to as "bi-pooling."

Compared to Kleiner, Moldovanu and Strack (2021), this paper characterizes the extreme points of distributions under the *first-order* stochastic dominance order, rather than the second-order stochastic dominance order. Moreover, our characterization applies to an *interval* of distributions: those that are dominated by a distribution and dominate another distribution at the same time. This contrasts with an orbit, which contains only distributions that are either dominated by one distribution or dominate another.[3] Furthermore, since any FOSD interval can be written as a convex polyhedron defined by finitely many linear inequalities when restricted to distributions supported on a common finite set, our characterization can be regarded as a continuum analogue of the well-known fact that extreme points of such an n-dimensional polyhedron are characterized by at least n binding linear constraints (see, for instance, proposition 15.2 of Simon 2011).

Several recent papers exploit properties of extreme points to derive economic implications. Bergemann, Brooks and Morris (2015) use the extreme points of the convex set of market segments that induce the same optimal monopoly price to construct the consumer-surplus-maximizing market segmentation. Lipnowski and Mathevet (2018) use the extreme points of posterior covers to reduce the support of optimal signals in a general persuasion framework. Kleiner, Moldovanu and Strack (2021) use the extreme points of majorization orbits to derive novel proofs of the celebrated Border's condition, the Bayesian-dominance equivalence result, optimality of bi-pooling signals in mean-based persuasion settings, as well as the equivalence of persuasion and a class of delegation problems. Finally, several works in mechanism design use the extreme points of feasible mechanisms to establish the optimality of rationing and randomized posted prices (e.g., Dworczak ⓡ Kominers ⓡ Akbapour 2021, Loertscher and Muir 2022, Kang 2022). These papers exploit the result of Winkler (1988), which characterizes the extreme points of convex subsets defined by finitely many linear inequalities.

The first application of this paper 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). Our application can be regarded as a complement, as it characterizes the distributions of posterior quantiles, instead of means. This characterization generalizes the results of Benoît and Dubra (2011), who identify the

[3]The qualitative structure of the extreme points of FOSD intervals shares some similarity with that given by Kleiner, Moldovanu and Strack (2021). In particular, any extreme point of an FOSD interval either must coincide with one of the bounds or must pool all states in an interval into one mass point.

Bayesian-rationalizable self-ranking data where subjects place themselves relative to the population according to a posterior quantile.

Our gerrymandering results are related to the literature on redistricting, particularly to Owen and Grofman (1988), Friedman and Holden (2008), Gul and Pesendorfer (2010), and Kolotilin and Wolitzky (2020), who also adopt the distribution-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, our result characterizes the *feasible compositions* of a legislative body that a district map can induce.

Our application to Bayesian 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). We complement this literature by providing a foundation for solving communication problems where only the posterior *quantiles* are payoff-relevant.

Finally, our reframing of security design using FOSD intervals 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. We generalize and unify results in those seminal papers under a common structure, revealing how security design problems can be solved using FOSD intervals when either type of asymmetric information is at play.

Outline. The remainder of the paper proceeds as follows. Section 2 gives the main result. Section 3 uses the main result to characterize the distributions of posterior quantiles. Economic applications to the psychology of judgment, gerrymandering, and Bayesian persuasion follow in that section. Section 4 illustrates FOSD intervals as a unifying framework for security design with limited liability when there is either moral hazard or adverse selection. Section 5 concludes.

2 Extreme Points of First-Order Stochastic Dominance Intervals

2.1 Notation

Let $\mathcal{F}$ be the collection of cumulative distribution functions (CDFs) on $\mathbb{R}$.[4] For any $F, G \in \mathcal{F}$ such that $G(x) \leq F(x)$ for all $x \in \mathbb{R}$, let

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

Namely, $\mathcal{I}(F, G)$ is the collection of distributions that dominate F and simultaneously are dominated by G in the sense of *first-order stochastic dominance* (FOSD). In other words, $\mathcal{I}(F, G)$ is the first-order stochastic dominance *interval* between G and F.

For any $F \in \mathcal{F}$ and for any $x \in \mathbb{R}$, let $F(x^-) := \lim_{y \uparrow x} F(x)$ denote the left-limit of F at x. Meanwhile, for any $F \in \mathcal{F}$ and for any $\tau \in (0, 1)$, let F^{-1} be the *quantile* function of F. Namely, $F^{-1}(\tau) := \inf\{x \in \mathbb{R} | F(x) \geq \tau\}$.[5]

2.2 Extreme Points of First-Order Stochastic Dominance Intervals

For any two distributions F and G, the FOSD interval $\mathcal{I}(F, G)$ is a convex set. Our main result characterizes the *extreme points* of this set, which are in turn useful for understanding properties of the FOSD interval.

Specifically, H is an extreme point of $\mathcal{I}(F, G)$ if H cannot be written as a convex combination of two distinct elements of $\mathcal{I}(F, G)$. A well-known equivalent definition is that H is an extreme point of $\mathcal{I}(F, G)$ if and only if, for any non-zero (measurable) function $\widehat{H}$ on $\mathbb{R}$, either $\widehat{H} + H \notin \mathcal{I}(F, G)$, or $H - \widehat{H} \notin \mathcal{I}(F, G)$. Theorem 1 characterizes the extreme points of $\mathcal{I}(F, G)$.

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

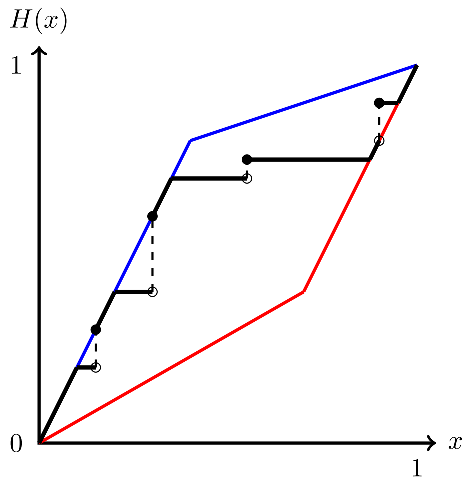
1. *$H(x) \in \{G(x), F(x)\}$ for all $x \notin \cup_{n=1}^{\infty} [\underline{x}_n, \overline{x}_n)$.*

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

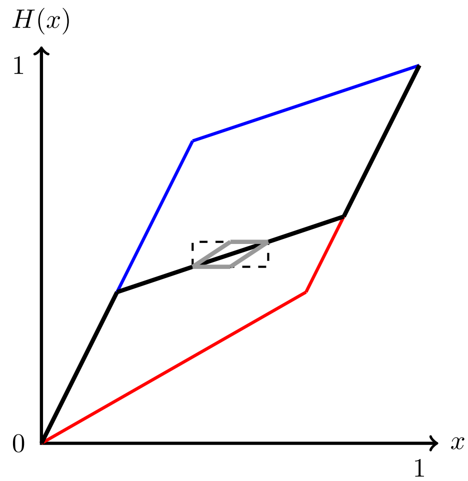
Figure IA depicts an extreme point of an FOSD interval $\mathcal{I}(F, G)$, where the blue CDF is the lower bound F, and the red CDF is the upper bound G. According to Theorem 1, any

[4] $\mathcal{F}$ is endowed with the weak-* topology and the induced Borel σ-algebra, unless otherwise specified.

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



(A) An Extreme Point



(B) Not an Extreme Point

Figure I
Extreme Points of $\mathcal{I}(F, G)$

extreme point H of this FOSD interval must either coincide with one of the bounds, or be constant on an interval, 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 below. For the sufficiency part, consider any H that satisfies conditions 1 and 2 of Theorem 1, and consider any non-zero function $\widehat{H}$. Clearly, if either $H + \widehat{H}$ or $H - \widehat{H}$ is not a CDF, then it is not an element of $\mathcal{I}(F, G)$. If both $H + \widehat{H}$ and $H - \widehat{H}$ are CDFs, then they must both be nondecreasing. Since $\widehat{H}$ is non-zero, there exists x_0 such that $\widehat{H}(x_0) \neq 0$. If $x_0 \notin \cup_{n=1}^{\infty}[\underline{x}_n, \overline{x}_n)$, then either $H(x_0) + |\widehat{H}(x_0)| = F(x_0) + |\widehat{H}(x_0)| > F_0(x_0)$ or $H(x_0) - |\widehat{H}(x_0)| = G(x_0) - |\widehat{H}(x_0)| < G_0(x_0)$. Alternatively, if $x \in [\underline{x}_n, \overline{x}_n)$ for some $n \in \mathbb{N}$, then since $H + \widehat{H}$ and $H - \widehat{H}$ are nondecreasing and since H is constant on $[\underline{x}_n, \overline{x}_n)$, $\widehat{H}$ must be constant on $[\underline{x}_n, \overline{x}_n)$ as well. This, in turn, implies that either $H(\underline{x}_n) + |\widehat{H}(\underline{x}_n)| = F(\underline{x}_n) + |\widehat{H}(x_0)| > F(\underline{x}_n)$, or $H(\overline{x}_n) - |\widehat{H}(\overline{x}_n)| = G(\overline{x}_n^-) - |\widehat{H}(x_0)| < G(\overline{x}_n^-)$. Therefore, either $H + \widehat{H}$ or $H - \widehat{H}$ is not in $\mathcal{I}(F, G)$.

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 IB, there exists a rectangle that lies in between the graphs of F and G, so that when restricted to this rectangle, the graph of H is not a step function. Then, since extreme points of uniformly bounded nondecreasing functions are exactly the step functions (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 F and G, this, in turn, implies that H can be written as a convex combination of two distinct distributions in $\mathcal{I}(F, G)$.

2.3 The Economics of the Extreme Points

In what follows, we demonstrate how the characterization of extreme points of FOSD 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 us to express any element H of $\mathcal{I}(F,G)$ as a mixture of its extreme points. As a result, if one wishes to establish some property for every element of $\mathcal{I}(F,G)$, and if this property is preserved under convex combinations, then it suffices to establish the property for all extreme points of $\mathcal{I}(F,G)$, which is a much smaller set.

In Section 3, we use this observation to characterize the distributions of posterior quantiles. This characterization is an analogue of the celebrated characterization of the distributions of posterior means that follows from Strassen's theorem (Strassen 1965). We also show how the characterization of distributions of posterior quantiles leads to several economic applications. The first among these is generalizing (and simplifying the proof of) a widely known result due to Benoît and Dubra (2011) in the literature on the psychology of judgment. The second application is to political redistricting, and the third application is to Bayesian persuasion.

The second property of extreme points that we rely on 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 it is convex and if the feasible set is related to an FOSD interval.

In Section 4, we use this property to generalize and unify several results in the literature on security design with limited liability. Feasible securities can be viewed as an FOSD interval bounded from below by zero and bounded from above by the cash flow of the asset. From this perspective, Theorem 1 sheds light on optimal securities and generalizes canonical results in various settings, including those with moral hazard (Innes 1990) and adverse selection (DeMarzo and Duffie 1999).

3 Distributions of Posterior Quantiles

In this section, we use Theorem 1 to characterize the distributions of posterior quantiles. We then demonstrate the economic significance of this characterization by applying it to topics in the psychology of judgement, gerrymandering, and Bayesian persuasion.

