Discussion Paper Series – CRC TR 224

Discussion Paper No. 564
Project B 02

The Comparative Statics of Persuasion

Gregorio Curello[1]
Ludvig Sinander[2]

June 2024

[1]University of Bonn
[2]University of Oxford

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

The comparative statics of persuasion*

Gregorio Curello
University of Bonn

Ludvig Sinander
University of Oxford

22 May 2024

Abstract

In the canonical persuasion model, comparative statics has been an open question. We answer it, delineating which shifts of the sender's interim payoff lead her optimally to choose a more informative signal. Our first theorem identifies a coarse notion of 'increased convexity' that we show characterises those shifts of the sender's interim payoff that lead her optimally to choose *no less* informative signals. To strengthen this conclusion to '*more* informative' requires further assumptions: our second theorem identifies the necessary and sufficient condition on the sender's interim payoff, which strictly generalises the 'S' shape commonly imposed in the literature. We identify conditions under which increased alignment of interests between sender and receiver lead to comparative statics, and study a number of applications.

1 Introduction

The persuasion model of Kamenica and Gentzkow (2011) is by now canonical, yet has proved intractable. Little is known about the qualitative properties of

*We are grateful for comments from Dağhan Carlos Akkar, Itai Arieli, Yunus Aybas, Gabe Carroll, Tommaso Denti, Piotr Dworczak, Michael Eldar, Pia Ennuschat, Matteo Escudé, Alkis Georgiadis-Harris, Ben Golub, Duarte Gonçalves, Olivier Gossner, Ian Jewitt, Paul Klemperer, Peter Klibanoff, Anton Kolotilin, Annie Liang, Elliot Lipnowski, Thomas Mariotti, Laurent Mathevet, Meg Meyer, Wojciech Olszewski, Paula Onuchic, Marco Ottaviani, Alessandro Pavan, Antonio Penta, Jacopo Perego, Philipp Strack, Bruno Strulovici, Alex Wolitzky, Kun Zhang and audiences at Arizona State, Berlin, Bonn, Caltech, Cambridge, Cergy, City, CREST, EUI, Glasgow, LSE, Mannheim, Michigan, Nottingham, Oxford, Pompeu Fabra, Surrey, Tilburg, UCL, VSET, the 2022 Warwick Theory Workshop, the 2022 & 2023 Transatlantic Theory Workshops, the Saarland Workshop in Economic Theory, the 2023 Workshop on the Design of Strategic Interactions, SAET 2022 & 2023, and the 2022 European Winter Meeting of the Econometric Society.

Support by the German Research Foundation (DFG) through CRC TR 224 (Project B02) is gratefully acknowledged.

optimal signals beyond very special cases, such as when the sender's interim payoff is 'S'-shaped or the state is binary.

In this paper, we advance our understanding of optimal signals in the canonical persuasion model by changing the question: rather than ask what optimal signals look like, we ask how they vary with economic primitives. Concretely, we pose and answer the comparative-statics question: what shifts of model primitives, specifically of the sender's interim payoff, lead her optimally to choose a more informative signal?

Recall that the persuasion model features an uncertain state of the world, whose distribution is called the *prior,* and a character called the *sender.* The sender flexibly designs what will and won't be revealed about the state, by choosing a *signal.* The model's primitives are the prior and the sender's interim payoff function, which maps each posterior belief into an expected payoff. (This interim payoff is a reduced-form description of a downstream interaction, typically involving other players called 'receivers'.)

Motivated by applications, we focus on the case in which the sender's interim payoff depends on only one moment of the posterior belief—without loss, the mean. This 'single-moment' assumption is maintained in much of the recent literature.

Our first theorem shows that a coarse notion of 'less convex than' characterises '*non-decreasing*' comparative statics: coarsely more convex interim payoffs are exactly those that lead *no less* informative signals to be chosen by the sender, whatever the prior.

Our main theorem characterises what more is needed to obtain *increasing* comparative statics: it identifies a property of interim payoffs that is necessary and sufficient for coarse-convexity shifts to cause *more* informative signals to be chosen, whatever the prior. This property, the *crater property,* is a simple geometric condition that strictly generalises the 'S' shape commonly assumed in the literature.

A string of further results shows that our main theorem is robust. We further show that shifts of the prior *cannot* produce robust comparative statics, and that relaxing the 'single-moment' assumption also yields impossibility.

The crater property is demanding. Nevertheless, we show that it is satisfied in a number of applications, permitting new comparative-statics conclusions to be drawn about the problems of persuading a privately informed receiver (Kolotilin, Mylovanov, Zapechelnyuk & Li, 2017), persuading voters (à la Alonso & Câmara, 2016), designing (health) risk warnings (Mariotti, Schweizer, Szech & von Wangenheim, 2023), costly information acquisition (e.g. Ravid, Roesler & Szentes, 2022), discretionary delegation (e.g. Xu, 2024), and persuasion with choice.

Finally, we ask whether and when increased alignment of interests between the sender and a *receiver* (who takes an action) yields a coarse-convexity shift, and thus potentially greater information-provision. We identify a simple condition that is sufficient and almost necessary.

1.1 Related literature

The persuasion model was introduced by Kamenica and Gentzkow (2011), with precedents in Aumann and Maschler (1968/1995), Brocas and Carrillo (2007) and Rayo and Segal (2010). A great deal of effort has been devoted to characterising optimal signals, yielding sharp descriptions of information-provision in a few special cases as well as some high-level general insights.[1]

Comparative statics has been an open question. It has been answered only in a handful of special cases, each of which concerns a particular sort of shift of interim payoffs and restricts attention either to 'S'-shaped interim payoffs (Kolotilin, Mylovanov & Zapechelnyuk, 2022), to binary priors (Yoder, 2022), or both (Gitmez & Molavi, 2023). Our theorems nest all of these cases, as we detail in §4.6 below.

We are informed and inspired by the general theory of monotone comparative statics (e.g. Topkis, 1978; Milgrom & Shannon, 1994; Quah & Strulovici, 2009). The results of that literature turn out to be of limited use for obtaining our main theorem, however—our proofs instead exploit the particular structure of the persuasion model. A detailed discussion of how our analysis relates to the comparative-statics literature is given in appendix J.

At a high level, our work bears a kinship with Anderson and Smith (2024). Like us, these authors consider a canonical model (Becker's (1973) marriage model) in which optimal/equilibrium outcomes have proved difficult to characterise outside of very special cases, and make progress by instead posing and answering a comparative-statics question.

1.2 Roadmap

We describe the persuasion model in the next section. In §3, we characterise 'non-decreasing' comparative statics in terms of a coarse notion of 'less convex than' (Theorem 1). We then (§4) give necessary and sufficient conditions for 'increasing' comparative statics (Theorem 2 and several propositions).

[1] See e.g. Kolotilin (2014, 2018), Dworczak and Martini (2019), Kleiner, Moldovanu and Strack (2021), Arieli, Babichenko, Smorodinsky and Yamashita (2023), Dworczak and Kolotilin (2022) and Kolotilin, Corrao and Wolitzky (2024).

In §5 we consider shifts of the prior, and in §6 we drop the 'single-moment' assumption. We conclude in §7 by studying alignment and applications.

2 The persuasion model

There is an uncertain state of the world, formally a random variable taking values in a bounded interval $[\underline{x}, \bar{x}]$. We assume without loss of generality that $\underline{x} = 0$ and $\bar{x} = 1$. We shall use the term *distribution* to refer to CDFs $[0,1] \to [0,1]$. We write F_0 for the distribution of the state, and refer to it as 'the prior (distribution)'. For two distributions F and G, recall that F is a mean-preserving contraction of G exactly if

$$\int_0^x F \leq \int_0^x G \quad \text{for every } x \in [0,1], \text{ with equality at } x = 1,$$

or equivalently iff $\int \psi \mathrm{d}F \leq \int \psi \mathrm{d}G$ for every convex $\psi : [0,1] \to \mathbf{R}$.[2]

A sender chooses a *signal*, i.e. a random variable jointly distributed with the state.[3] Given a signal, each signal realisation induces a posterior belief via Bayes's rule, whose expectation we call the *posterior mean*. Each signal thus induces a random posterior mean, with some distribution. Call a distribution *feasible (given F_0)* iff it is the posterior-mean distribution induced by some signal. Kolotilin (2014, Proposition 1) showed that the feasible distributions are precisely the mean-preserving contractions of the prior F_0.[4]

The sender's (interim) payoff at a given realised posterior belief is assumed to depend only on its mean: her payoff at posterior mean $m \in [0,1]$ is $u(m)$, where $u : [0,1] \to \mathbf{R}$ is upper semi-continuous. Her problem is to choose among the feasible distributions F to maximise her expected payoff $\int u \mathrm{d}F$.

Remark 1. Our assumption that only the mean matters is motivated by applications, where it is common for payoffs to depend on a single moment of the posterior distribution—without loss, the mean.[5] This 'single-moment' assumption is maintained in much of the recent persuasion literature. We relax it in §6 below.

[2] See e.g. Shaked and Shanthikumar (2007, §3.A).

[3] Formally, a signal is (M, π), where M is a compact metric space and π is a Borel measurable map $[0,1] \to \Delta(M)$, where $\Delta(M)$ is set of all the Borel probabilities on M, with the topology of weak convergence. The interpretation is that M is a set of messages, and that $\pi(x) \in \Delta(M)$ is the distribution of messages sent if the state is $x \in [0,1]$.

[4] This result may be traced to Hardy, Littlewood and Pólya (1929) and Blackwell (1951).

[5] This is without loss because if payoffs depend on the interim expectation of $f(X)$, where X is the state of the world and $f : [0,1] \to \mathbf{R}$ is continuous, then we may re-define the state of the world to be $Y := f(X)$.

2.1 Informativeness

Definition 1. For distributions F and G, we call F *less informative* than G exactly if F is a mean-preserving contraction of G.

This captures informativeness in the spirit of Blackwell: a less informative distribution is precisely one that is preferred ex-ante by every expected-utility decision-maker who cares about the state only through its mean.[6]

Since there need not be a unique optimal posterior-mean distribution, comparative statics requires comparing *sets* of distributions. We handle this in standard fashion by using the *weak set order:* for two sets S, S' of feasible distributions, we call S *lower* than S' exactly if for any $F \in S$ and $G \in S'$, there is a distribution in S' that is more informative than F, and there is a distribution in S that is less informative than G. We say that S is *strictly lower* than S' exactly if S is lower than S' and S' is not lower than S. Finally, we call S' *(strictly) higher* than S exactly if S is (strictly) lower than S'.

2.2 Interpretation

The interim payoff $u : [0, 1] \to \mathbf{R}$ is a reduced-form object, capturing the (expected) payoff consequences for the sender of whatever downstream interaction takes place after her chosen signal realises.

In the simplest case, the downstream interaction involves a (single) *receiver* taking an action. Formally, there is a non-empty set $\mathcal{A}$ of actions, and the sender's and receiver's interim payoffs $U_S(a, m)$ and $U_R(a, m)$ depend on the chosen action $a \in \mathcal{A}$ and on the mean $m \in [0, 1]$ of their (posterior) belief about the state.[7] When the posterior mean is $m \in [0, 1]$, the receiver chooses action $A(m) \in \arg\max_{a \in \mathcal{A}} U_R(a, m)$, so the sender's interim payoff is $u(m) \coloneqq U_S(A(m), m)$. The assumption that u is upper semi-continuous can be micro-founded by assuming that (A is such that) the receiver breaks ties in the sender's favour. This simple sender–receiver model of a downstream interaction nests some but not all of our applications in §7.

Our analysis will be robust to the details of the downstream interaction, giving conditions directly on the interim payoff u that are necessary and sufficient for comparative statics. These conditions may then be checked in particular applications; we give several examples in §7 below.

[6] Explicitly, F is less informative than G exactly if for any non-empty (action) set $\mathcal{A}$ and any (payoff) $U : \mathcal{A} \times [0, 1] \to \mathbf{R}$ such that $U(a, \cdot)$ is affine for each $a \in \mathcal{A}$, we have $\int \sup_{a \in \mathcal{A}} U(a, m) F(\mathrm{d}m) \leq \int \sup_{a \in \mathcal{A}} U(a, m) G(\mathrm{d}m)$. This is because a function $\psi : [0, 1] \to \mathbf{R}$ is convex iff it equals $m \mapsto \sup_{a \in \mathcal{A}} U(a, m)$ for some such $\mathcal{A}$ and U.

[7] Equivalently, ex-post payoffs $\bar{u}_S(a, x)$ and $\bar{u}_R(a, x)$ depend on the action $a \in \mathcal{A}$ and the state $x \in [0, 1]$, and $\bar{u}_S(a, \cdot)$ and $\bar{u}_R(a, \cdot)$ are affine for each $a \in \mathcal{A}$.

3 'Non-decreasing' comparative statics

In this section, we ask a preliminary 'non-decreasing' comparative-statics question: what shifts of the sender's interim payoff u ensure that she does *not* choose a *strictly less* informative distribution? Intuition suggests that (local) convexity should be decisive, since a 'more convex' u embodies a greater liking for informative distributions. We validate this intuition, by defining a new *coarse* notion of relative convexity and proving that it is the necessary and sufficient condition for 'non-decreasing' comparative statics.

The definition is as follows:

Definition 2. For functions $u, v : [0,1] \to \mathbf{R}$, we say that u is *coarsely less convex* than v exactly if for any $x < y$ in $[0,1]$ such that

$$u(\alpha x + (1-\alpha)y) \leq \alpha u(x) + (1-\alpha)u(y) \tag{$u:\alpha$}$$

holds for every $\alpha \in (0,1)$, we also have

$$v(\alpha x + (1-\alpha)y) \leq \alpha v(x) + (1-\alpha)v(y) \tag{$v:\alpha$}$$

for every $\alpha \in (0,1)$, and furthermore any $\alpha \in (0,1)$ at which the inequality $(u:\alpha)$ is strict is also one at which $(v:\alpha)$ is strict.

We call v *coarsely more convex* than u exactly if u is coarsely less convex than v. By inspection, the relation 'coarsely less convex than' is transitive and reflexive, but not anti-symmetric.

There is a simple sufficient condition:

Lemma 1. For functions $u, v : [0,1] \to \mathbf{R}$, if $v(x) = \Phi(u(x), x)$ for every $x \in [0,1]$, where $\Phi : \mathbf{R} \times [0,1] \to \mathbf{R} \cup \{\infty\}$ is convex with $\Phi(\cdot, x)$ strictly increasing for every $x \in (0,1)$, then u is coarsely less convex than v.

Proof. Fix $x < y$ in $[0,1]$ and $\alpha \in (0,1)$ such that $u(\alpha x + (1-\alpha)y) \leq (<)$ $\alpha u(x) + (1-\alpha)u(y)$. Since $\alpha x + (1-\alpha)y \in (0,1)$, $\Phi(\cdot, \alpha x + (1-\alpha)y)$ is strictly increasing, so

$$\begin{aligned} v(\alpha x + (1-\alpha)y) &\leq (<)\ \Phi\Big(\alpha u(x) + (1-\alpha)u(y), \alpha x + (1-\alpha)y\Big) \\ &\leq \alpha v(x) + (1-\alpha)v(y), \end{aligned}$$

where the latter inequality follows from the convexity of Φ. ∎

Thus u is coarsely less convex than v whenever u is less convex than v in the conventional sense: $v = \phi \circ u$ for some convex and strictly increasing function $\phi : \mathbf{R} \to \mathbf{R} \cup \{\infty\}$ (to see this, take $\Phi(k, x) \coloneqq \phi(k)$ in Lemma 1). A different sufficient condition for u to be coarsely less convex than v, which features in the literature on costly information acquisition (see §7.5 below), is that $v = u + \psi$ for some convex $\psi : [0, 1] \to \mathbf{R}$ (take $\Phi(k, x) \coloneqq k + \psi(x)$ in Lemma 1). In case u and v are twice continuously differentiable, the former sufficient condition is equivalent to $u'' \cdot |v'| \leq v'' \cdot |u'|$, and the latter to $u'' \leq v''$. For later reference, we summarise these findings in a corollary:

Corollary 1. For $u, v : [0, 1] \to \mathbf{R}$, u is coarsely less convex than v whenever either (i) $v = \phi \circ u$ for some convex and strictly increasing $\phi : \mathbf{R} \to \mathbf{R} \cup \{\infty\}$ or (ii) $v = u + \psi$ for some convex $\psi : [0, 1] \to \mathbf{R}$.

We show in appendix K that Lemma 1 is nearly tight, by giving a partial converse, as well as an exact characterisation of coarse-convexity-increasing transformations $\Phi : \mathbf{R} \times [0, 1] \to \mathbf{R}$.

The following result characterises 'non-decreasing' comparative statics.

Theorem 1. Let $u, v : [0, 1] \to \mathbf{R}$ be upper semi-continuous. If u is coarsely less convex than v, then for any distribution F_0,

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int u \mathrm{d}F \quad \begin{array}{c}\text{is not strictly}\\ \text{higher than}\end{array} \quad \operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int v \mathrm{d}F. \qquad (\star)$$

Conversely, if $(\star)$ holds for every distribution F_0, then u must be coarsely less convex than v.

The proof is in appendix B. The second half (necessity) is straightforward. For sufficiency, we prove that if u is coarsely less convex than v, then $F \mapsto \int u \mathrm{d}F$ is *interval-dominated* by $F \mapsto \int v \mathrm{d}F$; this requires a substantial argument, the key ingredients of which are Karr's (1983) theorem on extreme points and (a general version of) Blackwell's theorem (see e.g. Phelps, 2001, p. 94). Given this, a standard comparative-statics theorem due to Quah and Strulovici (2007) yields that $(\star)$ must hold for every distribution F_0.

4 'Increasing' comparative statics

In this section, we ask what is required for a shift of the sender's interim payoff to lead her optimally to choose a *more* informative distribution. By Theorem 1, it is necessary that the payoff become coarsely more convex.

This condition is not sufficient if all upper semi-continuous interim payoffs $u, v : [0,1] \to \mathbf{R}$ and all prior distributions F_0 are considered. (We will see this explicitly §4.2 below, in a sketch proof.) Our question is thus: on what restricted domains of interim payoffs u, v and/or priors F_0 are coarse-convexity shifts *sufficient* for 'increasing' comparative statics?

Our main result (Theorem 2) describes the maximal domain of interim payoffs on which 'increasing' comparative statics holds. Concretely, it identifies the condition on the interim payoff u that is necessary and sufficient for 'increasing' comparative statics to hold under any prior F_0 between u and any coarsely more convex v. This condition is called the *crater property*. (The result features some mild regularity conditions on payoffs and priors.)

We also exhibit a suitable domain of priors. A *binary* prior is one with a two-point support; under such a prior, the state is effectively binary. We show (Proposition 1) that for 'increasing' comparative statics between payoffs u and v to hold across all *binary* priors F_0, it is both necessary and sufficient that u be coarsely less convex than v.

We then show that Theorem 2 is robust (so Proposition 1 is tight): the crater property is indispensable for 'increasing' comparative statics even when only a restricted domain of priors is considered, so long as that domain contains at least one non-binary prior (Proposition 2). In other words, binary priors are special: they are the only ones for which 'increasing' comparative statics can be obtained without the crater property. The crater property remains indispensable when only more specific shifts of u are considered, and when the sender is subject to constraints (see appendices L and M).

The crater property is demanding. A key message of this section is therefore that comparative statics are often highly prior-sensitive. On the other hand, the crater property does often hold in applications, allowing comparative-statics conclusions to be drawn; we show this in §7 below.

We next ask the mirror image of the question answered by Theorem 2: what condition on an interim payoff v is necessary and sufficient for 'decreasing' comparative statics to hold under any prior F_0 between v and any coarsely *less* convex u? The answer (Proposition 3) is that v must be trivial: either concave or strictly convex. This finding reinforces the key message that comparative statics are prior-sensitive in the persuasion model.

Finally, we consider the special case of S-shaped interim payoffs, showing how our results generalise three known comparative-statics results (Kolotilin, Mylovanov & Zapechelnyuk, 2022; Yoder, 2022; Gitmez & Molavi, 2023).

4.1 Regularity

We shall mostly restrict attention to moderately well-behaved payoffs:

Definition 3. Call a function $u : [0, 1] \to \mathbf{R}$ *regular* iff (i) u is continuous and possesses a continuous and bounded derivative $u' : (0, 1) \to \mathbf{R}$, and (ii) $[0, 1]$ may be partitioned into finitely many intervals, on each of which u is either affine, strictly convex, or strictly concave.

Part (ii) of regularity merely rules out pathological functions whose curvature changes sign infinitely often; the same condition is imposed by Dworczak and Martini (2019).

For a regular function $u : [0, 1] \to \mathbf{R}$, we extend the derivative $u' : (0, 1) \to \mathbf{R}$ to a continuous map $[0, 1] \to \mathbf{R}$ by letting $u'(0)$ be the right-hand derivative of u at 0 and $u'(1)$ the left-hand derivative at 1.

4.2 Maximal domain of interim payoffs

The following property will be the key to comparative statics.

Definition 4. A regular function $u : [0, 1] \to \mathbf{R}$ satisfies the *crater property* exactly if for any $x < y < z < w$ in $[0, 1]$ such that u is concave on $[x, y]$ and $[z, w]$ and strictly convex on $[y, z]$, the tangents to u at x and at w cross at coordinates $(X, Y) \in \mathbf{R}^2$ satisfying $y \leq X \leq z$ and $Y \leq u(X)$.

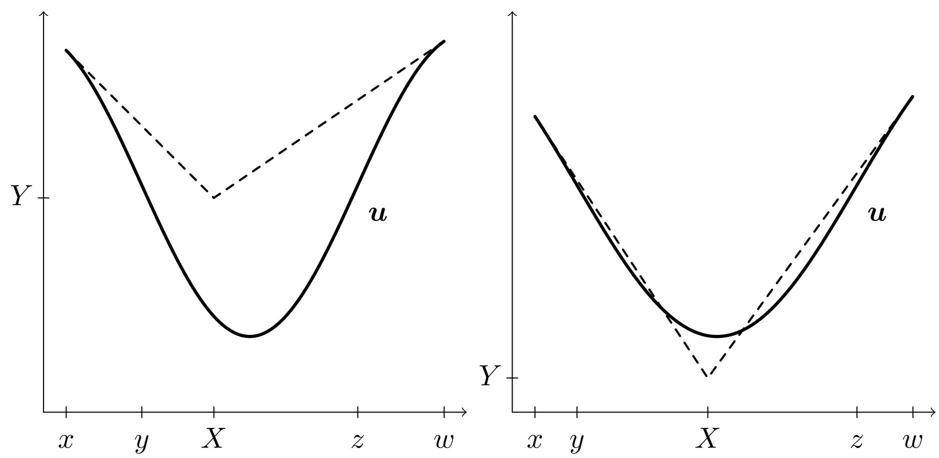
The property is illustrated in Figure 1. Loosely, it requires that any 'valley' of u be sufficiently steep-walled, wide and shallow—like a crater.

The crater property is demanding. It rules out multiple interior strict local maxima, for example. More generally, the crater property demands *affine-closedness* (as defined by Dworczak & Martini, 2019).

Nevertheless, there are important classes of interim payoffs which satisfy the crater property. Call a function *S-shaped* iff it is continuous and either strictly convex on $[0, x]$ and concave on $[x, 1]$ or concave on $[0, x]$ and strictly convex on $[x, 1]$, for some $x \in [0, 1]$. Examples include the logit and probit functions, and all unimodal CDFs. Much of the persuasion literature focusses on S-shaped interim payoffs u, as this allows for a sharp characterisation of optimal distributions.[8] All S-shaped functions satisfy the crater property.

More generally, the crater property is satisfied by all *W-shaped* functions, meaning continuous functions that are strictly convex on $[0, x]$ and on $[y, 1]$ and concave on $[x, y]$, for some $x \leq y$ in $[0, 1]$. Examples of such functions include the densities of the Beta, Normal, Laplace and Cauchy distributions.

[8] In particular, 'one-sided censorship' distributions are optimal in this case (Kolotilin, 2014, p. 14). See also Kolotilin, Mylovanov and Zapechelnyuk (2022).



(a) A violation.

(b) Not a violation.

Figure 1: Illustration of the crater property.

Theorem 2. Let $u : [0, 1] \to \mathbf{R}$ be regular. If u satisfies the crater property, then for every regular $v : [0, 1] \to \mathbf{R}$ that is coarsely more convex than u and every atomless convex-support distribution F_0,

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int u \mathrm{d}F \quad \text{is lower than} \quad \operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int v \mathrm{d}F. \qquad (\star\star)$$

Conversely, if $(\star\star)$ holds for every regular v that is coarsely more convex than u and every atomless convex-support distribution F_0, then u satisfies the crater property.

In short, the crater property is necessary and sufficient for coarse-convexity shifts to yield ‘increasing’ comparative statics. Since the crater property is demanding, this may be viewed as a negative result: comparative statics is prior-sensitive, so that conclusions often cannot be drawn robustly across all (atomless and convex-support) priors F_0. But there is a bright side: the crater property does hold in some applications, and in such cases Theorem 2 delivers comparative statics. We treat several such applications in §7 below.

We view the restriction to atomless and convex-support priors F_0 as a mild form of regularity.[9] It important to note that although the theorem focusses on prior distributions that have neither atoms nor support gaps, it

[9]A simple way of dropping this restriction is to replace it with the requirement that there be a unique distribution optimal given F_0 for u and for v. With this substitution, Theorem 2 remains true as stated. The first (sufficiency) half follows from Theorem 2 and the facts

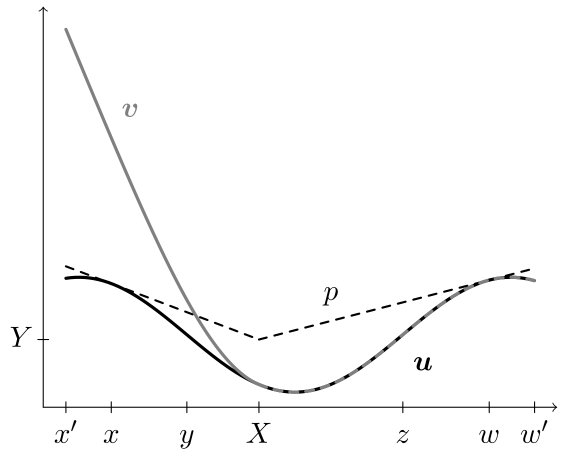


Figure 2: Sketch proof of the converse part of Theorem 2.

imposes no such restrictions on what distributions the sender may *choose:* any distribution that is less informative than the prior F_0 is permitted.

Theorem 2 is proved in appendix C. The proof of the first part (the sufficiency of the crater property) works with the dual of the persuasion problem (see Dworczak & Martini, 2019). To convey why the crater property is *necessary* for comparative statics, we now give a sketch of the proof of the converse part of Theorem 2.

Sketch proof of the converse part. Suppose that u is regular and violates the crater property; we shall find a regular and coarsely more convex $v : [0,1] \to \mathbf{R}$ and an atomless convex-support distribution F_0 such that ($\star\star$) fails.

