Adversarial forecasters, surprises and randomization

Roberto Corrao[*] Drew Fudenberg[†] David Levine[‡]
MIT MIT EUI and WUSTL

September 15, 2022

Abstract

In an *adversarial forecaster* model, utility over lotteries is the sum of an expected utility function and a "suspense function" measuring how surprising the outcome is given the forecast made by an adversarial forecaster who attempts to find the forecast that minimizes the surprise. We show that an adversarial forecaster model gives rise to preferences that are concave and satisfy a form of differentiability condition, and that any preference relation that has a concave representation that satisfies the differentiability condition arises from an adversarial forecaster model. Because of concavity, the agent typically prefers to randomize. We characterize the support size of optimally chosen lotteries, and how it depends on risk preferences. We also show that the induced preferences in some problems of Bayesian persuasion and resource allocation have an adversarial forecaster representation, so that our results apply there.

[*]Department of Economics, MIT, rcorrao@mit.edu
[†]Department of Economics, MIT, drew.fudenberg@gmail.com
[‡]Department of Economics, EUI and WUSTL, david@dklevine.com

We are grateful to Oğuzhan Çelebi, Simone Cerreia-Vioglio, Philip Dybvig, Joel Flynn, Mira Frick, Ryota Ijima, Giacomo Lanzani, Fabio Maccheroni, Massimo Marinacci, Stephen Morris, Pietro Ortoleva, Lorenzo Stanca, and Alexander Wolitzky for helpful comments and conversations, and NSF grant SES 1951056 and the Gordon B. Pye Dissertation Fellowship for financial support.

1 Introduction

Consider a specific type of lottery: an agent must choose one of their local sports team's matches to watch. They care only about whether their team wins or loses, and prefer to watch their team win for sure than lose for sure. Some theories of preferences over lotteries assume stochastic dominance or monotonicity, which implies that the agent's most preferred match is one where their team is guaranteed to win. But that would be a rather boring match, and the agent might prefer to watch a match where their team is favored but not guaranteed to win. For this reason, we might wish to reject the axiom of monotonicity. Similar considerations arise in political economy in the theory of expressive voting, in which people get utility from watching a political contest, and their utility is enhanced by participation. Just as in the theory of sports matches, some may prefer a more exciting contest, so even without strategic considerations turnout is likely to be higher when the polls show a close race (see for example Levine, Modica, and Sun [2021]).

Suppose, then, that the agent has a preference for being surprised, and that their overall utility is the sum of a utility from which team wins, and a measure of how surprising the outcome is. An outcome is surprising if it is difficult to forecast in advance, where a forecast is a probability distribution over outcomes that is chosen by an adversarial forecaster who attempts to minimize surprise. We refer to this minimized surprise as the suspense, and we assume that forecasting the true lottery always minimizes surprise. We refer to this as the *adversarial forecaster* representation.

We say that a preference has a *continuous local expected utility* if there is a linear functional, i.e. an expected utility, that continuously varies with the distribution considered and that "supports" the preference at each lottery. We show that an adversarial forecaster model gives rise to preferences that have continuous local expected utility. In fact this is the only restriction on preferences: any preference relation that has continuous local utility arises from an adversarial forecaster model. Notably, preferences with a local expected utility have a representation that is concave in probabilities, so a preference for surprise rationalizes stochastic choice.

The adversarial forecaster model lets us impose additional restrictions on preferences in a natural way through the surprise function. One important class of examples is when the surprise function corresponds to the widely used method of moments technique. We show that this results in a quadratic - and hence easy to analyze - utility function.

An important application of the idea of surprise to is to story-telling. We first consider one such application: a receiver who likes an exciting story. Our model of an exciting story is a more general version of Ely, Frankel, and Kamenica [2015]. In that model, the agent cared only about a particular kind of suspense and not at all about the outcome. In our model, the agent also cares about the state, so a sender that maximizes the receiver's total utility designs the initial distribution over states as well as how information is revealed. As in Ely, Frankel, and Kamenica [2015] we find that the optimal information policy for a given distribution over states does not depend on preferences over states. However, the optimal distribution over states does depend on the receiver's state preferences, and thus so does the chosen information policy.

To better understand preference for surprise and the extent of deliberate randomization of the agent, we turn to study the structure of optimal lotteries for the special case where the surprise function depends has a finite-dimensional parameterization, e.g. a function of a finite number of moments. We apply this to settings where the agent chooses a lottery subject to moment restrictions, such as that lottery's expected value equals to the endowment. We show that when the parameter space is finite dimensional, there is always an optimal lottery that has finite support. Specifically if the adversarial forecaster has k parameters with which to make forecasts, and there are m moment restrictions, then there is an optimal lottery with support of no more than $(k+1)(m+1)$ points. For example, in the sports case, suppose that preferences are not merely over which team wins or loses, but also over the score, where the latter can take on a continuum of values. If the forecaster is limited to predicting mean score and there are no moment constraints, then one most preferred choice is a binary lottery between two scores.

We then consider the more general class of preferences that are induced by a sequential game against an adversary with an arbitrary set of feasible actions. We show these preferences have an adversarial forecaster representation where the surprise function has weaker continuity properties. Moreover, we show that they admit an adversarial forecaster representation if and only if the adversary has a unique best response to each lottery.

We study the monotonicity properties of this more general model with respect to stochastic orders, and apply them to the question of how preference for surprise is reflected in attitudes towards risk. First we show that these preferences preserve a stochastic order if and only if, for every lottery, there is a best response of the adversary that induces a utility over outcomes that reflects the stochastic order. We then apply this result to stochastic orders capturing risk-aversion (i.e., the mean-preserving spread order) and higher-order risk aversion. In particular, we show how preferences for surprise may lead an agent with a risk-averse expected utility component to have preferences that are overall risk loving. We then show that an asymmetric version of the method of moments generates adversarial forecaster preferences that are consistent with *prudence* and, more in general, with experimental data on higher-order risk aversion. Our model of a persuasive story is based on the disclosure literature, as, for example, in Ben-Porath, Dekel, and Lipman [2018]. Our sender has some control over the environment, perhaps through the choice of the riskiness of a project. We show there is a preference for surprise in the choice of underlying riskiness, and give simple sufficient conditions for the optimal distribution to be binary, so the sender optimally induces a binary prior.

The disclosure model is an example of how adversarial forecasting preferences arise naturally when there is an initial stage in which the designer chooses a distribution that determines the feasible policies in the second stage. Another example is when a designer chooses a distribution of qualities of a good they then allocate to consumers. Here we give simple sufficient conditions for the planner to optimally produce only two qualities.

Related Work Our paper is related to three distinct types of decision theoretic models of the preferences over lotteries. Most directly it is related to models of agents with "as-if" adversaries, e.g. Maccheroni [2002], Cerreia-Vioglio [2009], Chatterjee and Krishna [2011], Cerreia-Vioglio, Dillenberger, and Ortoleva [2015], and Fudenberg, Iijima, and Strzalecki [2015] as well as to the Ely, Frankel, and Kamenica [2015] model where the adversary is left implicit. It is also related to models of agents with dual selves that are not directly opposed, as in Gul and Pesendorfer [2001] and Fudenberg and Levine [2006]. We also complement the analyses in Cerreia-Vioglio, Dillenberger, and Ortoleva [2020] and Loseto and Lucia [2021] of optimization problems with non expected utility preferences by characterizing the optimal lotteries with first-order conditions and bounding the size of their supports in more general environments.

Our work on induced preference is related to the study of induced preferences due to temporal risk, as in Machina [1984], where the agent chooses a lottery over outcomes and then chooses an action without observing the lottery's realization, which convex preferences over the first-stage choice. We show that when the lottery chosen in the first stage only affects the set of feasible policies in the second stage, such as in disclosure and allocation problems, the induced preference is concave instead of convex, which has very different implications for the designer's preferences for randomization. This lets us analyze first-stage choices in models inspired by Ben-Porath, Dekel, and Lipman [2018]'s analysis of disclosure and Loertscher and Muir [2022]'s work on nonlinear pricing.

Finally, our analysis of monotonicity is related to the work on stochastic orders that we discuss in Section 6.

2 The Adversarial Forecaster

This section introduces the *adversarial forecaster model*, in which the agent has preferences over lotteries of outcomes that depend on both the expected utility of the lottery's outcome and a measure of suspense.

2.1 The Model

The agent plays a sequential move game against an adversarial forecaster. The agent moves first, and chooses a lottery $F \in \mathcal{F}$, the set of Borel measures on a compact metric space X of outcomes. We endow $\mathcal{F}$ with the topology of weak convergence, which makes it metrizable and compact. Then the adversary observes F and chooses a *forecast* $\hat{F} \in \mathcal{F}$, that is, a probabilistic statement about how likely different outcomes are. We study the agent's preference over lotteries of outcomes that is induced by backward induction in this sequential game

Let δ_x denote the Dirac measure on x.

Definition 1. (i) We say that $\sigma : X \times \mathcal{F} \to \mathbb{R}_+$ is a *surprise function* if $\sigma(x, \delta_x) = 0$ for all $x \in X$, σ is jointly continuous, and if $\int \sigma(x, F) dF(x) \leqslant \int \sigma(x, \hat{F}) dF(x)$ for all $F, \hat{F} \in \mathcal{F}$.

(ii) The *suspense* of lottery F given the surprise function σ is $\Sigma(F) = \min_{\hat{F} \in \mathcal{F}} \int \sigma(x, \hat{F}) dF(x)$.

Definition 1 requires that there is no surprise when the realized outcome was predicted by the forecaster to have probability 1, and that the forecast F minimizes the expected surprise when the true lottery is F. A lottery's suspense is the least amount of expected surprise that an adversarial forecaster can attain given that lottery. For a standard example of surprise function take $X = \{0, 1\}$ and $\sigma(x, F) = \left(x - \int x dF(x)\right)^2$, so surprise is measured by mean-squared error. We illustrate this functional form in Example 1 below.

Let $C(X)$ denote the space of continuous real functions over X, endowed with the topology induced by the supnorm.