3.1 Characterization of the Distributions of Posterior Quantiles

Consider a one-dimensional variable $x \in \mathbb{R}$ that is drawn from a prior F_0. A *signal* for x is defined as a probability measure $\mu \in \Delta(\mathcal{F})$ such that

$$\int_{\mathcal{F}} F(x)\mu(\mathrm{d}F) = F_0(x), \tag{1}$$

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

For any distribution $F \in \mathcal{F}$ and for any $\tau \in (0,1)$, denote the set of τ-quantiles of F by $[F^{-1}(\tau), F^{-1}(\tau^+)]$.[7] Furthermore, we say that a transition probability $r : \mathcal{F} \times [0,1] \to \Delta(\mathbb{R})$ is a *quantile selection rule* if, for all $F \in \mathcal{F}$ and for all $\tau \in (0,1)$, $r(\cdot|F,\tau)$ assigns probability 1 to the set of τ-quantiles of F. In other words, a quantile selection rule r selects (possibly through randomization) a τ-quantile for every CDF F and for every $\tau \in (0,1)$, whenever it is not unique. Let $\mathcal{R}$ be the collection of all selection rules.

For any $\tau \in (0,1)$, for any signal $\mu \in \mathcal{M}$, and for any selection rule $r \in \mathcal{R}$, 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 that can be induced by some signal $\mu \in \mathcal{M}$ and selection rule $r \in \mathcal{R}$.

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 $\underline{F}_0^\tau$ and $\overline{F}_0^\tau$ as follows:

$$\underline{F}_0^\tau(x) := \min\left\{\frac{1}{\tau}F_0(x), 1\right\}, \quad \overline{F}_0^\tau(x) := \max\left\{\frac{F_0(x)-\tau}{1-\tau}, 0\right\}.$$

Note that $\overline{F}_0^\tau(x) \leq \underline{F}_0^\tau(x)$ for all $x \in \mathbb{R}$ and for all $\tau \in (0,1)$. In essence, $\underline{F}_0^\tau$ is the conditional distribution of F_0 in the event that x is smaller than a τ-quantile of F_0; whereas $\overline{F}_0^\tau$ is the conditional distribution of F_0 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 $\tau \in (0,1)$,*

$$\mathcal{H}_\tau = \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau).$$

Theorem 2 completely characterizes the distributions of posterior τ-quantiles by the FOSD

[6] From Blackwell's theorem (Blackwell 1953), given any $\mu \in \mathcal{M}$, each $F \in \mathrm{supp}(\mu)$ can be interpreted as a *posterior* for x obtained via Bayes' rule under a prior F_0, after observing the realization of a signal that is correlated with x. The marginal distribution of this signal is summarized by μ.

[7] $F^{-1}(\tau^+) := \lim_{q \downarrow \tau} F^{-1}(q)$ denotes the right-limit of F^{-1} at τ.

interval $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$. Figure II illustrates Theorem 2 for the case when $\tau = 1/2$. The distribution $\underline{F}_0^{1/2}$ is colored blue, whereas the distribution $\overline{F}_0^{1/2}$ is colored red. The green dotted curve represents the prior, F_0. According to Theorem 2, any distribution H bounded by $\underline{F}_0^{1/2}$ and $\overline{F}_0^{1/2}$ (for instance, the black curve in the figure) can be induced by a signal $\mu \in \mathcal{M}$ and a select rule $r \in \mathcal{R}$. 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.

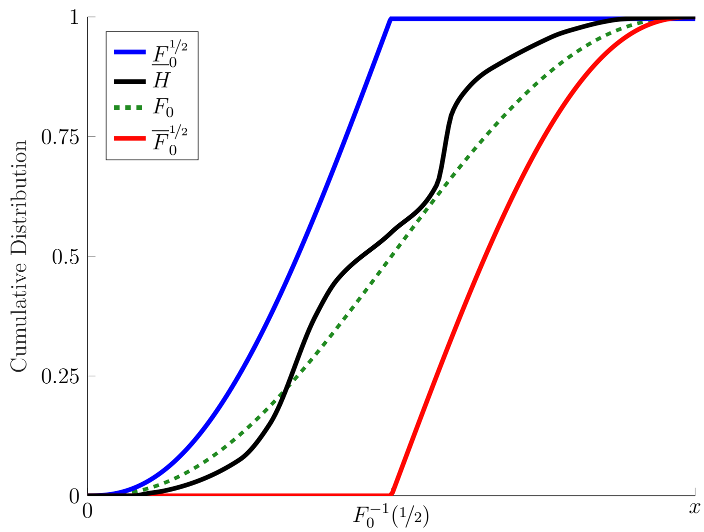


Figure II
Distributions of Posterior Quantiles

Theorem 2 can be regarded as a natural analogue of the well-known characterization of the distributions of posterior *means* that follows from Strassen (1965). Strassen's theorem implies that a CDF $H \in \mathcal{F}$ is a distribution of posterior means if and only if H is a mean-preserving contraction of the prior F_0 (i.e., H *majorizes* F_0). Instead of posterior means, Theorem 2 pertains to posterior quantiles. According to Theorem 2, H is a distribution of posterior quantiles if and only if H dominates the lower-truncated prior $\underline{F}_0^\tau$ and is dominated by the upper-truncated prior $\overline{F}_0^\tau$, in the sense of FOSD.

The necessity part of Theorem 2 is straightforward from the martingale property of posterior beliefs. Indeed, for any signal $\mu \in \mathcal{M}$ and for any $r \in \mathcal{R}$,

$$H^\tau(x|\mu, r) \leq \mu(\{F \in \mathcal{F} | F^{-1}(\tau) \leq x\}) = \mu(\{F \in \mathcal{F} | F(x) \geq \tau\}),$$

for all $x \in \mathbb{R}$, where the first inequality holds because the right-hand side corresponds to the distribution of posterior quantiles induced by μ when the lowest τ-quantile is selected with probability 1. Furthermore, for any $x \in \mathbb{R}$, if we regard $F(x) \in [0,1]$ as a random variable whose distribution is implied by μ, it then follows from (1) that its distribution must be a mean-preserving spread of $F_0(x)$. As a result, $\mu(\{F \in \mathcal{F} | F(x) \geq \tau\})$ can be at most $\min\{F_0(x)/\tau, 1\}$, since otherwise, the mean of $F(x)$ can never be $F_0(x)$. This implies that $H^\tau(x|\mu, r) \leq \underline{F}_0^\tau(x)$. A similar argument leads to the conclusion that $H^\tau(x|\mu, r) \geq \overline{F}_0^\tau(x)$.

The sufficiency part, however, 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 $H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$. Although it might be easier to construct a signal that induces some specific distribution of posterior quantiles, constructing a signal for any arbitrary distribution $H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ does not seem to be tractable.[8] Nonetheless, Theorem 1 allows us to bypass this challenge and 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 $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ 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 $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$. To illustrate the intuition, consider an extreme point H of $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ that takes the following form:

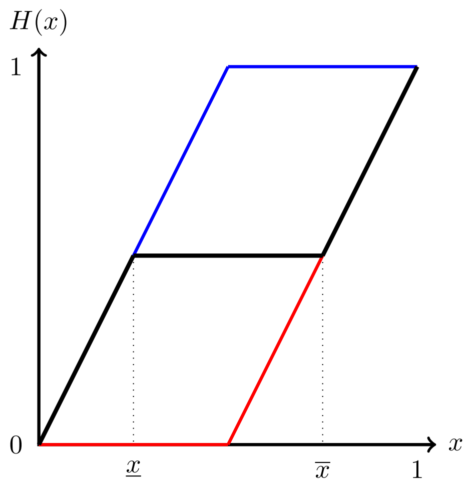
$$H(x) = \begin{cases} \underline{F}_0^\tau(x), & \text{if } x < \underline{x} \\ \underline{F}_0^\tau(\underline{x}), & \text{if } x \in [\underline{x}, \overline{x}) \\ \overline{F}_0^\tau(x), & \text{if } x \geq \overline{x} \end{cases},$$

for some $\underline{x}, \overline{x}$ such that $\underline{F}_0^\tau(\underline{x}) = \overline{F}_0^\tau(\overline{x}^-)$, as depicted by Figure IIIA. 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 $\underline{F}_0^\tau(\underline{x}) = \overline{F}_0^\tau(\overline{x}^-)$, by choosing α correctly, 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 IIIB.[9] Similarly, each $x > \overline{x}$ would become a τ-quantile of the posterior it belongs to as well. Together, by properly selecting the posterior quantiles, the induced distribution of posterior quantiles under this signal would indeed be H.

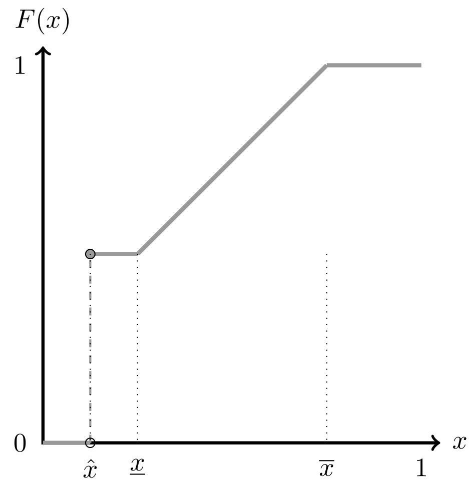
Although the characterization of Theorem 2 may seem to rely on selection rules $r \in \mathcal{R}$, the result remains (essentially) the same even when restricted to signals that always induce a unique posterior τ-quantile, provided that the prior F_0 has full support on an

[8] For example, $\overline{F}_0^\tau$ and $\underline{F}_0^\tau$ can be attained using a modified version of the "matching extreme" signal introduced by Friedman and Holden (2008). However, matching extreme signals would inevitably assign positive probability to posteriors whose quantiles are nearby the prior quantile, and hence, matching extremes cannot induce any distributions $H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ that assign probability zero to some interval containing $[F_0^{-1}(\tau), F_0^{-1}(\tau^+)]$.

[9] Specifically, $\alpha = \frac{1-\tau}{\tau} F_0(\underline{x}) / (\frac{\tau}{1-\tau}(1 - F_0(\overline{x}^-)) + \frac{1-\tau}{\tau} F_0(\underline{x}))$.



(A) An Extreme Point H



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

Figure III
CONSTRUCTING A SIGNAL THAT INDUCES H

interval. Theorem 3 below formalizes this statement. To this end, Let $\mathcal{H}_\tau^0$ 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 $\mathcal{H}_\tau^0$ relates to a family of perturbations of the set $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, denoted by $\{\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon})\}_{\varepsilon>0}$, where

$$\underline{F}_0^{\tau,\varepsilon}(x) := \begin{cases} \frac{1}{\tau+\varepsilon}F_0(x), & \text{if } x < F_0^{-1}(\tau) \\ 1, & \text{if } x \geq F_0^{-1}(\tau) \end{cases} ; \text{ and } \overline{F}_0^{\tau,\varepsilon}(x) := \begin{cases} 0, & \text{if } x < F_0^{-1}(\tau) \\ \frac{F_0(x)-(\tau-\varepsilon)}{1-(\tau-\varepsilon)}, & \text{if } x \geq F_0^{-1}(\tau) \end{cases},$$

for all $\varepsilon \geq 0$ and for all $x \in \mathbb{R}$. Note that $\mathcal{I}(\underline{F}_0^{\tau,0}, \overline{F}_0^{\tau,0}) = \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, and that $\{\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon})\}_{\varepsilon>0}$ 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,*

$$\bigcup_{\varepsilon>0} \mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon}) \subseteq \mathcal{H}_\tau^0 \subseteq \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau).$$

As an immediate corollary of Theorem 2 and Theorem 3, we now have an analogue of the celebrated law of iterated expectation, which we refer to as the *law of iterated quantiles.*

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

1. *For any $F_0 \in \mathcal{F}$ and for any closed interval $Q \subseteq \mathbb{R}$, $Q = [H^{-1}(\tau), H^{-1}(\tau^+)]$ for some $H \in \mathcal{H}_{\tilde{\tau}}$ if and only if $Q \subseteq [(\underline{F}_0^{\tilde{\tau}})^{-1}(\tau), (\overline{F}_0^{\tilde{\tau}})^{-1}(\tau^+)]$.*

[10] As a convention, let $\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon}) := \emptyset$ when $\varepsilon \geq \max\{\tau, 1-\tau\}$.