Since u violates the crater property (refer to Figure 2), there are $x' < x < y < z < w < w'$ in $[0,1]$ such that u is concave on $[x', y]$ and $[z, w']$ and strictly convex on $[y, z]$, and (assuming that u is affine on neither $[x', y]$ nor $[z, w']$, which is the interesting case) there is a function $p : [0,1] \to \mathbf{R}$ and an $X \in (x, w)$ such that p is affine on $[x', X]$ and on $[X, w']$, weakly exceeds u on $[x', w']$, strictly exceeds u at X, and is tangent to u at x and at w. Let F_0 be a distribution that is atomless with support $[x', w']$, and

$$\frac{1}{F_0(X)} \int_0^X \xi F_0(\mathrm{d}\xi) = x \quad \text{and} \quad \frac{1}{1 - F_0(X)} \int_X^1 \xi F_0(\mathrm{d}\xi) = w.$$

Since u' is bounded, we may choose a regular $v : [0,1] \to \mathbf{R}$ that coincides

that when the space of distributions has the topology of weak convergence, it is sequentially compact (by Prokhorov's theorem, e.g. Theorem 5.1 in Billingsley (1999)), the atomless convex-support distributions form a dense subset, $F_0 \mapsto \arg\max_{F \text{ feasible given } F_0} \int u \mathrm{d}F$ is upper hemi-continuous, and the binary relation 'is less informative than' is continuous.

with u on $[X, 1]$ and that weakly exceeds u and is strictly convex on $[0, X]$ (refer to Figure 2). It is easily seen that v is coarsely more convex than u.

As v is S-shaped, an 'upper censorship' distribution G is optimal by Kolotilin's (2014, p. 14) well-known result: for $a \in (0, 1)$ satisfying

$$\frac{v(b) - v(a)}{b - a} = v'(b), \quad \text{where} \quad b := \frac{1}{1 - F_0(a)} \int_a^1 \xi F_0(\mathrm{d}\xi),$$

this distribution G fully reveals $[0, a)$ and pools $[a, 1]$.[10] A simple graphical argument shows that a must be strictly smaller than X.[11] Thus the optimal distribution G pools some states to the left of X with states to its right.

For the payoff u, however, it is strictly sub-optimal to pool states on either side of X together. In particular, the distribution F that reveals (only) whether the state exceeds X is strictly better than any distribution that pools states on either side of X together, because p is kinked at X.[12]

Thus ($\star\star$) fails: no distribution optimal for u given F_0 is less informative than G, since the latter pools across X while the former do not. ■

Remark 2. The crater property is local in character: it can be checked by separately inspecting each maximal interval $[x, w]$ on which u is concave–strictly convex–concave. This is noteworthy since it contrasts with the global character of the persuasion problem, in which a change of u on an interval $I \subseteq [0, 1]$ can impact optimal information-provision about states far from I.

4.3 The domain of binary priors

Call a distribution F *binary* iff its support comprises at most two values: $F = p\mathbf{1}_{[x,1]} + (1-p)\mathbf{1}_{[y,1]}$ for some $p, x, y \in [0, 1]$. When the prior distribution F_0 is binary, the persuasion model is equivalent to a simpler model in which

[10] Explicitly, $G = F_0$ on $[0, a)$, $G = F_0(a)$ on $[a, b)$ and $G = 1$ on $[b, 1]$. G is optimal since for any distribution H feasible given F_0, letting $q : [0, 1] \to \mathbf{R}$ match v on $[0, a]$ and match $x \mapsto v(a) + (x - a)v'(b)$ on $[a, 1]$, $\int v\mathrm{d}G = \int q\mathrm{d}G = \int q\mathrm{d}F_0 \geq \int q\mathrm{d}H \geq \int v\mathrm{d}H$, where the steps hold because, respectively, $v = q$ G-a.e., q is affine on $[a, 1]$, q is convex and H is feasible given F_0, and $q \geq v$.

[11] We have $b < w$, since $b \geq w$ would imply both $a < X$ (for tangency, as $p > u = v$ at X) and $a \geq X$ (as b is the mean conditional on the event $[a, 1]$). Then since b (w) equals the mean conditional on the event $[a, 1]$ ($[X, 1]$), we must have $a < X$.

[12] Explicitly, $F = 0$ on $[0, x)$, $F = F_0(X)$ on $[x, w)$ and $F = 1$ on $[w, 1]$. F is (strictly) better than any distribution H (with $\int_0^X H < \int_0^X F_0$) that is feasible given F_0 since $\int u\mathrm{d}F = \int p\mathrm{d}F = \int p\mathrm{d}F_0 \geq (>) \int p\mathrm{d}H \geq \int u\mathrm{d}H$, where the steps hold because, respectively, $u = p$ F-a.e., p is affine on $[0, X]$ and on $[X, 1]$, p is convex and H is feasible given F_0 (and p is affine on no open interval $I \ni X$ and $\int_0^X H < \int_0^X F_0$), and $p \geq u$.

there are just *two* states, and the sender's interim payoff at posterior belief $(q, 1-q)$ is $u(q)$, for some upper semi-continuous function $u : [0,1] \to \mathbf{R}$.

Proposition 1. Let $u, v : [0,1] \to \mathbf{R}$ be upper semi-continuous. If u is coarsely less convex than v, then for any binary distribution F_0,

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int u \mathrm{d}F \quad \text{is lower than} \quad \operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int v \mathrm{d}F. \qquad (\star\star)$$

Conversely, if $(\star\star)$ holds for every binary distribution F_0, then u must be coarsely less convex than v.

Thus restricting attention to binary priors obviates the need for the crater property, or indeed for any condition at all on u. The proof is in appendix D.

4.4 Robustness and tightness

While Theorem 2 asserts that $u : [0,1] \to \mathbf{R}$ must satisfy the crater property if coarse-convexity shifts are to lead to greater information-provision under *any* prior, Proposition 1 shows that no condition on u is required if attention is restricted to binary priors. In this section, we show that Theorem 2 is robust (so Proposition 1 is tight): binary priors are the *only* priors under which the crater property can be dispensed with.

Call a function $u : [0,1] \to \mathbf{R}$ *M-shaped* iff it is continuous and is concave on $[0, x]$ and on $[y, 1]$ and strictly convex on $[x, y]$, for some $x \leq y$ in $[0,1]$. Unlike S and W shapes, M-shaped functions can violate the crater property.

Proposition 2. For any distribution F_0 that is not binary, there are regular $u, v : [0,1] \to \mathbf{R}$ such that u is coarsely less convex than v, and yet $(\star\star)$ fails. These u and v may be chosen to be M- and S-shaped, respectively.

In other words, binary distributions F_0 are the *only* ones for which $(\star\star)$ holds between any u and any coarsely more convex v, even if attention is restricted to very well-behaved $u, v : [0,1] \to \mathbf{R}$ (in particular regular and, respectively, M- and S-shaped). 'Increasing' comparative statics can thus be guaranteed *only* by either restricting attention to interim payoffs u that satisfy the crater property (as in Theorem 2) or by restricting attention to binary prior distributions F_0 (as in Proposition 1).

The proof of Proposition 2 is in appendix E. The logic is close to that of the sketch proof of the necessity part of Theorem 2 (§4.2 above).

4.5 'Decreasing' comparative statics

The question answered by Theorem 2 has a symmetric counterpart: what is the necessary and sufficient condition on an interim payoff v for every coarse-convexity *decrease* (to some u) to yield a *decrease* of informativeness, regardless of the prior distribution F_0? The answer is as follows.

Proposition 3. Let $v : [0,1] \to \mathbf{R}$ be regular. If v is either concave or strictly convex, then for every regular $u : [0,1] \to \mathbf{R}$ that is coarsely less convex than v and every atomless convex-support distribution F_0,

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int u \mathrm{d}F \quad \text{is lower than} \quad \operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int v \mathrm{d}F. \tag{$\star\star$}$$

Conversely, if ($\star\star$) holds for every regular u that is coarsely less convex than v and every atomless convex-support distribution F_0, then v is either concave or strictly convex.

In other words, 'decreasing' comparative statics are highly prior-sensitive: a coarse-convexity decrease from v yields decreased informativeness whatever the prior F_0 only in the trivial cases of a concave v (when full pooling is optimal) or a strictly convex v (when full revelation is uniquely optimal).

The proof is in appendix F. The first half is close to obvious. For (the contra-positive of) the second half, the key observation is that if v is neither concave nor strictly convex, then it must be S-shaped on some interval, in which case we may find a regular, M-shaped and coarsely less convex $u : [0,1] \to \mathbf{R}$ as in Figure 2 (p. 11), so that ($\star\star$) fails by the logic of the sketch proof of the necessity part of 2 (§4.2 above).

4.6 Special cases

Our results generalise the comparative-statics results of Kolotilin, Mylovanov and Zapechelnyuk (2022), Yoder (2022) and Gitmez and Molavi (2023). The first paper's Proposition 1 assumes that u is S-shaped[13] and less convex than v in the conventional sense ($v = \phi \circ u$ for some convex and strictly increasing $\phi : \mathbf{R} \to \mathbf{R} \cup \{\infty\}$). Theorem 2 shows that it suffices for u to be W-shaped, and more generally that the crater property is enough. Similarly, u need only be *coarsely* less convex than v, which admits e.g. convexity of $v - u$ as an alternative sufficient condition. Similarly for the authors' Proposition 2.

Call an interim payoff $u : [0,1] \to \mathbf{R}$ *forward S-shaped* iff it is regular and is strictly convex on $[0,x]$ and strictly concave on $[x,1]$, for some $x \in [0,1]$.

[13] Their definition of 'S-shaped' is slightly weaker than ours.

Suppose that u is forward S-shaped and that u' is more convex than v' in the conventional sense: $u' = \phi \circ v'$ for some convex and strictly increasing $\phi : \mathbf{R} \to \mathbf{R} \cup \{\infty\}$. These hypotheses imply that u is coarsely less convex than v.[14] Thus by Theorem 2, less information is provided under u than under v, whatever the prior. The same is true if u is *backward* S-shaped (i.e. $x \mapsto u(1-x)$ is forward S-shaped) and u' is *less* convex than v'. These findings generalise the main result of Gitmez and Molavi (2023), which draws the same conclusion under the additional assumption that the prior is *binary*.

Finally, Yoder (2022) likewise restricts attention to binary priors, and assumes that $v - u$ is convex. This is a special case of Proposition 1.

5 Shifts of the prior distribution

Our main results concerned comparative statics with respect to shifts of the sender's interim payoff u. In this section, we consider shifts of the other primitive of the persuasion model: the distribution F_0 of the state.

Shifts of F_0 may be interpreted as changes in the information available to the sender. In particular, if the sender secures better access to information about the latent state of the world (whose distribution is fixed), this manifests precisely as increased informativeness of F_0.

Proposition 4. There are no atomless distributions $F_0 \neq G_0$ such that

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int u \mathrm{d}F \quad \text{is lower than} \quad \operatorname*{arg\,max}_{F \text{ feasible given } G_0} \int u \mathrm{d}F \qquad (\dagger)$$

holds for every regular and S-shaped $u : [0,1] \to \mathbf{R}$.

In other words, the effect on optimal information-provision of a shift of the prior distribution F_0 depends finely on the interim payoff u: there are

[14] Clearly v is also forward S-shaped, with the same inflection point $\bar{x}$. For any $x \in [0,1)$, define $R_u^x : [x,1] \to \mathbf{R}$ by $R_u^x(y) := [u(y) - u(x)]/(y-x)$ for each $y \in (x,1]$ and $R_u^x(x) := \lim_{y \downarrow x} R_u^x(y)$. For any $y \in (x,1]$, since u is forward S-shaped, R_u^x is increasing on $[x,y]$ iff R_u^x is strictly increasing on $[x,y]$ iff $u(\alpha x + (1-\alpha)y) \leq \alpha u(x) + (1-\alpha)u(y)$ for every $\alpha \in [0,1]$ iff $u(\alpha x + (1-\alpha)y) < \alpha u(x) + (1-\alpha)u(y)$ for every $\alpha \in (0,1)$. The same applies to R_v^x, analogously defined. What must be shown is therefore that for any $x < y$ in $[0,1]$, if R_u^x is increasing on $[x,y]$, then so is R_v^x. So fix any $x < y$ in $[0,1]$. Since R_u^x and R_v^x are strictly quasi-concave, it suffices to show that their respective maximisers z and w satisfy $z \leq w$. This is immediate if $w = 1$, so assume that $w < 1$. The first-order conditions are $R_u^x(z) \leq u'(z)$, with equality if $z < 1$, and $R_v^x(w) = v'(w)$. Thus since $z, w \in [\bar{x}, 1]$, $z \leq w$ holds iff $R_u^x(w) \geq u'(w)$. And indeed $R_u^x(w) = (w-x)^{-1} \int_x^w \phi \circ v' \geq \phi((w-x)^{-1} \int_x^w v') = \phi(R_v^x(w)) = \phi(v'(w)) = u'(w)$ by Jensen's inequality, since ϕ is convex. This argument is adapted from Gitmez and Molavi (2023, appendix A).

no shifts which deliver 'increasing' comparative statics robustly across all possible interim payoffs, not even if attention is restricted to the (small and well-behaved) class of regular and S-shaped interim payoffs.

The proof of Proposition 4 is in appendix G. In the same appendix, we explain how the atomlessness hypothesis may be dropped.

6 Beyond the 'single-moment' case

Our analysis has focussed on the salient case in which interim payoffs depend on only a single moment of the posterior belief—without loss, the mean. In this section, we extend our theorems to the general case. We find that whereas Theorem 1 extends directly, yielding 'non-decreasing' comparative statics, the analogue of Theorem 2 is a negative result stating that there is no hope of 'increasing' comparative statics beyond the 'single-moment' case.

6.1 The general persuasion model

In the general 'multi-moment' persuasion model (e.g. Dworczak & Kolotilin, 2022, §4), the uncertain state of the world is a random *vector* drawn from a non-empty, compact and convex set $E \subseteq \mathbf{R}^n$, where $n \in \mathbf{N}$. By *distribution*, we shall mean a CDF $\mathbf{R}^n \to [0,1]$ concentrated on E. The distribution of the state ('the prior') is denoted by F_0. For distributions F and G, we call F *less informative* than G iff $\int \psi \mathrm{d}F \leq \int \psi \mathrm{d}G$ for every convex $\psi : E \to \mathbf{R}$.

A sender chooses a signal. Given a signal, each signal realisation induces a posterior belief via Bayes's rule, whose expectation (a vector) we call the *posterior mean*. Each signal thus induces a random posterior mean, with some distribution. Call a distribution *feasible (given F_0)* iff it is the posterior-mean distribution induced by some signal. The feasible distributions are exactly those that are less informative than the prior F_0 (e.g. Phelps, 2001, p. 94).

The sender's (interim) payoff at a given realised posterior belief is assumed to depend only on its mean: her payoff at posterior mean $m \in E$ is $u(m)$, where $u : E \to \mathbf{R}$ is upper semi-continuous. Her problem is to choose among the feasible distributions F to maximise her expected payoff $\int u \mathrm{d}F$.

Remark 3. The special case $E \subseteq \mathbf{R}$ is the one studied in the rest of this paper. The persuasion model of Kamenica and Gentzkow (2011) is the special case in which E is a simplex, i.e. the convex hull of an affinely independent set $\Omega \subseteq \mathbf{R}^n$,[15] and the prior F_0 is concentrated on Ω. The interpretation is

[15] A set $S \subseteq \mathbf{R}^n$ is called *affinely independent* iff it is finite and for any $\alpha : S \to \mathbf{R}$ such that $\sum_{x \in S} \alpha(x) = \sum_{x \in S} \alpha(x)x = 0$, we have $\alpha(x) = 0$ for each $x \in S$.

that $\Omega = \text{supp}(F_0)$ is the set of states of the world, the simplex $\Delta(\Omega) = E$ is the set of all possible beliefs about the state, and the interim payoff u is an arbitrary upper semi-continuous function of the posterior belief.

6.2 Comparative statics

For any non-empty and finite set $S \subseteq E$, let $\Delta(S)$ denote the set of all maps $\alpha : S \to [0,1]$ such that $\sum_{x \in S} \alpha(x) = 1$.

Definition 5. For functions $u, v : [0,1] \to \mathbf{R}$, we say that u is *coarsely less convex* than v exactly if for any affinely independent $S \subseteq E$ such that $u(\sum_{x \in S} \alpha(x)x) \leq \sum_{x \in S} \alpha(x)u(x)$ holds for every $\alpha \in \Delta(S)$, we also have $v(\sum_{x \in S} \alpha(x)x) \leq \sum_{x \in S} \alpha(x)v(x)$ for every $\alpha \in \Delta(S)$, and furthermore any $\alpha \in \Delta(S)$ at which the former inequality is strict is also one at which the latter inequality is strict.

'Coarsely less convex than' admits the same sufficient conditions as in the 'single-moment' case: Lemma 1 and Corollary 1 (pp. 6 and 7) remain true as stated, except with '$[0,1]$' and '$(0,1)$' replaced by 'E'.

Our 'non-decreasing' comparative-statics result, Theorem 1, remains true exactly as stated, except with '$[0,1]$' replaced by 'E':

Theorem 1′. Let $u, v : E \to \mathbf{R}$ be upper semi-continuous. If u is coarsely less convex than v, then for any distribution F_0,

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int u \mathrm{d}F \quad \begin{array}{c}\text{is not strictly}\\ \text{higher than}\end{array} \quad \operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int v \mathrm{d}F. \qquad (\star)$$

Conversely, if $(\star)$ holds for every distribution F_0, then u must be coarsely less convex than v.

The exact same proof (appendix B) applies, except with '$[0,1]$' replaced by 'E' and binary distributions replaced by distributions with affinely independent support, plus a few smaller changes (e.g. replacing '$\mathbf{R}$' by '$\mathbf{R}^n$').

Recall (p. 9) our definition of regularity for functions $u : [0,1] \to \mathbf{R}$. We call a function $u : E \to \mathbf{R}$ *strongly regular* iff it is twice continuously differentiable with bounded derivatives and for all distinct $x, y \in E$, the map $[0,1] \to \mathbf{R}$ given by $\alpha \mapsto u(\alpha x + (1-\alpha)y)$ is regular and not affine. (We insist on second derivatives and no affine regions in order to to rule out uninteresting complications.)

Theorem 2′. Suppose that E is not one-dimensional,[16] and let $u : E \to \mathbf{R}$ be strongly regular. If u is either strictly concave or strictly convex, then for every strongly regular $v : E \to \mathbf{R}$ that is coarsely more convex than u and every atomless convex-support distribution F_0,

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int u \mathrm{d}F \quad \text{is lower than} \quad \operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int v \mathrm{d}F. \qquad (\star\star)$$

Conversely, if $(\star\star)$ holds for every strongly regular v that is coarsely more convex than u and every atomless convex-support distribution F_0, then u is either strictly concave or strictly convex.

In other words, comparative statics are highly prior-sensitive outside of the 'single-moment' case: a coarse-convexity increase from u yields increased informativeness whatever the prior F_0 only in the trivial cases of a strictly concave u (when full pooling is uniquely optimal) or a strictly convex u (when full revelation is uniquely optimal). The proof, given in appendix H, utilises Dworczak and Kolotilin's (2022) duality techniques.

Proposition 1 (p. 13) similarly fails beyond the 'single-moment' case: there exist priors F_0 with affinely independent support (the Kamenica and Gentzkow (2011) special case) such that $(\star\star)$ fails for some strongly regular $u, v : E \to \mathbf{R}$ with u is coarsely less convex than v.

7 Applications

In this section, we apply our theorems to various economic environments.

In most applications, the sender's interim payoff u arises from a *receiver* choosing an action at the interim stage, informed by the realisation of the signal chosen by the sender. The shape of the reduced-form interim payoff u is then determined by the nature of the conflict of interest between the sender and receiver. Motivated by this, we begin (in the next section) by identifying when a closer alignment of interests makes u coarsely more convex.

In the remainder, we apply our results to derive novel comparative statics for the problems of persuading a privately informed receiver (Kolotilin, Mylovanov, Zapechelnyuk & Li, 2017), persuading voters (à la Alonso & Câmara, 2016), designing (health) risk warnings (Mariotti, Schweizer, Szech & von Wangenheim, 2023), costly information acquisition (e.g. Ravid, Roesler & Szentes, 2022), discretionary delegation (e.g. Xu, 2024), and persuasion with choice.

[16] The *dimension* of a convex set $E \subseteq \mathbf{R}^n$ is $\max|\{S \subseteq E : S \text{ affinely independent}\}| - 1$.

7.1 Alignment and coarse convexity

In this section, we ask whether and when an increased alignment of interests between the sender and receiver translates into coarse-convexity shifts of the sender's reduced-form interim payoff u.

Recall the sender–receiver interpretation from §2.2. There is a non-empty set $\mathcal{A}$ of actions, and the sender's and receiver's interim payoffs $U_S(a, m)$ and $U_R(a, m)$ depend on the chosen action $a \in \mathcal{A}$ and on the mean $m \in [0, 1]$ of their (posterior) belief about the state. When the posterior mean is $m \in [0, 1]$, the receiver chooses action $A(m) \in \mathcal{A}$, so the sender's reduced-form interim payoff is $u(m) := U_S(A(m), m)$. We assume that $A : [0, 1] \to \mathcal{A}$ is U_R-*optimal*, i.e. a selection from the correspondence $m \mapsto \arg\max_{a \in \mathcal{A}} U_R(a, m)$.

We consider shifts of the sender's interim payoff from $(a, m) \mapsto U_S(a, m)$ to $(a, m) \mapsto \Phi(U_S(a, m), U_R(a, m), m)$, where $\Phi : \mathbf{R}^2 \times [0, 1] \to \mathbf{R}$ is strictly increasing in its first argument—that is, Φ is a *utility transformation.* We are interested in *alignment-increasing* utility transformations Φ, meaning those that are increasing in their second argument (the receiver's payoff).

Proposition 5. Let $\Phi : \mathbf{R}^2 \times [0, 1] \to \mathbf{R}$ be convex with $\Phi(\cdot, \ell, x)$ strictly increasing and $\Phi(k, \cdot, x)$ increasing for all $k, \ell \in \mathbf{R}$ and $x \in [0, 1]$. Then for any action set $\mathcal{A}$, any sender's and receiver's payoffs $U_S, U_R : \mathcal{A} \times [0, 1] \to \mathbf{R}$, and any U_R-optimal $A : [0, 1] \to \mathbf{R}$, the map $x \mapsto U_S(A(x), x)$ is coarsely less convex than the map $x \mapsto \Phi(U_S(A(x), x), U_R(A(x), x), x)$.

In words, applying a *convex* alignment-increasing transformation Φ to the sender's payoff U_S always makes her reduced-form interim payoff u coarsely more convex. Convexity is satisfied by many natural alignment-increasing transformations, such as $(k, \ell, x) \mapsto k + \rho\ell$ for $\rho \geq 0$.

The proof is in appendix I. The convexity-of-Φ hypothesis is essential, indeed nearly necessary: Proposition 5 has a partial converse similar to that of Lemma 1 (see appendix K). It is therefore not *generally* true that increased alignment of interests leads to coarse-convexity shifts.

Example 1. Consider the alignment-increasing transformation Φ defined by $\Phi(k, \ell, x) := k + \phi(\ell)$ for all $k, \ell \in \mathbf{R}$ and $x \in [0, 1]$, where $\phi : \mathbf{R} \to \mathbf{R}$ is strictly increasing. It is natural for ϕ to be concave, as this captures inequality-aversion in the sender's evaluation of (distributions of interim) receiver welfare. But when ϕ is concave and not convex, $x \mapsto U_S(A(x), x)$ fails to be coarsely less convex than $x \mapsto \Phi(U_S(A(x), x), U_R(A(x), x), x)$ for some $U_S, U_R : \mathcal{A} \times [0, 1] \to \mathbf{R}$ and some U_R-optimal $A : [0, 1] \to \mathcal{A}$.[17]

[17] In particular, for any U_R and U_R-optimal A such that the (convex) function $x \mapsto$

7.2 Persuading a privately informed receiver

In the model of Kolotilin, Mylovanov, Zapechelnyuk and Li (2017), the receiver chooses whether to participate ($a = 1$) or not ($a = 0$). Participation may mean purchasing a good (at a fixed price), for example.

The receiver's inside option (i.e. her payoff from participating) is uncertain, with distribution F_0. Her outside option is privately known to her; from the sender's perspective, it is a random variable that is statistically independent of the inside option, with a distribution denoted by G. The sender values participation: her payoff is 1 if the receiver participates, and 0 otherwise.

The sender chooses a signal. No generality is lost by ruling out screening mechanisms that offer a *menu* of signals, even though the receiver has private information (Kolotilin, Mylovanov, Zapechelnyuk & Li, 2017, Theorem 1).

At the interim stage, the receiver participates iff $r \leq m$, where r is her outside option and $m \in [0, 1]$ is the mean of her posterior belief about the inside option. The sender's interim expected payoff is thus $u(m) \coloneqq G(m)$ when the posterior mean is $m \in [0, 1]$. The function $u : [0, 1] \to \mathbf{R}$ is S-shaped if the outside-option distribution G is unimodal.

Since a monotone-likelihood-ratio-higher distribution is exactly one that is more convex, Theorem 2 implies that the sender optimally provides more information whenever the outside-option distribution shifts from a unimodal G to a monotone-likelihood-ratio-higher distribution H. This result, due to Kolotilin, Mylovanov and Zapechelnyuk (2022, §4.2), may be refined using our theorems. The shift from G to H can be more general: assuming for simplicity that G, H admit densities g, h, it suffices e.g. for $h - g$ to be increasing (by Corollary 1, p. 7) or for G to be less diffuse than H in the sense of having a more convex density (see §4.6). Furthermore, unimodality may be weakened to W-shapedness.

Applying Proposition 5 and Theorem 2 yields that given unimodality, any *convex* increase of alignment leads the sender to provide more information. An example is when the sender's interim payoff shifts from G to $m \mapsto G(m) + \phi(W(m))$, where $\phi : \mathbf{R} \to \mathbf{R}$ is increasing and convex, and $W(m) \coloneqq \int \max\{r, m\} G(\mathrm{d}r)$ denotes the receiver's interim expected payoff (not conditioned on her realised outside option). This example nests Kolotilin, Mylovanov and Zapechelnyuk's (2022) Proposition 3(i), in which ϕ is assumed affine. Increases of alignment that are *not* convex may not produce

$U_R(A(x), x)$ is less convex than and not more convex than ϕ^{-1} in the conventional sense, the map $x \mapsto U_S(A(x), x) - \Phi(U_S(A(x), x), U_R(A(x), x), x) = -\phi(U_R(A(x), x))$ is convex and not concave, so by Corollary 1 (p. 7), $x \mapsto U_S(A(x), x)$ is coarsely more convex than and not coarsely less convex than $x \mapsto \Phi(U_S(A(x), x), U_R(A(x), x), x)$.

comparative statics: if ϕ is concave and not convex, then increased alignment may lead to *strictly less* information-provision, by Example 1 and Theorem 1.