Definition 2. Preferences $\succsim$ have an *adversarial forecaster representation* if they can be represented by a function V satisfying

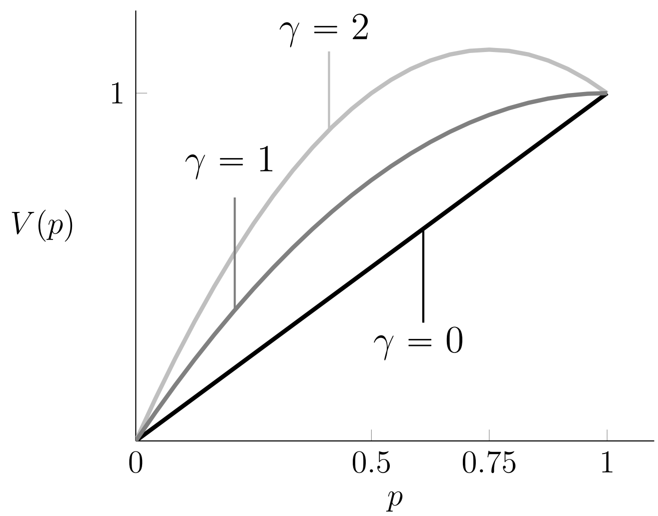
$$V(F) = \int v(x)dF(x) + \min_{\hat{F}\in\mathcal{F}} \int \sigma(x, \hat{F})dF(x), \tag{1}$$

where σ is a surprise function, and $v \in C(X)$ is an expected utility function.

This representation can be interpreted as follows: The agent has a baseline preference over outcomes described by the expected utility function utility v, and a preference for surprise captured by σ. Given a forecast $\hat{F}$ of the adversary, the agent's total utility is the sum of their expected baseline utility and their expected surprise. An outcome is surprising if it is hard for the adversarial forecaster to predict, so the adversary tries to minimize σ, hence the total utility from F is the sum of its baseline expected utility plus its suspense. Given this interpretation, we sometimes refer to σ as the *loss function* of the adversary. Equation 1 shows that V is continuous and concave, because it is the minimum over the collection of affine and continuous functionals.[1] Finally, note that the representation functional V must satisfy $V(\delta_x) = v(x)$ since $\min_{\hat{F}\in\mathcal{F}} \sigma(x, \hat{F}) = \sigma(x, \delta_x) = 0$.

Example 1. In a sports match, the outcome is $x = 1$ if the preferred team wins and $x = 0$ if it looses. Let p be the probability of winning, $\hat{F}$ be the forecast, and use $\gamma(x - \int \tilde{x} d\hat{F}(\tilde{x}))^2$ to measure the outcome's surprise given the forecast $\hat{F}$. The adversary tries to minimize expected surprise, as measured by the mean square error. The decision maker gets utility $v(x) = x$ plus γ times the squared error of the forecast, which induces expected utility preferences over (x, p) with utility function

[1] Continuity follows from the Berge maximum theorem since V is continuous and $\mathcal{F}$ is compact.



$V(p) = p + \gamma p(1-p)$

$x + \gamma(x-p)^2$. The adversary's optimal choice is to forecast p, and the expected surprise is just the variance $p(1-p)$, so the agent's preference over lotteries $\gtrsim$ is given by $V(p) = p + \gamma p(1-p)$. If $\gamma > 1$ and the agent can choose any value of p, the best lottery is $p = (1+\gamma)/(2\gamma)$, so that the preferred team might lose, while if $0 \leqslant \gamma \leqslant 1$ the best lottery is $p = 1$. $\triangle$

2.2 Continuous Local Expected Utility

Suppose that preferences can be represented by a continuous utility function V. We now define what it means for this function to be concave and to have a (jointly) continuous local expected utility, and show this is equivalent to the existence of an adversarial forecasting representation.

A concave function can be characterized by its supporting hyperplanes, and this is the approach we take. We say that $w \in C(X)$ is a *local expected utility* of V at F if it is a supporting hyperplane: that is, for every $\tilde{F} \in \mathcal{F}$, we have $\int w(x) d\tilde{F}(x) \geqslant V(\tilde{F})$ with $\int w(x) dF(x) = V(F)$. The function V has a *local expected utility* if there is at least one local expected utility at each F. Any function that has a local expected utility is concave.[2] Moreover, when V has a local expected utility w at F, if $\int w(x) dF(x) \geqslant$

[2]See e.g. Aliprantis and Border [2006] p. 264. Local utility, unlike concavity, requires there

$\int w(x)d\tilde{F}(x)$ (resp. $>$), then $F \gtrsim \tilde{F}$ (resp. $>$).[3] This property justifies the name we adopt for this supporting hyperplane.

We say that V has a *continuous local expected utility* if there is a local expected utility $w(x, F)$ that is jointly continuous. This does not imply that there is a unique local expected utility at a point: generally there will be a continuum of local expected utilities at boundary points.[4]

Theorem 1. *Let $\gtrsim$ be a preference over $\mathcal{F}$. The following are equivalent:*

(i) Preference $\gtrsim$ admits an adversarial forecaster representation.

(ii) Preference $\gtrsim$ has a representation V with a continuous local expected utility.

The proof of this and all other results is in the Appendix except where otherwise noted. Intuitively, if V has an adversarial forecaster representation, then $w(\cdot, F) = v + \sigma(\cdot, F)$ is a local expected utility of V. In turn, the joint continuity of σ implies that w is jointly continuous, yielding that V has a continuous local expected utility. Conversely, given a representation V, we can use its continuous local expected utility w_V to define the utility over outcomes and the surprise function by $v(x) = V(\delta_x)$ and $\sigma(x, F) = w(x, F) - v(x)$. Given that w is jointly continuous, it follows that V admits a representation as in Equation 1.

Preferences that admit an adversarial forecaster representation are concave, so typically the optimal lottery will be in the interior of the probability simplex and so exhibit a preference randomization. In particular, if X is an interval of real numbers, as in Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella [2019], and each local utility $w(\cdot, F)$ is strictly increasing, the induced stochastic choice satisfies their *Rational mixing* axiom. Also, Theorem 1 immediately implies that the Additive Perturbed Utility (APU) preferences of Fudenberg, Iijima, and Strzalecki [2015] (which are only defined for finite X) have an adversarial forecaster representation. The converse is not true, because choices generated by APU preferences satisfy Regularity, while adversarial preferences need not, as we show next.[5]

are supporting hyperplanes at boundary points. Machina [1982] uses a different definition that is neither weaker nor stronger than ours; see Online Appendix V.

[3]This follows from the concavity of V. See Online Appendix V for a formal proof.

[4]Boundary points are especially important in the infinite-dimensional case since with the topology of weak convergence all points are on the boundary.

[5]The stochastic choice function P satisfies *Regularity* if $P(x|\overline{X}) \leqslant P(x|\overline{X}')$ for all $x \in \overline{X}' \subseteq \overline{X}$.

Example 2. Suppose the agent's baseline utility v is concave and twice continuously differentiable, and that the agent has a preference for surprise given by $\sigma(x, F) = \left(x - \int \tilde{x} dF(\tilde{x})\right)^2$. The proof of Theorem 1 implies that this generates local utility $w(x, F) = v(x) + \left(x - \int_0^1 \tilde{x} dF(\tilde{x})\right)^2$. If $v(x) = x$, the stochastic choice rule induced by these preferences need not satisfy Regularity: The uniquely optimal choice for the agent from $\Delta(\{-1, 0\})$ is δ_0, while simple computations show that the optimal lottery in $\Delta(\{-1, 0, 1\})$ is $1/4\delta_{-1} + 3/4\delta_1$. Here outcome -1 is chosen with strictly positive probability only when it can be paired with outcome 1 so to generate enough suspense. For general v that are not too concave, i.e. when $v'' \geqslant -2$, the local utility is convex in x for all forecasts F. Theorem 5 below will show that this implies the agent weakly prefers any mean-preserving spread $\tilde{F}$ of F to F itself. We say more about the effect of surprise on risk aversion in Section 6. $\triangle$

2.3 Generalized Method of Moments

We now introduce a class of adversarial forecaster representations where the adversary's loss function only depends on a set of functions of the agent's chosen distribution. Suppose X is a closed bounded subset of $\mathbb{R}^m$, and let S be a compact metric space of *parameters* with the Borel sigma algebra. Given any integrable function $h : X \times S \to \mathbb{R}$, define $h(F, s) = \int h(x, s) dF(x)$ for all $s \in S$ and $F \in \mathcal{F}$. For a given h, we call the set $\{h(\cdot, s)\}_{s \in S} \subseteq C(X)$ the *generalized moments*. We assume here that the forecaster's objective is to choose a forecast $\hat{F}$ that minimizes a weighted sum of these generalized moments.

Definition 3. The function $\sigma : X \times \mathcal{F} \to \mathbb{R}_+$ is based on the *generalized method of moments* (GMM)[6] if there is a Borel probability space (S, μ) and a continuous function $h : X \times S \to \mathbb{R}$ such that

$$\sigma(x, \hat{F}) = \int \left(h(x, s) - h(\hat{F}, s)\right)^2 d\mu(s).$$

Proposition 1. *A function σ based on the generalized methods of moments forecast*

[6] We abuse terminology here; in econometrics, the generalized method of moments minimizes a quadratic loss function on the data under the constraint that a number of generalized moment restrictions are satisfied.

is a surprise function and the suspense is quadratic

$$\Sigma(F) = \int H(x,x)dF(x) - \int\int H(x,\tilde{x})dF(x)dF(\tilde{x})$$

where $H(x,\tilde{x}) = \int h(x,s)h(\tilde{x},s)d\mu(s)$. *If* μ *has full support and* $F \mapsto h(F,\cdot)$ *is one-to-one, then* $\Sigma(F)$ *and* $V(F)$ *are strictly concave.*

This shows that GMM surprise functions generate *quadratic utilities*. This is introduced in Machina [1982]. Chew, Epstein, and Segal [1991] shows that quadratic utilities satisfy *mixture symmetry*, a weakening of both independence and betweeness that is more consistent with experimental findings. As we show in Section 4, quadratic utilities are very tractable. Here are two classes of GMM surprise functions.