2. *For any continuous $F_0 \in \mathcal{F}$ that has a full support on an interval and for any $\hat{x} \in \mathbb{R}$, $\hat{x} \in [H^{-1}(\tau), H^{-1}(\tau^+)]$ for some $H \in \mathcal{H}^0_{\tilde{\tau}}$ if and only if $\hat{x} \in [(\underline{F}^{\tilde{\tau}}_0)^{-1}(\tau), (\overline{F}^{\tilde{\tau}}_0)^{-1}(\tau)]$.*

According to Corollary 1, while the expectation of posterior means under any signal is always the expectation under the prior, the possible τ-quantiles of posterior $\tilde{\tau}$-quantiles are exactly $[(\underline{F}^{\tilde{\tau}}_0)^{-1}(\tau), (\overline{F}^{\tilde{\tau}}_0)^{-1}(\tau^+)]$. For example, the collection of all possible *medians* of posterior *medians* is exactly the *interquartiles* $[F_0^{-1}(1/4), F_0^{-1}(3/4)]$ of the prior.

3.2 Economic Applications

Apparent Overconfidence

A key issue in the psychology of judgment is explaining why people rank themselves better or worse than others in certain tasks. By the 2000s, a consensus had emerged among researchers that most people commonly rank themselves as better than average on simple tasks and worse than average on difficult tasks (Moore 2007; Kruger, Windschitl, Burrus, Fessel and Chambers 2008). Up for debate, however, was whether this behavior was rational.

Here we show how Theorem 3 can speak to this debate. Consider the following setting of individual self-evaluation, a setting due to Benoît and Dubra (2011). There is a unit mass of individuals, and each one of them is attached to a "type" $x \in [0, 1]$, which is distributed according to a CDF $F_0 \in \mathcal{F}$. 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 of individuals who predict there own type is in $[z_{k-1}, z_k)$.

It is well-documented in the experimental literature that a prediction dataset can be very different from the population distribution F_0. One common explanation found in this literature is that individuals are truly overconfident or truly underconfident (Alicke, Klotz, Breitenbecher, Yurak and Vredenburg 1995; De Bondt and Thaler 1995; Camerer 1997). But Benoît and Dubra (2011) proposed an alternative explanation: This difference can simply be caused by noises in each individual's signal. People are only *apparently* misconfident. Individuals can still be fully Bayesian even if the prediction dataset is different from the population distribution. We show next how a general version of Benoît and Dubra (2011)'s insight follows immediately from Theorem 3.

Consider the following Bayesian framework: Each individual receives a signal $s \in S$ for their type x, which is drawn from a conditional distribution given each realized x. After observing their signal realizations, individuals then update their belief via Bayes' rule, and they predict their types according to their posterior medians (e.g., Hoelzl and Rustichini 2005). Given the distribution F_0 of types and a partition $0 = z_0 < z_1 < \ldots < z_K = 1$, a prediction dataset $(\theta_k)_{k \in K}$ is said to be *median rationalizable (τ-quantile rationalizable)*,[11] if there exists a signal for 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.[12]

Under this framework, theorem 1 (and theorem 4) of Benoît and Dubra (2011) characterizes the collection of median (τ-quantile) rationalizable datasets, under the assumption that $F_0(z_k) = {}^{k}/_{K}$ for all $k \in \{1, \ldots, K\}$. In other words, Benoît and Dubra (2011) characterize the collection of rationalizable datasets in the context of *self-ranking*, where individuals are asked to place themselves into a K-cile relative to the population according to their posterior medians (τ-quantiles).

Although relative self-ranking is one of the common types of experiments in the literature, as noted by Benoît and Dubra (2011), many other experiments involve some *absolute* scales. For example, a large overconfidence literature asks students to forecast their exam scores (e.g., Murstein 1965; Grimes 2002; Hossain and Tsigaris 2015), which are typically on an absolute scale of 0 to 100. Alternatively, Weinstein (1980) asks students to predict their employment probabilities after graduation, which are also on an absolute scale of 0 to 1. As an immediate corollary of Theorem 3, we generalize the result of Benoît and Dubra (2011) and characterize the collection of τ-quantile rationalizable datasets on an arbitrary scale.[13]

Corollary 2 (Rationalizable Apparent Misconfidence). *For any $\tau \in (0,1)$, for any $F_0 \in \mathcal{F}$ 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 $k \in \{1, \ldots, K\}$,*

$$\sum_{i=1}^{k} \theta_i < \frac{1}{\tau} F_0(z_k) \tag{2}$$

[11] Not all experiments would clearly instruct the individuals to use their posterior median when predicting their ability. Other statistics of a posterior could potentially be used by an individual when the instruction is not clear. When individuals use the posterior means to predict their types, the set of rationalizable data would be given by the mean-preserving contractions of the prior, as noted by Benoît and Dubra (2011).

[12] In other words, $(\theta_k)_{k=1}^{K}$ is τ-quantile rationalizable if there exists $H \in \mathcal{H}_\tau^0$ 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, our results generalize their conclusion even with this stringent requirement.

[13] Note that the scale on which the dataset lies is unrelated to the statistics that individuals use to predict their type. Therefore, it would be reasonable to assume that individuals predict their performance—both on a relative scale and an absolute scale—using either the median, a τ-quantile, or the mean of their posteriors. As Benoît and Dubra (2011) note: "Just considering medians and means, there are four ways to interpret answers to scale questions".

and

$$\sum_{i=k}^{K} \theta_i < \frac{1 - F_0(z_{k-1}^-)}{1-\tau} \tag{3}$$

Proof. The necessity part follows directly from the proof of theorem 4 of Benoît and Dubra (2011). For sufficiency, consider any prediction dataset $(\theta_k)_{k=1}^K$ such that (2) and (3) hold. Let $H(x)$ be the distribution that assigns probability θ_k at $(z_k + z_{k-1})/2$. Then, there exists $\varepsilon > 0$ such that $H \in \mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon})$. By Theorem 3, there exists a signal μ with $\mu(\{F \in \mathcal{F} | F^{-1}(\tau) < F^{-1}(\tau^+)\}) = 0$ 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 1. For comparison, when $z_k = {}^k/_K$ for all k, and when F_0 is uniform, Corollary 2 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 $\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon})$ are attained by explicitly constructed signals, as shown in the proof of Theorem 3. It is also noteworthy that, although theorem 4 of Benoît and Dubra (2011) can be used to prove Theorem 2 indirectly (by taking $K \to \infty$ and establishing proper continuity properties) when F_0 admits a density, the same argument cannot be used to prove Theorem 3, which is crucial for the proof of Corollary 2.[14]

Limits of Gerrymandering

Beyond the psychology of judgment, Theorem 2 and Theorem 3 can be applied to political redistricting. The study of redistricting ranges across many fields: Legal scholars, political scientists, mathematicians, computer scientists, and economists have all contributed to this vast literature.[15]

While existing economic theory on redistricting has largely 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 2020), another fundamental question is the scope of redistricting'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"?

[14] This is because of the the failure of upper-hemicontinuity when signals that induce multiple quantiles are excluded.

[15] See, for example, Shotts (2001); Besley and Preston (2007); Coate and Knight (2007); McCarty, Poole and Rosenthal (2009); Fryer Jr and Holden (2011); McGhee (2014); Stephanopoulos and McGhee (2015); Alexeev and Mixon (2018).

Theorem 2 and Theorem 3 describe the extent to which unrestrained gerrymandering can shape the composition of elected representatives. 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 $F_0 \in \mathcal{F}$. 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 $F \in \text{supp}(\mu)$ is described by the conditional distribution of the ideal positions of citizens who belong to it.[16] 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 F must have an ideal position that is a median of F. When there are multiple medians in a district, the representative's ideal position is determined by a selection rule $r \in \mathcal{R}$, which is either flexible or stipulated by election laws.[17]

Given any $\mu \in \mathcal{M}$ and any selection rule $r \in \mathcal{R}$, the induced distribution of posterior medians $H^{1/2}(\cdot|\mu, r)$ can be interpreted as a distribution of the ideal positions of the elected representatives. Meanwhile, the bounds $\underline{F}_0^{1/2}$ and $\overline{F}_0^{1/2}$ can be interpreted as distributions of representatives that only reflect one side of voters' political positions relative to the median of the population. Specifically, $\underline{F}_0^{1/2}$ describes an "all-left" legislature, which only reflects citizens' ideal positions that are left of the population median. Likewise, $\overline{F}_0^{1/2}$ 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 completely characterizes the set of possible compositions of the legislature across all election maps.

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

1. *$H \in \mathcal{I}(\underline{F}_0^{1/2}, \overline{F}_0^{1/2})$.*
2. *H is a distribution of the representatives' ideal positions under some map $\mu \in \mathcal{M}$ and some selection rule $r \in \mathcal{R}$.*

Furthermore, for any fixed selection rule $\hat{r} \in \mathcal{R}$, every $H \in \cup_{\varepsilon>0}\mathcal{I}(\underline{F}_0^{1/2,\varepsilon}, \overline{F}_0^{1/2,\varepsilon})$ is a distribution of the representatives' ideal positions under some map $\mu \in \mathcal{M}$ and selection $\hat{r}$.

Hence, *any* composition of the legislative body ranging from the "all-left" to the "all-right," and anything in between those two extremes, can be procured by some gerrymandered

[16]See Yang and Zentefis (2022a) for a formal link between segmenting a prior and partitioning a population on a physical, two dimensional map.

[17]Recall that 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).

map. Meanwhile, *any* composition that is more extreme than the "all-left" or the "all-right" bodies is not possible regardless of how the districts are drawn.[18]

If we specify the model for the legislature to enact legislation, we may further 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).[19] 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 3 below.

Corollary 3 (Limits of Legislative Outcomes)**.** *Suppose that the median voter property holds both at the district level and at the legislative level. Then an outcome $x \in \mathbb{R}$ can be enacted as legislation under some map if and only if $x \in [F_0^{-1}(1/4), F_0^{-1}(3/4)]$.*

According to Corollary 3, 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. Moreover, if the population is more polarized (i.e., the interquartile range is wider), more extreme legislation can pass. Conversely, Corollary 3 also suggests it is impossible to enact any legislative outcome *beyond* the interquartile range, regardless of how the districts are drawn.

Finally, Proposition 1 can help identify the citizenry's distribution of ideal positions. A common approach to identify that distribution is to map public opinion survey responses to an ideological spectrum. But a disadvantage of this approach is the absence of consistent questions asked over time to create a stable mapping and the lack of representativeness in some surveys (Lax and Phillips 2009). Identifying the ideal positions of elected officials has been more successful because of the abundance of roll-call voting records available in the estimation (Poole and Rosenthal 1985; Shor and McCarty 2011). Nonetheless, inferring the citizenry's distribution of ideal positions from that of elected officials is difficult, as the distribution of ideal positions of elected officials might be very different from that of the citizenry due to gerrymandering.

Using Proposition 1, one can identify the possible distributions of citizens' ideal positions from the observed distribution of representatives' ideal positions. Suppose that H is the observed distribution of representatives' ideal positions. Proposition 1 implies that the pop-

[18] Gomberg, Pancs and Sharma (2022) 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.

[19] 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.

ulation distribution F_0 must have H be dominated by $\overline{F}_0^{1/2}$ and dominate $\underline{F}_0^{1/2}$ at the same time. This leads to Corollary 4 below.