We may alternatively interpret this model as having a *population* of receivers whose outside options are heterogeneous, with cross-sectional distribution G. In this case, alignment should be defined in terms of individual receivers' payoffs $\max\{r, m\}$ rather than the average payoff $W(m)$. When alignment increases in the sense that the sender's payoff from (non-)participation changes from 1 (0) to $1 + \phi(\max\{r, m\})$ $(0 + \phi(\max\{r, m\}))$, where $\phi : \mathbf{R} \to \mathbf{R}$ is increasing, the sender's interim payoff shifts from $u = G$ to $v = G + \psi$, where $\psi(x) \coloneqq \int \phi(\max\{r, x\}) G(\mathrm{d}r)$ for each $x \in [0, 1]$. This is a coarse-convexity shift by Corollary 1 provided ϕ is 'not too concave' in the sense that $\phi''/\phi' \geq -g/G$, since then ψ is convex. Then given unimodality of G, the sender optimally provides more information by Theorem 2.

7.3 Persuading voters

Consider a generalisation of the model of the previous section featuring $n \in \mathbf{N}$ receivers, who each cast a vote ('yes' or 'no'). The receivers (collectively) participate iff at least $k \in \mathbf{N}$ of them voted 'yes', where $k \leq n$. The inside option is the same for all receivers, but outside options differ: from the sender's perspective, they are independent draws from a distribution G.

We restrict the sender to choosing a *public* signal, so that all receivers are symmetrically informed ex interim. It remains weakly dominant for each receiver to vote for participation whenever her outside option is less than the mean m of her posterior belief about the inside option. The sender's interim payoff at posterior mean $m \in [0, 1]$ is therefore $u(m) \coloneqq G^{k:n}(m)$, where $G^{k:n}$ denotes the distribution of the k^{th}-lowest of n independent draws from G.

This model is like that of Alonso and Câmara (2016), except that voters' preferences are not observed by the sender, and depend only on the mean. Sun, Schram and Sloof (2024) study a slight generalisation of this model.

If G admits a strictly log-concave and differentiable density, then the sender optimally provides more information (i) when the outside-option distribution improves in the monotone-likelihood-ratio sense, (ii) when the size n of the electorate falls, (iii) when the voting threshold k rises, and (iv) when both n and k increase by an equal amount. To see why, observe that in each of these cases, $G^{k:n}$ improves in the monotone-likelihood-ratio sense,[18] which by Corollary 1 (p. 7) implies that the sender's interim payoff u becomes coarsely more convex. Furthermore, $G^{k:n}$ admits a strictly log-

[18] By Corollary 1.C.34 and Theorem 1.C.31 in Shaked and Shanthikumar (2007).

concave density since G does; hence $G^{k:n}$ is unimodal, so u is S-shaped and thus satisfies the crater property. Theorem 2 is therefore applicable.

These results may be generalised to allow ex-ante heterogeneity, so long as the receivers $i \in \{1, \ldots, n\}$ are ordered: for all $i < j$, i's outside-option distribution G_i is worse in the monotone-likelihood-ratio sense than j's distribution G_j. The exact same argument applies.

7.4 (Health) risk warnings

Mariotti, Schweizer, Szech and von Wangenheim (2023) study welfare-maximising information-provision to present-biased consumers about the long-term risks of consuming products like tobacco, sugary drinks or alcohol. The authors describe optimal signals, but obtain no comparative-statics results about their informativeness; our theorems deliver such results.

In their model, a consumer (the receiver) chooses in each of two periods $t \in \{0, 1\}$ whether to consume ($a_t = 1$) or abstain ($a_t = 0$). If she consumes, she earns utility 1 immediately, but may bear a cost $C > 0$ two periods later.

The consumer is present-biased: if she believes consumption to be harmful with probability $m \in [0, 1]$, her period-0 and period-1 selves' payoffs are

$$\begin{aligned} &a_0 + \beta\delta a_1 - \beta\delta^2 C m a_0 - \beta\delta^3 C m a_1 \quad &&\text{and} \\ &a_0 + \delta a_1 - \beta\delta^2 C m a_0 - \beta\delta^3 C m a_1, \quad &&\text{respectively,} \end{aligned}$$

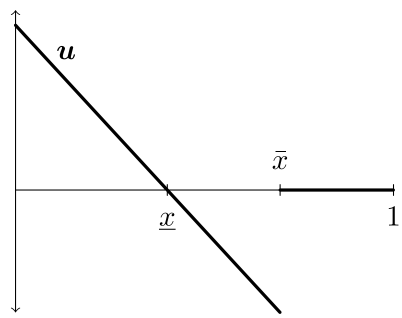
where $\beta, \delta \in (0, 1]$ are parameters. A lower value of β (of δ) captures greater present bias (impatience). Assume $\beta\delta^2 C > 1$ (abstaining is optimal if $m = 1$).

The consumer cannot commit: a_t is chosen by her period-t self, who (by inspection) consumes iff $m \leq \bar{x} := 1/\beta\delta^2 C$. Hence welfare, judged from the period-0 perspective, is

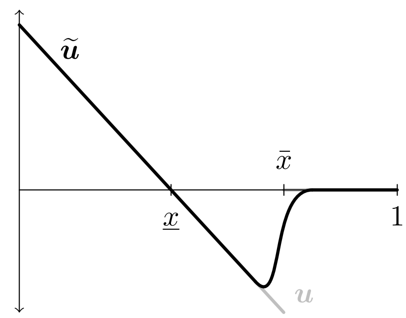
$$u(m) = \begin{cases} 1 + \beta\delta - (1 + \delta)\beta\delta^2 C m & \text{if } m < \bar{x} \\ 0 & \text{if } m \geq \bar{x}. \end{cases}$$

This is depicted in Figure 3a. Note that present bias ($\beta < 1$) engenders time-inconsistency: the period-0 self desires consumption in period 1 iff $m \leq \underline{x} := (1 + \beta\delta)\bar{x}/(1 + \delta)$, so whenever $m \in (\underline{x}, \bar{x})$, the consumer suffers ($u(m) < 0$) from her inability to commit today to abstain tomorrow.

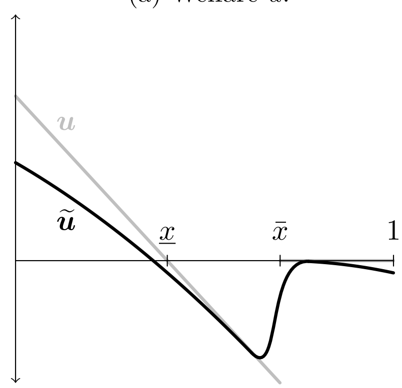
The consumer's risk (the probability with which consumption is harmful) is drawn from an atomless full-support distribution F_0. The authors study welfare-maximising information-provision about risk, e.g. via product labels.



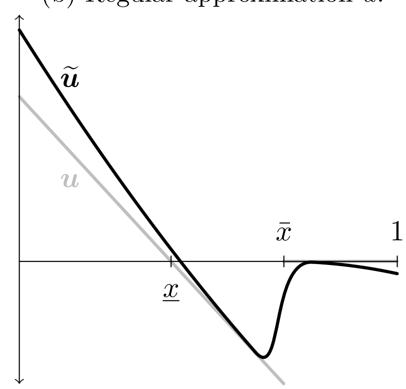
(a) Welfare u.



(b) Regular approximation $\widetilde{u}$.



(c) M-shaped regular approximation.



(d) S-shaped regular approximation.

Figure 3: Application to (health) risk warnings.

There are multiple optimal posterior-mean distributions. Welfare u may be approximated as in Figure 3b by a regular function $\widetilde{u}$ without changing the set of optimal distributions.

Other approximations select from among the set of optimal distributions. Approximating welfare u by a regular M-shaped $\widetilde{u}$, as in Figure 3c, amounts to selecting the least informative optimal distribution. Approximating by a regular S-shaped $\widetilde{u}$, as in Figure 3d, selects Kolotilin's (2014) 'upper censorship' distribution, which fully reveals $[0, a)$ and pools $[a, 1]$, where a is the least $x \in [\underline{x}, 1]$ such that $\frac{1}{1-F_0(x)} \int_x^1 y F_0(\mathrm{d}y) \geq \bar{x}$. The former kind of approximation $\widetilde{u}$ does not satisfy the crater property; the latter kind does.

Mariotti, Schweizer, Szech and von Wangenheim (2023) focus on the least informative optimal distribution, and they do not obtain comparative-statics results about its informativeness. Theorem 2 suggests why: the least informative optimum need not become more informative as parameters shift because this selection from the set of optima amounts to assuming that welfare is M-shaped as in Figure 3c, so that the crater property fails.

By contrast, the optimal upper-censorship distribution is monotone: it becomes more informative whenever any of the model's three parameters C, β, δ decrease. In other words, more information is optimally provided to consumers who are less vulnerable, more present-biased, or more impatient.

To derive this result, we apply Theorem 2. The crater property is satisfied since selecting the upper-censorship optimum amounts to approximating by an S-shaped function $\widetilde{u}$. It remains to show that any decrease of C, β or δ causes a coarse-convexity shift. This follows from two easily-verified facts: (i) that both $\underline{x}$ and $\bar{x}$ are decreasing in each of C, β and δ, and (ii) that any increase of either $\underline{x}$ or $\bar{x}$ produces a coarse-convexity shift.

7.5 Costly information acquisition ('rational inattention')

In the literature on costly information acquisition with mean-measurable costs (e.g. Ravid, Roesler & Szentes, 2022; Mensch & Ravid, 2022; Kreutzkamp, 2023; Thereze, 2023a, 2023b; Mensch & Malik, 2023), a decision-maker chooses flexibly how to learn before taking an action. Each posterior-mean distribution F has a cost $C(F)$ and a benefit $W(F)$. These are assumed to be posterior-mean-separable: $C(F) = \int c \mathrm{d}F - c(\mu_0)$ and $W(F) = \int w \mathrm{d}F$ for each feasible distribution F, where $c, w : [0,1] \to \mathbf{R}$ are convex and continuous, and $\mu_0 \coloneqq \int x F_0(\mathrm{d}x)$ denotes the prior mean. The interim benefit w is interpreted as arising from a decision problem: $w(x) = \sup_{a \in \mathcal{A}} U(a, x)$ for each $x \in [0, 1]$, where $U(a, m)$ denotes the interim payoff of action $a \in \mathcal{A}$ given posterior mean $m \in [0, 1]$. The decision-maker's flexible-learning problem is to choose among the feasible distributions F to maximise $W(F) - C(F)$. This is nested by the persuasion model, with $u \coloneqq w - c$.

Following the literature,[19] we say that *information becomes more valuable* when the interim benefit w shifts to $\widetilde{w} = w + \psi$, where $\psi : [0,1] \to \mathbf{R}$ is convex. Changes of the underlying decision problem $(\mathcal{A}, U)$ which cause information to become more valuable include raising the stakes and, sometimes, adding actions (Whitmeyer, 2024). When information becomes more valuable, the interim payoff $u = w - c$ becomes coarsely more convex by Corollary 1 (p. 7). Hence by Theorem 1, the agent optimally learns no less.[20] The same occurs when information becomes cheaper in the sense that the interim cost c shifts to $\widetilde{c} = c - \psi$, where $\psi : [0,1] \to \mathbf{R}$ is convex.

In case the prior F_0 is binary, Proposition 1 provides that when information becomes cheaper or more valuable, the decision-maker optimally learns *more*. This result directly applies to Denti's (2022, §IV) experimental test of the costly-information-acquisition model's comparative-statics predictions.

Beyond the binary-prior case, Theorem 2 suggests that results about the decision-maker optimally learning *more* will prove elusive. The mere fact that u is the difference of two convex functions implies almost nothing.[21] Rather, satisfaction by $u = w - c$ of the crater property (or sufficient conditions like W-shapedness) depends on the relative curvatures of the interim cost c and interim benefit w, requiring either strong assumptions or hard-to-interpret joint restrictions. Some examples are available: for instance, $u = w - c$ is W-shaped if $c(x) \coloneqq \kappa|x - \mu_0|$ for each $x \in [0,1]$, where $\kappa > 0$.

7.6 Discretionary delegation

Decision rights are often not set in stone, but instead granted or withdrawn as circumstances dictate. Delegating decision-making to an agent is principal-optimal only when its efficiency benefit (the agent has additional decision-relevant information, or a lower cost of action) outweighs its agency cost (the agent's preferences over actions are imperfectly aligned with the principal's), and this balance depends on the available information.

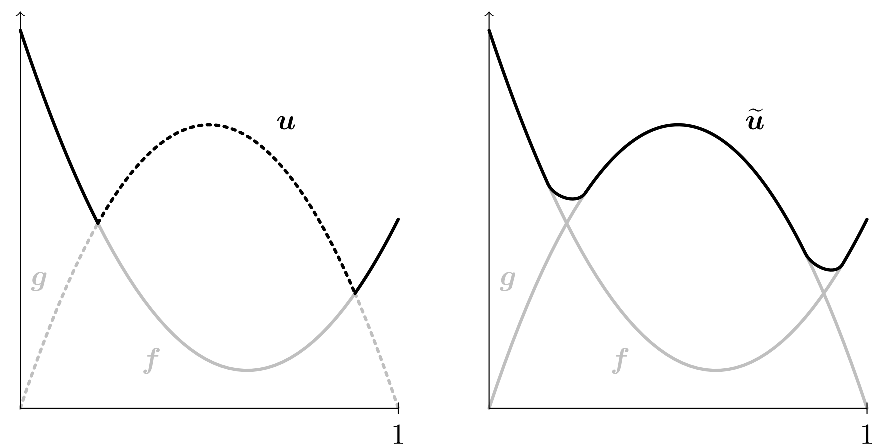
To study this trade-off, consider a simple reduced-form model.[22] After the realisation of the principal's (sender's) chosen signal is publicly observed, inducing some posterior mean $m \in [0,1]$, the principal chooses whether or

[19] E.g. Chambers, Liu and Rehbeck (2020), Denti (2022) and Whitmeyer (2024).

[20] This recovers part of Theorem 3.1 in Whitmeyer (2024).

[21] For any continuous $v : [0,1] \to \mathbf{R}$ and any $\varepsilon > 0$, there are convex $c, w : [0,1] \to \mathbf{R}$ such that $\sup_{x \in [0,1]} |v(x) - [w(x) - c(x)]| < \varepsilon$ (see e.g. Sinander, 2022, Lemma S.3).

[22] We use the notation of the sender–receiver interpretation from §2.2 (and §7.1).



(a) Interim payoff u. (b) W-shaped regular approximation $\widetilde{u}$.

Figure 4: Application to discretionary delegation.

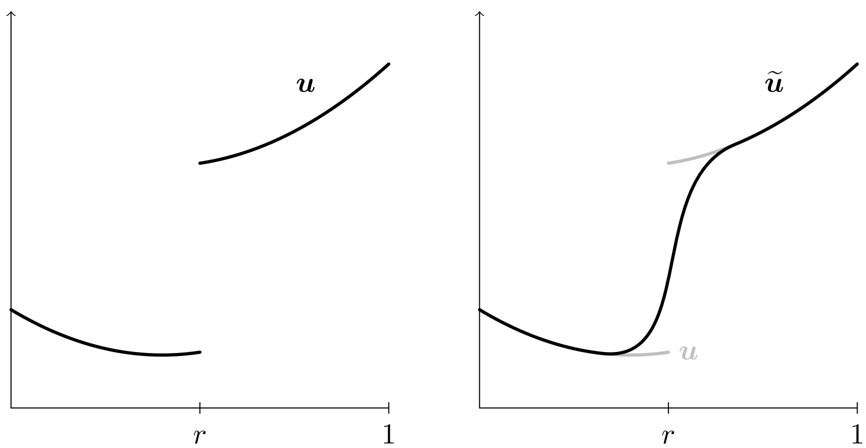
not to delegate. Her interim payoff is $f(m) \coloneqq \sup_{a \in \mathcal{A}} U_S(a, m)$ if she does not delegate and $g(m) \coloneqq B(m) + U_S(A(m), m)$ if she delegates, where A is U_R-optimal and $B \geq 0$ captures the efficiency benefit of delegating, arising e.g. from a cost saving or from information available only to the agent. We assume for simplicity that preferences are sufficiently misaligned that g is concave, and that the action set $\mathcal{A}$ is rich enough that f is strictly convex. The principal's interim payoff is $u \coloneqq \max\{f, g\}$, depicted in Figure 4a.

Xu (2024) studies the same trade-off using a different model, motivated by the problem of algorithm-assisted decision-making. One difference is that she gives an explicit micro-foundation for the efficiency benefit B of delegation; another is that she focusses on the binary-prior binary-action case.

The interim payoff u may be approximated as in Figure 4b by a regular W-shaped function $\widetilde{u}$ without affecting the set of optimal posterior-mean distributions. Hence the crater property is satisfied, so Theorem 2 is applicable.

When the efficiency benefit of delegation falls from B to $B - k$ where $k \geq 0$, the principal optimally acquires more information. To see why, observe that the sender's interim payoff after such a shift is $v = \max\{u, f + k\} - k$. The map $(a, x) \mapsto \max\{a, f + k\} - k$ does not quite satisfy the hypotheses of Lemma 1 (p. 6),[23] but it does satisfy those of its refinement Lemma 1* in appendix K. Hence u is coarsely less convex than v, so Theorem 2 applies.

[23] It is convex, and it is increasing in its first argument, but not strictly so.



(a) Interim payoff u.

(b) Regular approximation $\widetilde{u}$.

Figure 5: Application to persuasion with choice.

7.7 Persuasion with choice

Consider a hybrid of the models in §7.2 and §7.5, in which the sender's chosen signal informs both a participation decision by the receiver and an action choice by the sender herself. For simplicity, assume that receiver's outside option $r \in (0,1)$ is known to the sender, and that the sender's payoff is separable between the receiver's action and her own: $u \coloneqq \mathbf{1}_{[r,1]} + \alpha w$, where $w : [0,1] \to \mathbf{R}$ is convex and $\alpha \geq 0$. This is depicted in Figure 5a.

The interim payoff u may be approximated as in Figure 5b by a regular W-shaped function $\widetilde{u}$ without affecting the set of optimal posterior-mean distributions. Thus the crater property is satisfied, so Theorem 2 is applicable.

Regardless of the prior F_0, the sender provides more information whenever her own action becomes more important (α increases) or information becomes more valuable in the sense defined in §7.5 (a shift of w). This follows from Theorem 2 and Corollary 1 (p. 7), since both kinds of shift amount to adding a convex function to the interim payoff u.

Appendix A Product structure of 'more informative than'

In this appendix, we characterise the 'less informative than' order on distributions in terms of the product order on convex functions $[0,1] \to \mathbf{R}$. This result will be used in appendices B and C below.

Given a prior F_0, we write $\mathcal{F}$ for the space of all feasible distributions. For each $F \in \mathcal{F}$, let C_F denote the function $[0,1] \to \mathbf{R}$ given by $C_F(x) := \int_0^x F$ for each $x \in [0,1]$. Let $\mathcal{C}$ be the space of all convex functions $C : [0,1] \to \mathbf{R}$ whose right-hand derivative $C^+ : [0,1) \to \mathbf{R}$ satisfies $0 \leq C^+ \leq 1$ and which obey $C(x) \leq \int_0^x F_0$ for every $x \in [0,1]$, with equality at $x = 0$ and $x = 1$. Given any $C \in \mathcal{C}$, define $C^+(1) := 1$ by convention. The *product order* (or 'pointwise order') on $\mathcal{C}$ is the partial order in which C smaller than C' exactly if $C(x) \leq C'(x)$ for every $x \in [0,1]$.

The following extends Gentzkow and Kamenica's (2016) observation: not only do distributions F correspond one-to-one with convex functions C_F, but greater informativeness of F is equivalent to C_F being pointwise higher.

Lemma 2. Fix a prior F_0. The map $F \mapsto C_F$ is a bijection $\mathcal{F} \to \mathcal{C}$ (with inverse $C \mapsto C^+$), and is increasing when $\mathcal{F}$ is ordered by 'less informative than' and $\mathcal{C}$ has the product order. Thus $\mathcal{F}$ and $\mathcal{C}$ are order-isomorphic.

Proof. Clearly the map $F \mapsto C_F$ carries $\mathcal{F}$ into $\mathcal{C}$, and is increasing. The map $C \mapsto C^+$ similarly carries $\mathcal{C}$ into $\mathcal{F}$, and by inspection $F = C_F^+$ for every $F \in \mathcal{F}$; so we've found an inverse of $F \mapsto C_F$ defined on all of $\mathcal{C}$, meaning that $F \mapsto C_F$ is bijective. ■

Corollary 2. For any given prior F_0, the set $\mathcal{F}$ of all feasible distributions ordered by 'less informative than' is a complete lattice.

Proof. By Lemma 2, we need only show that when $\mathcal{C}$ has the product order, it holds for any family $\mathcal{C}' \subseteq \mathcal{C}$ that $C^\star := \sup_{C \in \mathcal{C}'} C$ is its least upper bound, and that the convex envelope of $\inf_{C \in \mathcal{C}'} C$, which we'll call $C_\star$, is its greatest lower bound. For the former, $C^\star$ clearly belongs to $\mathcal{C}$, is clearly an upper bound of $\mathcal{C}'$, and is clearly pointwise smaller than any other upper bound. For the latter, $C_\star$ is an element of $\mathcal{C}$, is clearly a lower bound of $\mathcal{C}'$, and exceeds every other lower bound by definition of the convex envelope. ■

Appendix B Proof of Theorem 1 (p. 7)

We shall prove the following generalisation of Theorem 1. Recall that for two distributions F and G, the *order interval* $[G, F]$ is the set of all distributions that are more informative than G and less informative than F.

Theorem 1*. For upper semi-continuous $u, v : [0,1] \to \mathbf{R}$, the following are equivalent:

(i) u is coarsely less convex than v.

(ii) For every distribution F_0, ($\star$) holds.

(iii) For all distributions G_0, F_0 such that G_0 is less informative than F_0 and $\int u\mathrm{d}F, \int v\mathrm{d}G > -\infty$ for some $F, G \in [G_0, F_0]$,

$$\operatorname*{arg\,max}_{F\in[G_0,F_0]} \int u\mathrm{d}F \quad \begin{array}{c}\text{is not strictly}\\ \text{higher than}\end{array} \quad \operatorname*{arg\,max}_{F\in[G_0,F_0]} \int v\mathrm{d}F.$$

In proving Theorem 1*, we shall write μ_F for the mean of a distribution F, and shall sometimes abbreviate 'F is less informative than G' to '$F \preceq G$'. For $x, y \in \mathbf{R}$ and $\alpha \in [0,1]$, we shall write $x_\alpha y \coloneqq \alpha x + (1-\alpha)y$.

In Theorem 1*, property (iii) implies property (ii) because a distribution is feasible given prior F_0 exactly if it belongs to $[\nu, F_0]$, where ν is the point mass concentrated on μ_{F_0}, and obviously $\int u\mathrm{d}\nu, \int v\mathrm{d}\nu > -\infty$. We shall prove that (ii) implies (i) and that (i) implies (iii).

B.1 Proof that (ii) implies (i)

Observe that given $u, v : [0,1] \to \mathbf{R}$, u is coarsely less convex than v iff for any $x < z$ in $[0,1]$ satisfying

$$u(x_\alpha z) \leq u(x)_\alpha u(z) \quad \text{for all } \alpha \in (0,1), \tag{$\triangle$}$$

it holds for each $\alpha \in (0,1)$ that

$$u(x_\alpha z) \leq (<)\ u(x)_\alpha u(z) \quad \text{implies} \quad v(x_\alpha z) \leq (<)\ v(x)_\alpha v(z). \tag{$\Rightarrow : \alpha$}$$

We prove the contra-positive. Assume that (i) fails, meaning there are $x < z$ in $[0,1]$ and an $\alpha \in (0,1)$ such that ($\triangle$) holds and ($\Rightarrow : \alpha$) fails; we seek a distribution F_0 such that

$$M_{F_0}(u) \coloneqq \operatorname*{arg\,max}_{F\in[\nu,F_0]} \int u\mathrm{d}F \quad \begin{array}{c}\text{is strictly}\\ \text{higher than}\end{array} \quad M_{F_0}(v) \coloneqq \operatorname*{arg\,max}_{F\in[\nu,F_0]} \int v\mathrm{d}F,$$

where ν denotes the point mass concentrated at μ_{F_0}. For this, it suffices that $F_0 \in M_{F_0}(u)$ and $\mu_{F_0} \in M_{F_0}(v)$ (so that $M_{F_0}(u)$ is higher than $M_{F_0}(v)$) and that either $F_0 \notin M_{F_0}(v)$ or $\mu_{F_0} \notin M_{F_0}(u)$ (so that $M_{F_0}(v)$ is not higher than $M_{F_0}(u)$). We shall use the standard 'concavification' reasoning (see Kamenica & Gentzkow, 2011). Consider two cases.

Case 1: $v(x_\alpha z) \leq v(x)_\alpha v(z)$. Let F_0 be the distribution assigning weight α to x and $1-\alpha$ to z, so that $\mu_{F_0} = x_\alpha z$. By ($\triangle$), F_0 belongs to $M_{F_0}(u)$. Since

$v(x_\alpha z) \leq v(x)_\alpha v(z)$ and $(\Rightarrow : \alpha)$ fails by hypothesis, it must be that $u(x_\alpha z) < u(x)_\alpha u(z)$ and $v(x_\alpha z) = v(x)_\alpha v(z)$, or equivalently $u(\mu_{F_0}) < \int u \mathrm{d}F_0$ and $v(\mu_{F_0}) = \int v \mathrm{d}F_0$. Then μ_{F_0} belongs to $M_{F_0}(v)$ but not to $M_{F_0}(u)$.

Case 2: $v(x_\alpha z) > v(x)_\alpha v(z)$. Let $\widehat{v}$ be the concave envelope (i.e. pointwise least majorant) of the restriction of v to $[x, z]$, and note that $\widehat{v}(x_\alpha z) \geq v(x_\alpha z) > v(x)_\alpha v(z)$ and (since v is upper semi-continuous) that $\widehat{v}(x) = v(x)$ and $\widehat{v}(z) = v(z)$. Then there is a $\beta \in (0, 1)$ such that $\widehat{v}$ is not affine on any neighbourhood of $x_\beta z$, and $\widehat{v}(x_\beta z) = v(x_\beta z)$ since v is upper semi-continuous. Let F_0 be the distribution assigning weight β to x and $1 - \beta$ to z, so that $\mu_{F_0} = x_\beta z$. Then $F_0 \notin M_{F_0}(v)$ since $\widehat{v}$ is not affine, and $\mu_{F_0} \in M_{F_0}(v)$ since $\widehat{v}(\mu_{F_0}) = v(\mu_{F_0})$. And F_0 belongs to $M_{F_0}(u)$ by $(\triangle)$. ■

B.2 Proof that (i) implies (iii), using lemmata

Definition 6. Let $u, v : [0, 1] \to \mathbf{R}$ be upper semi-continuous. Given distributions F, H such that $F \preceq H$, we say that u *is dominated by* v *on* $[F, H]$ iff

$$\int u \mathrm{d}H > -\infty, \quad \int v \mathrm{d}F > -\infty, \quad \text{and} \quad H \in \operatorname*{arg\,max}_{G \in [F,H]} \int u \mathrm{d}G$$

implies that $\int v \mathrm{d}H \geq \int v \mathrm{d}F$, with the inequality strict if $\int u \mathrm{d}H > \int u \mathrm{d}F$. We say that u *is interval-dominated by* v iff for all distributions $F \preceq H$, u is dominated by v on $[F, H]$.