Finite Moments If $S = \{s_1, ..., s_m\}$ is a finite set of non-negative integers, we can take $h(x,s) = \prod_{i=1}^{m} x_i^{s_i}$, the standard method of moments.[7] The simplest case is the one with only the first moment, $S = \{1\}$, such as in Examples 1 and 2.

Moment Generating Function If for some $\tau > 0$ the parameter space is $S = [-\tau,\tau]^m$ we may take $h(x,s) = e^{s\cdot x}$. Here $h(F,s)$ is the moment generating function of F, where the map $F \mapsto h(F,\cdot)$ is one-to-one, so that the forecaster aims to match the entire distribution chosen by the agent. As Section 4 shows, the forecaster's loss function is a key determinant of how much the agent will choose to randomize.

3 Suspense and Surprise: Writing a Suspenseful Novel

Ely, Frankel, and Kamenica [2015] consider how the writer of a novel, or a sports broadcaster who knows the ending, can best reveal information about that outcome over time. The designer's objective is to maximize the utility of the watcher, who in turn has preferences for suspense and surprise. Here we use an example to show that their model of suspense can be interpreted as method of moments forecasting, and extend it in two ways: we allow the watcher to have preferences over realized outcomes, and let the broadcaster design *both* the distribution over states and the

[7] See for example Chapter 18 in Greene [2003].

information revealed over time, for example both the ending of the story and how the story unfolds.

We restrict attention to a binary state space $S = \{0, 1\}$, and let $X = S \times \Delta(S)$. Let $p \in \Delta(S) = [0, 1]$ denote an arbitrary belief over states with p the probability that $s = 1$, so an outcome is a pair $x = (s, p)$. There are two time periods and two agents, a watcher (W) and a broadcaster (B). The broadcaster chooses an initial distribution over states from a closed interval $\overline{\Delta} \subseteq [0, 1]$ and a signal structure with the goal of maximizing the watcher's total utility, which depends on the realized state and on the realized surprise over two periods. Instead of working directly with the signals, we represent them with distributions over posteriors: the broadcaster chooses a joint distribution $F \in \mathcal{F}$ over states and conditional beliefs of the watcher. The set of feasible joint distributions are those such that, conditional on the realization of the belief p, the induced conditional belief over S is equal to p itself:

$$\overline{\mathcal{F}} = \left\{F \in \mathcal{F} : marg_S F \in \overline{\Delta}, \forall p \in \Delta(S), F(\cdot|p) = p\right\}.$$

For every, $F \in \overline{\mathcal{F}}$, we let $p_F \in \overline{\Delta}$ denote the induced probability that $s = 1$ and let $F_\Delta \in \Delta([0, 1])$ denote the induced distribution over beliefs.[8]

In Period 0, B chooses a distribution p_F over S, and commits to a distribution over beliefs F_Δ such that $\int p dF_\Delta(p) = p_F$. In Period 1, W observes the signal realization, forms a posterior belief p, and their first-period surprise is realized, and in Period 2, W observes the state realization s and their second-period surprise is realized.

The surprise in period 1 given $F \in \overline{\mathcal{F}}$ is

$$V_1(F) = g\left(\int \frac{1}{2}||p - p_F||^2 dF_\Delta(p)\right) = g\left(\int_0^1 p^2 dF_\Delta(p) - p_F^2\right),$$

where $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable, strictly increasing, concave function, with $g(0) = 0$, which implies that V_1 has continuous local expected utility. The surprise in period 2 given $F \in \overline{\mathcal{F}}$ is

$$V_2(F) = \int g\left(\sum_{s\in S} \frac{1}{2}||\delta_s - p||^2 p(s)\right) dF_\Delta(p) = \int_0^1 g(p - p^2) dF_\Delta(p)$$

where δ_s represents the degenerate belief over s and V_2 is now a linear function of F_Δ

[8] In Ely, Frankel, and Kamenica, $\overline{\Delta} = \{p_0\}$, that is, B cannot affect the probability over states.

that does not depend on the original prior p_F. Finally, W gets direct utility equal to $\tilde{v} \in \mathbb{R}$ when the realized state is $s = 1$ and direct utility 0 when $s = 0$; the case $\tilde{v} = 0$ yields the preferences in Ely, Frankel, and Kamenica.[9]

B wants to maximize the total utility of W, that is, to solve

$$\max_{F \in \overline{\mathcal{F}}} p_F \tilde{v} + (1-\beta) g\left(\int_0^1 p^2 dF_\Delta(p) - p_F^2\right) + \beta \int_0^1 g(p - p^2) dF_\Delta(p). \tag{2}$$

where $\beta \in [0,1]$ captures the relative importance of surprises across periods. Let $V_\beta(F)$ denote the total utility of W defined in Equation 2. The discussion above shows V_β has a continuous local expected utility, so by Theorem 1 it admits an adversarial forecaster representation. The local utilities of V_β are:

$$w_\beta(s,p,F) = s\tilde{v} + (1-\beta) g'(D_2(F))(p^2 - p_F^2) + \beta g(p - p^2), \tag{3}$$

where $D_2(F) = \int \tilde{p}^2 dF_\Delta(\tilde{p}) - p_F^2$, and the baseline utility of W is $v_\beta(s,p) = V_\beta(\delta_{(s,p)}) = s\tilde{v} + \beta g(p-p^2)$ yielding a surprise function $\sigma_\beta(s,p,F) = (1-\beta)g'(D_2(F))(p^2 - p_F^2)$.

Because the total payoff of the watcher depends only on the marginals of F, we can think to the broadcaster as choosing only p_F and F_Δ given the consistency constraint. Next, we describe how the optimal marginals (p_F^*, F_Δ^*) depend on β.

Proposition 2. *For every $\beta \in [0,1]$, there exists an optimal distribution F_Δ^* supported on no more than three beliefs. Moreover, there exist $\underline{\beta}, \overline{\beta} \in (0,1)$ with $\underline{\beta} \leqslant \overline{\beta}$ such that*

1. *When $\beta \geqslant \overline{\beta}$, no disclosure is uniquely optimal (i.e., $F_\Delta^* = \delta_{p_F^*}$) and p_F^* is optimal if and only if it solves $\max_{p \in \overline{\Delta}} \{p\tilde{v} + \beta g(p - p^2)\}$.*

2. *When $\beta \leqslant \underline{\beta}$, full disclosure is uniquely optimal (i.e., $F_\Delta^* = (1 - p_F^*)\delta_0 + p_F^* \delta_1$) and p_F^* is optimal if and only if it solves $\max_{p \in \overline{\Delta}} \{p\tilde{v} + (1-\beta) g'(p-p^2)(p-p^2)\}$.*

The proof of this result is in Online Appenidx I. It is derived by computing the local expected utility of V_β at the candidate solution (p_F^*, F_Δ^*) and verifying that F_Δ^* is indeed optimal for that local expected utility. Because the state is binary, each local utility is a linear combination $g'(D_2(F))p^2$ and $g(p-p^2)$, where the first term is strictly convex and the second is strictly concave For example, if $g(d) = \sqrt{d}$, then

[9] Ely, Frankel, and Kamenica have two different preference specifications for W, capturing preferences for suspense and surprise respectively. In their specification for suspense, flow utility at t depends on the expected surprise in period $t+1$ given the belief at period t.

$g'(D_2(F))$ is very high for F such that F_Δ is concentrated around P_F, since in this case $D_2(F)$ is close to 0. Thus revealing no information cannot maximize V_β, since the local expected utility $w_\beta(s, p, F)$ is strictly convex in p. More generally, because W has nonlinear preferences over F_Δ, B might want to induce more than 2 posteriors, unlike in Bayesian persuasion with a binary state. Section 4 derives a more general result on the support size of optimal distributions.

In the linear case $g(d) = d$ with $\overline{\Delta} = [0, 1]$ we can completely characterize the solution. For every $F \in \overline{\mathcal{F}}$, the total payoff of the watcher simplifies to $V_\beta(F) = p_F(\tilde{v} + \beta) - p_F^2(1 - \beta) + \int(1 - 2\beta)p^2 dF_\Delta(p)$. The utility over realized posteriors $(1 - 2\beta)p^2$ is strictly concave when $\beta > 1/2$, so non- disclosure is uniquely optimal. When $\beta < 1/2$, this term is strictly convex, so full disclosure is uniquely optimal, and when $\beta = 1/2$, W is indifferent over all the information structures. For every value of β and $\tilde{v}$, $p_F^* = \max\left\{0, \min\left\{1, \frac{\tilde{v}+\max\{\beta,1-\beta\}}{2\max\{\beta,1-\beta\}}\right\}\right\}$: the broadcaster assigns a probability p_F^* to $s = 1$ that depends on the baseline value v as well as on the surprise parameter β. The nature of the optimal information structure between the two periods is always extreme (full or no-disclosure) and depends only on β. Observe that the disclosure policy in not affected by the baseline value v, hence with EFK preferences, that is with $\tilde{v} \equiv 0$, the optimal disclosure policy would be the same. However, the optimal probability of $s = 1$ would be $p^* = 1/2$, which is independent of the weight β. We give a more detailed analysis of the linear case in Online Appendix I .

Measures of uncertainty and information For an adversarial forecaster representation, the suspense function simplifies to $\Sigma(F) = V(F) - \int v(x)dF(x)$. The properties of V imply that Σ is concave and that $\Sigma(\delta_x) = 0$ for all $x \in X$. Frankel and Kamenica [2019] show that these are the two properties characterizing a valid measure of uncertainty, that is, a function representing the cost of uncertainty in a given decision problem. In our model, the coupled decision problem they consider is the forecasting problem faced by the adversary whose prior over outcomes coincides with the lottery F chosen by the agent. Because the Bregman divergence of Σ coincides with the surprise function σ, their Theorem 3 implies that the surprise function is what they call a valid measure of information: the amount of surprise generated by x given lottery F coincides with ex-post value for the adversary of observing the realized outcome as opposed to receiving no additional information.