Corollary 4 (Identification Set of F_0)**.** *Suppose that $H \in \mathcal{F}$ is the distribution of ideal positions of a legislature. Then the distribution of citizens' ideal position F_0 must satisfy*

$$\frac{1}{2}H(x) \leq F_0(x) \leq \frac{1+H(x)}{2}, \tag{4}$$

for all $x \in \mathbb{R}$. Conversely, for any $F_0 \in \mathcal{F}$ satisfying (4), there exists a map $\mu \in \mathcal{M}$ and a selection rule $r \in \mathcal{R}$, such that H is the distribution of ideal positions of the legislature.

According to Corollary 4, the distribution of citizens' ideal positions can be identified by (4), even when only the distribution of the representatives' ideal positions can be observed.[20]

Quantile-Based Persuasion

Theorem 2 and Theorem 3 also lead to applications in Bayesian persuasion. Consider the Bayesian persuasion problem formulated by Kamenica and Gentzkow (2011): A state $x \in \mathbb{R}$ is distributed according to a common prior F_0. A sender chooses a signal $\mu \in \mathcal{M}$ to inform the 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 $\hat{v} : \mathcal{F} \to \mathbb{R}$, where $\hat{v}(F) := \mathbb{E}_F[u_S(x, a^*(F))]$ is the reduced-form value function of the sender, and $a^*(F) \in A$ is the optimal action of the receiver under posterior $F \in \mathcal{F}$.[21]

When $|\text{supp}(F_0)| \geq 2$, this "concavafication" method requires finding the concave closure of a multi-variate function, which is known to be computationally challenging, especially when $|\text{supp}(F_0)| = \infty$. For tractability, many papers have restricted attention to preferences where the only payoff-relevant statistic of a posterior is its mean (i.e., $\hat{v}(F)$ is measurable with respect to $\mathbb{E}_F[x]$). See, for example, Gentzkow and Kamenica (2016); Kolotilin, Li, Mylovanov and Zapechelnyuk (2017); Dworczak and Martini (2019); Arieli, Babichenko, Smorodinsky and Yamashita (forthcoming) and Kolotilin, Mylovanov and Zapechelnyuk (2022b).

A natural analogue 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

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

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

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 sender's and receiver's payoffs are such that there exists $\tau \in (0,1)$, a selection rule $r \in \mathcal{R}$, and a measurable function $v_S : \mathbb{R} \to \mathbb{R}$ in which $\hat{v}(F) = \int_{\mathbb{R}} v_S(x) r(\mathrm{d}x|F,\tau)$, for all $F \in \mathcal{F}$. Then*

$$\mathrm{cav}(\hat{v})[F_0] = \sup_{H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)} \int_{\mathbb{R}} v_S(x) H(\mathrm{d}x). \tag{5}$$

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

$$\mathrm{cav}(\hat{v})[F_0] \leq \mathrm{cav}(\bar{v})[F_0] = \sup_{H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)} \int_{\mathbb{R}} v_S(x) H(\mathrm{d}x).$$

Meanwhile, by Theorem 3,

$$\sup_{H \in \cup_{\varepsilon > 0} \mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon})} \int_{\mathbb{R}} v_S(x) H(\mathrm{d}x) \leq \mathrm{cav}(\hat{v})[F_0].$$

Together, since $\mathrm{cl}(\{\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon})\}) = \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, (5) then follows. ∎

By Proposition 2, any τ-quantile-based persuasion problem can be solved by simply choosing a distribution in $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ to maximize the expected value of $v_S(x)$, rather than concavafying the infinite-dimensional functional $\hat{v}$. Furthermore, since the objective function of (5) is affine, Theorem 1 further reduces the search for the solution to only distributions that satisfy its conditions 1 and 2.[22]

For example, consider the canonical setting where the receiver chooses an action to match the state and minimizes some loss function, while the sender's payoff is state-independent. To fix ideas, we can let the sender be an investment advisor and the receiver be a client. The investment 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.

[22] A recent elegant contribution by Kolotilin, Corrao and Wolitzky (2022a) provides a tractable method that simplifies persuasion problems with a one-dimensional state and a one-dimensional action in certain cases. One of these cases 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 within this case has the sender's payoff being state-independent and increasing in the receiver's action, while the receiver's optimal action for each posterior is quantile-measurable. When one further assumes that the sender's payoff is increasing, the conditions of Proposition 4 lead to the same example. Since we allow for arbitrary (state-independent) sender payoffs, Proposition 4 generalizes this example in an orthogonal direction and complements their method.

A standard assumption in this setting is that the receiver's loss function is quadratic, so that $u_R(x,a) := -(x-a)^2$. Under this assumption, the receiver's optimal action $a^*(F)$, given a posterior F, equals the posterior expected value $\mathbb{E}_F[x]$, and hence, the sender's problem is mean-measurable. This leads to a tractable problem since the distributions of the receiver's actions are equivalent to mean-preserving contractions of the prior.[23] With Proposition 4, we are now able to completely solve the sender's problem when the receiver's loss function is *absolute* rather than quadratic. That is, when $u_R(x,a) := -|x-a|$, or more generally, when $u_R(x,a) := -\rho_\tau(x-a)$, with $\rho_\tau(y) := y(\tau - \mathbf{1}\{y < 0\})$ being the "pinball" loss function. For any $\tau \in (0,1)$, when the receiver's payoff is given by $u_R(x,a) = -\rho_\tau(x-a)$ and the sender's payoff is $u_S(x,a) = v_S(a)$, since any $a \in [F^{-1}(\tau), F^{-1}(\tau^+)]$ is optimal for the receiver when the posterior is F, Proposition 2 applies, and the sender's problem can be rewritten via (5).

For instance, if the sender's payoff v_S is nondecreasing, then $\overline{F}_0^\tau$ is optimal, whereas if v_S is nonincreasing, $\underline{F}_0^\tau$ is optimal. Or, as in many settings, v_S may be non-monotonic. In the example of the investment 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 in the client's chosen portfolio weight, with a peak at some $a_0 \in (0,1)$ that has the client put some wealth in stocks and the remainder in bonds, rather than all wealth in either asset class alone. In this case, assuming that $a_0 < F_0^{-1}(\tau)$, the solution to (5) is given by

$$H^*(x) := \begin{cases} 0, & \text{if } x < a_0 \\ \underline{F}_0^\tau(x), & \text{if } x \geq a_0 \end{cases} .$$

Notice that if v_S is concave, then the sender's optimal signal is always the null signal if the receiver's loss function is quadratic. In contrast, when the receiver's loss function is absolute, the sender would optimally reveal some information about the state. In other words, the shape of the receiver's loss function has substantive implications for the type of information the sender optimally discloses.

4 Security Design with Limited Liability

In this second class of applications, we show how FOSD 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. We generalize and unify seminal results in this literature under a common framework when either type

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

of asymmetric information is at play. Section 4.1 addresses security design problems under moral hazard, whereas Section 4.2 handles those under adverse selection.

4.1 Security Design with Moral Hazard

Consider the following setting of security design in the presence of moral hazard, a setting due to Innes (1990). 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 $e \geq 0$, the project's profit is distributed according to $\Phi(\cdot|e) \in \mathcal{F}$, and the (additively separable) effort cost to the entrepreneur is $C(e) \geq 0$.

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.[24] Given a security H, the entrepreneur chooses an effort level to solve

$$\sup_{e \geq 0} \int_0^\infty (x - H(x))\Phi(\mathrm{d}x|e) - C(e). \tag{6}$$

For simplicity, we make the following technical assumptions: 1) The supports of the profit distributions $\{\Phi(\cdot|e)\}_{e\geq 0}$ 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 $e \geq 0$. 3) $\{\Phi(\cdot|e)\}_{e\geq 0}$ and C are such that (6) admits a solution and every solution to (6) can be characterized by the first-order condition.[25]

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

[24] 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 can contribute additional funds to the project so that only monotone profits would be observed.

[25] For example, we may assume that C is strictly increasing and strictly convex and that $\frac{\partial}{\partial e}\phi(x|e) > 0$, $\frac{\partial^2}{\partial e^2}\phi(x|e) \leq 0$ for all x and for all e.

solves

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

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

Innes (1990) characterizes the optimal security in this setting using an additional crucial assumption: The project profit distributions $\{\phi(\cdot|e)\}_{e \geq 0}$ satisfy the monotone likelihood ratio property (Milgrom 1981). Under this assumption, he shows that every optimal security must be a standard debt contract $H^d(x) := \min\{x, d\}$ for some $d > 0$. 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 $H \in \mathcal{I}(F,G)$ is a *contingent debt contract*, if there exists an interval partition $\{I_n\}$ of $[0,1]$ and a sequence $\{d_n\} \subseteq (0,1]$ such that $H(x) = H^{d_n}(x)$ for all $x \in I_n$. Figure IV illustrates a contingent debt contract $\widehat{H}$ with $I_1 = [0, 1/2)$, $I_2 = [1/2, 1]$, $d_1 = 1/4$, 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.[26]

Clearly, every standard debt contract with face value d is a contingent debt contract where $I_1 = [0,1]$ and $d_1 = d$. Moreover, a contingent debt contract never involves the entrepreneur and investor sharing in the equity of the project. To see how the cash flow is split between parties, 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, instead, 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

[26] Contingent debt contracts share some similarity with state-contingent debt instruments (SCDIs) from the sovereign debt literature, which tie a country's principal or interest payments to its nominal GDP (Lessard and Williamson 1987; Shiller 1994; Borensztein and Mauro 2004).

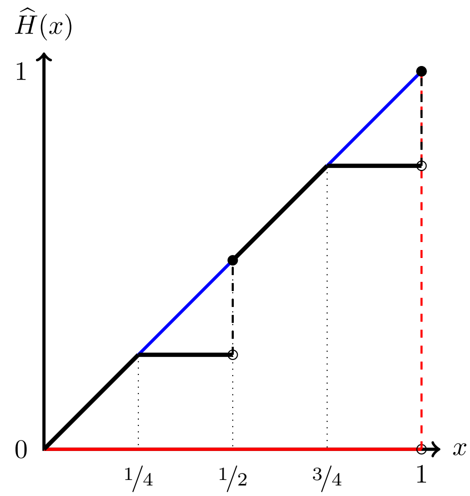


Figure IV
A contingent Debt Contract

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.

From Theorem 1, we show that a portfolio of at most three contingent debt contracts is optimal.

Proposition 3. *There exists contingent debt contracts* $\{H_i^*\}_{i=1}^3$ *and* $\{\lambda_i\}_{i=1}^3 \subseteq [0,1]$*, with* $\lambda_1 + \lambda_2 + \lambda_3 = 1$*, such that* $H^* := \lambda_1 H_1^* + \lambda_2 H_2^* + \lambda_3 H_3^*$ *is a solution to the entrepreneur's problem* (7)*.*

Proof. For any fixed $e \geq 0$, the objective function of the entrepreneur's security design problem (7) is linear, and the two constraints are linear. Thus, for any fixed e, (7) must have a solution that is an extreme point of the feasible set. By proposition 2.1 of Winkler (1988), extreme points of the feasible set must take the form of a convex combination of at most three extreme points of $\mathcal{I}(F,G)$. The proof is then completed by noticing that H is an extreme point of $\mathcal{I}(F,G)$ if and only if H is a contingent debt contract. ■

According to Proposition 3, it is sufficient for the entrepreneur to use a portfolio of contingent debt contracts without sharing the equity of the project with the investor. The nature of standard debt contracts, which grant the entrepreneur only residual rights, is preserved even without the monotone likelihood ratio assumption. The only difference is that the entrepreneur may be liable for more when the project earns more.

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 an individually-rational standard debt contract with the same expected payment. Meanwhile, the monotone likelihood ratio property ensures that this debt contract inventivizes the entrepreneur to exert higher effort, thus relaxing the incentive-compatibility constraint. 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 simple portfolios of 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.