Interval-dominance is a standard concept in the comparative-statics literature, due to Quah and Strulovici (2007, 2009). Our definition is slightly adapted from the standard one in order to deal with the $-\infty$ case; this adaptation ensures that standard results remain applicable.

Our proof will use some measure-theoretic concepts and lemmata. Recall that a *distribution* is a CDF $[0, 1] \to [0, 1]$. A *distribution family* is a collection $\lambda = (\lambda_x)_{x \in [0,1]}$, where λ_x is a distribution for each $x \in [0, 1]$, and $x \mapsto \int w \mathrm{d}\lambda_x$ is Borel measurable for any continuous $w : [0, 1] \to \mathbf{R}$. For any distribution family λ and any distribution F, define $F^\lambda : [0, 1] \to [0, 1]$ by

$$F^\lambda(x) := \int \lambda_y(x) F(\mathrm{d}y) \quad \text{for each } x \in [0, 1].$$

It follows from the next result that F^λ is well-defined. (Specifically, part (a) yields that $y \mapsto \int \mathbf{1}_{[0,x]} \mathrm{d}\lambda_y = \lambda_y(x)$ is Borel measurable, hence F-integrable.)

Lemma 3. Let λ be a distribution family, let F be a distribution, and let $u : [0, 1] \to \mathbf{R}$ be upper semi-continuous. Then

(a) $x \mapsto \int u \mathrm{d}\lambda_x$ is Borel measurable, and

(b) F^λ is a distribution, and $\int u \mathrm{d}F^\lambda = \int \int u \mathrm{d}\lambda_x F(\mathrm{d}x)$.

Moreover, for any distribution family ν such that $\nu_x \preceq \lambda_x$ for F-a.e. $x \in [0,1]$,

(c) $F^\nu \preceq F^\lambda$, and

(d) there exists a distribution family $(\rho_x)_{x\in[0,1]}$ such that

$$\rho_x \in \underset{G\in[\nu_x,\lambda_x]}{\arg\max} \int u \mathrm{d}G \quad \text{for } F\text{-a.e. } x \in [0,1]. \tag{1}$$

Lemma 4. Fix a distribution F and distribution families λ, ν such that $\nu_x \preceq \lambda_x$ for all $x \in [0,1]$. Let $u, v : [0,1] \to \mathbf{R}$ be upper semi-continuous, and suppose that u is dominated by v on $[\nu_x, \lambda_x]$ for all $x \in [0,1]$. Then u is dominated by v on $\left[F^\nu, F^\lambda\right]$.

We relegate the proofs of Lemmata 3 and 4 to appendix B.3 below.

Proof that (i) implies (iii). Let $u, v : [0,1] \to \mathbf{R}$ be upper semi-continuous, with u coarsely less convex than v. We shall show that u is interval-dominated by v. This suffices by Proposition 5 in Quah and Strulovici (2007).[24]

So fix any distributions $F \preceq H$; we must show that u is dominated by v on $[F,H]$. We consider three cases of increasing generality.[25] Recall that we call a distribution *binary* iff its support comprises at most two values.

Case 1: F is a point mass and H is binary. If H is a point mass, then there is nothing to prove. Assume for the remainder that $\text{supp}(H) = \{x, z\}$ where $x < z$. By the standard 'concavification' reasoning (see Kamenica & Gentzkow, 2011), $H \in \arg\max_{G\in[F,H]} \int u \mathrm{d}G$ iff ($\triangle$) holds. Moreover, choosing $\alpha \in (0,1)$ so that F is the point mass at $x_\alpha z$, $\int u \mathrm{d}H \geq (>) \int u \mathrm{d}F$ holds iff $u(x_\alpha z) \leq (<) u(x)_\alpha u(z)$, and $\int v \mathrm{d}H \geq (>) \int v \mathrm{d}F$ holds iff $v(x_\alpha z) \leq (<) v(x)_\alpha v(z)$. Since u is coarsely less convex than v, it follows that u is dominated by v on $[F,H]$.

Case 2: F is a point mass. By Lemma 4 and the previous case, it suffices to exhibit distribution families ν, λ and a distribution G such that $F = G^\nu$, $H = G^\lambda$, and for all $x \in [0,1]$, $\nu_x \preceq \lambda_x$, ν_x is a point mass, and λ_x is binary.

[24] Our definition of interval-dominance is adapted from the standard one so as to allow for the possibility that some integrals may be $-\infty$. Under our definition, Proposition 5 in Quah and Strulovici (2007) remains valid, with the same proof.

[25] We thank Ian Jewitt for suggesting this argument.

For each $x \in [0,1]$, let ν_x be the point mass at μ_H; clearly $\nu = (\nu_x)_{x\in[0,1]}$ is a distribution family. Toward constructing λ and G, let $\mathcal{F}$ be the set of all distributions with mean μ_H, and let $\mathcal{B}$ be the set of all elements of $\mathcal{F}$ that are binary. By Theorem 2.1 in Karr (1983), $\mathcal{B}$ is precisely the set of extreme points of $\mathcal{F}$. Moreover, the topology of weak convergence makes $\mathcal{F}$ compact and metrisable by Prokhorov's theorem (e.g. Billingsley, 1999, Theorems 5.1 and 6.8). Hence $\mathcal{B}$ is a Borel subset of $\mathcal{F}$ (e.g. Phelps, 2001, Proposition 1.3) and, by Choquet's theorem (e.g. Phelps, 2001, p. 14), there is a Borel probability measure π on $\mathcal{F}$ such that $\pi(\mathcal{B}) = 1$ and

$$\int w \mathrm{d}H = \int \int w \mathrm{d}L \pi(\mathrm{d}L) \quad \text{for any continuous } w : [0,1] \to \mathbf{R}.$$

Since $\mathcal{F}$ is compact and metrisable, it is a standard Borel space. Hence by the Borel isomorphism theorem (e.g. Srivastava, 1998, Theorem 3.3.13), there exists a Borel measurable bijection $\phi : [0,1] \to \mathcal{F}$ with Borel measurable inverse ϕ^{-1}. Let G be the CDF of the pushforward of π by ϕ^{-1}. Since $\pi(\mathcal{B}) = 1$, there exists a Borel measurable $\lambda : [0,1] \to \mathcal{B}$ such that $\lambda = \phi$ G-a.e.[26] Since λ is Borel measurable, it is a distribution family (e.g. Warga, 1972, Theorem IV.1.6). We have $G^\nu = F$ since F is the point mass at μ_H (as $F \preceq H$), and for all $x \in [0,1]$, $\nu_x \preceq \lambda_x$ and λ_x is binary. Finally, to show that $G^\lambda = H$, observe that π is the pushforward by ϕ of the Borel measure $A \mapsto \int_A \mathrm{d}G$, and thus π equals the pushforward by λ of $A \mapsto \int_A \mathrm{d}G$. Hence

$$\int w \mathrm{d}H = \int \int w \mathrm{d}L \pi(\mathrm{d}L) = \int \int w \mathrm{d}\lambda_x G(\mathrm{d}x) = \int w \mathrm{d}G^\lambda$$

for all continuous $w : [0,1] \to \mathbf{R}$, where the last equality follows from Lemma 3(b). It follows that $H = G^\lambda$.

Case 3: F and H are arbitrary. By Lemma 4 and the previous case, it suffices to exhibit distribution families ν, λ such that $F = F^\nu$, $H = F^\lambda$, and for all $x \in [0,1]$, $\nu_x \preceq \lambda_x$ and ν_x is a point mass. To that end, for each $x \in [0,1]$, let ν_x be the point mass at x; clearly $\nu = (\nu_x)_{x\in[0,1]}$ is a distribution family, and $F = F^\nu$. By Blackwell's theorem (e.g. Phelps, 2001, p. 94), there exists a distribution family $\lambda = (\lambda_x)_{x\in[0,1]}$ such that $\mu_{\lambda_x} = x$ for all $x \in [0,1]$ and $H = F^\lambda$; clearly $\nu_x \preceq \lambda_x$ for each $x \in [0,1]$. ■

B.3 Proofs of the lemmata

Proof of Lemma 3. For (a), recall that since u is upper semi-continuous, it is the pointwise limit of a pointwise decreasing sequence $(u_n)_{n\in\mathbf{N}}$ of

[26] For example, $\lambda \coloneqq \mathbf{1}_{\phi^{-1}(\mathcal{B})} \times \phi + \mathbf{1}_{[0,1]\setminus\phi^{-1}(\mathcal{B})} \times L$, where $L \in \mathcal{B}$.

continuous functions. By the monotone convergence theorem, $x \mapsto \int u \mathrm{d}\lambda_x$ is the pointwise limit of the (pointwise decreasing) sequence $x \mapsto \int u_n \mathrm{d}\lambda_x$ of Borel measurable functions. Hence $x \mapsto \int u \mathrm{d}\lambda_x$ is Borel measurable.

For (b), note that $w \mapsto \int \int w \mathrm{d}\lambda_x F(\mathrm{d}x)$ defines a continuous linear functional on the space of continuous functions $w : [0,1] \to \mathbf{R}$ endowed with the supremum norm, mapping positive functions to positive values and constant functions to their images. Hence by the Riesz–Markov representation theorem (e.g. Aliprantis & Border, 2006, Theorem 14.12), there exists a unique distribution G such that $\int w \mathrm{d}G = \int \int w \mathrm{d}\lambda_x F(\mathrm{d}x)$ for every continuous $w : [0,1] \to \mathbf{R}$. Moreover,

$$\begin{aligned}\int u \mathrm{d}G = \lim_{n\to\infty} \int u_n \mathrm{d}G = \lim_{n\to\infty} \int \int u_n \mathrm{d}\lambda_x F(\mathrm{d}x) \\ = \int \lim_{n\to\infty} \int u_n \mathrm{d}\lambda_x F(\mathrm{d}x) = \int \int u \mathrm{d}\lambda_x F(\mathrm{d}x),\end{aligned}$$

where the first, third and fourth equalities follow from the monotone convergence theorem. For any $x \in [0,1]$, the above argument with u replaced by $\mathbf{1}_{[0,x]}$ yields $G(x) = F^\lambda(x)$, showing that $G = F^\lambda$; hence (b) holds.

For (c), given any convex function $\phi : [0,1] \to \mathbf{R}$, we have $\int \phi \mathrm{d}\nu_x \leq \int \phi \mathrm{d}\lambda_x$ for F-a.e. $x \in [0,1]$ since $\nu_x \preceq \lambda_x$ for F-a.e. $x \in [0,1]$, so that

$$\int \phi \mathrm{d}F^\nu = \int \int \phi \mathrm{d}\nu_x F(\mathrm{d}x) \leq \int \int \phi \mathrm{d}\lambda_x F(\mathrm{d}x) = \int \phi \mathrm{d}F^\lambda$$

where the equalities follow from part (b) since ϕ is convex and thus upper semi-continuous.

For (d), let $\mathcal{G}$ be the space of all distributions endowed with the topology of weak convergence, and let $\mathcal{D}$ be the set of all pairs $(G,H) \in \mathcal{G}^2$ that satisfy $G \preceq H$, equipped with the product topology. $\mathcal{G}$ is separable and metrisable by Prokhorov's theorem (e.g. Billingsley, 1999, Theorem 6.8). Hence by the measurable maximum theorem (e.g. Aliprantis & Border, 2006, Theorem 18.19), for each $n \in \mathbf{N}$, the correspondence $\mathcal{D} \rightrightarrows \mathcal{G}$ given by

$$(G,H) \mapsto \operatorname*{arg\,max}_{L \in [G,H]} \int u_n \mathrm{d}L$$

admits a Borel measurable selection $R^n : \mathcal{D} \to \mathcal{G}$, since the correspondence $(G,H) \mapsto [G,H]$ is continuous with non-empty and compact values, and the map $L \mapsto \int u_n \mathrm{d}L$ is continuous.

A collection $(\pi_x)_{x\in[0,1]} \subseteq \mathcal{G}$ is a distribution family if and only if $x \mapsto \pi_x$ is a Borel measurable map $[0,1] \to \mathcal{G}$ (e.g. Warga, 1972, Theorem IV.1.6).

Hence $x \mapsto \lambda_x$ and $x \mapsto \nu_x$ are Borel measurable maps $[0,1] \to \mathcal{G}$. Since $\nu_x \preceq \lambda_x$ for F-a.e. $x \in [0,1]$, it follows (possibly after modifying $x \mapsto \lambda_x$ and $x \mapsto \nu_x$ on an F-null set) that $x \mapsto (\nu_x, \lambda_x)$ is a Borel measurable map $[0,1] \to \mathcal{D}$. Then for each $n \in \mathbf{N}$, $x \mapsto R^n(\nu_x, \lambda_x) =: \rho_x^n$ is a Borel measurable map $[0,1] \to \mathcal{G}$, so $(\rho_x^n)_{x \in [0,1]}$ is a distribution family.

By Theorem IV.2.1 in Warga (1972), we may assume (passing to a subsequence is necessary) that there is a distribution family $(\rho_x)_{x \in [0,1]}$ such that

$$\lim_{n \to \infty} \int \int w(x,y) \rho_x^n(\mathrm{d}y) F(\mathrm{d}x) = \int \int w(x,y) \rho_x(\mathrm{d}y) F(\mathrm{d}x) \tag{2}$$

for any $w : [0,1]^2 \to \mathbf{R}$ with $w(x,\cdot)$ continuous for each $x \in [0,1]$, $w(\cdot,y)$ Borel measurable for each $y \in [0,1]$, and $x \mapsto \max_{y \in [0,1]} |w(x,y)|$ F-integrable.

It remains to establish (1). To this end, note that $W : \mathcal{D} \to \mathbf{R}$ defined by

$$W(G,H) \coloneqq \max_{L \in [G,H]} \int u \mathrm{d}L \quad \text{for each } (G,H) \in \mathcal{D}$$

is upper semi-continuous (e.g. Aliprantis & Border, 2006, Lemma 17.30), so that the map $U : [0,1] \to \mathbf{R}$ given by

$$U(x) \coloneqq \max_{G \in [\nu_x, \lambda_x]} \int u \mathrm{d}G \quad \text{for each } x \in [0,1]$$

is Borel measurable, being the composition of the Borel measurable map $x \mapsto (\nu_x, \lambda_x)$ with W. For each $n \in \mathbf{N}$, the map $U_n : [0,1] \to \mathbf{R}$ defined by

$$U_n(x) \coloneqq \int u_n \mathrm{d}\rho_x^n = \max_{G \in [\nu_x, \lambda_x]} \int u_n \mathrm{d}G \quad \text{for each } x \in [0,1]$$

is Borel measurable since $(\rho_x^n)_{x \in [0,1]}$ is a distribution family, and satisfies $U_n \geq U$ since $u_n \geq u$. Hence for any Borel $A \subseteq [0,1]$,

$$\begin{aligned} \int_A \int u \mathrm{d}\rho_x F(\mathrm{d}x) &= \int_A \lim_{m \to \infty} \int u_m \mathrm{d}\rho_x F(\mathrm{d}x) = \lim_{m \to \infty} \int_A \int u_m \mathrm{d}\rho_x F(\mathrm{d}x) \\ &= \lim_{m \to \infty} \lim_{n \to \infty} \int_A \int u_m \mathrm{d}\rho_x^n F(\mathrm{d}x) \geq \lim_{n \to \infty} \int_A U_n \mathrm{d}F \geq \int_A U \mathrm{d}F \end{aligned}$$

where the first two equalities follow from the monotone convergence theorem, the third equality follows from (2) above since u_m is continuous for each $m \in \mathbf{N}$, the first inequality holds since $(u_m)_{m \in \mathbf{N}}$ is pointwise decreasing,[27]

[27] For any $m \leq n$, we have $u_m \geq u_n$, hence $\int u_m \mathrm{d}\rho_x^n \geq \int u_n \mathrm{d}\rho_x^n = U_n(x)$ for every $x \in [0,1]$, hence $\int_A \int u_m \mathrm{d}\rho_x^n F(\mathrm{d}x) \geq \int_A U_n \mathrm{d}F$. Now let $n \to \infty$, then $m \to \infty$.

and the final inequality holds since $U_n \geq U$ for all $n \in \mathbf{N}$. Thus $\int u \mathrm{d}\rho_x \geq U(x)$ for F-a.e. $x \in [0,1]$.

Hence to establish (1), it suffices to show that $\nu_x \preceq \rho_x \preceq \lambda_x$ for F-a.e. $x \in [0,1]$. We shall prove that $\nu_x \preceq \rho_x$ for F-a.e. $x \in [0,1]$, omitting the analogous argument for the other half. To this end, note that there exists a countable set Φ of continuous convex functions $[0,1] \to \mathbf{R}$ such that any convex function $[0,1] \to \mathbf{R}$ is the pointwise limit of a pointwise decreasing sequence of functions in Φ.[28] Moreover, it holds for any $\phi \in \Phi$ that $\int \phi \mathrm{d}\nu_x \leq \int \phi \mathrm{d}\rho_x$ for F-a.e. $x \in [0,1]$, since if this inequality were to fail for all $x \in A$ where $A \subseteq [0,1]$ is F-non-null, then

$$\int_A \int \phi \mathrm{d}\rho_x F(\mathrm{d}x) < \int_A \int \phi \mathrm{d}\nu_x F(\mathrm{d}x) \leq \int_A \int \phi \mathrm{d}\rho_x^n F(\mathrm{d}x) \quad \text{for all } n \in \mathbf{N},$$

(where the second inequality holds since $\nu_x \preceq \rho_x^n$ for all $x \in [0,1]$,) which would contradict (2) with $w(x,y) \coloneqq \mathbf{1}_A(x)\phi(y)$. Since Φ is countable, it follows that there is an F-null set $A \subseteq [0,1]$ such that $\int \phi \mathrm{d}\nu_x \leq \int \phi \mathrm{d}\rho_x$ for every $x \in [0,1] \setminus A$ and every $\phi \in \Phi$. Hence by the monotone convergence theorem, $\int \phi \mathrm{d}\nu_x \leq \int \phi \mathrm{d}\rho_x$ holds for every $x \in [0,1] \setminus A$ and every convex $\phi : [0,1] \to \mathbf{R}$. Equivalently (since A is F-null), $\nu_x \preceq \rho_x$ for F-a.e. $x \in [0,1]$. ■

Proof of Lemma 4. Suppose that

$$\int u \mathrm{d}F^\lambda > -\infty, \quad \int v \mathrm{d}F^\nu > -\infty, \quad \text{and} \quad F^\lambda \in \operatorname*{arg\,max}_{G \in [F^\nu, F^\lambda]} \int u \mathrm{d}G. \tag{3}$$

We must show that $\int v \mathrm{d}F^\lambda \geq \int v \mathrm{d}F^\nu$, and that the inequality is strict if $\int u \mathrm{d}F^\lambda > \int u \mathrm{d}F^\nu$.

By Lemma 3(d), we may choose a distribution family $(\rho_x)_{x \in [0,1]}$ such that

$$\rho_x \in \operatorname*{arg\,max}_{G \in [\nu_x, \lambda_x]} \int u \mathrm{d}G \quad \text{for } F\text{-a.e. } x \in [0,1].$$

Define

$$A \coloneqq \left\{ x \in [0,1] : \int u \mathrm{d}\lambda_x > -\infty \text{ and } \int v \mathrm{d}\nu_x > -\infty \right\}$$

$$\text{and} \quad B \coloneqq \left\{ x \in A : \lambda_x \in \operatorname*{arg\,max}_{G \in [\nu_x, \lambda_x]} \int u \mathrm{d}G \right\}.$$

[28] For example, the set of all maps of the form $x \mapsto \max_{k \in \{1,\dots,K\}} [\alpha(k)x + \beta(k)]$ where $K \in \mathbf{N}$ and $\alpha, \beta : \{1,\dots,K\} \to \mathbf{Q}$. (Here $\mathbf{Q} \subseteq \mathbf{R}$ denotes the rational numbers.)

We have $\int_A \mathrm{d}F = 1$ by (3) and Lemma 3(b). We further claim that $\int_B \mathrm{d}F = 1$. Suppose toward a contradiction that $\int_B \mathrm{d}F < 1$; then

$$\int u \mathrm{d}F^\lambda = \int \int u \mathrm{d}\lambda_x F(\mathrm{d}x) < \int \int u \mathrm{d}\rho_x F(\mathrm{d}x) = \int u \mathrm{d}F^\rho,$$

where the equalities follow from Lemma 3(b), and the inequality is strict since $\int u \mathrm{d}F^\lambda > -\infty$. But F^ρ belongs to $\left[F^\nu, F^\lambda\right]$ by Lemma 3(c) since $\rho_x \in [\nu_x, \lambda_x]$ for F-a.e. $x \in [0,1]$, so $\int u \mathrm{d}F^\lambda \geq \int u \mathrm{d}F^\rho$ by (3); a contradiction.

Since u is dominated by v on $[\nu_x, \lambda_x]$ for all $x \in [0,1]$, we have $\int v \mathrm{d}\lambda_x \geq \int v \mathrm{d}\nu_x$ for every $x \in B$. This together with $\int_B \mathrm{d}F = 1$ implies that

$$\int v \mathrm{d}F^\lambda = \int \int v \mathrm{d}\lambda_x F(\mathrm{d}x) \geq \int \int v \mathrm{d}\nu_x F(\mathrm{d}x) = \int v \mathrm{d}F^\nu,$$

where the equalities follow from Lemma 3(b).

It remains only to show that if $\int u \mathrm{d}F^\lambda > \int u \mathrm{d}F^\nu$, then $\int v \mathrm{d}F^\lambda > \int v \mathrm{d}F^\nu$. So assume that $\int u \mathrm{d}F^\lambda > \int u \mathrm{d}F^\nu$, and let

$$C^w \coloneqq \left\{ x \in B : \int w \mathrm{d}\lambda_x > \int w \mathrm{d}\nu_x \right\} \quad \text{for } w \in \{u, v\}.$$

Note that C^u is F-non-null, since otherwise

$$\int u \mathrm{d}F^\lambda = \int \int u \mathrm{d}\lambda_x F(\mathrm{d}x) \leq \int \int u \mathrm{d}\nu_x F(\mathrm{d}x) = \int u \mathrm{d}F^\nu,$$

where the inequality holds since $\int_B \mathrm{d}F = 1$, and the equalities follow from Lemma 3(b). We have $C^u \subseteq C^v$ since u is dominated by v on $[\nu_x, \lambda_x]$ for all $x \in [0,1]$. Thus C^v is F-non-null, so

$$\int v \mathrm{d}F^\lambda = \int \int v \mathrm{d}\lambda_x F(\mathrm{d}x) > \int \int v \mathrm{d}\nu_x F(\mathrm{d}x) = \int v \mathrm{d}F^\nu,$$

where the inequality holds since $\int v \mathrm{d}\lambda_x \geq \int v \mathrm{d}\nu_x$ for F-a.e. $x \in [0,1]$ and $\int v \mathrm{d}F^\nu > -\infty$, and the equalities follow from Lemma 3(b). ■

Appendix C Proof of Theorem 2 (p. 10)

We prove the necessity of the crater property for comparative statics in §C.1, following the logic of the sketch proof in the text (§4.2), and then prove sufficiency in §C.2. The latter proof relies on a lemma whose proof is relegated to a separate section, §C.3.

C.1 Proof of necessity

We shall rely on the following lemma.

Lemma 5. Suppose there are $x' < x < X < w < w'$ in $[0,1]$ and a function $p : [0,1] \to \mathbf{R}$ that is affine on $[x', X]$ and on $[X, w']$, with $p \geq u$ on $[x', w']$ with equality on $\{x, w\}$, and $p'(x) < p'(w)$. Let F_0 be any atomless distribution with support $[x', w']$ such that

$$\frac{1}{F_0(X)} \int_0^X \xi F_0(\mathrm{d}\xi) = x \quad \text{and} \quad \frac{1}{1 - F_0(X)} \int_X^1 \xi F_0(\mathrm{d}\xi) = w.$$

Then $\int_0^X F = \int_0^X F_0$ for any distribution F optimal for u given prior F_0.

See Figure 2 (p. 11) for a graphical illustration of the hypotheses. The lemma asserts that under these hypotheses, it is strictly sub-optimal to pool states on either side of the kink point X: optimal distributions F may pool within $[0, X]$ and within $[X, 1]$, but not across.

Proof. Fix a distribution F that is feasible given F_0 and that satisfies $\int_0^X F < \int_0^X F_0$. Let G be the distribution with support $\{x, w\}$ and mean μ_{F_0}, and note that G is less informative than F_0 and that $\int_0^X G = \int_0^X F_0$. We will show that G is strictly better than F for u given F_0. Note that $F(x') = 0$ and $F(w') = 1$ since F_0 is atomless with support $[x', w']$.[29] We have

$$\begin{aligned}
\int u \mathrm{d}F \leq \int p \mathrm{d}F &= p(w') - \int_0^{w'} p' F \\
&= p(w') - p'(w)\left(\int_0^{w'} F\right) + [p'(w) - p'(x)] \int_0^X F \\
&< p(w') - p'(w)\left(\int_0^{w'} G\right) + [p'(w) - p'(x)] \int_0^X G \\
&= \int p \mathrm{d}G = \int u \mathrm{d}G,
\end{aligned}$$

where the weak inequality holds since $u \leq p$ on $[x', w']$, the first equality is obtained by integrating by parts,[30] the second equality holds since p is affine on $[x', X]$ and on $[X, w']$, the strict inequality holds since $\int_0^{w'} F =$

[29] For the former, using the notation from appendix A, we have $C_F \leq C_{F_0}$ and $C_{F_0}(x') = 0 \leq C_F(x')$, whence $[C_F(x) - C_F(x')]/(x - x') \leq [C_{F_0}(x) - C_{F_0}(x')]/(x - x')$ for every $x \in (x', w']$, so that letting $x \downarrow x'$ yields $F(x') \leq F_0(x') = 0$. The latter is analogous.

[30] Invoking e.g. Theorem 18.4 in Billingsley (1995).

$w' - \mu_{F_0} = \int_0^{w'} G$, $\int_0^X F < \int_0^X F_0 = \int_0^X G$ and $p'(x) < p'(w)$, the penultimate equality holds for the same reasons as the first two equalities (recalling that $G(w') = 1$), and the final equality holds because $p = u$ G-a.e. ■

Proof of the converse (necessity) part of Theorem 2. Let $u : [0,1] \to \mathbf{R}$ be regular, and suppose that it does not satisfy the crater property. That means that there are $x' < y < z < w'$ in $[0,1]$ such that u is concave on $[x', y]$ and $[z, w']$ and strictly convex on $[y, z]$, we have $u'(x') \neq u'(w')$,[31] and the tangents to u at x' and at w' cross at coordinates $(X', Y') \in \mathbf{R}^2$ that either violate $y \leq X' \leq z$ or satisfy $Y' > u(X')$. It cannot be that u is affine on both $[x', y]$ and $[z, w']$, since that would imply $y \leq X' \leq z$ and $Y' < u(X')$. Assume that u is not affine on $[x', y]$; the other case is analogous.