4 Parametric adversarial forecasters

We turn now to the study of optimization problems with support restrictions and moment constraints, e.g. that the expected outcome must be constant across lotteries, as is the case with fair insurance.As a tool for this analysis, we consider a broader class of surprise functions. Recall that a GMM representation is defined by a probability space (S, μ) and a continuous function $h(F, s) = \int h(x, s) dF(x)$. We may define the space of moments as the image of the map $P(F) \equiv h(F, \cdot)$, that is, $Y \equiv P(\mathcal{F}) \subseteq \mathbb{R}^S$. When S is finite, Y is a subset of a Euclidean space. If we then define a *parametric surprise function* on Y by $\hat{\sigma}(x, y) = \int (h(x, s) - y(s))^2 \, d\mu(s)$, we see that the GMM surprise function $\sigma(x, F) = \hat{\sigma}(x, P(F))$ depends on F only through $P(F)$. This lets us work with the function $\hat{\sigma}(x, y)$ instead of $\sigma(x, F)$, which is easier to study since it is strictly concave and differentiable in $P(F)$. Parametric adversarial forecaster representations generalize these properties to other settings where surprise depends on the lottery only through a space Y of parameters.

Definition 4. A surprise function σ is *parametric* if there exist a set $Y \subseteq \mathbb{R}^m$, a continuous map $P : \mathcal{F} \to \mathbb{R}^m$, and a continuous function $\hat{\sigma} : X \times Y \to \mathbb{R}_+$ that is strictly concave and differentiable in y, such that $Y = P(\mathcal{F})$ and $\sigma(x, F) = \hat{\sigma}(x, P(F))$ for all $(x, F) \in X \times \mathcal{F}$.

When $\gtrsim$ has an adversarial forecaster representation with a parametric surprise function σ, we say that it has a *parametric representation.* In this case

$$V(F) = \min_{y \in Y} \int v(x) + \hat{\sigma}(x, y) dF(x).$$

and we let $\hat{y}(F)$ denote the parameter attaining the minimum.

We now apply the parametric method to the study of optimization problems with a *moment restriction*. We fix some closed (possibly finite) subset $\overline{X} \subseteq X$ and a finite collection of k continuous functions $\Gamma = \{g_1, ..., g_k\} \subseteq C(X)$ together with the feasibility set

$$\mathcal{F}_\Gamma \left(\overline{X}\right) = \left\{ F \in \Delta\left(\overline{X}\right) : \forall g_i \in \Gamma, \int g_i(x) \, dF(x) \leqslant 0 \right\}, \tag{4}$$

which we assume is non-empty. If x is money, then $\int x dF(x) = 0$ is the budget constraint that the agent may choose any fair lottery. Alternatively we can consider an agent that must guarantee a minimum level of realized utility for k external observers whose utilities depend on the outcome (e.g., investors on a project). Define $g_i(x) = m_i - v_i(x)$, where v_i is the utility of the i-th external observer and $m_i \geqslant 0$. Here the agent trades off preferences for surprise with the utility guarantees. Whenever Γ is empty, the feasibility set reduces to $\Delta(\overline{X})$, and the agent is able to choose all the lotteries over a restricted subset of outcomes.

Given a preference $\succsim$ with a parametric adversarial forecaster representation, a domain of outcomes $\overline{X} \subseteq X$, and a collection of relevant moments Γ, we study the optimization problem

$$\max_{F \in \mathcal{F}_\Gamma(\overline{X})} \min_{y \in Y} \int v(x) + \hat{\sigma}(x, y) dF(x). \tag{5}$$

The next result shows that when an adversarial forecaster representation is parametric, there is always a solution of this optimization problem whose support is a finite set of outcomes. Moreover, the upper bound on this finite number of outcomes only depends on the dimension of Y and on the number of moment restrictions defining the feasible set of lotteries.

Theorem 2. *Let $\succsim$ have a parametric representation with $Y \subseteq \mathbb{R}^m$, $\overline{X} \subseteq X$ be a closed set, and Γ a collection of k functions. There exists an optimal lottery for 5 that has finite support on no more than $(k+1)(m+1)$ points of $\overline{X}$.*

The theorem implies that with parametric preferences, for any lottery F there is a lottery $\tilde{F}$ with support of at most $m+1$ points with $\operatorname{supp} \tilde{F} \subseteq \operatorname{supp} F$ and $\tilde{F} \succsim F$. An implication is that if the adversary is a GMM forecaster who uses only m moments, the agent need place weight on no more than $m+1$ points. Its proof has several steps. Recall that each $F \in \mathcal{F}_\Gamma\left(\overline{X}\right)$ admits an integral representation with respect to the extreme points of $\mathcal{F}_\Gamma\left(\overline{X}\right)$.[10] The first step of the proof, Theorem 7 in the Appendix, establishes that each of the extreme points representing an optimal lottery F must satisfy a linear equation with parameters given by the optimal value $V^* \in \mathbb{R}$ of Problem 5 and the optimal action of the adversary $\hat{y}(F)$. The second step of the proof, Theorem 8 in the Appendix, involves an approximation argument. We start by arguing via the transversality theorem that, for every set of finitely many extreme

[10]Given any convex set $\overline{\mathcal{F}} \subseteq \mathcal{F}$, Choquet's theorem says that, for every $F \in \overline{\mathcal{F}}$, there exists $\lambda \in \Delta\left(ext\left(\overline{\mathcal{F}}\right)\right)$ such that $F = \int \tilde{F} d\lambda(\tilde{F})$.

points of $\mathcal{F}_\Gamma\left(\overline{X}\right)$, the previous necessary optimality condition cannot be satisfied by more than $m+1$ extreme points. Next, we consider an increasing sequence of finite sets of extreme points converging to entire set of extreme points of $\mathcal{F}_\Gamma\left(\overline{X}\right)$ and argue via the Berge maximum theorem that the bounds on the number of extreme points is inherited by at least one optimal lottery of the original problem. Finally, we invoke a result of Winkler [1988] to show that each extreme point of $\mathcal{F}_\Gamma\left(\overline{X}\right)$ is supported in no more that $k+1$ points of $\overline{X}$. Because there is an optimal lottery that is the combination of no more than $m+1$ extreme points, the result follows. Moreover, strict concavity (see Proposition 1) implies that there is only one optimum that has finite support.[11]

When Y is infinite dimensional, the choice set can have a thicker support.

Theorem 3. *Assume that $X=[0,1]$, $\Gamma=\varnothing$, the kernel of a generalized methods of moments forecasting surprise function $H(x,\tilde{x})=\int h(x,s)h(\tilde{x},s)d\mu(s)=G(x-\tilde{x})$ is positive definite, and $H(0,\tilde{x})$ is non-negative, strictly decreasing (when positive) and strictly convex in $\tilde{x}$. Then there exists a unique maximum F and it has full support over X.*

To prove this, we first invoke Proposition 1 to obtain strict concavity of the function V, which implies that the unique optimal distribution F for V over $\mathcal{F}$ is characterized by first-order conditions. Then the complementary slackness condition, together with the assumptions on H, imply that there cannot be an open set in X to which F assigns probability zero.

Theorem 2 shows that when the adversary is confined to a small set of forecasts the support of the optimum is thin. Theorem 3 gives a sufficient condition for the support of the optimum to be thick. For this condition to be satisfied the adversary must have a "large"set of forecasts. Example 7 in Online Appendix IV shows that the set of stochastic processes with continuous sample paths on a unit interval is large enough to generate thick support for the optimum.

We close this section with a corollary that follows from Theorems 2 and 3.

Corollary 1. *Maintain the assumptions of Theorem 3, and let F denote the unique fully supported solution. There exists a sequence of method-of-moments representations V^n with $|S^n|=m^n\in\mathbb{N}$, and a sequence of lotteries F^n such that:*

[11] In particular, Theorem 8 in the Appendix shows that, whenever $\overline{X}$ is finite, the bound on the support stated in Theorem 2 holds generically for every optimal lottery. This genericity result is extended to the infinite case in Online Appendix III.B, but only for "robust" optimal lotteries.

1. *For all $n \in \mathbb{N}$, F^n is optimal for V^n and supported on up to $m^n + 1$ points.*

2. *$F^n \to F$ weakly, with* $\operatorname{supp} F^n \to \operatorname{supp} F = X$ *in the Hausdorff topology.*

Intuitively, as the number of moments that the adversary matches increases, the agent randomizes over more and more outcomes, up to the point that each outcome is in the support of the optimal lottery.[12]

5 Adversarial expected Utility and weakly adversarial forecasters

We now generalize the adversarial forecaster representation to adversaries with other objectives than minimizing surprise. This provides a link between the adversarial forecaster representation and the induced preferences of an expected utility agent in a zero-sum sequential game (see Theorem 4). Moreover, it lets us deal with some preference representations that do not satisfy differentiability, such as in Example 3 below where the loss function of the adversary is the absolute value of the error. Finally, it clarifies the relation of adversarial preferences to other risk preferences that admit a maxmin representation, such as those in Maccheroni [2002], Cerreia-Vioglio [2009], and Cerreia-Vioglio, Dillenberger, and Ortoleva [2015], (See Proposition 3 and Theorem 4). Let X be a compact metric space of outcomes, Y a compact metric space of choices for the adversary, and $G \in \mathcal{G}$ the space of probability measures on the Borel sets of $X \times Y$, endowed with the topology of weak convergence. Given $F \in \mathcal{F}$ and $y \in Y$, we let $(F, y) \in \mathcal{G}$ denote the joint distribution over $X \times Y$ that assigns mass 1 to y.