At a more technical level, Proposition 3 is reminiscent of mechanism design problems whose solutions feature rationing or randomized posted prices. (See, for example, Samuelson 1984; Dworczak ⓡ Kominers ⓡ Akbapour 2021; Loertscher and Muir 2022; Kang 2022). The common structure of these problems is that the objective function is affine, the feasible set is the collection of uniformly bounded monotone functions, and the constraints are affine in the choice variables. Proposition 2.1 of Winkler (1988) implies that there must be at least one solution that can be represented as a convex combination of at most $n+1$ extreme points of the feasible set, where n is the number of constraints. Just as rationing and randomized posted-price mechanisms are mixtures of posted-price mechanisms—which are extreme points of the feasible set—portfolios of contingent debt contracts are mixtures of extreme points of the feasible set $\mathcal{I}(F, G)$ in problem (7) as well.

4.2 Security Design with Adverse Selection

Consider the following setting of security design in the presence of adverse selection, a setting due to DeMarzo and Duffie (1999). 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 $x \geq 0$. The cash flow is distributed according to $\Phi_0 \in \mathcal{F}$, which is supported on a compact interval normalized to $[0,1]$. Because $\delta < 1$, the issuer has demand for liquidity 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 $F(x) := x$ and $G(x) := \mathbf{1}\{x = 1\}$ for all $x \in [0,1]$. The set of securities can again be written as $\mathcal{I}(F, G)$.

Given any security $H \in \mathcal{I}(F, G)$, 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

$$\delta\left(\mathbb{E}[x - H(x)|s] + (1-q)\mathbb{E}[H(x)|s]\right) + qP(q) = q(P(q) - \delta\mathbb{E}[H(x)|s]) + \delta\mathbb{E}[x|s].$$

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,[27] 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

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

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 FOSD under some s_0, conditional on every interval I of $[0,1]$.[28] 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. As in Section 4.1, we say that a security is

[27] 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.

[28] 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 that conditional distribution given signal realization s_0, denoted ${}^{\Phi|_I(\cdot|s_0)}/_{\Phi(I|s_0)}$, in the sense of first-order stochastic dominance.

a contingent debt contract if there exists an interval partition $\{I_n\}$ of $[0,1]$ and $\{d_n\} \subseteq (0,1]$ such that $H(x) = H^{d_n}(x)$ for all $x \in I_n$. 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

$$\sup_{H \in \mathcal{I}(F,G), \underline{z} \geq 0} \left[(1-\delta)\underline{z}^{\frac{1}{1-\delta}} \int_S \left(\int_0^1 H(x)\Phi(\mathrm{d}x|s) \right)^{-\frac{\delta}{1-\delta}} \Psi(\mathrm{d}s) \right]$$
$$\text{s.t.} \int_0^1 H(x)\Phi(\mathrm{d}x|s_0) = \underline{z}. \tag{8}$$

As shown by Proposition 4 below, there always exists an optimal security in this setting that is a portfolio of at most two contingent debt contracts.

Proposition 4. *There exists contingent debt contracts H_1^*, H_2^* and $\lambda \in [0,1]$ such that $H^* := \lambda H_1^* + (1-\lambda)H_2^*$ is a solution to the issuer's problem (8). Furthermore, if $\Phi(\cdot|s)$ has full support on $[0,1]$ for all $s \in S$, this solution is unique.*

Proof. For any fixed $\underline{z} \geq 0$, the objective function of the issuer's problem (8) is convex, and the constraint is linear. Thus, an extreme point of the feasible set must be a solution to (8). By proposition 2.1 of Winkler (1988), such an extreme point can be written as a convex combination of at most two extreme points of $\mathcal{I}(F,G)$, as desired, since H is an extreme point of $\mathcal{I}(F,G)$ if and only if H is a contingent debt contract. For uniqueness, notice that when $\Phi(\cdot|s)$ has full support for all s, the objective function of (8) is strictly convex in H. Therefore, every solution must be an extreme point of the feasible set. This completes the proof. ■

Overall, this section showcases the unifying role of extreme points of FOSD intervals in security design. Rationalizing the existence of different financial securities observed in practice has been a crowning achievement of this literature. The literature has done this under a variety of economic environments and assumptions, which punctuates 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 feasible securities as an FOSD interval is that it strips the problem down to its basic elements. Whether the setting has hidden action or hidden information, and whether the asset's cash flow distributions exhibit MLRP, are not defining. Limited liability, monotone contracts, and convexity of the issuer's objective function 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 of the debt 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 FOSD 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. Well-known results directly follow, but so do new insights that are straightforward to uncover using this framework.

5 Conclusion

We characterize the extreme points of first-order stochastic dominance (FOSD) intervals, and we reveal how these intervals are at the heart of many distinct topics in economics. We show that any extreme point of an FOSD interval must either coincide with one of the FOSD interval's bounds, or be constant on an interval, where at least one end of the interval reaches one of the bounds. FOSD intervals describe the distributions of posterior quantiles. We apply this insight to topics in the psychology of judgment, political economy, and Bayesian persuasion. We also use this insight to prove the law of iterated quantiles. Finally, FOSD 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.

Other applications involving FOSD intervals undoubtedly exist. For instance, their link to the distributions of posterior quantiles opens many potential research avenues. When consumers' values or firms' marginal costs follow distributions, different points on the inverse supply and demand curves are quantiles, which might contain further applications in consumer or firm theory. 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 an FOSD interval, such as R&D investments and screening problems with stochastic inventories, are yet other directions for future work.

References

Alexeev, B. and D. G. Mixon (2018) "An Impossibility Theorem for Gerrymandering," *American Mathematical Monthly*, 125 (10), 878–884.

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

Alicke, M. D., M. L. Klotz, D. L. Breitenbecher, T. J. Yurak, and D. S. Vredenburg (1995) "Personal Contact, Individuation, and the Better-Than-Average Effect.," *Journal of Personality and Social Psychology*, 68 (5), 804–825.

Allen, F. and A. Barbalau (2022) "Security Design: A Review," Working Paper. Imperial College Business School, London, UK.

Arieli, I., Y. Babichenko, R. Smorodinsky, and T. Yamashita (forthcoming) "Optimal Persuasion via Bi-Pooling," *Theoretical Economics*.

Bawa, V. S. (1982) "Research Bibliography—Stochastic Dominance: A Research Bibliography," *Management Science*, 28 (6), 698–712.

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.

Besley, T. and I. Preston (2007) "Electoral Bias and Policy Choice: Theory and Evidence," *Quarterly Journal of Economics*, 122 (4), 1473–1510.

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.

Camerer, C. F. (1997) "Progress in Behavioral Game Theory," *Journal of Economic Perspectives*, 11 (4), 167–188.

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

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.

Coate, S. and B. Knight (2007) "Socially Optimal Districting: A Theoretical and Empirical Exploration," *Quarterly Journal of Economics*, 122 (4), 1409–1471.

De Bondt, W. F. and R. H. Thaler (1995) "Financial Decision-Making in Markets and Firms: A Behavioral Perspective," *Handbooks in Operations Research and Management Science*, 9, 385–410.

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

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

Dworczak, P. ⓡ S. D. Kominers ⓡ M. Akbapour (2021) "Redistribution through Markets," *Econometrica*, 89 (4), 1665–1698.

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," January, Working Paper. Northwestern University, Evanston, IL.

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.

Fryer Jr, R. G. and R. Holden (2011) "Measuring the compactness of political districting plans," *Journal of Law and Economics*, 54 (3), 493–535.

Gentzkow, M. and E. Kamenica (2016) "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review: Papers and Proceedings*, 597–601.

Gomberg, A., R. Pancs, and T. Sharma (2022) "Electoral Maldistricting," June, Working Paper. ITAM, Mexico.

Grimes, P. (2002) "The Overconfident Principles of Economics Student: An Examination of a Metacognitive Skill," *Journal of Economic Education*, 33 (1), 15–30.

Gul, F. and W. Pesendorfer (2010) "Strategic Redistricting," *American Economic Review*, 100 (4), 1616–41.

Hadar, J. and W. R. Russell (1969) "Rules for Ordering Uncertain Prospects," *American Economic Review*, 59 (1), 25–34.

Hanoch, G. and H. Levy (1969) "The Efficiency Analysis of Choices Involving Risk," *Review of Economic Studies*, 36 (3), 335–346.

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.

Hoelzl, E. and A. Rustichini (2005) "Overconfident: Do you Put Your Money in It?" *Economic Journal*, 115 (503), 305–318.

Hossain, B. and P. Tsigaris (2015) "Are Grade Expectations Rational? A Classroom Experiment," *Education Economics*, 23 (2), 199–212.

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*, 2560–2615.

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

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," June, Working Paper. Massachusetts Institute of Technology, Cambridge, MA.

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 (2020) "The Economics of Partisan Gerrymandering," September, Working Paper. Massachusetts Institute of Technology, Cambridge, MA.

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

Kroll, Y. and H. Levy (1980) "Stochastic Dominance: A Review and Some New Evidence," *Research in Finance*, 2.

Kruger, J., P. D. Windschitl, J. Burrus, F. Fessel, and J. R. Chambers (2008) "The Rational Side of Egocentrism in Social Comparisons," *Journal of Experimental Social Psychology*, 44 (2), 220–232.

Lax, J. R. and J. H. Phillips (2009) "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science*, 53 (1), 107–121.

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

Levy, H. (1990) *Stochastic Dominance*, 251–254, London: Palgrave Macmillan UK.

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.

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

——— (2009) "Does Gerrymandering Cause Polarization?" *American Journal of Political Science*, 53 (3), 666–680.

McGhee, E. (2014) "Measuring Partisan Bias in Single-Member District Electoral Systems," *Legislative Studies Quarterly*, 39 (1), 55–85.

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

Moore, D. A. (2007) "Not So Above Average After All: When People Believe They Are Worse than Average and Its Implications for Theories of Bias in Social Comparison," *Organizational Behavior and Human Decision Processes*, 102 (1), 42–58.

Murstein, B. (1965) "The Relationship of Grade Expectations and Grades Believed to be Deserved to Actual Grades Received," *Journal of Experimental Education*, 33 (4), 357–362.

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

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," October, Working Paper. Carnegie Mellon University, Pittsburgh, PA.

Poole, K. T. and H. Rosenthal (1985) "A Spatial Model for Legislative Roll Call Analysis," *American Journal of Political Science*, 357–384.

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

Rothschild, M. and J. Stiglitz (1970) "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 225–243.

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.

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

Shor, B. and N. McCarty (2011) "The Ideological Mapping of American Legislatures," *American Political Science Review*, 105 (3), 530–551.

Shotts, K. W. (2001) "The Effect of Majority-Minority Mandates on Partisan Gerrymandering," *American Journal of Political Science*, 120–135.

Simon, B. (2011) *Convexity: An Analytic Viewpoint.* Cambridge Tracts in Mathematics 187: Cambridge University Press.

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

Stephanopoulos, N. O. and E. M. McGhee (2015) "Partisan Gerrymandering and the Efficiency Gap," *U. Chi. L. Rev.*, 82, 831.

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

Weinstein, N. (1980) "Unrealistic Optimism about Future Life Events," *Journal of Personality and Social Psychology*, 39 (5), 806–20.

Whitmore, G. A. (1970) "Third-Degree Stochastic Dominance," *American Economic Review*, 60 (3), 457–459.

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

Yang, K. H. and A. K. Zentefis (2022a) "Gerrymandering and the Limits of Representative Democracy," April, Working Paper. Yale University, New Haven, CT.