We seek a regular $v : [0,1] \to \mathbf{R}$ that is coarsely more convex than u and an atomless convex-support distribution F_0 such that ($\star\star$) fails. We shall first construct a distribution F_0 and $x < X < w$ in (x', w') such that the hypotheses of Lemma 5 are satisfied, and then construct a regular $v : [0,1] \to \mathbf{R}$ that is coarsely more convex than u and a distribution F that is optimal for v given prior F_0 and has $\int_0^X F < \int_0^X F_0$. Then by Lemma 5, every distribution G that is optimal for u given F_0 satisfies $\int_0^X F < \int_0^X F_0 = \int_0^X G$, so fails to be less informative than F—thus ($\star\star$) fails.

We consider separately the cases in which u is not and is affine on $[z, w']$. (The sketch proof in the text corresponds to the first case.)

Case 1: u is not affine on $[z, w']$. In this case, there are $x \in (x', y)$ and $w \in (z, w')$ such that $u'(x) < u'(w)$ and the tangents to u at x and at w intersect at coordinates $(X, Y) \in \mathbf{R}^2$ satisfying $y \leq X \leq z$ and $Y > u(X)$.[32] Let $p : [0,1] \to \mathbf{R}$ be the pointwise maximum of the two tangents (refer to Figure 2 on p. 11); it is affine on $[x', X]$ and on $[X, w']$, satisfies $p \geq u$ on $[x', w']$ with equality on $\{x, w\}$, and $p'(x) < p'(w)$. Let F_0 be a distribution that is atomless with support $[x', w']$,

$$\frac{1}{F_0(X)} \int_0^X \xi F_0(\mathrm{d}\xi) = x \quad \text{and} \quad \frac{1}{1 - F_0(X)} \int_X^1 \xi F_0(\mathrm{d}\xi) = w.$$

Observe that the hypotheses of Lemma 5 are satisfied.

Since u' is bounded, we may choose a regular $v : [0,1] \to \mathbf{R}$ that coincides with u on $[X, 1]$ and that weakly exceeds u and is strictly convex on $[0, X]$

[31] This is without loss since if $u'(x') = u'(w')$ then we may choose x' and w' differently (in particular, closer together) such that $u'(x') \neq u'(w')$.

[32] If the tangent to u at x' (at w') strictly exceeds u at z (at y), choose x (w) such that the tangent to u at x (at w) crosses u at z (at y); if not, choose $x := x' + \varepsilon$ ($w := w' - \varepsilon$) for a sufficiently small $\varepsilon \in (0, (w' - x')/2)$.

(refer to Figure 2 on p. 11). It is easily seen that v is coarsely more convex than u. By Lemma 5, it suffices to exhibit a distribution F that is optimal for v given prior F_0 and that satisfies $\int_0^X F < \int_0^X F_0$.

It is easily verified (see footnote 11 on p. 12) that there are $a \in [0, X)$ and $b \in (z, w)$ which satisfy

$$\frac{v(b) - v(a)}{b - a} = v'(b) \quad \text{and} \quad b \coloneqq \frac{1}{1 - F_0(a)} \int_a^1 \xi F_0(\mathrm{d}\xi).$$

Define a distribution F by $F \coloneqq F_0$ on $[0, a)$, $F \coloneqq F_0(a)$ on $[a, b)$ and $F \coloneqq 1$ on $[b, 1]$. Clearly F is feasible given F_0, and $\int_0^X F < \int_0^X F_0$ since $a < X \leq z < b$.

It remains to prove that F is optimal for v given F_0. To this end, let $q : [0, 1] \to \mathbf{R}$ match v on $[0, a] \cup \{b\}$ and be affine on $[a, 1]$. By inspection, q exceeds v on $[x', w']$, is convex on $[x', w']$, has $\int q\mathrm{d}F_0 = \int q\mathrm{d}F$, and satisfies $q = v$ F-a.e. Using each of these facts in turn,[33] we obtain for any distribution G that is feasible given F_0 (i.e. any G less informative than F_0) that

$$\int v\mathrm{d}G \leq \int q\mathrm{d}G \leq \int q\mathrm{d}F_0 = \int q\mathrm{d}F = \int v\mathrm{d}F.$$

Case 2: u is affine on $[z, w']$. In this case, there is an $x \in (x', y)$ such that the tangent to u at x crosses u at some $X \in (z, w')$. Fix any $w \in (X, w')$, and let $p : [0, 1] \to \mathbf{R}$ match the aforementioned tangent on $[x', X]$ and match u on $[X, w']$; it is affine on $[x', X]$ and on $[X, w']$, satisfies $p \geq u$ on $[x', w']$ with equality on $\{x, w\}$, and $p'(x) < p'(w)$. Let F_0 be a distribution that is atomless with support $[x', w']$,

$$\frac{1}{F_0(X)} \int_0^X \xi F_0(\mathrm{d}\xi) = x \quad \text{and} \quad \frac{1}{1 - F_0(X)} \int_X^1 \xi F_0(\mathrm{d}\xi) = w.$$

The hypotheses of Lemma 5 are satisfied.

Let $v : [0, 1] \to \mathbf{R}$ be regular, affine on $[z, w']$, and strictly convex on $[0, z]$ and $[w', 1]$. Clearly v is coarsely more convex than u. Thus by Lemma 5, it suffices to exhibit a distribution F that is optimal for v given prior F_0 and that satisfies $\int_0^X F < \int_0^X F_0$ (i.e. states on either side of X are pooled).

To that end, let

$$b \coloneqq \frac{1}{1 - F(z)} \int_z^1 \xi F_0(\mathrm{d}\xi),$$

[33] We could instead appeal to the duality theorem of Dworczak and Martini (2019).

and define a distribution F by $F := F_0$ on $[0, z)$, $F := F_0(z)$ on $[z, b)$ and $F := 1$ on $[b, 1]$. Then $\int_0^X F < \int_0^X F_0$,[34] and F is optimal for v given F_0. ■

C.2 Proof of sufficiency

Given any distribution F, let $C_F : [0,1] \to \mathbf{R}$ be given by $C_F(x) := \int_0^x F$ for each $x \in [0,1]$. We shall make free use of the order isomorphism described in appendix A between distributions F ordered by informativeness and convex functions C_F ordered by pointwise inequality.

The sufficiency proof relies on three lemmata. The first is a version of Dworczak and Martini's (2019) duality theorem. Given any regular $u : [0,1] \to \mathbf{R}$, let $\mathcal{M}(u)$ denote the space of all convex and Lipschitz continuous functions $p : [0,1] \to \mathbf{R}$ satisfying $p \geq u$.

Lemma 6. Let $u : [0,1] \to \mathbf{R}$ be regular, and let F_0 be an atomless distribution. Then

$$\min_{p \in \mathcal{M}(u)} \int p \mathrm{d}F_0 = \max_{F \text{ feasible given } F_0} \int u \mathrm{d}F,$$

where both sides are well-defined. Moreover, for $p \in \mathcal{M}(u)$ and a distribution F feasible given F_0 to solve (respectively) the minimisation and maximisation problems, it is necessary and sufficient that both

(a) p is affine on any interval on which $C_F < C_{F_0}$, and

(b) $p = u$ on $\operatorname{supp}(F)$.

Proof of Lemma 6. Fix a distribution F_0. The result is trivial if F_0 is degenerate, so suppose not. Since u is regular, for any convex and continuous $q : [0,1] \to \mathbf{R}$ such that $q \geq u$, there is a $p \in \mathcal{M}(u)$ such that $p \leq q$. Thus the first part follows from Theorem 1(ii) in Dizdar and Kováč (2020) applied to the restriction of u to $\operatorname{supp}(F_0)$, since u is regular.

For the second part, fix any $p \in \mathcal{M}(u)$ and any distribution F that is feasible given F_0. Since F_0 is atomless, we have $F_0(0) = 0$ and thus $F(0) = 0$.[35] Because p is convex and Lipschitz, we may extend its derivative $p' : (0,1) \to \mathbf{R}$ continuously to $[0,1]$ by letting $p'(0)$ and $p'(1)$ be the right-

[34] If $b \geq X$, then $\int_0^X (F_0 - F) = \int_z^X [F_0 - F_0(z)] > 0$ as $F_0 > F_0(z)$ on $(z, 1]$, while if $b \leq X$ we have $\int_0^X (F_0 - F) = \int_X^1 (F - F_0) = \int_X^1 (1 - F_0) > 0$ as $F_0 < 1$ on $[0, w')$.

[35] We have $C_F \leq C_{F_0}$ and $C_{F_0}(0) = 0 \leq C_F(0)$, whence $[C_F(x) - C_F(0)]/x \leq [C_{F_0}(x) - C_{F_0}(0)]/x$ for every $x \in (0,1]$, so that letting $x \downarrow 0$ yields $F(0) \leq F_0(0) = 0$.

and left-hand derivatives at 0 and at 1, respectively. Then for any distribution G with $G(0) = 0$, integrating by parts twice,[36]

$$\int p \mathrm{d}G = p(1) - \int p' G = p(1) - p'(1) C_G(1) + \int C_G \mathrm{d}p',$$

where the last term is to be understood in the Lebesgue–Stieltjes sense. Thus

$$\int p \mathrm{d}F_0 \geq \int p \mathrm{d}F \geq \int u \mathrm{d}F,$$

where the first inequality is strict unless (a) holds, while the second is strict unless (b) holds since p and u are continuous. ■

Lemma 7. Let $u : [0, 1] \to \mathbf{R}$ be regular and satisfy the crater property, and suppose there are $x < z$ in $[0, 1]$ such that the tangent to u at x (at z) weakly exceeds u on $[x, z]$. Then there is a $y \in (x, z]$ (a $y \in [x, z)$) such that u is concave on $[x, y]$ (on $[y, z]$) and strictly convex on $[y, z]$ (on $[x, y]$).

Proof of Lemma 7. Suppose that the tangent to u at x weakly exceeds u on $[x, z]$; the other case is analogous. Let y be the largest $y' \in [x, z]$ such that u is concave on $[x, y']$. We have $y > x$ since u is regular. It remains to show that u is strictly convex on $[y, z]$. This is immediate if $y = z$, so suppose for the remainder that $y < z$.

Let $\widehat{z}$ be the largest $w \in [y, 1]$ such that u is strictly convex on $[y, w]$; clearly $\widehat{z} > y$ by the regularity of u. We must show that $\widehat{z} \geq z$, so suppose toward a contradiction that $\widehat{z} < z$. Then by regularity, u is concave on $[\widehat{z}, w]$ for some $w \in (\widehat{z}, z]$. But then u violates the crater property, since the tangent to u at x strictly exceeds u on $[y, \widehat{z}]$ (as u is strictly convex on $[y, \widehat{z}]$). ■

Lemma 8. Let $u, v : [0, 1] \to \mathbf{R}$ be regular, and suppose that u satisfies the crater property and is coarsely less convex than v. Let F_0 be an atomless convex-support distribution. Then for any

$$p \in \operatorname*{arg\,min}_{r \in \mathcal{M}(u)} \int r \mathrm{d}F_0 \quad \text{and} \quad q \in \operatorname*{arg\,min}_{r \in \mathcal{M}(v)} \int r \mathrm{d}F_0,$$

if q is affine on an interval $[x, y] \subseteq \operatorname{supp}(F_0)$, then so is p.

Lemma 8 is proved in the next section.

[36]This is licensed by e.g. Theorem 18.4 in Billingsley (1995).

Proof of the first (sufficiency) part of Theorem 2. Fix regular $u, v : [0,1] \to \mathbf{R}$ such that u satisfies the crater property and is coarsely less convex than v, let F_0 be an atomless convex-support distribution, and fix

$$G' \in \operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int u \mathrm{d}F \quad \text{and} \quad H' \in \operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int v \mathrm{d}F.$$

We shall construct

$$G'' \in \operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int u \mathrm{d}F \quad \text{and} \quad H'' \in \operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int v \mathrm{d}F$$

such that G'' is less informative than H' and G' is less informative than H''.

We derive G'' from G' by fully pooling signal realisations over each concavity interval of u, in the following precise sense. Assume without loss of generality that u is not strictly convex, and enumerate the maximal proper intervals on which u is concave as $([x_k, z_k])_{k=1}^K$ (where $K \in \mathbf{N}$). For each k, let y_k denote the mean of G' conditional on the event $[x_k, z_k]$. (In case $[x_k, z_k]$ is G'-null, let y_k be an arbitrary element of $[x_k, z_k]$.) Define a distribution G'' by

$$G''(w) \coloneqq \begin{cases} G'(x_k-) & \text{if } w \in [x_k, y_k) \text{ for some } k \in \{1, \dots, K\} \\ G'(z_k) & \text{if } w \in [y_k, z_k] \text{ for some } k \in \{1, \dots, K\} \\ G'(w) & \text{otherwise,} \end{cases}$$

where '$G(x-)$' is shorthand for $\lim_{y \uparrow x} G'(y)$. For any G'-non-null $[x_k, z_k]$, the distribution 'G'' conditional on $[x_k, z_k]$' is less informative than the distribution 'G' conditional on $[x_k, z_k]$',[37] so $\int_{[x_k,z_k]} u \mathrm{d}G'' \geq \int_{[x_k,z_k]} u \mathrm{d}G'$. And we have $G'' = G'$ on $\mathcal{X} \coloneqq [0,1] \setminus \bigcup_{k=1}^K [x_k, z_k]$, so that $\int_{\mathcal{X}} u \mathrm{d}G'' = \int_{\mathcal{X}} u \mathrm{d}G'$ since $\mathcal{X}$ is open. Thus $\int u \mathrm{d}G'' \geq \int u \mathrm{d}G'$, which since G' optimal for u given prior F_0 implies that G'' is, too.

We similarly derive H'' from H' by spreading signal realisations over each convexity interval of v as much as possible subject keeping H'' less informative than the prior F_0. Formally, assume without loss of generality that v is not strictly concave, enumerate the maximal proper intervals on which v is convex as $(I_\ell)_{\ell=1}^L$ (where $L \in \mathbf{N}$), and define $I \coloneqq \bigcup_{\ell=1}^L I_\ell$. Let C be the convex envelope of $\mathbf{1}_I C_{F_0} + \mathbf{1}_{[0,1] \setminus I} C_{H'}$, and let the distribution H'' be be defined by $C_{H''} = C$. We have $H'' = H'$ off I, and clearly 'H'' conditional on I_ℓ' is more informative than 'H' conditional on I_ℓ' for each H'-non-null

[37] Explicitly: the distribution $\mathbf{1}_{(z_k,1]} + \mathbf{1}_{[x_k,z_k]} \times [G'' - G''(x_k-)]/[G''(z_k) - G''(x_k-)]$ is less informative than the distribution $\mathbf{1}_{(z_k,1]} + \mathbf{1}_{[x_k,z_k]} \times [G' - G'(x_k-)]/[G'(z_k) - G'(x_k-)]$.

I_ℓ, so $\int v \mathrm{d}(H'' - H') = \sum_{\ell=1}^{L} \int_{I_\ell} v \mathrm{d}(H'' - H') \geq 0$, which since H' is optimal for v given prior F_0 implies that H'' is, too.

It remains to prove that G'' is less informative than H' and that G' is less informative than H''. We shall rely on the following claim, whose proof (relegated to the end) hinges on Lemmata 7 and 8.

Claim. Let G and H be optimal (given prior F_0) for u and v, respectively. Then for any $a < b$ in $[0,1]$ such that $C_H < C_{F_0}$ on (a,b) and $C_H = C_{F_0}$ on $\{a,b\}$, there are $c \leq d$ in $\mathrm{supp}(G)$ such that $C_G \leq C_H$ on $[a,b] \setminus (c,d)$ and u is affine on $[c,d]$.

To prove that G'' is less informative than H', it suffices to show that for any $a < b$ in $[0,1]$ such that $C_{H'} < C_{F_0}$ on (a,b) and $C_{H'} = C_{F_0}$ on $\{a,b\}$, we have $C_{G''} \leq C_{H'}$ on (a,b). So fix such a pair $a < b$. By the claim, there are $c \leq d$ in $\mathrm{supp}(G'')$ such that $C_{G''} \leq C_{H'}$ on $[a,b] \setminus (c,d)$ and u is affine on $[c,d]$. And (c,d) is empty, since $\mathrm{supp}(G'') \cap [c,d]$ must be a singleton by definition of G'' and the fact that u is concave on $[c,d]$.

Similarly, to prove that G' is less informative than H'', it suffices to show that for any $a < b$ in $[0,1]$ such that $C_{H''} < C_{F_0}$ on (a,b) and $C_{H''} = C_{F_0}$ on $\{a,b\}$, we have $C_{G'} \leq C_{H''}$ on (a,b). So fix such a pair $a < b$. By the claim, there are $c \leq d$ in $\mathrm{supp}(G')$ such that $C_{G'} \leq C_{H''}$ on $[a,b] \setminus (c,d)$ and u is affine on $[c,d]$. If $[a,b]$ and $[c,d]$ are disjoint, then we are done. Suppose for the remainder that $[a,b] \cap [c,d]$ is non-empty. We must show that $C_{G'} \leq C_{H''}$ on $[a',b'] \coloneqq [a,b] \cap [c,d]$.

v is convex on $[a',b']$ since $[a',b'] \subseteq [c,d]$, so by definition of H'', the restriction of $C_{H''}$ to $[a',b']$ equals the convex envelope of $\mathbf{1}_{(a',b')} C_{F_0} + \mathbf{1}_{\{a',b'\}} C_{H''}$. We have $C_{G'} \leq \mathbf{1}_{(a',b')} C_{F_0} + \mathbf{1}_{\{a',b'\}} C_{H''}$ on $[a',b']$ by hypothesis and the fact that G' is less informative than the prior F_0.[38] Thus since $C_{G'}$ is convex, it must satisfy $C_{G'} \leq C_{H''}$ on $[a',b']$.

Proof of the claim. Fix $a < b$ in $[0,1]$ such that $C_H < C_{F_0}$ on (a,b) and $C_H = C_{F_0}$ on $\{a,b\}$. Note that $[a,b] \subseteq \mathrm{supp}(F_0)$ since the latter is convex. Since u and v are regular, Lemma 6 provides that there exist

$$p \in \operatorname*{arg\,min}_{r \in \mathcal{M}(u)} \int r \mathrm{d}F_0 \quad \text{and} \quad q \in \operatorname*{arg\,min}_{r \in \mathcal{M}(v)} \int r \mathrm{d}F_0,$$

and that q is affine on $[a,b]$. By Lemma 8, it follows that p is also affine on $[a,b]$. Write $[a',b']$ for the maximal interval I such that p is affine on I and

[38] At a', we have if $a' = c$ that $C_{G'}(a') \leq C_{H''}(a')$, and if not then $a' = a$, in which case $C_{G'}(a') \leq C_{F_0}(a') = C_{H''}(a')$ since G' is less informative than F_0. Similarly at b'.

$[a,b] \subseteq I \subseteq \operatorname{supp}(F_0)$. We have $C_G = C_{F_0}$ on $\{a', b'\}$ by Lemma 6, which since $\operatorname{supp}(F_0)$ is convex and contains $[a', b']$ implies that $(a', b') \cap \operatorname{supp}(G)$ is non-empty. Define

$$c \coloneqq \inf\bigl[(a', b') \cap \operatorname{supp}(G)\bigr] \quad \text{and} \quad d \coloneqq \sup\bigl[(a', b') \cap \operatorname{supp}(G)\bigr].$$

We first show that $C_G \leq C_H$ on $[a,b] \setminus (c,d)$. This is trivial if $c \leq a$ and $b \leq d$, so suppose not. Assume that $a < c$; we will show that $C_G \leq C_H$ on $[a, \min\{b,c\}]$. (We omit the analogous argument that $C_G \leq C_H$ on $[\max\{a,d\}, b]$ when $d < b$.) By definition of c, C_G is affine on $[a', c]$. Since $C_G \leq C_{F_0}$ with equality at a', where C_{F_0} is convex and differentiable at a' (F_0 being atomless), C_G coincides on $[a', c]$ with the tangent to C_{F_0} at a'. Similarly, since $C_H \leq C_{F_0}$ with equality at a and C_H is convex, we have on $[a,1]$ that C_H exceeds the tangent to C_{F_0} at a. Since the latter tangent exceeds the former on $[a,1]$, it follows that $C_G \leq C_H$ on $[a', c] \cap [a,1] = [a,c] \supseteq [a, \min\{b,c\}]$.

It remains to show that u is affine on $[c,d]$. Since u is regular, it suffices to show that u is affine on $[x,w]$ for any $x < w$ in $(a', b') \cap \operatorname{supp}(G)$. Fix such a pair $x < w$, and note that by Lemma 6, p is tangent to u at x and at w. Then since $p \geq u$ and u satisfies the crater property, Lemma 7 provides that there are $y \in (x,w]$ and $z \in [x,w)$ such that u is concave on $[x,y]$ and on $[z,w]$ and strictly convex on $[x,z]$ and on $[y,w]$. Clearly it must be that $y = w$ and $z = x$, so that u is concave on $[x,w]$. Since p is convex and $p \geq u$ on $[x,w]$ with equality on $\{x,w\}$, it follows that u is affine on $[x,w]$. □

With the claim established, the proof is complete. ■

C.3 Proof of Lemma 8

We rely on the following result, which follows from Lemmata 6 and 7.

Corollary 3. Let $u : [0,1] \to \mathbf{R}$ be regular, let F_0 be an atomless convex-support distribution, and let p minimise $\int p \mathrm{d}F_0$ over $\mathcal{M}(u)$. Then

(i) for any $x < z$ such that $[x,z]$ is maximal among the intervals of affineness of p within $\operatorname{supp}(F_0)$, there are

$$x < y \leq \frac{\int_x^z \xi F_0(\mathrm{d}\xi)}{F_0(z) - F_0(x)} \leq y' < z$$

such that $p(y) = u(y)$ and $p(y') = u(y')$, and

(ii) if $p(y) > u(y)$ for some $y \in \operatorname{supp}(F_0)$ such that $F_0(y) > 0$ ($F_0(y) < 1$), then $y > 0$ and there is $x \in [0,y)$ ($y < 1$ and there is $z \in (y,1]$) such that p is affine on $[x,y]$ (on $[y,z]$).

Moreover, if u satisfies the crater property, then

(iii) given $x < y$ such that $[x, y]$ is maximal among the intervals of affineness of p within $\operatorname{supp}(F_0)$, and $F_0(x) > 0$ ($F_0(y) < 1$), it holds that $p(x) = u(x)$ ($p(y) = u(y)$), that u is convex and not affine on some open interval I containing x (y), and that

$$u' <(>) \frac{p(y) - p(x)}{y - x} \quad \text{on } (0, x) \cap I \text{ (on } (y, 1) \cap I).$$

Proof of Corollary 3. Fix F maximising $\int u \mathrm{d}F$ among distributions feasible given F_0. For (i), fix $x < z$ such that $[x, z]$ is maximal among intervals of affineness of p within $\operatorname{supp}(F_0)$. Then $C_F = C_{F_0}$ on $\{x, z\}$ by Lemma 6.[39] Then (x, z) is F-non-null since F_0 has convex support,[40] and thus there are $y, y' \in \operatorname{supp}(F)$ such that

$$x < y \leq \frac{\int_{(x,z)} \xi F(\mathrm{d}\xi)}{\int_{(x,z)} \mathrm{d}F} \leq y' < z.$$

By (b), $p(y) = u(y)$ and $p(y') = u(y')$. Finally, since $C_F = C_{F_0}$ on $\{x, z\}$ and F_0 is atomless, $F = F_0$ on $\{x, z\}$ and F is continuous at x and z, so that

$$\frac{\int_x^z \xi F(\mathrm{d}\xi)}{\int_{(x,z)} \mathrm{d}F} = \frac{zF(z) - xF(x) - [C_F(z) - C_F(x)]}{F(z) - F(x)} = \frac{\int_x^z \xi F_0(\mathrm{d}\xi)}{F_0(z) - F_0(x)}.$$

This proves (i).

For (ii), suppose that $p(y) > u(y)$ for some $y \in \operatorname{supp}(F_0)$ such that $F_0(y) > 0$ (the case $F_0(y) < 1$ is analogous). Then $y \notin \operatorname{supp}(F)$ by (b), so that C_F is affine on a neighbourhood of y. Moreover, $y > \min \operatorname{supp}(F_0)$ since F_0 is atomless. Then, $y > 0$ and, since $\operatorname{supp}(F_0)$ is convex and C_{F_0} is strictly convex on $\operatorname{supp}(F_0)$, there is $x \in [0, y)$ such that $C_F < C_{F_0}$ on $[x, y)$. Hence, p is affine on $[x, y]$ by (a), as p is continuous.

For (iii), fix $x < y$ such that $[x, y]$ is maximal among intervals of affineness of p within $\operatorname{supp}(F_0)$, and $F_0(x) > 0$ (the case $F_0(y) < 1$ is analogous). By (i), there is $w \in (x, y)$ such that $p(w) = u(w)$, so that p is tangent to u at w. Then, there is $z \in [x, w)$ such that u is strictly convex on $[x, z]$ and

[39] If e.g. $C_F(x) < C_{F_0}(x)$, then x lies in the interior of $\operatorname{supp}(F_0)$, $C_F < C_{F_0}$ on a neighbourhood of x, and p is affine on this neighbourhood by (a), contradicting the definition of $[x, z]$.

[40] Since F_0 has convex support, C_{F_0} is not affine on $[x, z]$. Then, neither is C_F, and thus $\operatorname{supp}(F) \cap (x, z)$ is not empty.

concave on $[z, w]$, by Lemma 7. Let $b \coloneqq \min \operatorname{supp}(F_0)$ and a be the smallest $a' \in [b, x]$ such that p is affine on $[a', x]$. We consider two cases.

Case 1. $a = x$. Note that $x > b$ since F_0 is atomless and $F_0(x) > 0$. Then, by the hypothesis of this case, there exists an increasing sequence $(x_k)_{k \in \mathbf{N}} \subseteq (b, x)$ such that $\lim_k x_k = x$ and on which $C_F = C_{F_0}$, by (a). Then, there exists an increasing sequence $(y_k)_{k \in \mathbf{N}} \subseteq (b, x) \cap \operatorname{supp}(F)$ such that $\lim_k y_k = x$, since C_{F_0} is strictly convex on $\operatorname{supp}(F_0)$. By (b), $p(y_k) = u(y_k)$ for each $k \in \mathbf{N}$. Then, since p is convex and u is regular, by the hypothesis of this case, u is convex and not affine on $[y_{k'}, x]$ for some $k' \in \mathbf{N}$, and

$$u' < \frac{p(y) - p(x)}{y - x} \quad \text{on } (y_{k'}, x).$$[41]

Moreover, $p(x) = u(x)$ and thus u is affine on $[x, w]$ if $z = x$, since $p \geq u$ with equality on $\{x, w\}$ and u is concave on $[z, w]$. The result follows by choosing $I = (y_{k'}, z)$ if $z > x$, and $I = (y_{k'}, w)$ otherwise.