We suppose that the agent has expected utility preferences $\overline{\succsim}$ over $\mathcal{G}$ and the adversary has the opposite preferences: it prefers what is least liked by the agent.[13] Under these assumptions, for any F there is at least one $y \in Y$ such that $(F, y)\overline{\precsim}(F, \tilde{y})$ for all $\tilde{y} \in Y$. We let $\hat{Y}(F)$ denote the set of all such y.[14]

Definition 5. Preference $\succsim$ over $\mathcal{F}$ has an *adversarial expected utility* representation if there exists a compact metric space Y and an expected utility preference $\overline{\succsim}$ over $\mathcal{G}$

[12]Note that, in general, weak convergence does not imply Hausdorff convergence of the supports.
[13]That is, $\overline{\succsim}$ is complete, transitive, continuous, and satisfies the independence axiom.
[14]Theorem 4 below shows that each $\hat{Y}(F)$ is nonempty.

such that

$$F \succsim \tilde{F} \iff (F, y)\overline{\succsim}(\tilde{F}, \tilde{y}) \quad \text{for some} \quad y \in \hat{Y}(F), \tilde{y} \in \hat{Y}(\tilde{F}). \tag{6}$$

If in addition $\hat{Y}(F)$ is a singleton for all $F \in \mathcal{F}$, we say that $\succsim$ satisfies *uniqueness*.

Our next result shows that this is a special case of Maccheroni [2002]'s maxmin model under risk. The additional condition that the adversary's action space is compact is a key ingredient of Theorem 5.

Proposition 3. *Preference $\succsim$ over $\mathcal{F}$ has an adversarial expected utility representation if and only if there exists a compact metric space Y and a* jointly continuous *utility function $u : X \times Y \to \mathbb{R}$ such that $\succsim$ is represented by*

$$V(F) = \min_{y \in Y} \int u(x, y) dF(x).^{15} \tag{7}$$

This proposition shows that if a preference $\succsim$ has an adversarial expected utility representation, oops it is the preference over (mixed) strategies $\mathcal{F}$ induced by backward induction in a sequential game played against an adversary. It also implies that if $\succsim$ has an adversarial forecaster representation, it has an adversarial expected utility representation with $Y = \mathcal{F}$ and $u(x, F) = v(x) + \sigma(x, F)$.

When the preference $\succsim$ has an adversarial expected utility representation, we let Y and u denote an arbitrary representation of $\succsim$. Next we characterize these preferences in terms of the properties of the representation V.

Definition 6. A continuous functional $V : \mathcal{F} \to \mathbb{R}$ has a *compact local expected utility* if there exists a function $w : X \times \mathcal{F} \to \mathbb{R}$ such that $\{w(\cdot, F)\}_{F \in \mathcal{F}}$ is equicontinuous and $w(\cdot, F)$ is a local expected utility of V at F, for all $F \in \mathcal{F}$.

We say that preference $\succsim$ over $\mathcal{F}$ has a continuous local expected utility if it can be represented by a functional V with compact local expected utility. Note that if V has a continuous expected utility then it has a compact local expected utility, but the converse does not hold in general, as for example in (7) when Y is finite.[16]

[15]Maccheroni [2002] does not assume that Y is compact, and claims the resulting representation involves a minimum that is attained, which is not correct. Machina [1984] and Frankel and Kamenica [2019] make the same mistake, See Corrao, Fudenberg, and Levine [2022].

[16]Chatterjee and Krishna [2011] axiomatizes a particular case where $X = [0, 1]$ and the set of local utilities of the agent is equi-Lipschitz continuous (hence compact).

Now we relax joint continuity in the definition of surprise functions.

Definition 7. We say that $\tilde{\sigma} : X \times \mathcal{F} \to \mathbb{R}_+$ is a *weak surprise function* if it $\{\tilde{\sigma}(\cdot, F)\}_{F \in \mathcal{F}}$ is equicontinuous in X, $\tilde{\sigma}(x, \delta_x) = 0$ for all $x \in X$, and if $\int \tilde{\sigma}(x, F) dF(x) \leqslant \int \tilde{\sigma}(x, \hat{F}) dF(x)$ for all $F, \hat{F} \in \mathcal{F}$.

When a preference $\gtrsim$ can be represented as in Equation 1 by using a weak surprise function $\tilde{\sigma}$, we say that it has a *weakly adversarial forecaster representation.* The generalization of differentiability that we considered characterizes both the adversarial expected utility representation and the weakly adversarial forecaster representation.

Theorem 4. *Let $\gtrsim$ be a preference over $\mathcal{F}$. The following conditions are equivalent*

(i) The preference $\gtrsim$ has a weakly adversarial forecaster representation.

(ii) The preference $\gtrsim$ is has an adversarial expected utility representation.

(iii) The preference $\gtrsim$ has a compact local expected utility.

Moreover, $\gtrsim$ has an adversarial expected utility representation that satisfies uniqueness if and only if it has an adversarial forecaster representation.

In the next example, the agent's preferences have an adversarial expected utility representation but cannot be represented by an adversarial forecaster representation.

Example 3. Consider the same setting of Example 1, but suppose the adversary's objective is to minimize the absolute deviation, so

$$V(F) = \int_0^1 v(x) dF(x) + \min_{c \in [0,1]} \int |c - x| dF(x)$$

In this case, the relevant statistic for the adversary is the median of the chosen distribution. The median need not be unique for some F. As Example 6 in Online Appendix IV shows, this implies that V does not have a continuous local expected utility. Nevertheless, the existence of a median $\hat{c}(F)$ for every F implies that V has a weakly continuous local expected utility, so from Theorem 4 it has a weakly adversarial forecaster representation. △

Theorem 2 can be also applied to the case where the adversary has only m actions. It is enough to linearly extend $u(x,\cdot)$ to $\hat{u}(x,\cdot) : \Delta(Y) \to \mathbb{R}$ over mixed adversary's actions, and then $\Delta(Y)$ and $\hat{u}$ satisfy all the assumptions of Theorem 2.[17]

6 Monotonicity and behavior

This section characterizes monotonicity with respect to integral stochastic orders (e.g. first-order stochastic dominance, second-order stochastic dominance, and the mean-preserving spread order) in the adversarial expected utility representation, and uses it to analyze higher-order risk aversion and correlation aversion.

6.1 Integral stochastic orders and monotonicity

We start with the definition of the stochastic order induced by a set $\mathcal{W} \subseteq C(X)$.

Definition 8. Given a set $\mathcal{W} \subseteq C(X)$, the stochastic order $\succsim_{\mathcal{W}}$ is defined as:

$$F \succsim_{\mathcal{W}} \tilde{F} \iff \int w(x)dF(x) \geqslant \int w(x)d\tilde{F}(x) \quad \forall w \in \mathcal{W}. \tag{8}$$

For every set $\mathcal{W} \subseteq C(X)$, let $\langle \mathcal{W} \rangle$ denote smallest closed convex cone containing $\mathcal{W}$ and all the constant functions. The elements of $\langle \mathcal{W} \rangle$ are all the expected utility functions that are equivalent to functions in $\mathcal{W}$, so $\succsim_{\mathcal{W}} = \succsim_{\langle \mathcal{W} \rangle}$. Next, we give our notion of monotonicity with respect to the stochastic order induced by a set $\mathcal{W}$.

Definition 9. Given $\mathcal{W} \subseteq C(X)$, preference $\succsim$ *preserves* $\succsim_{\mathcal{W}}$ if, for all $F, \tilde{F} \in \mathcal{F}$, $F \succsim_{\mathcal{W}} \tilde{F}$ implies $F \succsim \tilde{F}$.

Notice that if $\succsim$ preserves $\succsim_{\mathcal{W}}$ then it also does so for any larger set $\tilde{\mathcal{W}} \supseteq \mathcal{W}$. From Theorem 2 in Castagnoli and Maccheroni [1999], for every $v \in C(X)$, the expected utility preference $\succsim_v$ preserves $\succsim_{\mathcal{W}}$ if and only if $v \in \langle \mathcal{W} \rangle$. Intuitively, if v is not parallel to some function in $\mathcal{W}$ then it crosses every function in $\mathcal{W}$, and so has a less preferred point that is preferred by $\succsim_{\mathcal{W}}$.

Given a compact metric space Y, for every Borel set $\hat{Y} \subseteq Y$, we let $\mathcal{H}(\hat{Y})$ denote the space of Borel probability measures over $\hat{Y}$. Moreover, given a continuous $u : X \times$

[17] Theorem 11 in Online Appendix III.A uses a different proof to show that when Y is finite with m elements the upper bound on the support can be improved to $k + m$.

$Y \to \mathbb{R}$ and a probability measure $H \in \mathcal{H}$, we define $u(\cdot, H) = \int u(\cdot, y) dH(y) \in C(X)$. In an adversarial expected utility representation, we can associate the utility function u with the set $\mathcal{W}_{u,Y} = \{u(\cdot, y) : y \in \hat{Y}(F), F \in \mathcal{F}\}$ and a stochastic order $\gtrsim_{u,Y}$ on $\mathcal{F}$. It is clear that the preference $\gtrsim$ induced by u preserves $\gtrsim_{u,Y}$, and more generally, preserves any stochastic order $\gtrsim_{\tilde{\mathcal{W}}}$ generated by a set $\tilde{\mathcal{W}} \supseteq \mathcal{W}_{u,Y}$. Theorem 5 provides a converse and extends it to the weakly adversarial forecaster model.

Theorem 5. *Let $\gtrsim$ have an adversarial expected utility representation (Y, u) and fix a set $\mathcal{W} \subseteq C(X)$. The following conditions are equivalent:*

(i) The preference $\gtrsim$ preserves $\gtrsim_{\mathcal{W}}$.

(ii) For all $F \in \mathcal{F}$, there exists $H \in \mathcal{H}(\hat{Y}(F))$ such that $u(\cdot, H) \in \langle \mathcal{W} \rangle$.

(iii) The preference $\gtrsim$ has a weakly adversarial forecaster representation such that $v + \sigma(\cdot, F) \in \langle \mathcal{W} \rangle$ for all $F \in \mathcal{F}$.