——— (2022b) "Distributions of Posterior Quantiles and Economic Applications," July, Working Paper. Yale University, New Haven, CT.

Appendix

A.1 Proof of Theorem 1

Consider any $F, G, H \in \mathcal{F}$ such that $G(x) \leq H(x) \leq F(x)$ 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 $\mathcal{I}(F,G)$. To this end, first note that $\mathcal{I}(F,G) \subseteq \mathcal{F}$ 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 $H + \widehat{H} \notin \mathcal{I}(F,G)$ or $H - \widehat{H} \notin \mathcal{I}(F,G)$. Clearly, if $H + \widehat{H} \notin \mathcal{F}$ or $H - \widehat{H} \notin \mathcal{F}$, then it must be that either $H + \widehat{H} \notin \mathcal{I}(F,G)$ or $H - \widehat{H} \notin \mathcal{I}(F,G)$. 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 $H(x_0) \in \{G(x_0), F(x_0)\}$ and hence both $H(x_0) + |\widehat{H}(x_0)| > F(x_0)$ and $H(x_0) - |\widehat{H}(x_0)| < G(x_0)$. Thus, it must be that either $H + \widehat{H} \notin \mathcal{I}(F,G)$ or $H - \widehat{H} \notin \mathcal{I}(F,G)$. 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 $[\underline{x}_n, \overline{x}_n)$ as both $H + \widehat{H}$ and $H - \widehat{H}$ are nondecreasing. Thus, either $H(\underline{x}_n) + |\widehat{H}(\underline{x}_n)| = F(\underline{x}_n) + |\widehat{H}(x_0)| > F(\underline{x}_n)$, or $H(\overline{x}_n^-) - |\widehat{H}(\overline{x}_n^-)| = G(\overline{x}_n^-) - |\widehat{H}(x_0)| < G(\overline{x}_n^-)$, and hence either $H + \widehat{H} \notin \mathcal{I}(F,G)$ or $H - \widehat{H} \notin \mathcal{I}(F,G)$, as desired.

Conversely, suppose that H is an extreme point of $\mathcal{I}(F,G)$. 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 $G(x_0^-) < H(x_0) := \eta < F(x_0)$ for some $x_0 \in \mathbb{R}$, then it must be that either $H(x) = H(x_0)$ for all $x \in [F^{-1}(\eta^+), x_0]$ or $H(x) = H(x_0)$ for all $x \in [x_0, G^{-1}(\eta))$. Indeed, suppose the contrary, so that there exists $\underline{x} \in [F^{-1}(\eta^+), x_0)$ and $\overline{x} \in (x_0, G^{-1}(\eta))$ 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 $H^{-1}(\eta) > F(\eta^+)$ and $H^{-1}(\eta^+) < G(\eta)$. Moreover, since $x \mapsto F(x^-)$ is left-continuous, $H^{-1}(\eta) > \underline{x} \geq F^{-1}(\eta^+)$ implies $F(H^{-1}(\eta)^-) > \eta$. Likewise, $H^{-1}(\eta^+) < \overline{x} < G^{-1}(\eta)$ implies that $G(H^{-1}(\eta^+)) < \eta$. Now define a function $\Phi : [0,1]^2 \to \mathbb{R}^2$ as

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

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

$$G(H^{-1}(\underline{\eta}^+)^-) \leq G(H^{-1}(\overline{\eta}^+)) < \underline{\eta} < \eta < \overline{\eta} < F(H^{-1}(\underline{\eta})^-) \leq F(H^{-1}(\underline{\eta})). \tag{A.9}$$

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 (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 $H_1, H_2 : [H^{-1}(\underline{\eta}), H^{-1}(\overline{\eta}^+)] \to [\underline{\eta}, \overline{\eta}]$ and 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 $\widehat{H}_1, \widehat{H}_2 \in \mathcal{I}(F,G)$. Indeed, for any $i \in \{1,2\}$ and for any $x, y \in \mathbb{R}$ with $x < y$, if either $x, y \notin [H^{-1}(\underline{\eta}), H^{-1}(\overline{\eta}^+)]$, then $\widehat{H}_i(x) = H(x) \leq H(y) \notin \widehat{H}_i(y)$. Meanwhile, if $x, y \in [H^{-1}(\underline{\eta}), H^{-1}(\overline{\eta}^+)]$, then $\widehat{H}_i(x) = h_i(x) \leq h_i(y) = \widehat{H}_i(x)$. If $x < H^{-1}(\underline{\eta})$ and $y \in [H^{-1}(\underline{\eta}), H^{-1}(\overline{\eta}^+)]$, then $\widehat{H}_i(x) = H(x) \leq \underline{\eta} \leq h_i(y) = \widehat{H}_i(y)$. Likewise, if $y > H^{-1}(\overline{\eta}^+)$ and $x \in [H^{-1}(\underline{\eta}), H^{-1}(\overline{\eta}^+)]$, then $\widehat{H}_i(x) = h_i(x) \leq \overline{\eta} \leq H(y) = \widehat{H}_i(y)$. 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.9), we have

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

Together with $H \in \mathcal{I}(F,G)$, it then follows that $G(x) \leq \widehat{H}_i(x) \leq F(x)$ for all $x \in \mathbb{R}$, and hence $\widehat{H}_i \in \mathcal{I}(F,G)$ for all $i \in \{1,2\}$. Consequently, there exists distinct $\widehat{H}_1, \widehat{H}_2 \in \mathcal{I}(F,G)$ 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 $\mathcal{I}(F,G)$, as desired.

As a result, for any extreme point H of $\mathcal{I}(F,G)$, the set $\{x \in \mathbb{R} | G(x) < H(x) < F(x)\}$ can be partitioned into three classes of open intervals: I^F, I^G, and $I^{F,G}$ such that for any open interval $(\underline{x}, \overline{x}) \in I^F$, H is a constant on $[\underline{x}, \overline{x})$ and $H(\underline{x}) = F(\underline{x})$; for any open interval $(\underline{x}, \overline{x}) \in I^G$, H is a constant on $[\underline{x}, \overline{x})$ and $H(\overline{x}^-) = G(\overline{x}^-)$; and for any open interval $(\underline{x}, \overline{x}) \in I^{F,G}$, H is a constant on $[\underline{x}, \overline{x})$ and $F(\underline{x}) = H(\underline{x}) = H(\overline{x}^-) = G(\overline{x}^-)$. Note that since F, G, H are nondecreasing and since $H \in \mathcal{I}(F,G)$, every interval in I^F and I^G 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^F and I^G must be countable. Meanwhile, any distinct intervals $(\underline{x}_1, \overline{x}_1), (\underline{x}_2, \overline{x}_2) \in I^{F,G}$ 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, $I^{F,G}$ must be countable as well.

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

A.2 Proof of Theorem 2

To show that $\mathcal{H}_\tau \subseteq \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, consider any $H \in \mathcal{H}_\tau$. Let $\mu \in \mathcal{M}$ and any $r \in \mathcal{R}$ 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 $x \in \mathbb{R}$,

$$H(x|\mu, r) \leq \mu(\{F \in \mathcal{F} | F^{-1}(\tau) \leq x\}) = \mu(\{F \in \mathcal{F} | F(x) \geq \tau\}).$$

Now consider any $x \in \mathbb{R}$. Clearly, $\mu(\{F \in \mathcal{F} | F(x) \geq \tau\}) \leq 1$, since μ is a probability measure. Moreover, let $M_x^+(q) := \mu(\{F \in \mathcal{F} | F(x) \geq q\})$ 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 $F_0(x)$. Therefore, whenever $\tau \geq F_0(x)$, then $M_x^+(\tau)$ can be at most $F_0(x)/\tau$ to have a mean of $F_0(x)$.[29] Together, this implies that $\mu(\{F \in \mathcal{F} | F(x) \geq \tau\}) \leq \underline{F}_0^\tau(x)$ 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 $x \in \mathbb{R}$,

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

Now consider any $x \in \mathbb{R}$. Since μ is a probability measure, it must be that $\mu(\{F \in \mathcal{F} | F(x) > \tau\}) \geq 0$. Furthermore, let $M_x^-(q) := \mu(\{F \in \mathcal{F} | F(x) > q\})$ 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 $F_0(x)$. Therefore, whenever $\tau \leq F_0(x)$, then $M_x^-(\tau)$ must be at least $(F_0(x) - \tau)/(1 - \tau)$ to have a mean of $F_0(x)$.[30] Together, this implies that $\mu(\{F \in \mathcal{F} | F(x) > \tau\}) \geq \overline{F}_0^\tau$ for all $x \in \mathbb{R}$, which, in turn, implies that $\overline{F}_0^\tau(x) \leq H^\tau(x^-|\mu, r) \leq H^\tau(x|\mu, r) \leq \underline{F}_0^\tau(x)$ for all $x \in \mathbb{R}$, as desired.

To prove that $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau) \subseteq \mathcal{H}_\tau$, we first show that for any extreme point H of $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, there exists a signal $\mu \in \mathcal{M}$ and a selection rule $r \in \mathcal{R}$ such that $H(x) = H^\tau(x|\mu, r)$ for all $x \in \mathbb{R}$. Consider any extreme point H of $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$. 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 $(1 - \underline{F}_0^\tau(x))\overline{F}_0^\tau(x) = 0$ for all $x \notin [F_0^{-1}(\tau), F_0^{-1}(\tau^+)]$, there exists at most one $n \in \mathbb{N}$ such that $0 < H(\underline{x}_n) = \underline{F}_0^\tau(\underline{x}_n) = \overline{F}_0^\tau(\overline{x}_n^-) = H(\overline{x}_n^-) < 1$. Therefore, for $\underline{x}$ and $\overline{x}$ defined as

$$\underline{x} := \sup\{\underline{x}_n | n \in \mathbb{N},\, H(\underline{x}_n) = \underline{F}_0^\tau(\underline{x}_n)\},$$

and

$$\overline{x} := \inf\{\overline{x}_n | n \in \mathbb{N},\, H(\overline{x}_n^-) = \overline{F}_0^\tau(\overline{x}_n^-)\},$$

respectively, it must be that $\overline{x} \geq \underline{x}$, and that for all $n \in \mathbb{N}$, either $\overline{x}_n \leq \overline{x}$ and $H(\underline{x}_n) = \underline{F}_0^\tau(\underline{x}_n)$; or $\underline{x}_n \geq \underline{x}$ and $H(\overline{x}_n^-) = \overline{F}_0^\tau(\overline{x}_n^-)$. Henceforth, let $\mathbb{N}_1$ be the collection of $n \in \mathbb{N}$ such that $\overline{x}_n \leq \overline{x}$ and $H(\underline{x}_n) = \underline{F}_0^\tau(\underline{x}_n)$, and let $\mathbb{N}_2$ be the collection of $n \in \mathbb{N}$ such that $\underline{x}_n \geq \underline{x}$ and $H(\overline{x}_n^-) = \overline{F}_0^\tau(\overline{x}_n^-)$. 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 $r \in \mathcal{R}$ 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, $\underline{F}_0^\tau(\hat{x}) \geq H(\hat{x}) = \underline{F}_0^\tau(\underline{x}) + \eta$, and hence $F_0(\hat{x}) - \tau\eta \geq F_0(\underline{x})$. Likewise, $F_0(\hat{x}) + (1 - \tau)\eta \leq F_0(\overline{x}^-)$. Let

$$\underline{y} := F_0^{-1}([F_0(\hat{x}) - \tau\eta]^+), \quad \text{and} \quad \overline{y} := F_0^{-1}(F_0(\hat{x}) + (1 - \tau)\eta)).$$