Case 2. $a < x$. In this case, there is $\widehat{x} \in (a, x)$ such that $p(\widehat{x}) = u(\widehat{x})$, by (i). Then, p is tangent to u at $\widehat{x}$, and thus there is $\widehat{y} \in (\widehat{x}, x]$ such that u is concave on $[\widehat{x}, \widehat{y}]$, and strictly convex on $[\widehat{y}, x]$, by Lemma 7. Define

$$I \coloneqq \begin{cases} (\widehat{y}, z) & \text{if } \widehat{y} < x < z \\ (\widehat{x}, z) & \text{if } \widehat{y} = x \\ (\widehat{y}, w) & \text{if } x = z. \end{cases}$$

Note that $\widehat{y} < z$, for otherwise u would be concave on $[\widehat{x}, w]$ and thus p would be affine on $[\widehat{x}, w]$ (since $p = u$ on on $\{\widehat{x}, w\}$), contradicting $\widehat{x} < x$. Then I contains x, since $\widehat{x} < \widehat{y} \leq x \leq z < w$.

To show that u is convex and not affine on I, note that u is strictly convex on $[\widehat{y}, z]$, as it is regular and strictly convex on $[\widehat{y}, x]$ and $[x, z]$. Then $p(x) = u(x)$, since u satisfies the crater property and, clearly, the tangents to u at $\widehat{x}$ and w intersect at $(x, p(x))$. Hence u is affine on $[\widehat{x}, x]$ (on $[x, w]$) if $\widehat{y} = x$ ($x = z$), since u is concave on $[\widehat{x}, \widehat{y}]$ with $u(\widehat{x}) = p(\widehat{x})$ (on $[z, w]$ with $u(w) = p(w)$). Since $\widehat{y} < z$ and u is strictly convex on $[\widehat{y}, z]$, u is convex and not affine on I.

It remains to show that

$$u' < \frac{p(y) - p(x)}{y - x} \quad \text{on } (0, x) \cap I.$$

[41] To see why this last property must hold, suppose it were to fail. Then there is a sequence $(z_k)_{k=1}^{\infty} \subseteq (b, x)$ with $\lim_{k \to \infty} z_k = x$ such that $u'(z_k) \geq [p(y) - p(x)]/(y - x)$. Since u is regular, it follows that $u' \geq [p(y) - p(x)]/(y - x)$ on $[y_{k'}, x]$ for some $k' \in \mathbf{N}$. But then p is affine on $[y_{k'}, x]$ since $u(y_{k'}) = p(y_{k'})$, contradicting the hypothesis of this case.

To this end, since u is convex on I, we may assume without loss of generality that

$$u'(x) \geq \frac{p(y) - p(x)}{y - x}.$$

Then $x = z$ and equality holds, since $p \geq u$ with equality at x and u is strictly convex on $[x, z]$. The result follows since $\widehat{y} < z$ and u is strictly convex on $[\widehat{y}, x]$. ■

Proof of Lemma 8. Fix F_0, p and q. Suppose toward a contradiction that there exist $\widetilde{x} < \widetilde{z}$ in $\operatorname{supp}(F_0)$ such that q is affine on $[\widetilde{x}, \widetilde{z}]$, but p is not. Assume without loss of generality that $[\widetilde{x}, \widetilde{z}]$ is maximal among the intervals of affineness of q within $\operatorname{supp}(F_0)$. We consider two cases.

Case 1. u is convex on $[\widetilde{x}, \widetilde{z}]$. We shall construct $a \in [0, \widetilde{x}]$ such that u is concave on $[a, \widetilde{z}]$ and $p(a) = u(a)$. A similar argument yields $b \in [\widetilde{z}, 1]$ such that u is concave on $[\widetilde{x}, b]$ and $p(b) = u(b)$. Then u is concave on $[a, b]$ and thus p is affine on $[a, b]$, contradicting the fact that p is not affine on $[\widetilde{x}, \widetilde{z}] \subseteq [a, b]$.

To construct a, note that v is convex on $[\widetilde{x}, \widetilde{z}]$ by the hypothesis of this case, since u is coarsely less convex than v. Then v is affine on $[\widetilde{x}, \widetilde{z}]$ by (i) (since (i) implies that $q(y) = v(y)$ for some $y \in (\widetilde{x}, \widetilde{z})$). Then so is u, as it is coarsely less convex than v. Then, if $p(\widetilde{x}) = u(\widetilde{x})$, we may take $a = \widetilde{x}$. Hence, assume without loss of generality that $p(\widetilde{x}) > u(\widetilde{x})$.

Let $\bar{z}$ be the largest $z \in [\widetilde{x}, 1]$ such that p is affine on $[\widetilde{x}, z]$. Then $\bar{z} < \widetilde{z}$ by hypothesis, and $\bar{z} > \widetilde{x}$ by (ii) (which is applicable since $F_0(\widetilde{x}) < 1$). Let $\bar{x}$ be the smallest $x \in [0, \widetilde{x}] \cap \operatorname{supp}(F_0)$ such that p is affine on $[x, \bar{z}]$. By (i), there is $a \in (\bar{x}, \bar{z})$ such that $p(a) = u(a)$. And a belongs to $[0, \widetilde{x}]$ since u and p are affine on $[\widetilde{x}, \bar{z}]$ and since $p \geq u$, with strict inequality at $\widetilde{x}$.

It remains to prove that u is concave on $[a, \widetilde{z}]$. As u is affine on $[\widetilde{x}, \widetilde{z}]$ and regular, and $\widetilde{x} < \bar{z} < \widetilde{z}$, it suffices to show that u is concave on $[a, \bar{z}]$. Note that p is tangent to u at a as $\bar{x} < a < \bar{z}$ and $p(a) = u(a)$. Then u is concave on $[a, \bar{z}]$ by Lemma 7, as $p \geq u$ on $[a, \bar{z}]$, and u and p are affine on $[\widetilde{x}, \bar{z}]$.[42]

Case 2. u is not convex on $[\widetilde{x}, \widetilde{z}]$. In this case, since u is regular, there are $\widetilde{x} \leq c < d \leq \widetilde{z}$ such that $[c, d]$ is maximal among the intervals in $[\widetilde{x}, \widetilde{z}]$ on which u is strictly concave. Then p and u differ somewhere in (c, d) and thus, by (ii), p is not strictly convex on (c, d). Hence there are $\bar{x} < \bar{z}$ such that $[\bar{x}, \bar{z}]$ is maximal among the intervals of affineness of p within $\operatorname{supp}(F_0)$,

[42] Indeed, Lemma 7 yields $y \in (a, \bar{z}]$ such that u is concave on $[a, y]$ and strictly convex on $[y, \bar{z}]$. And $y = \bar{z}$ since u is affine on $[\widetilde{x}, \bar{z}]$.

and $[\bar{x}, \bar{z}] \cap (c, d)$ is not empty. Since p is not affine on $[\widetilde{x}, \widetilde{z}]$, either $\widetilde{x} < \bar{x}$ or $\bar{z} < \widetilde{z}$. We consider the case $\widetilde{x} < \bar{x}$; the other is analogous.

Note that $\bar{x} < d \leq \widetilde{z}$, where the strict inequality holds as $[\bar{x}, \bar{z}] \cap (c, d)$ is not empty. We shall exhibit a $w \in (\bar{x}, \widetilde{z}]$ such that

$$u(\bar{x})_\alpha u(w) \geq u(\bar{x}_\alpha w) \quad \text{for all } \alpha \in (0,1), \tag{4}$$

a $\widetilde{y} \in (\bar{x}, w)$ such that $q(\widetilde{y}) = v(\widetilde{y})$, and show that $v(\bar{x}) < q(\bar{x})$. To see why this suffices, note that it implies that given $\alpha \in (0,1)$ such that $\bar{x}_\alpha w = \widetilde{y}$,

$$v(\bar{x})_\alpha v(w) < q(\bar{x})_\alpha q(w) = q(\widetilde{y}) = v(\widetilde{y}),$$

where the strict inequality holds since $\alpha \in (0,1)$, $v(\bar{x}) < q(\bar{x})$ and $v(w) \leq q(w)$, and the first equality holds as q is affine on $[\widetilde{x}, \widetilde{z}] \supseteq [\bar{x}, w]$. Together with (4), this contradicts the fact that u is coarsely less convex than v.

To construct w note that, by (i), there is

$$\bar{x} < \frac{\int_{\bar{x}}^{\bar{z}} \xi F_0(\mathrm{d}\xi)}{F_0(\bar{z}) - F_0(\bar{x})} \leq \bar{y} < \bar{z}$$

such that $p(\bar{y}) = u(\bar{y})$. Define $w \coloneqq \min\{\bar{y}, \widetilde{z}\}$ and note that $w \in (\bar{x}, \widetilde{z}]$. To establish (4), note p is tangent to u at $\bar{y}$, so that there is $\gamma \in [\bar{x}, \bar{y})$ such that u is strictly convex on $[\bar{x}, \gamma]$ and concave on $[\gamma, \bar{y}]$, by Lemma 7. Then (4) holds since $p(\bar{x}) = u(\bar{x})$ by (iii) (which is applicable since $F_0(\bar{x}) > 0$ and $\widetilde{x} < \bar{x}$).[43]

To construct $\widetilde{y} \in (\bar{x}, w)$ such that $q(\widetilde{y}) = v(\widetilde{y})$, let $[a, b]$ be the maximal interval of convexity of u containing $\bar{x}$. (This is well-defined since u is regular). Note that if $\bar{x} \in (c, d)$ then $\gamma = \bar{x}$, as u is concave on (c, d) and on $[\gamma, \bar{y}]$, and strictly convex on $[\bar{x}, \gamma]$. But then u would be affine on $[\bar{x}, \bar{y}]$ since $p = u$ on $\{\bar{x}, \bar{y}\}$, contradicting the fact that u is strictly concave on $[c, d]$. Hence $\bar{x} < c$ as $\bar{x} < d$. Then $b \leq c$, and by (iii) (applicable since $F_0(\bar{x}) > 0$ and $\widetilde{x} < \bar{x}$) we have that $a < \bar{x} < b$, that u is not affine on $[a, b]$, and that

$$u' < \frac{p(\bar{z}) - p(\bar{x})}{\bar{z} - \bar{x}} \quad \text{on } (a, \bar{x}). \tag{5}$$

We rely on the following claim, proved at the end.

Claim. $a \leq \widetilde{x}$ and $\widetilde{z} \leq \bar{z}$.

[43]This is easily seen graphically. It follows from the facts that p is affine on $[\bar{x}, \bar{y}]$, that $p \geq u$ on $[\bar{x}, \bar{y}]$ with equality on $\{\bar{x}, \bar{y}\}$, that u is convex on $[\bar{x}, \widehat{z}]$ and concave on $[\widehat{z}, \bar{y}]$ for some $\widehat{z} \in [\bar{x}, \bar{y}]$, and that $\bar{x} < w \leq \bar{y}$.

By (i), we may choose

$$\widetilde{x} < y \leq \frac{\int_{\widetilde{x}}^{\widetilde{z}} \xi F_0(\mathrm{d}\xi)}{F_0(\widetilde{z}) - F_0(\widetilde{x})} < \widetilde{z}$$

such that $q(y) = v(y)$. Note that $y < \min\{\bar{y}, \widetilde{z}\} = w$ since $y < \widetilde{z}$ and

$$y \leq \frac{\int_{\widetilde{x}}^{\widetilde{z}} w \mathrm{d}F_0(w)}{F_0(\widetilde{z}) - F_0(\widetilde{x})} < \frac{\int_{\bar{x}}^{\bar{z}} w \mathrm{d}F_0(w)}{F_0(\bar{z}) - F_0(\bar{x})} \leq \bar{y},$$

where the strict inequality holds as F_0 has convex support, $\widetilde{x} < \bar{x}$ and, by the claim, $\widetilde{z} \leq \bar{z}$. Thus we may take $\widetilde{y} \coloneqq y$ if $y > \bar{x}$. If instead $y \leq \bar{x}$, note that v is convex on $[a, b]$, as u is coarsely less convex than v and convex on $[a, b]$. Moreover, q is affine on $[\widetilde{x}, \widetilde{z}]$ and $q \geq v$ with equality at y. Since $a \leq \widetilde{x} < y \leq \bar{x} < b \leq c \leq \widetilde{z}$, it follows that $v = q$ on $[\widetilde{x}, b] = [a, b] \cap [\widetilde{x}, \widetilde{z}]$. As $\bar{x} < w$, we may then choose any $\widetilde{y} \in (\bar{x}, \min\{b, w\})$.

It remains to prove that $v(\bar{x}) < q(\bar{x})$. Note that, by (4) and (5),

$$u(\widetilde{x})_\alpha u(w) > u(\widetilde{x}_\alpha w) \quad \text{for all } \alpha \in (0, 1),$$

since $a \leq \widetilde{x} < \bar{x}$, and u is convex on $[a, \bar{x}]$.[44] Hence, choosing $\alpha \in (0, 1)$ such that $\bar{x} = \widetilde{x}_\alpha w$,

$$q(\bar{x}) = q(\widetilde{x})_\alpha q(w) \geq v(\widetilde{x})_\alpha v(w) > v(\bar{x}),$$

where the equality holds since q is affine on $[\widetilde{x}, \widetilde{z}] \supseteq [\widetilde{x}, w]$, the weak inequality as $q \geq v$, and the strict inequality holds since u is less convex than v.

Proof of the claim. We begin by exhibiting $\widetilde{x} \leq c' < d' \leq \widetilde{z}$ such that u is strictly convex on $[\widetilde{x}, c']$ and $[d', \widetilde{z}]$, and concave on $[c', d']$. By (i),

$$v(\widetilde{x}_\alpha \widetilde{z}) = q(\widetilde{x}_\alpha \widetilde{z}) = q(\widetilde{x})_\alpha q(\widetilde{z}) \geq v(\widetilde{x})_\alpha v(\widetilde{z}) \quad \text{for some } \alpha \in (0, 1),$$

where the second equality holds since q is affine on $[\widetilde{x}, \widetilde{z}]$, and the inequality holds since $q \geq v$. Then

$$u(\widetilde{x}_\alpha \widetilde{z}) \geq u(\widetilde{x})_\alpha u(\widetilde{z}) \quad \text{for some } \alpha \in (0, 1), \tag{6}$$

since u is coarsely less convex than v. Hence the tangent to u at some $a_\star \in (\widetilde{x}, \widetilde{z})$ weakly exceeds u on $[\widetilde{x}, \widetilde{z}]$, as u is regular. Therefore, by Lemma 7,

[44] In detail, on $(a, \bar{x})$, $u' > [p(\bar{z}) - p(\bar{x})]/(\bar{z} - \bar{x}) = [p(w) - p(\bar{x})]/(w - \bar{x}) \geq [u(w) - u(\bar{x})]/(w - \bar{x})$, and thus the continuous map that matches u on $[\widetilde{x}, \bar{x}] \cup \{w\}$ and is affine on $[\bar{x}, w]$, is convex and not affine on $[\widetilde{x}, w]$. Then the result follows from (4).

there are $c' \in [\widetilde{x}, a_\star)$ and $d' \in (a_\star, \widetilde{z}]$ such that u is strictly convex on $[\widetilde{x}, c']$ and $[d', \widetilde{z}]$, and concave on $[c', a_\star]$ and $[a_\star, d']$. As u is regular, it is concave on $[c', d']$, as desired.

Note that $b \leq c < d \leq \widetilde{z}$. Then $a \leq \widetilde{x}$ since u is regular. Indeed, if $\widetilde{x} < a$ then, by definition of a and b, there would exist $\widetilde{x} \leq a' < a$ and $b < b' \leq \widetilde{z}$ such that u is strictly concave on $[a', a]$ and $[b, b']$. But then $c' \leq a'$ and $b' \leq d'$, contradicting the fact that u is convex and not affine on $[a, b]$.

It remains to show that $\widetilde{z} \leq \bar{z}$. Suppose this fails and seek a contradiction. Then $p(\bar{z}) = u(\bar{z})$ by (iii), and thus

$$u(\bar{x})_\alpha u(\bar{z}) = p(\bar{x})_\alpha p(\bar{z}) = p(\bar{x}_\alpha \bar{z}) \geq u(\bar{x}_\alpha \bar{z}) \quad \text{for all } \alpha \in (0,1), \tag{7}$$

where the first equality holds since $p(\bar{x}) = u(\bar{x})$, and the second since p is affine on $[\bar{x}, \bar{z}]$. Moreover, u is convex and not affine on some open interval I containing $\bar{z}$, by (iii). Then

$$c' \leq c < d \leq \bar{z},$$

where the first inequality holds since $\widetilde{x} \leq c < d$ and u is strictly convex on $[\widetilde{x}, c']$ and strictly concave on $[c, d]$, and the last inequality holds since $[\bar{x}, \bar{z}] \cap (c, d) \neq \emptyset$ and u is strictly concave on $[c, d]$ and convex on $I \ni \bar{z}$. Then u is convex on $[\bar{z}, \widetilde{z}]$, as it is convex and not affine on $I \ni \bar{z}$, concave on $[c', d']$, and strictly convex on $[d', \widetilde{z}]$. Then (5) and (7) contradict (6), since $a \leq \widetilde{x} < \bar{x}$ and u is convex on $[a, \bar{x}]$.[45] □

With the claim established, the proof is complete. ■

Appendix D Proof of Proposition 1 (p. 13)

The second (converse) part of Proposition 1 follows from the proof in appendix B.1 of the second (converse) part of Theorem 1*.[46]

To prove the first part, let $u, v : [0,1] \to \mathbf{R}$ be upper semi-continuous, assume that u is coarsely less convex than v, and let F_0 be a binary distribution. Write μ for the mean of F_0. Assume without loss of generality that F_0 is supported on $\{0,1\}$ (that is, $F_0 = 1 - \mu + \mu \mathbf{1}_{\{1\}}$).[47] Given $x, y \in \mathbf{R}$ and $\alpha \in [0,1]$, let us write $x_\alpha y \coloneqq \alpha x + (1-\alpha) y$.

[45] To see why, note that the map $\mathbf{1}_{[\tilde{x},\bar{x}] \cup [\bar{z},\tilde{z}]} u + \mathbf{1}_{(\bar{x},\bar{z})} p$ is convex and not affine on $[\widetilde{x}, \widetilde{z}]$.

[46] The argument there shows that if u is not coarsely less convex than v, then we can construct a prior F_0 such that $\arg\max_F \int u \mathrm{d}F$ is strictly higher than (a fortiori not lower than) $\arg\max_F \int v \mathrm{d}F$. And the constructed prior is, in fact, binary.

[47] If F_0 is degenerate ($F_0 = \mathbf{1}_{[\mu,1]}$) then the result is trivial. If not, then F_0 is supported on $\{x, y\}$ with $x < \mu < y$, all feasible distributions have support in $[x, y]$, and $u|_{[x,y]}$ is coarsely less convex than $v|_{[x,y]}$; so the interval $[x, y]$ may as well be $[0,1]$.

Write $\operatorname{cav} u$ for the concave envelope of u. Let $[x, w]$ be the maximal interval containing μ on which $\operatorname{cav} u$ is affine. Define

$$\mathcal{U} \coloneqq \{u = \operatorname{cav} u\} \cap [x, w],$$

and note that $x, w \in \mathcal{U}$ since u is upper semi-continuous. Further define

$$y \coloneqq \sup(\mathcal{U} \cap [0, \mu]) \quad \text{and} \quad z \coloneqq \inf(\mathcal{U} \cap [\mu, 1]),$$

and note that $y, z \in \mathcal{U}$ by upper semi-continuity. Clearly $x \leq y \leq \mu \leq z \leq w$.

Let

$$M(u) \coloneqq \underset{F \text{ feasible given } F_0}{\arg\max} \int u \mathrm{d}F.$$

Kamenica and Gentzkow (2011) showed that $M(u)$ is the set of all mean-μ distributions F such that $\int u \mathrm{d}F = (\operatorname{cav} u)(\mu)$. Thus $M(u)$ is the set of all mean-μ distributions supported on $\mathcal{U}$. It follows that the distribution G (H) with mean μ and support $\{y, z\}$ ($\{x, w\}$) is the least (most) informative distribution in $M(u)$.

For the function v, analogously define $\mathcal{V} \subseteq [0, 1]$, $x', y', z', w' \in \mathcal{V}$, and distributions G', H' in $M(v)$. We must show that H is less informative than H' and that G is less informative than G'. The former requires precisely that $x' \leq x$ and $w \leq w'$, while the latter requires that $y' \leq y$ and $z \leq z'$.

We first show that $x' \leq x$ and $w \leq w'$. Since $x, w \in \mathcal{U}$, we have $u(x_\alpha w) \leq u(x)_\alpha u(w)$ for every $\alpha \in (0, 1)$. As u is coarsely less convex than v, it follows that $v(x_\alpha w) \leq v(x)_\alpha v(w)$ for each $\alpha \in (0, 1)$, implying that $[x, w] \subseteq [x', w']$.

Claim. $\mathcal{V} \cap [x, w] \subseteq \mathcal{U}$.

Proof. Take any $\widehat{y} \in \mathcal{V} \cap [x, w]$. The result is trivial if $\widehat{y} = x$ or $\widehat{y} = w$, so suppose not: $\widehat{y} = x_\alpha w$ for some $\alpha \in (0, 1)$. Then

$$v(x_\alpha w) = (\operatorname{cav} v)(x_\alpha w) \geq (\operatorname{cav} v)(x)_\alpha (\operatorname{cav} v)(w) \geq v(x)_\alpha v(w)$$

since $x_\alpha w \in \mathcal{V}$ (the equality), $\operatorname{cav} v$ is concave (first inequality), and $\operatorname{cav} v \geq v$ (second inequality), whence $u(x_\alpha w) \geq u(x)_\alpha v(w)$ because u is coarsely less convex than v. So $u(x_\alpha w) = u(x)_\alpha v(w)$, and thus $\widehat{y} = x_\alpha w \in \mathcal{U}$. □

We now show that $y' \leq y$; the argument for $z \leq z'$ is analogous. If $y' < x$, then $y' < x \leq y$ since $y \in \mathcal{U} \subseteq [x, w]$. Suppose instead that $x \leq y'$. Then since $y' \leq \mu \leq w$, we have $y' \in [x, w]$. As $y' \in \mathcal{V}$, it follows from the claim that y' belongs to $\mathcal{U}$. So $y' \in \mathcal{U} \cap [0, \mu]$, and thus $y' \leq \sup(\mathcal{U} \cap [0, \mu]) = y$. ■

Appendix E Proof of Proposition 2 (p. 13)

The argument is close to the proof of the converse (necessity) half of Theorem 2, which we sketched in §4.2 and which is given in full in appendix C. Fix a distribution F_0 that is not binary. Choose an $X \in (0,1)$ such that $0 < \lim_{z \uparrow X} F_0(z) \leq F_0(X) < 1$. Define

$$x \coloneqq \frac{1}{F_0(X)} \int_{[0,X]} \xi F_0(\mathrm{d}\xi) \quad \text{and} \quad w \coloneqq \frac{1}{1-F_0(X)} \int_{(X,1]} \xi F_0(\mathrm{d}\xi),$$

and note that $x < X < w$. Fix a convex $p : [0,1] \to \mathbf{R}$ that is affine on $[0,X]$ and on $[X,1]$, but not affine on $[0,1]$. Clearly we may choose a regular and M-shaped $u : [0,1] \to \mathbf{R}$ such that $p = u$ on $\{x,w\}$ and $p > u$ on $[0,1] \setminus \{x,w\}$, and such that u is convex on $[X,y]$ and concave on $[y,1]$ for some $y \in (X,1)$. Let G be the distribution supported on $\{x,w\}$ whose mean is the same as that of F_0. Then G is uniquely optimal for u given F_0, since any other feasible distribution F has $\int u \mathrm{d}F < \int p \mathrm{d}F \leq \int p \mathrm{d}F_0 = \int p \mathrm{d}G = \int u \mathrm{d}G$, where the weak inequality holds since p is convex and F is feasible given F_0, the first equality holds since p is affine on $[0,X]$ and on $[X,1]$, and the final equality holds since $p = u$ G-a.e.

Since u' is bounded, we may choose a regular $v : [0,1] \to \mathbf{R}$ that coincides with u on $[X,1]$ and that weakly exceeds u and is strictly convex on $[0,X]$. Then v is S-shaped and coarsely more convex than u. Let $\delta \coloneqq F_0(a) - \lim_{z \uparrow a} F_0(z)$, and observe that there are $a \in [0,X]$ and $\pi \in [0,1]$ such that

$$\frac{v(b)-v(a)}{b-a} = v'(b), \quad \text{where} \quad b \coloneqq \frac{\pi\delta a + \int_a^1 \xi F_0(\mathrm{d}\xi)}{\pi\delta + 1 - F_0(a)} > 0.$$

Define F by $F \coloneqq F_0$ on $[0,a)$, $F \coloneqq F_0(a) - \pi\delta$ on $[a,b)$, and $F \coloneqq 1$ on $[b,1]$. (That is, F reveals $[0,a)$, pools $(a,1]$, reveals a with probability $1-\pi$, and otherwise pools it with $(a,1]$.) Let $q : [0,1] \to \mathbf{R}$ be affine on $[X,1]$ and satisfy $q \geq v$, with equality on $[0,a] \cup \{b\}$. The distribution F is optimal for v given F_0 since for any (other) feasible distribution H, we have $\int v \mathrm{d}H \leq \int q \mathrm{d}H \leq \int q \mathrm{d}F_0 = \int q \mathrm{d}F = \int v \mathrm{d}F$, where the second inequality holds since q is convex and H is feasible given F_0, the first equality holds since q is affine on $[a,1]$, and the final equality holds since $q = v$ F-a.e.

Since $p(X) > u(X)$, it must be either that $a < X$ or that $a = X$ and $\pi\delta > 0$. Thus F is not more informative than G, so ($\star\star$) fails. ■

Appendix F Proof of Proposition 3 (p. 14)

For the first half (sufficiency), fix an atomless convex-support prior distribution F_0, and let $u, v : [0,1] \to \mathbf{R}$ be regular with u coarsely less convex than v. If v is concave, then so is u, and thus u satisfies the crater property, so that ($\star\star$) holds by Theorem 2. If instead v is strictly convex, then F_0 is uniquely optimal for v given F_0, so ($\star\star$) holds.

For the second half (necessity), fix a regular $v : [0,1] \to \mathbf{R}$ that is neither concave nor strictly convex; we shall exhibit a regular $u : [0,1] \to \mathbf{R}$ that is coarsely less convex than v, an atomless convex-support prior distribution F_0, and a distribution F that is optimal for v given F_0 such that no distribution optimal for u given F_0 is less informative than F. The argument will be similar to the proof in appendix C.1 of the converse (necessity) part of Theorem 2, which we sketched in §4.2.