The equivalence between (ii) and (iii) follows from Theorem 4, and the implication from (ii) to (i) only formalizes the discussion before the theorem. The implication from (ii) to (i) is more involved. To show this, we first observe that the preference $\gtrsim$ preserves $\gtrsim_{\mathcal{W}}$ if and only if for all $F, G, \hat{G} \in \mathcal{F}$ such that $G \gtrsim_{\mathcal{W}} \hat{G}$, there exists $H \in \mathcal{H}(\hat{Y}(F))$ such that $\int u(x, H) dG(x) \geqslant \int u(x, H) d\hat{G}(x)$. By the Sion minmax theorem, this assertion is equivalent to the statement that there exists $H \in \mathcal{H}(\hat{Y}(F))$ such that $\gtrsim_{u(\cdot,H)}$ preserves $\gtrsim_{\mathcal{W}}$. Finally, because $u(\cdot, \hat{y}(F))$ is continuous, Theorem 2 in Castagnoli and Maccheroni [1999]) shows that $u(\cdot, \hat{y}(F)) \in \langle \mathcal{W} \rangle$.

Theorem 5 differs from other monotonicity results in the literature for preferences with concave representations because it characterizes monotonicity for a given representation, instead of constructing a representation with the desired properties.[18]

As an immediate corollary of Theorem 5, when $\gtrsim$ has an adversarial forecaster representation, it preserves an integral stochastic order $\gtrsim_{\mathcal{W}}$ if and only if $v + \sigma(\cdot, F) \in \langle \mathcal{W} \rangle$ for all $F \in \mathcal{F}$.[19] Theorem 5 underlies our proof of characterizations of the

[18]For example, Proposition 22 in Cerreia-Vioglio [2009] (for preferences with a quasiconcave representation), Theorem 4.2 in Chatterjee and Krishna [2011] (for preferences with a concave and Lipschitz continuous representation), and Theorem S.1 in Sarver [2018] (for preferences with a concave representation) assume that the underline preference preserves an integral stochastic order.

[19]When X is a compact interval in the real line, this last statement directly follows from Proposition 1 in Cerreia-Vioglio, Maccheroni, and Marinacci [2017] because, as we show in Proposition 8 in Online Appendix V, if V has an adversarial forecaster representation then it is Gateaux differentiable with derivative $v + \sigma(\cdot, F)$. However, the proof of Theorem 5 is quite different as it does not rely on Gateaux differentiability.

support size in the example of writing a novel (Proposition 2), allocation problems (Corollary 3) and disclosure (Corollary 2). We also used Theorem 5 in Section 2.3's discussion of how the effect of a preference for surprise on risk aversion. The next section extends this discussion.

6.2 Application: risk aversion and adversarial forecasters

Now we use the monotonicity result to show how a preference for surprise can alter the higher-order risk preferences of the agent. We consider an asymmetric version of the method of moments representation, where the forecaster is asymmetrically concerned about the direction of deviations of the realized moment from the forecast. For simplicity, we let $X = [0, 1]$ and consider only the first moment.[20]

Fix a strictly convex and twice continuously differentiable function $\rho : [-1, 1] \to \mathbb{R}_+$ such that $\rho(0) = 0$, $\rho'(z) < 0$ if $z < 0$, and $\rho'(z) > 0$ if $z > 0$, and consider the preferences over lotteries induced by

$$V(F) = \int_0^1 v(x)dF(x) + \min_{\hat{x}\in X} \int_0^1 \rho(x - \hat{x})dF(x).$$

Here the adversary's actions are restricted to X rather than $\mathcal{F}$, but nevertheless V does correspond to the adversarial forecaster representation

$$V(F) = \int_0^1 v(x)dF(x) + \min_{\hat{F}\in\mathcal{F}} \int_0^1 \rho(x - \hat{x}_\rho(\hat{F}))dF(x),$$

where $\hat{x}_\rho(F) = \operatorname{argmin}_{\hat{x}\in X} \int \rho(x - \hat{x})dF(x)$, and $\sigma(x, F) = \rho(x - \hat{x}_\rho(\hat{F}))$.[21] Thus the function $\int \rho(x - \hat{x}_\rho(\hat{F}))dF(x)$ can be interpreted as an index of the dispersion of F, without requiring symmetry. The local expected utility of the agent is $w(x, F) = v(x) + \rho\left(x - \hat{x}(F)\right)$, with second derivative $w''(x, F) = v''(x) + \rho''(x - \hat{x}(F))$ that now also depends on the lottery F.

In particular, consider the asymmetric loss function $\rho(z) = \lambda(\exp(z) - z)$, $\lambda \geqslant 0$. The relevant statistic is $\hat{x}(F) = \log\left(\int_0^1 \exp(x)dF(x)\right)$, that is, the (normalized) cumulant generating function evaluated at 1. With this loss function the agent prefers a positive surprise $x > \hat{x}(F)$ to a negative surprise $x < \hat{x}(F)$ of the same absolute

[20] It is easy to generalize this to finite or infinite numbers of moments as we did for the quadratic GMM in Section 2.3.

[21] To see why, note that, since $\hat{x}_\rho(\hat{F})$ is the unique minimizer, it varies continuously with F.

value. The second derivative of the local expected utility at an arbitrary lottery F is $w''(x,F) = v''(x) + \lambda \exp(x - \hat{x}(F))$, so the agent is more risk adverse over outcomes that are concentrated around $\hat{x}(F)$, with a relative preferences for higher outcomes. The n-th order derivative of each local utility is $w^{(n)}(x,F) = v^{(n)}(x) + \lambda \exp(x - \hat{x}(F))$, so for λ high enough, $w^{(n)} > 0$. From Theorem 5, this implies that higher enjoyment for surprise induces preferences over lotteries that are monotone with respect to the stochastic orders induced by smooth functions whose derivatives are positive. For example, as formalized in Menezes, Geiss, and Tressler [1980], aversion to downside risk (i.e., *prudence*), is equivalent to preserving the order $\gtrsim_{\mathcal{W}_3^+}$ induced by the smooth functions with positive third derivative $\mathcal{W}_3^+$, which is the case here whenever λ is high.[22] Here asymmetric preference for surprise is crucial: if the third derivatives of all the local expected utilities of V coincide with those of v, preferences for surprise do not affect higher-order risk aversion.

Eeckhoudt and Schlesinger [2006] formalize the idea that an agent is averse to higher-order risks through the comparison of pairs of lotteries that only differ for their n-th order risk. If at any wealth level the agent prefers the lottery with less n-th order risk, they say the preferences exhibit *risk apportionment* of order n. In our setting with general continuous preferences, a sufficient condition for risk apportionment of order n is monotonicity with respect to the n-th order stochastic dominance relation $\gtrsim_{\mathcal{W}_{SD_n}}$ where

$$\mathcal{W}_{SD_n} = \left\{ u \in C^n(X) : \forall m \leqslant n, sgn(u^{(m)}) = (-1)^{m-1} \right\}.$$

Agents with risk apportionment of order n for all n are called *mixed risk averse*. Most participants in the experiment of Deck and Schlesinger [2014], make choice that are consistent with mixed risk aversion (at their current wealth levels), but almost 20% make risk-loving choices. These participants are *mixed risk loving*, which means they are consistent with risk apportionment of order for odd n but not even n.

As an example, suppose $v(x) = 1 - \exp(-ax)/a$ for $a > 0$. If there is no preference for surprise, that is $\lambda = 0$, the agent is mixed risk averse, as most of the risk averse subjects in Deck and Schlesinger [2014]. However, as λ increases the sign of the even derivatives of the local expected utilities switches from negative to positive, while the

[22] A sufficient condition for all the local expected utilities to have strictly positive n-th derivative is that $\lambda > \tilde{v}^{(n)} exp(1)$, where $\tilde{v}^{(n)} = \max_{x \in X} |v^{(n)}(x)|$.

sign of the odd derivatives remains positive, so the agent shifts from mixed risk averse to mixed risk loving. Moreover, if the agent is very risk averse, that is, $a > 1$, then higher-order derivatives will be more affected by an increased taste for surprise, while the opposite is true if the agent is not very risk averse, that is, $a < 1$.

6.3 Application: repeated choices and correlation aversion

Our model also covers the case where the adversary can observe one realization of a lottery before choosing the next one. Consider $X = X_0 \times X_1$ where X_0 is finite and X_1 is an arbitrary compact subset of Euclidean space. Assume that the adversary takes two actions $(y_0, y_1) \in Y = Y_0 \times Y_1$, where the adversary takes the first action y_0 with no additional information about F, and then takes the second action after observing the realization of x_0. Assume that both Y_0 and Y_1 are compact subsets of Euclidean space. Here the set of strategies of the adversary is $Y = Y_0 \times Y_1^{X_0}$, which is compact. Therefore, the induced preferences

$$V(F) = \min_{y \in Y} \int u(x, y_0, y_1(x_0)) dF(x)$$

still admit an adversarial expected utility representation. These preferences capture the idea of aversion to correlation between x_0 and x_1, which is well documented in experiments (see for example Andersen et al. [2018]). Intuitively, the agent would tend to avoid lotteries with high correlation between x_0 and x_1 since this would imply that the adversary is better informed about the residual distribution of x_1 when choosing y_1. The next example formalizes this idea by applying Theorem 5.

Example 4. Let $X_0 = \{0, 1\}$, $X_1 = [0, 1]$, $v(x_0, x_1) = v_0(x_0) + v_1(x_1)$, and assume that the adversary tries to minimize mean squared error. We have $\sigma_0(x_0, \hat{F}_0) = \left(x_0 - \int \tilde{x_0} dF_0(\tilde{x_0})\right)^2$ and $\sigma_1(x_1, \hat{F}_1 | x_0) = \left(x_1 - \int \tilde{x_1} dF_1(\tilde{x_1} | x_0)\right)^2$, where F_0 and $F_1(\cdot | x_0)$ respectively denote the marginal and the conditional distributions of F, so the local expected utility of the agent is

$$\frac{\partial}{\partial x_1} w(1, x_1, F) - \frac{\partial}{\partial x_1} w(0, x_1, F) = -2 \left(\int \tilde{x_1} dF_1(\tilde{x_1} | 1) - \int \tilde{x_1} dF_1(\tilde{x_1} | 0) \right).$$

By Theorem 5, the preference of the agent preserves the submodular (resp. supermodular) order for all F such that $\int \tilde{x_1} dF_1(\tilde{x_1}|1) > \int \tilde{x_1} dF_1(\tilde{x_1}|0)$ (resp. $<$).[23] Thus, the agent has higher utility with distributions such that $\int \tilde{x_1} dF_1(\tilde{x_1}|1) = \int \tilde{x_1} dF_1(\tilde{x_1}|0)$: the best conditional forecast is independent of x_0. $\triangle$

We leave a more detailed analysis of correlation aversion under the adversarial expected utility model for future research.[24]

7 Induced preferences

This section augments the Bayesian persuasion and the optimal assignment problems with an initial stage where the designer chooses priors or quality distributions. We show that the induced preferences have an adversarial expected utility representation, and can also have the stronger adversarial forecaster form.