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 $\widehat{F}_0$

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

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

as follows: $\widehat{F}_0 \equiv 0$ if $\eta = 0$; and

$$\widehat{F}_0(x) := \begin{cases} 0, & \text{if } x < \underline{y} \\ \frac{F_0(x) - (F_0(\hat{x}) - \tau\eta)}{\eta}, & \text{if } x \in [\underline{y}, \overline{y}) \\ 1, & \text{if } x \geq \overline{y} \end{cases},$$

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

$$\widetilde{F}_0(x) := \frac{F_0(x) - \eta \widehat{F}_0(x)}{1 - \eta}.$$

By construction, $\eta\widehat{F}_0 + (1-\eta)\widetilde{F}_0 = F_0$. From the definition of $\underline{y}$ and $\overline{y}$, it can be shown that $\widetilde{F}_0 \in \mathcal{F}$ as well. Furthermore,

$$\widetilde{F}_0(\overline{x}^-) - \widetilde{F}_0(\underline{x}) = \frac{F_0(\overline{x}^-) - F_0(\underline{x}) - \eta}{1 - \eta} = \frac{1}{1-\eta}\left[\frac{\tau}{1-\tau}(1 - F_0(\overline{x}^-)) + \frac{1-\tau}{\tau}F_0(\underline{x})\right].$$

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

$$\widetilde{F}_1(x) := \begin{cases} \frac{F_0(x)}{F_0(\underline{x}) + \alpha(F_0(\overline{x}^-) - F_0(\underline{x}) - \eta)}, & \text{if } x < \underline{x} \\ \frac{F_0(\underline{x})\alpha(F_0(x) - F_0(\underline{x}) - \eta)}{F_0(\underline{x}) + \alpha(F_0(\overline{x}^-) - F_0(\underline{x}) - \eta)}, & \text{if } x \in [\underline{x}, \overline{x}) \\ 1, & \text{if } x \geq \overline{x} \end{cases};$$

and

$$\widetilde{F}_2(x) := \begin{cases} 0, & \text{if } x < \underline{x} \\ \frac{(1-\alpha)(F_0(x) - F_0(\underline{x}) - \eta)}{1 - F_0(\overline{x}^-) + (1-\alpha)(F_0(\overline{x}^-) - F_0(\underline{x}) - \eta)}, & \text{if } x \in [\underline{x}, \overline{x}) \\ \frac{F_0(x) - F_0(\underline{x}) + (1-\alpha)(F_0(\overline{x}^-) - F_0(\underline{x}) - \eta)}{1 - F_0(\overline{x}^-) + (1-\alpha)(\widetilde{F}_0(\overline{x}^-) - \widetilde{F}_0(\underline{x}) - \eta)}, & \text{if } x \geq \overline{x} \end{cases},$$

where

$$\alpha := \frac{\frac{1-\tau}{\tau}F_0(\underline{x})}{\frac{\tau}{1-\tau}(1 - F_0(\overline{x}^-)) + \frac{1-\tau}{\tau}F_0(\underline{x})}.$$

By construction, $\widetilde{\alpha}\widetilde{F}_1 + (1-\widetilde{\alpha})\widetilde{F}_2 = \widetilde{F}_0$, where $\widetilde{\alpha} \in (0,1)$ is given by $\widetilde{\alpha} := [F_0(\underline{x}) + \alpha(F_0(\overline{x}^-) - F_0(\underline{x}) - \eta)]/(1-\eta)$. 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:

$$\widetilde{F}_1^x(z) := \begin{cases} 0, & \text{if } z < x \\ \widetilde{F}_0(\underline{x}), & \text{if } z \in [x, \underline{x}) \\ \widetilde{F}_0(z), & \text{if } z \geq \underline{x} \end{cases}; \text{ and } \widetilde{F}_2^x(z) := \begin{cases} \widetilde{F}_0(z), & \text{if } z < \overline{x} \\ \widetilde{F}_0(\overline{x}^-), & \text{if } z \in [\overline{x}, x) \\ 1, & \text{if } z \geq x \end{cases}.$$

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

$$\widetilde{F}_1^n(z) := \frac{1}{\widetilde{F}_0(\overline{x}_n) - \widetilde{F}_0(\underline{x}_n)} \int_{\underline{x}_n}^{\overline{x}_n} \widetilde{F}_1^x(z)\widetilde{F}_0(\mathrm{d}x),$$

and

$$\widetilde{F}_2^m(z) := \frac{1}{\widetilde{F}_0(\overline{x}_m) - \widetilde{F}_0(\underline{x}_m)} \int_{\underline{x}_m}^{\overline{x}_m} \widetilde{F}_2^x(z)\,\mathrm{d}\widetilde{F}_0(\mathrm{d}x),$$

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

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

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

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

$$\int_{\mathcal{F}} F(z)\tilde{\mu}(\mathrm{d}F) = \int_{\mathbb{R}} \widetilde{G}^x(z)\widetilde{H}(\mathrm{d}x) = \widetilde{F}_0(z). \tag{A.10}$$

Moreover, let $\tilde{r} : \mathcal{F} \times (0,1) \to \Delta(\mathbb{R})$ be defined as

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

for all $F \in \mathcal{F}$ and for all $\hat{\tau} \in (0,1)$. It then follows that $H^\tau(x|\tilde{\mu}, \tilde{r}) = \widetilde{H}(x)$ for all $x \in \mathbb{R}$. Next, let $\mu \in \Delta(\mathcal{F}), r \in \mathcal{R}$ together be defined as

$$\mu := (1-\eta)\tilde{\mu} + \eta\delta_{\{\widehat{F}_0\}},$$

and

$$r(F, \hat{\tau}) := \begin{cases} \delta_{\{\hat{x}\}}, & \text{if } F = \widehat{F}_0, \tau = \hat{\tau} \\ \tilde{r}(F, \hat{\tau}), & \text{otherwise} \end{cases},$$

for all $F \in \mathcal{F}$ and for all $\hat{\tau} \in (0,1)$. Since $F_0 = \eta\widehat{F}_0 + (1-\eta)\widetilde{F}_0$, together with (A.10), 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 $\gamma \in \Delta(\mathbb{R} \times \mathcal{F})$ such that $\gamma(\{(x, F) \in \mathbb{R} \times \mathcal{F} | x \in [F^{-1}(\tau), F^{-1}(\tau^+)]\}) = 1$ and

$$\int_{\mathbb{R}\times\mathcal{F}} F(x)\gamma(\mathrm{d}x, \mathrm{d}F) = F_0(x),$$

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

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

for all $\gamma \in \Gamma$ and for all $x \in \mathbb{R}$. Then, since for any $\widehat{H}$ in the set of extreme points $\text{ext}(\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau))$ of $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, there exists $\hat{\mu} \in \mathcal{M}$ and $\hat{r} \in \mathcal{R}$ such that $H^\tau(x|\hat{\mu}, \hat{r}) = \widehat{H}(x)$ for all $x \in \mathbb{R}$, it must be that $\text{ext}(\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)) \subseteq \Xi(\Gamma)$.

Now consider any $H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$. Since $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ is a compact and convex set of a metrizable, locally convex topological space,[31] Choquet's theorem implies that there exists a probability measure $\Lambda_H \in \Delta(\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau))$ with $\Lambda_H(\text{ext}(\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau))) = 1$ such that

$$\int_{\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)} \widehat{H}(x)\Lambda_H(\mathrm{d}\widehat{H}) = H(x),$$

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

$$\int_\mathcal{F} F(x)\mu(\mathrm{d}F) = \int_\Gamma \left(\int_{\mathbb{R}\times\mathcal{F}} F(x)\tilde{\gamma}(\mathrm{d}\tilde{x}, \mathrm{d}F)\right)\widetilde{\Lambda}_H(\mathrm{d}\tilde{\gamma}) = F_0(x),$$

and hence $\mu \in \mathcal{M}$. Furthermore, by the disintegration theorem (c.f., Çinlar 2010, theorem 2.18), there exists a transition probability $\xi : \mathcal{F} \to \Delta(\mathbb{R})$ such that $\gamma(\mathrm{d}x, \mathrm{d}F) = \xi(\mathrm{d}x|F)\mu(\mathrm{d}F)$. Let $r(F, \hat{\tau}) := \xi(F)$ for all $F \in \mathcal{F}$ and for all $\hat{\tau} \in (0, 1)$. Since $\widetilde{\Lambda}_H(\Gamma) = 1$, we have $r \in \mathcal{R}$. Finally, for any $x \in \mathbb{R}$, since Ξ is affine,

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

as desired. This completes the proof. ■

[31] To see this, recall that for any sequence $\{H_n\} \subseteq \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, 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 $H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$.

A.3 Proof of Theorem 3

By Theorem 2, for any $\varepsilon > 0$,

$$\mathcal{H}_\tau^0 \subseteq \mathcal{H}_\tau = \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau).$$

It remains to show that

$$\bigcup_{\varepsilon>0} \mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon}) \subseteq \mathcal{H}_\tau^0.$$

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

$$\underline{x} := \sup\{\underline{x}_n | n \in \mathbb{N},\, H(\underline{x}_n) = \underline{F}_0^{\tau,\varepsilon}(\underline{x}_n)\} \quad \text{and} \quad \overline{x} := \inf\{\overline{x}_n | n \in \mathbb{N},\, H(\overline{x}_n^-) = \overline{F}_0^{\tau,\varepsilon}(\overline{x}_n^-)\},$$

respectively, it must be that $\overline{x} \geq \underline{x}$, and that, for all $n \in \mathbb{N}$, either $\overline{x}_n \leq \overline{x}$ and $H(\underline{x}_n) = \underline{F}_0^{\tau,\varepsilon}(\underline{x}_n)$, or $\underline{x}_n \geq \underline{x}$ and $H(\overline{x}_n^-) = \overline{F}_0^{\tau,\varepsilon}(\overline{x}_n^-)$. Henceforth, let $\mathbb{N}_1$ be the collection of $n \in \mathbb{N}$ such that $\overline{x}_n \leq \overline{x}$ and $H(\underline{x}_n) = \underline{F}_0^{\tau,\varepsilon}(\underline{x}_n)$, and let $\mathbb{N}_2$ be the collection of $n \in \mathbb{N}$ such that $\underline{x}_n \geq \underline{x}$ and $H(\overline{x}_n^-) = \overline{F}_0^{\tau,\varepsilon}(\overline{x}_n^-)$. 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 $\mu \in \mathcal{M}_\tau^0$ 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, $\underline{F}_0^{\tau,\varepsilon}(\hat{x}) \geq H(\hat{x}) = \underline{F}_0^{\tau,\varepsilon}(\underline{x}) + \eta$, and hence $F_0(\hat{x}) - (\tau + \varepsilon)\eta \geq F_0(\underline{x})$. Likewise, $F_0(\hat{x}) + (1 - \tau + \varepsilon)\eta \leq F_0(\overline{x}^-)$. Now let

$$\underline{y} := F_0^{-1}([F_0(\hat{x}) - \tau\eta]), \quad \text{and} \quad \overline{y} := F_0^{-1}(F_0(\hat{x}) + (1 - \tau)\eta).$$

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 $\widehat{F}_0$ as follows: $\widehat{F}_0 \equiv 0$ if $\eta = 0$; and

$$\widehat{F}_0(x) := \begin{cases} 0, & \text{if } x < \underline{y} \\ \frac{F_0(x) - (F_0(\hat{x}) - \tau\eta)}{\eta}, & \text{if } x \in [\underline{y}, \overline{y}) \\ 1, & \text{if } x \geq \overline{y} \end{cases},$$