By hypothesis (and using regularity), there are $x' < z < w'$ in $[0,1]$ such that either v is strictly convex on $[x', z]$ and concave on $[z, w']$, or v is concave on $[x', z]$ and strictly convex on $[z, w']$. We consider the former case (the latter is analogous), and distinguish two (sub-)cases.

Case 1: v is not affine on $[z, w']$. In this case, we may choose $w \in (z, w')$ such that the tangent to v at w crosses v on $[x', w)$ exactly once, at some $a' \in (x', z)$. Since v' is bounded, we may choose a regular $u : [0,1] \to \mathbf{R}$ such that $u - v$ is concave (so u is coarsely less convex than v), u is strictly concave on $[0, a']$ and on $[w, 1]$, and $u \leq v$ on $[x', w']$, with equality on $[a', w]$. Then since u is strictly concave on $[x', a']$ and strictly convex on $[a', z]$, we may choose an $x \in (x', a')$ such that the tangent to u at x lies strictly above (below) u at a' (at z). It follows that there is a convex $p : [0,1] \to \mathbf{R}$ and an $X \in (a', z)$ such that p is affine on $[x', X]$ and on $[X, w']$, and $u \geq p$ on $[x', w']$, with equality on $\{x, w\}$ and with strict inequality at X.

Let F_0 be a distribution that is atomless with support $[x', w']$,

$$\frac{1}{F_0(X)} \int_0^X \xi F_0(\mathrm{d}\xi) = x \quad \text{and} \quad \frac{1}{1 - F_0(X)} \int_X^1 \xi F_0(\mathrm{d}\xi) = w.$$

As v is S-shaped on $[x', w']$, an 'upper censorship' distribution F is optimal by Kolotilin's (2014, p. 14) well-known result: for $a \in (0,1)$ satisfying

$$\frac{v(b) - v(a)}{b - a} = v'(b), \quad \text{where} \quad b \coloneqq \frac{1}{1 - F_0(a)} \int_a^1 \xi F_0(\mathrm{d}\xi),$$

this distribution F fully reveals $[0, a)$ and pools $[a, 1]$.[48] It is easy to see graphically (in Figure 2 on p. 11, paying attention to p) that a must be

[48]Explicitly, $F = F_0$ on $[0, a)$, $F = F_0(a)$ on $[a, b)$ and $F = 1$ on $[b, 1]$.

strictly smaller than X. Thus the optimal distribution F pools some states to the left of X with states to its right. For the payoff u, however, it is strictly sub-optimal to pool states on either side of X together. This is reasonably intuitive given the shape of u; formally, it follows from Lemma 5 in appendix C.1. Thus ($\star\star$) fails: no distribution optimal for u given F_0 is less informative than F, since the latter pools across X while the former do not.

Case 2: v is affine on $[z, w']$. In this case, since v' is bounded, we may choose $x < y$ in (x', z) and a regular $u : [0,1] \to \mathbf{R}$ such that $u = v$ on $[y, w']$, u is strictly concave on $[0, y]$ and on $[w', 1]$, and the tangent to u at x crosses u exactly once on (z, w'), at some $X \in (z, w')$. Fix some $w \in (X, w')$.

Let $p : [0,1] \to \mathbf{R}$ match the aforementioned tangent on $[x', X]$ and match u on $[X, w']$. Clearly p is affine on $[x', X]$ and on $[X, w']$ with $p'(x) < p'(w)$ (so p is convex). We furthermore have $p \geq u$ on $[x', w']$, with equality on $\{x, w\}$. Let F_0 be a distribution that is atomless with support $[x', w']$,

$$\frac{1}{F_0(X)} \int_0^X \xi F_0(\mathrm{d}\xi) = x \quad \text{and} \quad \frac{1}{1 - F_0(X)} \int_X^1 \xi F_0(\mathrm{d}\xi) = w.$$

The hypotheses of Lemma 5 in appendix C.1 are satisfied. Thus to show that ($\star\star$) fails, it suffices to exhibit a distribution F that is optimal for v given F_0 and that satisfies $\int_0^X F < \int_0^X F_0$ (i.e. states on either side of X are pooled). To that end, let

$$b \coloneqq \frac{1}{1 - F(z)} \int_z^1 \xi F_0(\mathrm{d}\xi),$$

and define a distribution F by $F \coloneqq F_0$ on $[0, z)$, $F \coloneqq F_0(z)$ on $[z, b)$ and $F \coloneqq 1$ on $[b, 1]$. Then $\int_0^X F < \int_0^X F_0$,[49] and clearly F is optimal for v given F_0. ∎

Appendix G Proof of Proposition 4 (p. 15)

Fix any atomless $F_0 \neq G_0$; we shall find a regular and S-shaped $u : [0,1] \to \mathbf{R}$ for which (†) fails. If F_0 is not less informative than G_0, then (†) fails for any strictly convex $u : [0,1] \to \mathbf{R}$, since F_0 (G_0) is uniquely optimal for u given F_0 (G_0). Assume for the remainder that F_0 is less informative than G_0.

[49] If $b \geq X$, then $\int_0^X (F_0 - F) = \int_z^X [F_0 - F_0(z)] > 0$ as $F_0 > F_0(z)$ on $(z, 1]$, while if $b \leq X$ we have $\int_0^X (F_0 - F) = \int_X^1 (F - F_0) = \int_X^1 (1 - F_0) > 0$ as $F_0 < 1$ on $[0, w')$.

For any atomless distribution F, integration by parts[50] yields

$$\frac{1}{1-F(y)}\int_y^1 xF(\mathrm{d}x) = \frac{1-yF(y)-\int_y^1 F}{1-F(y)} = 1+\frac{(1-y)F(y)-\int_y^1 F}{1-F(y)}$$

for each $y \in (0,1)$. We have $\int_y^1 F_0 \geq \int_y^1 G_0$ for every $y \in (0,1)$ since F_0 is less informative than G_0. Since in addition $F_0 \neq G_0$, it cannot be that F_0 is first-order stochastically dominated by G_0, and thus $F_0(a) < G_0(a)$ for some $a \in (0,1)$. It follows that

$$b \coloneqq \frac{1}{1-F_0(a)}\int_a^1 xF_0(\mathrm{d}x) < \frac{1}{1-G_0(a)}\int_a^1 xG_0(\mathrm{d}x). \tag{8}$$

Choose a regular and S-shaped $u : [0,1] \to \mathbf{R}$ such that $(u(b)-u(a))/(b-a) = u'(b)$. Let F be the distribution given by $F \coloneqq F_0$ on $[0,a)$, $F \coloneqq F_0(a)$ on $[a,b)$ and $F \coloneqq 1$ on $[b,1]$. Write a' for the unique $y \in (0,1)$ satisfying

$$\frac{u(\beta(y))-u(y)}{\beta(y)-y} = u'(\beta(y)), \quad \text{where} \quad \beta(y) \coloneqq \frac{1}{1-G_0(y)}\int_y^1 xG_0(\mathrm{d}x),$$

define $b' \coloneqq \beta(a')$, and let G be the distribution given by $G \coloneqq G_0$ on $[0,a')$, $G \coloneqq G_0(a')$ on $[a',b')$ and $G \coloneqq 1$ on $[b',1]$. By Kolotilin's (2014, p. 14) well-known result, F (G) is uniquely optimal for u given F_0 (G_0). By (8), we have $a > a'$, so F is not less informative than G. Thus (†) fails. ■

The atomlessness hypothesis in Proposition 4 can be dropped: it suffices to assume that F_0 is not degenerate. Then there are $a, \alpha \in [0,1]$ such that

$$\lim_{x\uparrow a} F_0(x) + \alpha\left[F_0(a) - \lim_{x\uparrow a} F_0(x)\right] < \lim_{x\uparrow a} G_0(x) + \alpha\left[G_0(a) - \lim_{x\uparrow a} G_0(x)\right] < 1,$$

and thus the proof above remains applicable, with minor modifications along the lines of the proof of Proposition 2 (appendix E) to take care of atoms.

Appendix H Proof of Theorem 2′ (p. 17)

For the first half (sufficiency), fix an atomless convex-support prior distribution F_0, and let $u, v : E \to \mathbf{R}$ be strongly regular with u coarsely less convex than v. If u is strictly concave, then the point mass at $\mu_0 \coloneqq \int xF_0(\mathrm{d}x)$ is uniquely optimal for u given F_0, so ($\star\star$) holds. If u is strictly convex, then so is v, in which case F_0 is uniquely optimal for v given F_0, so ($\star\star$) holds.

[50] Licensed by e.g. Theorem 18.4 in Billingsley (1995).

For the second half (necessity), say that a strongly regular $u : E \to \mathbf{R}$ satisfies the crater property iff for all distinct $x, y \in E$, the map $[0,1] \to \mathbf{R}$ given by $\alpha \mapsto u(\alpha x + (1-\alpha)y)$ satisfies the crater property.

Lemma 9. Let $u : E \to \mathbf{R}$ be strongly regular and satisfy the crater property, and let ℓ be the Lebesgue measure on a two-dimensional affine subspace of $\mathbf{R}^n$. Then

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0} \int u \mathrm{d}F$$

is a singleton for any distribution F_0 admitting a density with respect to ℓ.

Proof of Lemma 9. Since u is strongly regular, it is Lipschitz continuous. Hence by Theorem 7 in Dworczak and Kolotilin (2022), it suffices to show that there exists no $\varepsilon > 0$ and distinct $x, y \in E$ such that $\nabla u(x) = \nabla u(y)$, $\alpha x + (1-\alpha)y \in E$ for all $\alpha \in [-\varepsilon, 1+\varepsilon]$, and

$$\alpha u(x) + (1-\alpha)u(y) \geq u(\alpha x + (1-\alpha)y) \quad \text{for all } \alpha \in [-\varepsilon, 1+\varepsilon].$$

So suppose toward a contradiction that some $\varepsilon > 0$ and $x, y \in E$ have these properties. Define $w : [0,1] \to \mathbf{R}$ by $w(\alpha) := u(\alpha x + (1-\alpha)y)$ for each $\alpha \in [0,1]$. By hypothesis, the tangent to w at 0 lies above the graph of w, and is tangent to w also at 1. Since u is strongly regular, w is not affine. Hence w violates the crater property by Lemma 7 (appendix C.2, p. 41), so u violates the crater property—a contradiction. ∎

Fix a strongly regular $u : E \to \mathbf{R}$ that is neither strictly concave nor strictly convex; we shall find a strongly regular $v : E \to \mathbf{R}$ that is coarsely more convex than u and an atomless convex-support distribution F_0 such that ($\star\star$) fails. If u violates the crater property, then such v and F_0 exist by Theorem 2. Assume for the remainder that u satisfies the crater property.

Assume without loss that E has dimension n, and note that $n \geq 2$ by hypothesis. For any $S \subseteq E$, let $\operatorname{int}(S)$ denote its relative interior. For each $x \in \operatorname{int}(E)$, let H_x denote the Hessian matrix of u at x. We consider separately the case in which u has a saddle point, i.e. an $x \in \operatorname{int}(E)$ at which H_x is indefinite,[51] and the case in which it does not.

Case 1: H_x is indefinite at some $x \in \operatorname{int}(E)$. Assume without loss that $x = 0$. Since H_0 is indefinite, it admits eigenvalues λ_1, λ_2 such that $\lambda_2 < 0 < \lambda_1$. As H_0 is symmetric, its eigenvectors (appropriately rescaled) form an orthonormal basis of $\mathbf{R}^n$. We henceforth express elements of E in coordinates

[51] Recall that an $n \times n$ matrix M is *positive (negative) semi-definite* iff $x \cdot (Mx) \geq (\leq) 0$ for all $x \in \mathbf{R}^n$, and *indefinite* iff it is neither positive nor negative semi-definite.

relative to this basis, with the eigenvectors associated with λ_1 and λ_2 as (respectively) the first and second basis vectors. Then $u_{11}(0) = \lambda_1$, $u_{12}(0) = 0$ and $u_{22}(0) = \lambda_2$, where subscripts denote partial derivatives. Assume without loss that $u(0) = u_1(0) = u_2(0) = 0$. Let

$$S \coloneqq \{x \in \mathbf{R}^n : \|x\| \leq 1 \text{ and } x_i = 0 \text{ for } i > 2\},$$

and note that since $0 \in \operatorname{int}(E)$, we may assume without loss that $S \subseteq E$.

Let $u^\star : S \to \mathbf{R}$ be given by

$$u^\star(x) \coloneqq \tfrac{1}{2}\left(\lambda_1 x_1^2 + \lambda_2 x_2^2\right) \quad \text{for each } x \in S.$$

A second-order Taylor expansion of u around 0 yields that

$$|u(x) - u^\star(x)| \,/\, \|x\|^2 \to 0 \quad \text{as } x \to 0 \text{ in } S. \tag{9}$$

Since u is strongly regular, we may choose a convex and twice differentiable $\psi : E \to \mathbf{R}$ with $|\psi(x)| \,/\, \|x\|^3 \to 0$ as $x \to 0$ such that $v : E \to \mathbf{R}$ given by

$$v(x) \coloneqq u(x) - \tfrac{1}{2}\lambda_2(x_1 + x_2)^2 + \psi(x) \quad \text{for each } x \in E$$

is strongly regular. Since $v - u$ is convex, v is coarsely more convex than u. By a second-order Taylor expansion of v around 0,

$$|v(x) - v^\star(x)| \,/\, \|x\|^2 \to 0 \quad \text{as } x \to 0 \text{ in } S, \tag{10}$$

where $v^\star : S \to \mathbf{R}$ is given by

$$v^\star(x) \coloneqq u^\star(x) - \tfrac{1}{2}\lambda_2(x_1 + x_2)^2 \quad \text{for each } x \in S.$$

Let $F_0^\star$ be the uniform distribution on S. Note that there are no distinct $x, y \in \operatorname{int}(S)$ such that either $\nabla u^\star(x) = \nabla u^\star(y)$ or $\nabla v^\star(x) = \nabla v^\star(y)$. Hence by Theorem 7 in Dworczak and Kolotilin (2022),

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0^\star} \int u^\star \mathrm{d}F = \{F^\star\} \quad \text{and} \quad \operatorname*{arg\,max}_{F \text{ feasible given } F_0^\star} \int v^\star \mathrm{d}F = \{G^\star\}$$

for some distributions $F^\star$ and $G^\star$. We shall (a) show that $G^\star$ is not more informative than $F^\star$, and then (b) deduce that ($\star\star$) fails for some atomless convex-support prior distribution F_0.

For part (a), let F be the posterior-mean distribution induced (given prior $F_0^\star$) by a signal that reveals the first coordinate of the state and nothing else. The map $p : S \to \mathbf{R}$ given by $p(x) \coloneqq \frac{1}{2}\lambda_1 x_1^2$ for each $x \in S$ is convex and Lipschitz with $p \geq u^\star$, and it satisfies $\int (p - u^\star)\mathrm{d}F = 0$ since F assigns probability 1 to $\{x \in S : x_2 = 0\}$. Hence $F^\star = F$ by Theorem 5 in Dworczak and Kolotilin (2022). Thus if $G^\star$ were more informative than $F^\star$, then *any* distribution more informative than $F^\star$ would also be optimal for $v^\star$, since $F^\star = F$ reveals the first coordinate of the state and $v^\star(x_1, \cdot)$ is affine for each $x_1 \in [-1, 1]$. As $G^\star$ is uniquely optimal for $v^\star$ given $F_0^\star$, it therefore cannot be more informative than $F^\star$.

For part (b), define $u^\varepsilon, v^\varepsilon : S \to \mathbf{R}$ by $u^\varepsilon(x) \coloneqq u(\varepsilon x)/\varepsilon^2$ and $v^\varepsilon(x) \coloneqq v(\varepsilon x)/\varepsilon^2$ for each $x \in S$ and $\varepsilon \in (0,1)$. Since u, v and thus $u^\varepsilon, v^\varepsilon$ are strongly regular and satisfy the crater property, Lemma 9 implies that

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0^\star} \int u^\varepsilon \mathrm{d}F = \{F^\varepsilon\} \quad \text{and} \quad \operatorname*{arg\,max}_{F \text{ feasible given } F_0^\star} \int v^\varepsilon \mathrm{d}F = \{G^\varepsilon\}$$

for some distributions F^ε and G^ε. Write F_0^ε for the pushforward of $F_0^\star$ by $x \mapsto \varepsilon x$. Since rescaling interim payoffs (by $1/\varepsilon^2$) and the prior (by $1/\varepsilon$) affects neither feasibility nor the sender's preferences,[52]

$$\operatorname*{arg\,max}_{F \text{ feasible given } F_0^\varepsilon} \int u \mathrm{d}F \quad \text{and} \quad \operatorname*{arg\,max}_{F \text{ feasible given } F_0^\varepsilon} \int v \mathrm{d}F$$

are equal to the pushforward by $x \mapsto \varepsilon x$ of (respectively) F^ε and G^ε. Since $F^\star$ ($G^\star$) is uniquely optimal for $u^\star$ ($v^\star$) given $F_0^\star$ and $u^\varepsilon \to u^\star$ ($v^\varepsilon \to v^\star$) uniformly as $\varepsilon \downarrow 0$ by (9) (by (10)), $F^\varepsilon \to F^\star$ ($G^\varepsilon \to G^\star$) weakly as $\varepsilon \downarrow 0$.[53] Since $G^\star$ is not more informative than $F^\star$, it follows there is an $\varepsilon > 0$ such that G^ε fails to be more informative than F^ε, so that ($\star\star$) fails for $F_0 = F_0^\varepsilon$.

Case 2: H_x is indefinite at no $x \in \operatorname{int}(E)$. Say that u is *locally (strictly) concave* at $x \in \operatorname{int}(E)$ iff u is (strictly) concave on an open convex neighbourhood of x. Analogously define local (strict) convexity.

Claim. For any $x \in \operatorname{int}(E)$, if H_x is not positive (negative) semi-definite, then u is locally strictly concave (convex) at x.

[52] Writing $F \circ \varepsilon^{-1}$ for the pushforward by $x \mapsto \varepsilon x$ of a distribution F, (i) a distribution F is feasible given $F_0^\star$ iff $F \circ \varepsilon^{-1}$ is feasible given F_0^ε, and (ii) for F, G concentrated on S, $\int u^\varepsilon \mathrm{d}F \geq (>) \int u^\varepsilon \mathrm{d}G$ iff $\int u \mathrm{d}(F \circ \varepsilon^{-1}) \geq (>) \int u \mathrm{d}(G \circ \varepsilon^{-1})$, and similarly for v^ε and v.

[53] We have $u^\varepsilon \to u^\star$ uniformly since $\sup_{x \in S} |u^\varepsilon(x) - u^\star(x)| = \sup_{x \in S} |u(\varepsilon x) - u^\star(\varepsilon x)|/\varepsilon^2 \to 0$ as $\varepsilon \downarrow 0$ by definition of $u^\star$ and (9). To conclude that $F^\varepsilon \to F^\star$ weakly, note first that by Prokhorov's theorem (e.g. Billingsley, 1999, Theorem 5.1), $(F^\varepsilon)_{\varepsilon > 0}$ converges weakly along a subsequence to some distribution F. Hence $\int u^\star \mathrm{d}(F - F^\varepsilon)$ and $\int |u^\star - u^\varepsilon| \mathrm{d}F^\varepsilon$ vanish as $\varepsilon \downarrow 0$, so that $\int u^\star \mathrm{d}F = \lim_{\varepsilon \downarrow 0} \int u^\varepsilon \mathrm{d}F^\varepsilon \geq \lim_{\varepsilon \downarrow 0} \int u^\varepsilon \mathrm{d}F^\star = \int u^\star \mathrm{d}F^\star$, where the inequality follows from the definition of F^ε, since $F^\star$ is feasible given $F_0^\star$. Since $F^\star$ is uniquely optimal for $u^\star$ given $F_0^\star$, it follows that $F = F^\star$. Similarly for v^ε and G^ε.

Proof of the claim. If H_x is not positive (negative) semi-definite, then the same is true of H_y for all y in an open convex neighbourhood of x, as $y \mapsto H_y$ is continuous. By the case-2 hypothesis, H_y is negative (positive) semi-definite for all y in this neighbourhood. So u is locally concave (convex) at x. By strong regularity, u must be locally *strictly* concave (convex) at x. ■

Since u is strongly regular and (by hypothesis) not strictly convex, it is not convex, so there is an $x \in \text{int}(E)$ at which H_x is not positive semi-definite. By the claim, u is locally strictly concave at x. Let T be the hyperplane in $\mathbf{R}^{n+1}$ tangent to the graph of u at x. Since u is not concave (it is strongly regular, and by hypothesis not strictly concave), we may choose x so that T intersects the graph of u at some $y \in \text{int}(E) \setminus \{x\}$. Since u is locally strictly concave at x and continuous, we may choose y so that T does not intersect the graph of u on $\text{co}(\{x, y\}) \setminus \{x, y\}$, where '$\text{co}(\cdot)$' denotes the convex hull.

Define $w : [0, 1] \to \mathbf{R}$ by $w(\alpha) \coloneqq u(\alpha x + (1 - \alpha)y)$ for each $\alpha \in [0, 1]$. By Lemma 7 (appendix C.2, p. 41), w is strictly convex on an open interval that contains 0. Hence, after replacing x with a nearby point if necessary, we may assume without loss that $w''(0) > 0$. Since w coincides with the restriction of u to $\text{co}(\{x, y\})$, it follows that H_y is not negative semi-definite, so that u is locally strictly convex at y by the claim.

Let $t : E \to \mathbf{R}$ be the map having graph T, and let $p \coloneqq \max\{u, t\}$. Assume that $n = 2$; this is without loss, as it amounts to replacing E by its intersection E' with a two-dimensional affine space containing x and y, and the v and F_0 constructed below (with domain E') can easily be extended to E. Since u is locally strictly concave (convex) at x (at y) and $t \geq u$ on $\text{co}(\{x, y\}) \setminus \{x, y\}$, replacing E by a convex two-dimensional subset containing $\text{co}(\{x, y\})$ if necessary, we may without loss assume that p is convex and that there is a convex open set $A \ni x$ such that p is affine on A, $p = u$ on $E \setminus A$, and both A and $E \setminus A$ are Lebesgue-non-null. Clearly we may choose a strongly regular $v : E \to \mathbf{R}$ that is coarsely more convex than u and an $x' \in \text{int}(E)$ such that, letting $p' \coloneqq \max\{v, t'\}$ where t' is the map $E \to \mathbf{R}$ whose graph equals the plane tangent to v at x', both of the following hold:

- p' is convex, p' is affine on an open convex set $A' \ni x'$ such that $A' \setminus A$ is Lebesgue-non-null, and $p' = v$ on $E \setminus A'$.
- There exists a distribution F_0 with full support, a density with respect to the Lebesgue measure on $\mathbf{R}^2$, and $\int_A z F_0(\mathrm{d}z) / \int_A F_0(\mathrm{d}z) = x$ and $\int_{A'} z F_0(\mathrm{d}z) / \int_{A'} F_0(\mathrm{d}z) = x'$.

Let F (F') pool states in A (in A') and reveal all other states. By Theorem 5 in Dworczak and Kolotilin (2022), F (F') is optimal for u (for v)

given F_0; by Lemma 9, *uniquely* optimal. Since $A' \setminus A$ is F_0-non-null, F' is not more informative than F. Hence ($\star\star$) fails. ■

Appendix I Proof of Proposition 5 (p. 19)

For $x, y \in \mathbf{R}$ and $\alpha \in [0,1]$, write $x_\alpha y \coloneqq \alpha x + (1-\alpha)y$. Define $u, v : [0,1] \to \mathbf{R}$ by $u(x) \coloneqq U_S(A(x), x)$ and $v(x) \coloneqq U_R(A(x), x)$ for each $x \in [0,1]$. Choose any $x < y$ in $[0,1]$ such that $u(x_\beta y) \leq u(x)_\beta u(y)$ for every $\beta \in (0,1)$, and fix an $\alpha \in (0,1)$. Note that $v(x_\alpha y) \leq v(x)_\alpha v(y)$ since v is convex (as A is U_R-optimal). Thus

$$\begin{aligned}\Phi\big(u(x_\alpha y), v(x_\alpha y), x_\alpha y\big) &\leq \Phi\big(u(x_\alpha y), v(x)_\alpha v(y), x_\alpha y\big) \\ &\leq \Phi\big(u(x)_\alpha u(y), v(x)_\alpha v(y), x_\alpha y\big) \\ &\leq \Phi\big(u(x), v(x), x\big)_\alpha \Phi\big(u(y), v(y), y\big),\end{aligned}$$

where the first inequality holds since $\Phi\big(u(x_\alpha y), \cdot, x_\alpha y\big)$ is increasing, the second holds since $\Phi\big(\cdot, v(x)_\alpha v(y), x_\alpha y\big)$ is (strictly) increasing, and the final inequality holds since Φ is convex. Moreover, the second inequality is strict if $u(x_\alpha y) < u(x)_\alpha u(y)$, as $\Phi\big(\cdot, v(x)_\alpha v(y), x_\alpha y\big)$ is strictly increasing. ■

Appendix J Relation to the theory of monotone comparative statics

In this appendix, we discuss in detail how our results relate to the general theory of monotone comparative statics.

J.1 The literature

The general comparative-statics literature (e.g. Topkis, 1978; Milgrom & Shannon, 1994; Quah & Strulovici, 2009) asks, for any problem in which an agent chooses an action a from a partially ordered set $\mathcal{A}$, what shifts of the agent's objective function $U : \mathcal{A} \to \mathbf{R}$ lead her optimally to choose a higher action.[54] This framework nests the persuasion problem, in which the sender's action is a distribution F drawn from the set of all distributions feasible given the prior F_0, ordered by 'less informative than', and her objective function is

[54] A detail: the literature actually restricts attention to action sets $\mathcal{A}$ whose partial order has a *lattice* structure. This proviso is satisfied in the persuasion model (see appendix A). This technicality has no bearing on the discussion in this appendix.

$U(F) := \int u \mathrm{d}F$. When u shifts, the sender's objective function U changes, and our question is which such shifts lead to more informative choices.

Topkis (1978) and Milgrom and Shannon (1994) ask what shifts of U yield an increase of the optimal choices $\arg\max_{a \in A} U(a)$ for *any* constraint set $A \subseteq \mathcal{A}$.[55] Quah and Strulovici (2009, 2007) describe weaker conditions which ensure that $\arg\max_{a \in A} U(a)$ increases for any *order interval* constraint set $A = [\underline{a}, \bar{a}] \subseteq \mathcal{A}$.[56] The persuasion model has an order interval constraint set, namely $[\delta, F_0]$, where F_0 is the prior distribution and δ denotes the point mass concentrated at the mean of F_0.

To allow for the possibility of multiple optimal actions, one must extend the notion of 'lower than' from actions to *sets* of actions. In this paper, we have used the natural extension (defined at the end of §2.1), which is called the *weak set order (WSO)* in the literature. Most of the literature concerns itself with the *strong set order (SSO)*. The *sense* in which the SSO is stronger is difficult to interpret, suggesting that the gap between the two set orders is a technical artefact without economic substance.[57] Our reading of the literature is that the SSO is used *despite* its uninterpretable extra strength, on the (reasonable) grounds that it yields a fruitful theory.[58] The WSO/SSO distinction will become important toward the end of our discussion below.