7.1 Coice and (Bayesian) persuasion

There are two agents, a sender (S) and a receiver (R); their payoffs depend on a state variable $x \in X = [0,1]$. The receiver takes an action $a \in A = [0,1]$ after they observe a signal about the state that was chosen (with commitment) by the sender. Equivalently we will think of the sender as choosing the receiver's posterior subject to the constraint that the expected posterior equals the prior, that is, in

$$\Delta^2(F) = \left\{ T \in \Delta(\mathcal{F}) : \int \tilde{F} dT(\tilde{F}) = F \right\}$$

The payoff functions of the sender and the receiver are given respectively by the continuous functions $S : X \times A \to \mathbb{R}$ and $R : X \times A \to \mathbb{R}$. For every posterior $\tilde{F}$, we let $a^\star(\tilde{F}) = \text{argmax}_{a \in A} \int R(x,a) d\tilde{F}(x)$ denote the receiver's best responses.

Unlike in the usual persuasion problem, we assume that the sender has some control over the initial distribution of the state. We let $\overline{\mathcal{F}} \subseteq \mathcal{F}$ denote the set of feasible initial distributions for S. In other words, the sender can choose $F \in \overline{\mathcal{F}}$ together with the information structure T. Importantly, while the receiver observes

[23] Recall that $\gtrsim$ preserves the submodular (resp. supermodular) order if $F \gtrsim G$ whenever $\int w(x) dF(x) \geqslant \int w(x) dG(x)$ for all functions $w \in C(X)$ such that $\frac{\partial}{\partial x_1} w(1, x_1) - \frac{\partial}{\partial x_1} w(0, x_1) \leqslant 0$ (resp. $\geqslant$). See Shaked and Shanthikumar [2007] for more details.

[24] Stanca [2021] analyzes correlation aversion for sequential decisions under uncertainty as opposed to risk.

the chosen distribution F, the chosen information structure T, and the realization $\tilde{F}$, the state x drawn from F is not observed. With this, we can solve the sender's problem by backward induction. For every $F \in \overline{\mathcal{F}}$ we solve a standard Bayesian persuasion problem with common prior F. This in turn induces a value function V_S over $\overline{\mathcal{F}}$ which the sender maximizes in the first stage.

Let $S^\star(\tilde{F}) = \max_{a \in a^\star(\tilde{F})} \int S(x,a) d\tilde{F}(x)$ be the interim expected payoff of the sender, where as usual we select the sender-preferred action. The value function for the sender over feasible initial distributions is then

$$V_S(F) = \max_{T \in \Delta^2(F)} \int S^\star(\tilde{F}) dT(\tilde{F}), \tag{9}$$

so the problem of the sender in the first stage is $\max_{F \in \overline{\mathcal{F}}} V_S(F)$.

Example 5 (Disclosure and choice)**.** Consider an agent (the sender) that chooses among risky projects $F \in \overline{\mathcal{F}}$ and commits to an information structure about the outcome x of the project. His payoff is increasing in the outcome x and in an observer's (the receiver) expectation of the outcome. With some probability, the agent will be able to disclose the signal realized from the information structure chosen before. In the language of this section, we have that the action $a \in A$ corresponds to the conditional expectation of the receiver: $R(x,a) = -(x-a)^2$. The payoff of the sender is given by

$$S(x,a) = \alpha x + (1-\alpha)\left(\beta \max\{a, \hat{a}\} + (1-\beta)\hat{a}\right)$$

where $\alpha \in (0,1)$ is a parameter describing the relative preference of the sender between the outcome and the receiver's beliefs, $\beta \in (0,1)$ is the probability with which the sender can disclose the realized signal, and $\hat{a}$ is the conditional expectation of the receiver if they do not observe a message. This model generalizes Ben-Porath, Dekel, and Lipman [2018] (BDL) by allowing the agent to commit to any information structure rather than just full disclosure or no disclosure, and by considering potentially infinite choice sets $\overline{\mathcal{F}}$.[25] $\triangle$

The $V_S(F)$ defined in Equation 9 corresponds to the concave closure of $S^\star$ evaluated at F. This not sufficient to obtain an adversarial expected utility representation,

[25] In BDL $a = x$, the set of feasible lotteries $\overline{\mathcal{F}}$ is finite, and the payoff of the sender is $S(x) = \alpha x + (1-\alpha)\left(\beta \max\{x, \hat{a}\} + (1-\beta)\hat{a}\right)$. In their general model, BDL also consider an adversary whose objective is to minimize the agent's payoff.

since the minimum in equation 7 is not necessarily attained. However, the preference induced by V_S does admit an adversarial expected utility representation under some mild regularity assumptions. Let $L(X)$ denote the set of Lipschitz continuous over X, $B(A)$ denote the set of bounded measurable functions over A, and define

$$\mathcal{W}(S^*) = \left\{ w \in L(X) : \forall \tilde{F} \in \mathcal{F}, \int_0^1 w(x) d\tilde{F} \geqslant S^*(\tilde{F}) \right\}$$

We next show that when S^* is Lipschitz continuous the function V_S admits an adversarial expected utility representation with local expected utilities in $\mathcal{W}(S^*)$.[26]

Proposition 4. *Let S^* be Lipschitz continuous. The preference V_S over $\mathcal{F}$ induced by the Bayesian persuasion problem admits a adversarial expected utility representation (u, Y) with Y a compact subset of $\mathcal{W}(S^*)$ and $u(x, y) = y(x)$. In addition, if R is strictly concave in a and continuously differentiable, then Y can be chosen such that*

$$Y \subseteq \left\{ \max_{a \in A} \{ S(\cdot, a) + q(a) R_a(\cdot, a) \} \in C(X) : q \in B(A) \right\}$$

The proofs of the results in this and all of the rest of the results in the main text are in Online Appendix I.

One case where S^* is Lipschitz continuous is when $a^*(F) = \int x dF(x)$ (as in Example 5) and S is Lipschitz continuous in a. In general, the adversarial expected utility representation in this proposition corresponds to the dual of the persuasion problem in Dworczak and Kolotilin [2022].[27] The proof of Proposition 4 shows that the compact set $\mathcal{W}_L(S^*)$ satisfies the properties of Definition 6, so V has a continuous local expected utility, and Theorem 4 then implies that V has an adversarial EU representation, where the set of adversarial actions Y corresponds to the set of restricted dual variables $\mathcal{W}_L(S^*)$.

We next use Proposition 4 and Theorem 5 to characterize the solution of the first-stage choice problem when the payoff of the sender is convex in the outcome and the feasible distributions are all those whose expectation is at least some lower bound. Define $\tilde{S}(x) = (1 - x) S(0, x) + x S(1, x)$ and let $\hat{S}(x)$ denote its concave closure.

[26] Here we use the metric for $\mathcal{F}$ induced by the Kantorovich-Rubinstein norm.

[27] Dworczak and Kolotilin [2022] shows that duality holds when S^* is Lipschitz continuous, and that S^* is Lipschitz continuous when $a^*(F) = \int x dF(x)$ and S is Lipschitz continuous in a.

Corollary 2. *Assume that S is convex in x, $R(x,a) = -(x-a)^2$, and that*

$$\overline{\mathcal{F}} = \left\{ F \in \mathcal{F} : \int_0^1 x dF(x) \geqslant x_0 \right\}$$

for some $x_0 \in X$. Then the distribution $F^* = (1-x^*)\delta_0 + x^*\delta_1$ solves the first-stage problem, where $x^* \in \operatorname{argmax}_{x \geqslant x_0} \hat{S}(x)$.

Given the assumptions on S and R, Proposition 4 implies that all the local expected utilities of the sender are convex in x. In turn, Theorem 5 implies that, given a feasible average state $x^* \geqslant x_0$, the sender will always pick the most dispersed prior, which here is binary. Theorem 1 in Kamenica and Gentzkow [2011] then implies that the induced utility over priors is $\hat{S}(x^*)$.

The model in Example 5 satisfies all the assumptions of Corollary 2, regardless of the value of $\hat{a}$. Our result here implies that, rather than choosing the prospect with the highest expected reward, the agent chooses an highly dispersed prospect whose average maximizes $\hat{S}(x)$. This result is in line with Theorem 1 of Ben-Porath, Dekel, and Lipman [2018] with the advantage of allowing for a much richer space of available information structures and risky investments.

7.2 Production and allocation

Consider an economy with a continuum of consumers parametrized by $\theta \in \Theta = [0,1]$ with distribution $Q \in \Delta(\Theta)$. There is a single good that comes in differentiated varieties/qualities $x \in X = [0,1]$. Type θ receives utility is $g(\theta, x)$ from quality x, where g is continuous and strictly supermodular. The cost of producing one unit of quality x is given by a continuous function $c(x)$.

Suppose a social planner want to choose a distribution of qualities $F \in \mathcal{F}$ and an allocation mechanism to maximize total surplus $\pi = g - c$.[28] In standard assignment problems (e.g. Santambrogio [2015] Section 1.7.3), the distribution over qualities F_0 is fixed, whereas in standard nonlinear pricing models (e.g. Mussa and Rosen [1978]) the designer can choose any distribution $F \in \mathcal{F}$. Ee consider the case $F \in \overline{\mathcal{F}} \subseteq \mathcal{F}$, where where the feasible distributions are partially but not fully constrained.[29] For

[28]The case where the designer maximizes instead of total surplus can be similarly analyzed.