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

$$\widetilde{F}_0(x) := \frac{F_0(x) - \eta\widehat{F}_0(x)}{1 - \eta}.$$

By construction, $\eta\widehat{F}_0 + (1 - \eta)\widetilde{F}_0 = F_0$. From the definition of $\underline{y}$ and $\overline{y}$, it can be shown that $\widetilde{F}_0 \in \mathcal{F}$ as well. Furthermore,

$$\widetilde{F}_0(\overline{x}^-) - \widetilde{F}_0(\underline{x}) = \frac{F_0(\overline{x}^-) - F_0(\underline{x}) - \eta}{1 - \eta} = \frac{1}{1 - \eta}\left[\frac{\tau - \varepsilon}{1 - (\tau - \varepsilon)}(1 - F_0(\overline{x}^-)) + \frac{1 - (\tau + \varepsilon)}{\tau + \varepsilon}F_0(\underline{x})\right].$$

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

$$\widetilde{F}_1(x) := \begin{cases} \frac{F_0(x)}{F_0(\underline{x})+\alpha(F_0(\overline{x}^-)-F_0(\underline{x})-\eta)}, & \text{if } x < \underline{x} \\ \frac{F_0(\underline{x})\alpha(F_0(x)-F_0(\underline{x})-\eta)}{F_0(\underline{x})+\alpha(F_0(\overline{x}^-)-F_0(\underline{x})-\eta)}, & \text{if } x \in [\underline{x}, \overline{x}) \\ 1, & \text{if } x \geq \overline{x} \end{cases} ;$$

and

$$\widetilde{F}_2(x) := \begin{cases} 0, & \text{if } x < \underline{x} \\ \frac{(1-\alpha)(F_0(x)-F_0(\underline{x})-\eta)}{1-F_0(\overline{x}^-)+(1-\alpha)(F_0(\overline{x}^-)-F_0(\underline{x})-\eta)}, & \text{if } x \in [\underline{x}, \overline{x}) \\ \frac{F_0(x)-F_0(\underline{x})+(1-\alpha)(F_0(\overline{x}^-)-F_0(\underline{x})-\eta)}{1-F_0(\overline{x}^-)+(1-\alpha)(\widetilde{F}_0(\overline{x}^-)-\widetilde{F}_0(\underline{x})-\eta)}, & \text{if } x \geq \overline{x} \end{cases} ,$$

where

$$\alpha := \frac{\frac{1-(\tau+\varepsilon)}{\tau+\varepsilon} F_0(\underline{x})}{\frac{\tau-\varepsilon}{1-(\tau-\varepsilon)}(1 - F_0(\overline{x}^-)) + \frac{1-(\tau+\varepsilon)}{\tau+\varepsilon} F_0(\underline{x})}.$$

By construction, $\widetilde{\alpha}\widetilde{F}_1+(1-\widetilde{\alpha})\widetilde{F}_2 = \widetilde{F}_0$, where $\widetilde{\alpha} \in (0,1)$ is given by $\widetilde{\alpha} := [F_0(\underline{x})+\alpha(F_0(\overline{x}^-)-F_0(\underline{x})-\eta)]/(1-\eta)$. 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:

$$\widetilde{F}_1^x(z) := \begin{cases} 0, & \text{if } z < x \\ \widetilde{F}_0(\underline{x}), & \text{if } z \in [x, \underline{x}) \\ \widetilde{F}_0(z), & \text{if } z \geq \underline{x} \end{cases} ; \text{ and } \widetilde{F}_2^x(z) := \begin{cases} \widetilde{F}_0(z), & \text{if } z < \overline{x} \\ \widetilde{F}_0(\overline{x}^-), & \text{if } z \in [\overline{x}, x) \\ 1, & \text{if } z \geq x \end{cases} .$$

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

$$\widetilde{F}_1^n(z) := \frac{1}{\widetilde{F}_0(\overline{x}_n) - \widetilde{F}_0(\underline{x}_n)} \int_{\underline{x}_n}^{\overline{x}_n} \widetilde{F}_1^x(z) \widetilde{F}_0(\mathrm{d}x),$$

and

$$\widetilde{F}_2^m(z) := \frac{1}{\widetilde{F}_0(\overline{x}_m) - \widetilde{F}_0(\underline{x}_m)} \int_{\underline{x}_m}^{\overline{x}_m} \widetilde{F}_2^x(z) \, \mathrm{d}\widetilde{F}_0(\mathrm{d}x),$$

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

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

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

$$\int_{\mathcal{F}} F(z)\tilde{\mu}(\mathrm{d}F) = \int_{\mathbb{R}} \widetilde{G}^x(z)\widetilde{H}(\mathrm{d}x) = \widetilde{F}_0(z). \tag{A.11}$$

Furthermore, $H^\tau(x|\tilde{\mu}) = \widetilde{H}(x)$ for all $x \in \mathbb{R}$. As a result, from (A.11), for $\mu \in \Delta(\mathcal{F})$ defined as

$$\mu := (1-\eta)\tilde{\mu} + \eta\delta_{\{\widehat{F}_0\}},$$

since $F_0 = \eta\widehat{F}_0 + (1-\eta)\widetilde{F}_0$, it must be that $\mu \in \mathcal{M}^0_\tau$. 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 $H \in \mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon})$. Since $\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon})$ is a convex and compact set in a metrizable space, Choquet's theorem implies that there exists a probability measure $\Lambda_H \in \Delta(\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon}))$ that assigns probability 1 to $\mathrm{ext}(\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon}))$ such that

$$H(x) = \int_{\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon})} \widetilde{H}(x)\Lambda_H(\mathrm{d}\widetilde{H}).$$

Meanwhile, define the linear functional $\Xi : \mathcal{M}^0_\tau \to \mathcal{F}$ as

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

for all $\tilde{\mu} \in \mathcal{M}^0_\tau$ and for all $x \in \mathbb{R}$. Now define a probability measure $\widetilde{\Lambda}$ on $\mathcal{M}^0_\tau$ by

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

for all $A \subseteq \mathcal{M}^0_\tau$. Then, since $\Lambda_H(\mathrm{ext}(\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon}))) = 1$ and since, for any $\widetilde{H} \in \mathrm{ext}(\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon}))$, there exists $\tilde{\mu} \in \mathcal{M}^0_\tau$ such that $H(x) = H^\tau(x|\tilde{\mu})$, it must be that $\widetilde{\Lambda}_H(\mathcal{M}^0_\tau(U)) = 1$, and hence $\widetilde{\Lambda}_H$ is a probability measure on $\mathcal{M}^0_\tau$. Let $\tilde{\mu} \in \mathcal{M}^0_\tau(U)$ be defined as

$$\tilde{\mu}(A) := \int_{\mathcal{M}^0_\tau} \mu(A)\widetilde{\Lambda}_H(\mathrm{d}\mu),$$

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

$$\begin{aligned} H(x) = \int_{\mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau+\varepsilon})} \widetilde{H}(x)\Lambda_H(\mathrm{d}\widetilde{H}) = &\int_{\mathcal{M}^0_\tau} \Xi(\mu)[x]\widetilde{\Lambda}_H(\mathrm{d}\mu) \\ =&\Xi(\tilde{\mu})[x] \\ =&H^\tau(x|\tilde{\mu}), \end{aligned}$$

and therefore, $H \in \mathcal{H}^0_\tau$. Together, for any $\varepsilon > 0$, any $H \in \mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon})$ must be in $\mathcal{H}^0_\tau$. In other words,

$$\bigcup_{\varepsilon>0} \mathcal{I}(\underline{F}_0^{\tau,\varepsilon}, \overline{F}_0^{\tau,\varepsilon}) \subseteq \mathcal{H}_0^\tau.$$

This completes the proof. ■

A.4 Proof of Corollary 1

For 1, consider any $H \in \mathcal{H}_{\tilde{\tau}}$. By Theorem 2, $H \in \mathcal{I}(\underline{F}_0^{\tilde{\tau}}, \overline{F}_0^{\tilde{\tau}})$. Thus, $(\underline{F}_0^{\tilde{\tau}})^{-1}(\tau) \leq H^{-1}(\tau) \leq H^{-1}(\tau^+) \leq (\overline{F}_0^{\tilde{\tau}})^{-1}(\tau^+)$, and therefore $[H^{-1}(\tau), H^{-1}(\tau^+)] \subseteq [(\underline{F}_0^{\tilde{\tau}})^{-1}(\tau), (\overline{F}_0^{\tilde{\tau}})^{-1}(\tau^+)]$. Conversely, consider any interval $Q = [\underline{x}, \overline{x}] \subseteq [(\underline{F}_0^{\tilde{\tau}})^{-1}(\tau), (\overline{F}_0^{\tilde{\tau}})^{-1}(\tau^+)]$. Then, let H be defined as

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

for all $x \in \mathbb{R}$. Then $H \in \mathcal{I}(\underline{F}_0^{\tilde{\tau}}, \overline{F}_0^{\tilde{\tau}})$ and $Q = [H^{-1}(\tau), H^{-1}(\tau^+)]$. Moreover, by Theorem 2, $H \in \mathcal{H}_{\tilde{\tau}}$, as desired.

For 2, consider any $H \in \mathcal{H}^0_{\tilde{\tau}}$. By Theorem 2, $H \in \mathcal{I}(\underline{F}_0^{\tilde{\tau}}, \overline{F}_0^{\tilde{\tau}})$. Thus, it must be that $[H^{-1}(\tau), H^{-1}(\tau^+)] \subseteq [(\underline{F}_0^{\tilde{\tau}})^{-1}(\tau), (\overline{F}_0^{\tilde{\tau}})^{-1}(\tau)]$. Conversely, for any $\hat{x} \in ((\underline{F}_0^{\tilde{\tau}})^{-1}(\tau), (\overline{F}_0^{\tilde{\tau}})^{-1}(\tau^+))$, note that since $\hat{x} > (\underline{F}_0^{\tilde{\tau}})^{-1}(\tau)$ and since F_0 is continuous, we have $F_0(\hat{x})/\tau > \tilde{\tau}$. Similarly, we also have $(F_0(\hat{x}) - \tau)/(1-\tau) < \tilde{\tau}$. Let $\varepsilon := \min\{F_0(\hat{x})\tau - \tilde{\tau}, \tilde{\tau} - (F_0(\hat{x}) - \tau)/(1-\tau)\}$. Then, either $\hat{x} = (\underline{F}_0^{\tilde{\tau},\varepsilon})^{-1}(\tau)$ or $\hat{x} = (\overline{F}_0^{\tilde{\tau},\varepsilon})^{-1}(\tau)$. Since both $\underline{F}_0^{\tilde{\tau},\varepsilon}$ and $\overline{F}_0^{\tilde{\tau},\varepsilon}$ are in $\mathcal{I}(\underline{F}_0^{\tilde{\tau},\varepsilon}, \overline{F}_0^{\tilde{\tau},\varepsilon})$, Theorem 3 implies that $\hat{x} = H^{-1}(\tau)$ for some $H \in \mathcal{H}^0_{\tilde{\tau}}$. Lastly, note that under a signal $\mu \in \mathcal{M}$ such that μ assigns probability τ to $\underline{F}_0^\tau$ and probability $1-\tau$ to $\overline{F}_0^\tau$, we have $\mu \in \mathcal{M}^0_{\tilde{\tau}}$ and$H^{\tilde{\tau}}(x|\mu) = \tau$ for all $x \in [(\underline{F}_0^{\tilde{\tau}})^{-1}(\tau), (\overline{F}_0^{\tilde{\tau}})^{-1}(\tau)]$. Hence, $[(\underline{F}_0^{\tilde{\tau}})^{-1}(\tau), (\overline{F}_0^{\tilde{\tau}})^{-1}(\tau)] \subseteq [H^{-1}(\tau), H^{-1}(\tau^+)]$ for some $H \in \mathcal{H}^0_{\tilde{\tau}}$, as desired.

■