The general theory features two classes of results and corresponding properties. The first concern 'encouragement' properties such as increasing/single-crossing differences and interval dominance, which capture the idea that one objective function is relatively more keen on higher actions than another objective function. Such 'encouragement' properties characterise 'non-decreasing' comparative statics: 'encouraging' shifts of the objective function *do not strictly decrease* optimal choices, and there is a converse.[59]

The second kind of result in the literature introduces 'complementarity' assumptions such as (quasi-)supermodularity and I-quasi-supermodularity. 'Complementary' objective functions U are those such that increasing one 'dimension' of the action makes the agent keener to increase other 'dimensions'.[60] When the objective shifts in an 'encouraging' way *and* either the old or the new objective exhibits 'complementarity', optimal choices *increase;*

[55] Actually, precisely: any non-empty *sublattice* $A \subseteq \mathcal{A}$.

[56] Recall that $[\underline{a}, \bar{a}] := \{a \in \mathcal{A} : \underline{a} \lesssim a \lesssim \bar{a}\}$, where $\lesssim$ denotes the partial order on $\mathcal{A}$.

[57] See Theorem 1 in Che, Kim and Kojima (2021) for a characterisation of the gap.

[58] In particular, it yields *necessity* as well as sufficiency in the comparative-statics theorems of Milgrom and Shannon (1994) and Quah and Strulovici (2009).

[59] This fact is dimly known in the literature, but rarely written down. Exceptions include Quah and Strulovici (2007, Proposition 5) and Anderson and Smith (2024).

[60] This gloss is exact if actions are ordered by a product order, such as the usual inequality on $\mathbf{R}^n$. Beyond product orders, the 'dimensions' language is purely analogical.

and there is a converse.

J.2 Our theorems in context

Our Theorem 1 is a result of the first kind: it identifies the correct 'encouragement' property for u, namely 'coarsely less convex than', which characterises 'non-decreasing' comparative statics for the persuasion model. The proof of the sufficiency half of Theorem 1 is the one place where we are able to use a result from the literature: we (i) show that if u is coarsely less convex than v, then $U(F) = \int u \mathrm{d}F$ is interval-dominated by $V(F) = \int v \mathrm{d}F$, and then (ii) invoke Quah and Strulovici's (2007) Proposition 5 to conclude that choices under u are not strictly higher than choices under v. (It turns out, however, that most of the action is in part (i): the argument there is fairly intricate, and exploits the special structure of the persuasion problem.) The necessity half of Theorem 1 is straightforward.

Having obtained Theorem 1, we next seek a result of the second kind, which identifies a further condition on u or v under which if u is coarsely less convex than v, then *less* informative choices are made under u than under v. The literature is of no help here, because the objective $U(F) = \int u \mathrm{d}F$ satisfies no 'complementarity' property except in trivial cases (e.g. if u is concave).

We must therefore strike out on our own, by asking for comparative statics in the (more natural) weak set order. This turns out to be fruitful: it delivers our Theorem 2, which describes a non-trivial property whose satisfaction by u is necessary and sufficient for comparative statics to hold between u and any v that is coarsely more convex. Both the sufficiency and necessity parts of our proof rely on novel arguments that exploit the structure of the persuasion model.

From the perspective of the comparative-statics literature, Theorem 2 may be viewed as a proof of concept: 'increasing' comparative statics in the *weak* set order can sometimes be had even though the literature's 'complementarity' assumptions fail. This matters because in our experience, 'complementarity' tends to fail in economic models, severely limiting the applicability of the existing theory. (There is one important exception: in applications with totally ordered actions (e.g. real numbers), 'complementarity' holds automatically.)

Appendix K Tightness of Lemma 1 (p. 6)

Lemma 1 is nearly tight, in the following sense:

Partial converse of Lemma 1. If $\Phi : \mathbf{R} \times [0,1] \to \mathbf{R}$ is such that for every upper semi-continuous $u : [0,1] \to \mathbf{R}$, u is coarsely less convex than $x \mapsto \Phi(u(x), x)$, then Φ must be convex on $\mathbf{R} \times (0,1)$ with $\Phi(\cdot, x)$ increasing for every $x \in (0,1)$.

This partial converse is implied by the following result, which closes the small gap between Lemma 1 and its converse by giving an exact characterisation of coarse-convexity-increasing transformations Φ. This result has other useful consequences, such as the fact (used in §7.6) that u is coarsely less convex than $\max\{u, \psi\}$ whenever $\psi : [0,1] \to \mathbf{R}$ is strictly convex.

Lemma 1*. For a map $\Phi : \mathbf{R} \times [0,1] \to \mathbf{R}$, the following are equivalent:

(i) For every $u : [0,1] \to \mathbf{R}$, u is coarsely less convex than $x \mapsto \Phi(u(x), x)$.

(ii) For every upper semi-continuous $u : [0,1] \to \mathbf{R}$, u is coarsely less convex than $x \mapsto \Phi(u(x), x)$.

(iii) For any $x < y$ in $[0,1]$, $\alpha \in (0,1)$ and $a, b, c \in \mathbf{R}$ such that $c \leq (<) \alpha a + (1-\alpha) b$, we have $\Phi(c, \alpha x + (1-\alpha) y) \leq (<) \alpha \Phi(a, x) + (1-\alpha) \Phi(b, y)$.

For the proof, we write $a_\alpha b \coloneqq \alpha a + (1-\alpha) b$ for $a, b \in \mathbf{R}$ and $\alpha \in [0,1]$.

Proof of the partial converse of Lemma 1. By Lemma 1*, it suffices to show that property (iii) implies that Φ is convex on $\mathbf{R} \times (0,1)$ and that $\Phi(\cdot, x)$ is increasing for each $x \in (0,1)$. So let Φ satisfy (iii), and note that it follows that for each $c \in \mathbf{R}$, $\Phi(c, \cdot)$ is convex, hence continuous on $(0,1)$.

For convexity, property (iii) immediately implies that $\Phi(\alpha(a, x) + (1-\alpha)(b, y)) \leq \Phi(a, x)_\alpha \Phi(b, y)$ for any $\alpha \in (0,1)$ and any $(a, x), (b, y) \in \mathbf{R} \times [0,1]$ such that $x \neq y$. To show that the same holds when $x = y = z \in (0,1)$, (in other words, that $\Phi(\cdot, z)$ is convex for each $z \in (0,1)$) observe that for any $x \in (0, z)$ and $y \in (z, 1)$ such that $x_\alpha y = z$, we have $\Phi(a_\alpha b, z) \leq \Phi(a, x)_\alpha \Phi(b, y)$, so letting $x, y \to z$ yields $\Phi(a_\alpha b, z) \leq \Phi(a, z)_\alpha \Phi(b, z)$ by continuity.

For monotonicity, take any $z \in (0,1)$ and $c < a$ in $\mathbf{R}$; we must show that $\Phi(c, z) \leq \Phi(a, z)$. For any $x \in (0, z)$ and $y \in (z, 1)$ such that $\frac{1}{2}x + \frac{1}{2}y = z$, property (iii) implies $\Phi(c, z) < \frac{1}{2}\Phi(a, x) + \frac{1}{2}\Phi(a, y)$, which as $x, y \to z$ yields $\Phi(c, z) \leq \Phi(a, z)$ by continuity. ■

Proof of Lemma 1.* (iii) implies (i) since for any $u : [0,1] \to \mathbf{R}$ and any $x < y$ in $[0,1]$ such that $u(x_\beta y) \leq u(x)_\beta u(y)$ for every $\beta \in (0,1)$, property (iii) (with $a \coloneqq u(x)$, $b \coloneqq u(y)$ and $c \coloneqq u(x_\alpha y)$) implies for each $\alpha \in (0,1)$ that $\Phi(u(x_\alpha y), x_\alpha y) \leq \Phi(u(x), x)_\alpha \Phi(u(y), y)$, with strict inequality if $u(x_\alpha y) <$

$u(x)_\alpha u(y)$. (i) immediately implies (ii). To show that (ii) implies (iii), we prove the contra-positive: let Φ violate (iii), meaning that there are $x < y$ in $[0,1]$, $\alpha \in (0,1)$ and $a, b, c \in \mathbf{R}$ such that either

(1) $c \leq a_\alpha b$ and $\Phi(c, x_\alpha y) > \Phi(a, x)_\alpha \Phi(b, y)$, or

(2) $c < a_\alpha b$ and $\Phi(c, x_\alpha y) \geq \Phi(a, x)_\alpha \Phi(b, y)$.

To show that (ii) fails, define $u : [0,1] \to \mathbf{R}$ by $u \coloneqq a$ on $[0, x]$, $u(x_\alpha y) \coloneqq c$, $u \coloneqq b$ on $[y, 1]$ and $u \coloneqq \min\{a, b, c\} - 1$ on $(x, x_\alpha y) \cup (x_\alpha y, y)$. Clearly u is upper semi-continuous. We have $u(x_\beta y) \leq u(x)_\beta u(y)$ for every $\beta \in (0,1)$, with strict inequality at $\beta = \alpha$ in case (2), and furthermore $\Phi(u(x_\alpha y), x_\alpha y) \geq \Phi(u(x), x)_\alpha \Phi(u(y), y)$, with strict inequality in case (1). Thus u is not coarsely less convex than $x \mapsto \Phi(u(x), x)$. ■

Appendix L Extension: specific shifts

In this appendix, we show that the crater property remains necessary for 'increasing' comparative statics when attention is confined to shifts of the sender's interim payoff u that are more specific than coarse-convexity shifts: in particular, conventional increased convexity and adding a convex function.

Proposition 6. Let $u : [0,1] \to \mathbf{R}$ be regular. The following are equivalent:

(i) u satisfies the crater property.

(ii) For any regular $v : [0,1] \to \mathbf{R}$ such that $v = \phi \circ u$ for some convex and strictly increasing $\phi : \mathbf{R} \to \mathbf{R} \cup \{\infty\}$, ($\star\star$) holds for every atomless convex-support distribution F_0.

(iii) For any regular $v : [0,1] \to \mathbf{R}$ such that $v = u + \psi$ for some convex $\psi : [0,1] \to \mathbf{R}$, ($\star\star$) holds for every atomless convex-support distribution F_0.

Proof. (i) implies (ii) and (iii) by Corollary 1 and Theorem 2 (pp. 7 and 10).

To show that (iii) implies (i), we shall prove the contra-positive, by arguing that in the proof of the necessity half of Theorem 2 (appendix C.1), v can be chosen so that $v - u$ is convex. We shall focus on Case 1 (the argument for Case 2 is similar). Since u is regular, we may choose a regular $w : [0,1] \to \mathbf{R}$ such that $w = u$ on $[X, 1]$ and, on each sub-interval of $[0, X]$ on which u is convex (concave), $w - u$ is affine (w is affine). Note that $w - u$

is convex, and that w is convex on $[0, X]$. Fix any $\chi : [0,1] \to \mathbf{R}$ that is continuously differentiable with bounded derivative, is strictly convex on $[0, X]$, and vanishes on $[X, 1]$. Then $v \coloneqq w + \chi$ weakly exceeds u, is strictly convex on $[0, X]$, and coincides with u on $[X, 1]$; and evidently $v - u$ is convex.

To show that (ii) implies (i), we shall modify the proof in appendix C.1 of the necessity half of Theorem 2. We again focus on Case 1 (Case 2 is similar). By replacing x and x' (w and w') with larger (smaller) values if necessary, we can ensure that $u(x) \neq u(w)$, without loss $u(x) < u(w)$, that $X \in (y, z)$, and that for some $z' \in (z, w)$, u is strictly increasing and strictly concave on $[z', w']$ and $\max_{[x',z']} u = u(z')$. Fix an $\varepsilon \in (0, \min\{u(w) - u(z'), 1\})$, and choose a $\phi : \mathbf{R} \to \mathbf{R}$ that is strictly increasing, continuously differentiable, equal to the identity on $(-\infty, u(w) - \varepsilon^2)$, affine on $(u(w) - \varepsilon^2/2, \infty)$, and satisfies $\phi(u(w) - \varepsilon^2/2) = \phi(u(w) - \varepsilon^2) + \varepsilon$. Then $v \coloneqq \phi \circ u$ equals u on $[x', z']$, and satisfies $v(w) > u(w)$ and $v'(w) > u'(w)$. Moreover, $[v(w) - u(w)]/[v'(w) - u'(w)]$ vanishes as $\varepsilon \downarrow 0$. Hence for sufficiently small ε, the tangent to v at w is steeper than the tangent to u at w, and the tangents cross in (z, w). Moreover, the former tangent approaches the latter as ε vanishes. Thus (recalling the properties of p and F_0) for sufficiently small ε, there exists a function $q : [0,1] \to \mathbf{R}$, an $x^\star \in (x, y)$, a $X^\star \in (X, z)$ and a $w^\star \in (z, w)$ such that q is affine on $[x', X^\star]$ and on $[X^\star, w']$, weakly exceeds v on $[x', w']$, is tangent to v at $x^\star$ and at $w^\star$, and satisfies

$$\frac{1}{F_0(X)} \int_0^{X^\star} \xi F_0(\mathrm{d}\xi) = x^\star \quad \text{and} \quad \frac{1}{1 - F_0(X)} \int_{X^\star}^1 \xi F_0(\mathrm{d}\xi) = w^\star.$$

Then the distribution F that reveals only whether the state exceeds $X^\star$ is optimal for v (by the argument in footnote 12, p. 12). Since $X^\star \neq X$, F pools states on either side of X, so ($\star\star$) fails. ■

Appendix M Extension: constrained persuasion

In this appendix, we extend our analysis to encompass constraints on the sender's choice of signal, following the small but growing literature on constrained (or costly) persuasion.[61] We focus on two important types of constraint: *monotonicity* and *coarseness.* In the former case, the sender can use only monotone partitional signals; in the latter, she can use only signals that send at most K messages, for some $K \geq 2$.

[61] See e.g. Gentzkow and Kamenica (2014), le Treust and Tomala (2019) and Doval and Skreta (2024). Some of this work is surveyed by Kamenica, Kim and Zapechelnyuk (2021).

We ask whether comparative-statics conclusions can be drawn under assumptions weaker than those identified by Theorem 2 (p. 10). For both constraint types, the answer is 'no': the crater property remains necessary.

M.1 Monotone partitional signals

In many applied settings, information is provided via scores: the state space $[0,1]$ is partitioned into intervals, and all that is revealed about the realisation of the state is which interval is belongs to. Examples include ratings in online commerce, grades in academic settings, and credit scores. Such signals are called *monotone partitional*.

We call a distribution F *M-feasible (given F_0)* iff it is the posterior-mean distribution induced by some monotone partitional signal. As is well-known, a distribution F is M-feasible given an atomless F_0 iff it is feasible for F_0 and $[0,1)$ may be partitioned into intervals $[x,y)$ such that either (i) $F = F_0$ on $[x,y)$ or (ii) $F = F_0(x)$ on $[x,\mu)$ and $F = F_0(y)$ on $[\mu,y)$ where $\mu := \left[\int_x^y z F_0(\mathrm{d}z)\right] / \left[F_0(y) - F_0(x)\right]$. In other words, states are either fully revealed (case (i)) or pooled with *adjacent* states (case (ii)).

Proposition 7. Let $u : [0,1] \to \mathbf{R}$ be regular. If

$$\operatorname*{arg\,max}_{F \text{ M-feasible given } F_0} \int u \mathrm{d}F \quad \text{is lower than} \quad \operatorname*{arg\,max}_{F \text{ M-feasible given } F_0} \int v \mathrm{d}F$$

for every regular $v : [0,1] \to \mathbf{R}$ that is coarsely more convex than u and every atomless convex-support distribution F_0, then u satisfies the crater property.

Thus restricting the sender to using only M-feasible distributions does not permit comparative-statics conclusions to be drawn under any weaker assumptions on the interim payoff u: the crater property remains necessary.

Proposition 7 follows directly from the proof of the necessity half of Theorem 2 (sketched in §4.2 above) since by inspection, the feasible distributions F and G which appear in that argument are in fact M-feasible.

M.2 Coarse signals

In practice, communication is often coarse, with only a finite number of messages in use. This may be due to bounded rationality or information-processing costs, for example. Such coarseness can be modelled by constraining the sender to use only signals that send at most K messages, for some exogenous K (Aybas & Turkel, 2024; Lyu, Suen & Zhang, 2023).

A distribution F is the posterior-mean distribution induced by a signal satisfying this constraint if and only if F is feasible given F_0 and has $|\text{supp}(F)| \leq K$. We call such distributions *K-feasible (given F_0).*

Proposition 8. Let $u : [0,1] \to \mathbf{R}$ be regular, and fix any $K \geq 2$. If

$$\underset{F\ K\text{-feasible given } F_0}{\arg\max} \int u \mathrm{d}F \quad \text{is lower than} \quad \underset{F\ K\text{-feasible given } F_0}{\arg\max} \int v \mathrm{d}F \qquad (\star_K)$$

for every regular $v : [0,1] \to \mathbf{R}$ that is coarsely more convex than u and every atomless convex-support distribution F_0, then u satisfies the crater property.

Sketch proof. We focus on the generic case in which optimal distributions are unique. We will show that with a small addition, the proof of the necessity half of Theorem 2 (sketched in §4.2 above) remains applicable. The argument there shows that if a regular $u : [0,1] \to \mathbf{R}$ violates the crater property, then there is a prior distribution F_0 and a coarsely more convex, regular and S-shaped $v : [0,1] \to \mathbf{R}$ such that the distribution G that is uniquely optimal for u given F_0 is binary, and is not less informative than the distribution F that is uniquely optimal for v given F_0. Since G is binary, it is K-feasible, so

$$\underset{H\ K\text{-feasible given } F_0}{\arg\max} \int u \mathrm{d}H = \{G\}.$$

Since v is S-shaped, we have by Proposition 8 in Lyu, Suen and Zhang (2023) that

$$\underset{H\ K\text{-feasible given } F_0}{\arg\max} \int u \mathrm{d}H = \left\{F^\dagger\right\}$$

for a distribution $F^\dagger$ that is less informative than F. Then G is not less informative than $F^\dagger$, so $(\star_K)$ fails. ■

References

Aliprantis, C. D., & Border, K. C. (2006). *Infinite dimensional analysis: A hitchhiker's guide* (3rd). Springer.

Alonso, R., & Câmara, O. (2016). Persuading voters. *American Economic Review*, *106*(11), 3590–3605. https://doi.org/10.1257/aer.20140737

Anderson, A., & Smith, L. (2024). The comparative statics of sorting. *American Economic Review*, *114*(3), 709–51. https://doi.org/10.1257/aer.20210890

Arieli, I., Babichenko, Y., Smorodinsky, R., & Yamashita, T. (2023). Optimal persuasion via bi-pooling. *Theoretical Economics*, *18*(1), 15–36. https://doi.org/10.3982/TE4663

Aumann, R. J., & Maschler, M. B. (1968/1995). *Repeated games with incomplete information* [circulated 1966–68, published 1995]. MIT Press.

Aybas, Y. C., & Turkel, E. (2024). *Persuasion with coarse communication* [working paper, 9 May 2024]. https://doi.org/10.48550/arXiv.1910.13547

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political Economy*, *81*(4), 813–846. https://doi.org/10.1086/260084

Billingsley, P. (1995). *Probability and measure* (3rd). Wiley.

Billingsley, P. (1999). *Convergence of probability measures* (2nd). Wiley.

Blackwell, D. (1951). Comparison of experiments. In J. Neyman (Ed.), *Berkeley symposium on mathematical statistics and probability* (pp. 93–102, Vol. 2). University of California Press.

Brocas, I., & Carrillo, J. D. (2007). Influence through ignorance. *RAND Journal of Economics*, *38*(4), 931–947. https://doi.org/10.1111/j.0741-6261.2007.00119.x

Chambers, C. P., Liu, C., & Rehbeck, J. (2020). Costly information acquisition. *Journal of Economic Theory*, *186*. https://doi.org/10.1016/j.jet.2019.104979

Che, Y.-K., Kim, J., & Kojima, F. (2021). *Weak monotone comparative statics* [working paper, 25 Nov 2021]. https://doi.org/10.48550/arXiv.1911.06442

Denti, T. (2022). Posterior separable cost of information. *American Economic Review*, *112*(10), 3215–3259. https://doi.org/10.1257/aer.20211252

Dizdar, D., & Kováč, E. (2020). A simple proof of strong duality in the linear persuasion problem. *Games and Economic Behavior*, *122*, 407–412. https://doi.org/10.1016/j.geb.2020.04.017

Doval, L., & Skreta, V. (2024). Constrained information design. *Mathematics of Operations Research*, *49*(1), 78–106. https://doi.org/10.1287/moor.2022.1346

Dworczak, P., & Kolotilin, A. (2022). *The persuasion duality* [working paper, 2 Sep 2022]. https://doi.org/10.48550/arXiv.1910.11392

Dworczak, P., & Martini, G. (2019). The simple economics of optimal persuasion. *Journal of Political Economy*, *127*(5), 1993–2048. https://doi.org/10.1086/701813

Gentzkow, M., & Kamenica, E. (2014). Costly persuasion. *American Economic Review: Papers & Proceedings*, *104*(5), 457–462. https://doi.org/10.1257/aer.104.5.457

Gentzkow, M., & Kamenica, E. (2016). A Rothschild–Stiglitz approach to Bayesian persuasion. *American Economic Review: Papers & Proceedings*, *106*(5), 597–601. https://doi.org/10.1257/aer.p20161049

Gitmez, A. A., & Molavi, P. (2023). *Informational autocrats, diverse societies* [working paper, 29 Aug 2023]. https://doi.org/10.48550/arXiv.2203.12698

Hardy, G. H., Littlewood, J. E., & Pólya, G. (1929). Some simple inequalities satisfied by convex functions. *Messenger of Mathematics*, *58*, 145–152.

Kamenica, E., & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, *101*(6), 2590–2615. https://doi.org/10.1257/aer.101.6.2590

Kamenica, E., Kim, K., & Zapechelnyuk, A. (2021). Bayesian persuasion and information design: Perspectives and open issues. *Economic Theory*, *72*, 701–704. https://doi.org/10.1007/s00199-021-01383-4

Karr, A. F. (1983). Extreme points of certain sets of probability measures, with applications. *Mathematics of Operations Research*, *8*(1), 74–85. https://doi.org/10.1287/moor.8.1.74

Kleiner, A., Moldovanu, B., & Strack, P. (2021). Extreme points and majorization: Economic applications. *Econometrica*, *89*(4), 1557–1593. https://doi.org/10.3982/ECTA18312

Kolotilin, A. (2014). *Optimal information disclosure: Quantity vs. quality* [working paper, Mar 2014; nominally superceded by Kolotilin (2018)]. https://doi.org/10.2139/ssrn.2302448

Kolotilin, A. (2018). Optimal information disclosure: A linear programming approach. *Theoretical Economics*, *13*(2), 607–635. https://doi.org/10.3982/TE1805

Kolotilin, A., Corrao, R., & Wolitzky, A. (2024). *Persuasion and matching: Optimal productive transport* [working paper, 7 Nov 2023]. https://doi.org/10.48550/arXiv.2311.02889

Kolotilin, A., Mylovanov, T., & Zapechelnyuk, A. (2022). Censorship as optimal persuasion. *Theoretical Economics*, *17*(2), 561–585. https://doi.org/10.3982/TE4071

Kolotilin, A., Mylovanov, T., Zapechelnyuk, A., & Li, M. (2017). Persuasion of a privately informed receiver. *Econometrica*, *85*(6), 1949–1964. https://doi.org/10.3982/ECTA13251

Kreutzkamp, S. (2023). *Endogenous information acquisition in cheap-talk games* [working paper, 18 Jan 2023].

le Treust, M., & Tomala, T. (2019). Persuasion with limited communication capacity. *Journal of Economic Theory*, *184*, 1–44. https://doi.org/10.1016/j.jet.2019.104940

Lyu, Q., Suen, W., & Zhang, Y. (2023). *Coarse information design* [working paper, 9 Oct 2023]. https://doi.org/10.48550/arXiv.2305.18020

Mariotti, T., Schweizer, N., Szech, N., & von Wangenheim, J. (2023). Information nudges and self-control. *Management Science*, *69*(4), 1935–2545. https://doi.org/10.1287/mnsc.2022.4428

Mensch, J., & Malik, K. (2023). *Posterior-mean separable costs of information acquisition* [working paper, 12 Dec 2023]. https://doi.org/10.48550/arXiv.2311.09496

Mensch, J., & Ravid, D. (2022). *Monopoly, product quality, and flexible learning* [working paper, 8 Jun 2022]. https://doi.org/10.48550/arXiv.2202.09985

Milgrom, P., & Shannon, C. (1994). Monotone comparative statics. *Econometrica*, *62*(1), 157–180. https://doi.org/10.2307/2951479

Phelps, R. R. (2001). *Lectures on Choquet's theorem* (2nd). Springer.

Quah, J. K.-H., & Strulovici, B. (2007). *Comparative statics with the interval dominance order: Some extensions* [working paper].

Quah, J. K.-H., & Strulovici, B. (2009). Comparative statics, informativeness, and the interval dominance order. *Econometrica*, *77*(6), 1949–1992. https://doi.org/10.3982/ECTA7583

Ravid, D., Roesler, A.-K., & Szentes, B. (2022). Learning before trading: On the inefficiency of ignoring free information. *Journal of Political Economy*, *130*(2), 346–387. https://doi.org/10.1086/717350

Rayo, L., & Segal, I. (2010). Optimal information disclosure. *Journal of Political Economy*, *118*(5), 949–987. https://doi.org/10.1086/657922

Shaked, M., & Shanthikumar, J. G. (2007). *Stochastic orders*. Springer.

Sinander, L. (2022). The converse envelope theorem. *Econometrica*, *90*(6), 2795–2819. https://doi.org/10.3982/ECTA18119

Srivastava, S. M. (1998). *A course on Borel sets*. Springer.

Sun, J., Schram, A. J. H. C., & Sloof, R. (2024). *Publicly persuading voters: The single-crossing property and the optimality of interval revelation* [working paper, 15 Apr 2024]. https://doi.org/10.2139/ssrn.4028840

Thereze, J. (2023a). *Adverse selection and endogenous information* [working paper, 13 Jan 2023].

Thereze, J. (2023b). *Screening costly information* [working paper, 31 Dec 2023].

Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations Research*, *26*(2), 305–321. https://doi.org/10.1287/opre.26.2.305

Warga, J. (1972). *Optimal control of differential and functional equations*. Academic Press.

Whitmeyer, M. (2024). *Making information more valuable* [working paper, 18 Dec 2024]. https://doi.org/10.48550/arXiv.2210.04418

Xu, R. (2024). *Persuasion, delegation, and private information in algorithm-assisted decisions* [working paper, 21 Feb 2024]. https://doi.org/10.48550/arXiv.2402.09384

Yoder, N. (2022). Designing incentives for heterogeneous researchers. *Journal of Political Economy*, *130*(8), 2018–2054. https://doi.org/10.1086/720072