[29]Recently, Loertscher and Muir [2022] studied a particular case of the model presented here where $g(\theta, x) = \theta x$, $c = 0$, and $\overline{\mathcal{F}} = \{F_0\}$ for a fixed distribution of qualities.

example, $\overline{\mathcal{F}}$ might defined by moment restrictions as in Section 4 that bound the average quality in the economy or its variance. In these cases we can reduce the allocation problem to the choice of a (stochastic) assignment function $\Xi : \Theta \to \mathcal{F}$ such that $F(x) = \int \Xi(x|\theta)dQ(\theta)$. For every production choice $F \in \overline{\mathcal{F}}$, let $\mathcal{I}(F)$ denote the set of assignment functions Ξ that induce F.

The designer chooses a production plan F and an allocation ξ to solve

$$\max_{F \in \overline{\mathcal{F}}} \max_{\Xi \in \mathcal{I}(F)} \int_{\Theta} \int_{X} \pi(\theta, x) d\Xi(x|\theta) dQ(\theta).$$

We are interested in their induced preference F, which is

$$V_{\pi,Q}(F) = \max_{\Xi \in \mathcal{I}(F)} \int_{X} \pi(\theta, x) d\Xi(x|\theta) dQ(\theta).$$

Unlike in Machina [1984], the initial choice of the distribution F here only affects the set of feasible allocations in the second stage. This lets us we can use duality theory on the second-stage problem to show that $V_{\pi,Q}$ admits an adversarial expected utility representation.

Let $\mathcal{C}_{\pi} \subseteq C(X)$ denote the set of functions $\phi \in C(X)$ such that

$$\phi(x) = \max_{\theta \in \Theta} \left\{ \pi(\theta, x) - \hat{\phi}(\theta) \right\} \tag{10}$$

for some $\hat{\phi} \in C(\Theta)$.[30] When $\phi \in \mathcal{C}_{\pi}$, there is a minimal (in the pointwise order) function ϕ^{π} that satisfies Equation 10 for ϕ.[31] Also define the set

$$\mathcal{W}_{\pi,Q} = \left\{ \phi + \int \phi^{\pi}(\theta) dQ(\theta) \in C(X) : \phi \in \mathcal{C}_{\pi} \right\}.$$

Let π_x denote the partial derivative of π with respect to x, and, for every non-decreasing real function ψ, let ψ^{-1} denote its generalized inverse function.[32]

Proposition 5. *The function $V_{\pi,Q}$ admits an adversarial expected utility representation (u, Y) with Y a compact subset of $\mathcal{W}_{\pi,Q}$ and $u(x, y) = y(x)$. In addition, if Q is absolutely continuous and π is continuously differentiable, then $V_{\pi,Q}$ admits an*

[30] In optimal transport theory, these functions are called π-*convex* and used to solve the dual Kantorovich problem. See for example Santambrogio [2015] for detailed analysis.

[31] Specifically $\phi^{\pi}(\theta) = \max_{x \in X} \{\pi(\theta, x) - \phi(x)\}$.

[32] Recall that this is defined as $\psi^{-1}(t) = \inf\{x \in [0, 1] : \psi(x) > t\}$.

adversarial forecaster representation with local expected utility:

$$w(x, F) = \int_{\xi_F(0)}^{x} \pi_x \left((\xi_F)^{-1}(z), z\right) dz, \tag{11}$$

where the point mass on $\xi_F = F^{-1} \circ Q$ is the unique solution of the allocation problem.

Corollary 3. *Assume that π is convex in x, that Q is absolutely continuous, and that*

$$\overline{\mathcal{F}} = \left\{ F \in \mathcal{F} : \int_0^1 x dF(x) \geqslant x_0 \right\}$$

for some $x_0 \in X$. Then the distribution $F^* = (1 - x^*)\delta_0 + x^*\delta_1$ and the allocation function

$$\xi_{F^*}(\theta) = \begin{cases} 0 & \text{if } \theta \leqslant Q^{-1}(1 - x^*) \\ 1 & \text{if } \theta > Q^{-1}(1 - x^*) \end{cases}$$

solve the planner problem, where

$$x^* \in \underset{x \geqslant x_0}{\operatorname{argmax}} \int_{1-x}^{1} \pi(Q^{-1}(t), 1) dt.$$

The conditions of the corollary are satisfied in the standard linear case where $g(x, \theta) = x\theta$ and $c(x) = kx$ for some $k \in [0, 1]$, so it shows that binary production plans are optimal there. The intuition of this corollary is similar to that of Corollary 2: When π is convex in x, all the local expected utilities of the designer are convex, so for every given feasible average quality $x^* \geqslant x_0$, the designer picks the distribution with expectation x^* that is maximal in the convex order, that is, $F^* = (1 - x^*)\delta_0 + x^*\delta_1$.[33] Given that each of these distributions is binary and π is strictly supermodular, the unique optimal allocation function is obtained by assigning types below the threshold $Q^{-1}(1 - x^*)$ to quality 0 and assigning the remaining types to 1, so the problem reduces to finding the optimal average quality, that is, Equation (3). For example, under the linear specification above, we have $x^* = \max\{x_0, Q^{-1}(1 - k)\}$ where the optimal average quality is decreasing in the unitary production cost k.[34]

[33] If π is convex in x, then each function $\phi \in \mathcal{C}_\pi$ is convex, which implies that each $w \in \mathcal{W}_{\pi,Q}$ is convex, so by Theorem 5, $V_{\pi,Q}$ is monotone in the convex order.

[34] Bergemann, Heumann, and Morris [2022] shows that finite-support quality distributions are optimal when the buyer's utility is linear and the planner can jointly design production, allocation, and the distribution of buyers' types. Their result crucially relies on the fact that planer chooses a

8 Conclusion

We have shown that the adversarial forecaster representation arises naturally in many settings. It allows the interpretation of random choice as a preference for surprise, and also allows sharp characterizations of the optimal "amount" (i.e., support size) of randomization and of various monotonicity properties. The more general weakly adversarial forecaster representation inherits many of the tractability and optimality properties of the adversarial forecaster representation, and applies to induced preferences arising from other settings, such as optimal allocation and Bayesian persuasion.

This paper focuses on an adversary who wants to minimize the surprise of a lottery. One can also consider an adversary that tries to maximize the surprise of a lottery, because the agent suffers anxiety from lotteries that are hard to be predict, as in Caplin and Leahy [2001] and Battigalli, Corrao, and Dufwenberg [2019]. These preferences are the flipped version of our model, and generate a preference for deterministic outcomes. We leave a detailed analysis of this extension for future research.

Appendix I: Sections 2 and 5

This section proves the results in Section 5 and then shows that Theorem 1 of Section 2 follows. It then proves Proposition 1 on the GMM representation. We start with some preliminary results whose proofs are relegated to Online Appendix II.

Preliminaries

We will make use of the *Bregman divergence*, which is closely related to local expected utility. Fix a continuous V that has a local expected utility. For each $F \in \mathcal{F}$, let $\mathcal{W}_V(F) \subseteq C(X)$ denote the (nonempty) set of local expected utilities of V at F.

Definition 10. Let V be continuous and have a local expected utility. We say that $B : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_+$ is a Bregman divergence for V is

$$B(\tilde{F}, F) = V(F) - V(\tilde{F}) - \int w_F(x) d(F - \tilde{F})(x) \qquad \forall F \in \mathcal{F}$$

for some $w_F \in \mathcal{W}_V(F)$.

discrete distribution of buyers' types, while our result holds even if the distribution of types is fixed and absolutely continuous with respect to Lesbegue measure.

Definition 11. We say that $\sigma : X \times \mathcal{F} \to \mathbb{R}_+$ is a *pseudo surprise function* if $\sigma(\cdot, F)$ is continuous for all $F \in \mathcal{F}$, $\sigma(x, \delta_x) = 0$ for all $x \in X$, and if $\int \sigma(x, F) dF(x) \leqslant \int \sigma(x, \hat{F}) dF(x)$ for all $F, \hat{F} \in \mathcal{F}$.

Theorem 6. *Let V be a continuous functional. The following are equivalent:*

(i) V has a local expected utility.

(ii) There exist $v \in C(X)$ and a pseudo surprise function σ such that

$$V(F) = \int v(x) dF(x) + \min_{\hat{F} \in \mathcal{F}} \int \sigma(x, \hat{F}) dF(x) \qquad \forall F \in \mathcal{F}. \tag{12}$$

(iii) There is a separable metric space Y and a jointly continuous function $u : X \times Y \to \mathbb{R}$ such that

$$V(F) = \min_{y \in Y} \int u(x, y) dF(x) \qquad \forall F \in \mathcal{F}.$$

If any of these conditions holds, then

1. *v is uniquely defined by $v(x) = V(\delta_x)$;*

2. *σ satisfies 12 if and only if $\sigma(x, F) = B(\delta_x, F)$ for some Bregman divergence.*

This result implies that even if multiple surprise functions are consistent with (12), the induced suspense function Σ is uniquely defined by $\Sigma(F) = V(F) - \int v(x) dF(x)$.

Lemma 1. *Suppose $F^n \to F$ and that $w^n \to w$. Then $\int w^n(x) dF^n(x) \to \int w(x) dF(x)$. Moreover, if V is continuous with continuous local expected utility and if each w^n is a local expected utility for F^n, then w is a local expected utility for F.*

Lemma 2. *Let V have a continuous local expected utility w. For all $F, \tilde{F}, \overline{F} \in \mathcal{F}$ such that there exists $\mu > 0$ with $F + \mu(\tilde{F} - \overline{F}) \in \mathcal{F}$, we have*

$$DV(\tilde{F} - \overline{F}) := \int w(x, F) d\tilde{F}(x) - \int w(x, F) d\overline{F}(x) = \lim_{\lambda \downarrow 0} \frac{V(F + \lambda(\tilde{F} - \overline{F})) - V(F)}{\lambda}$$

Under the assumptions of Lemma 2, we call $\tilde{F} - \overline{F}$ a *relevant direction* for F and $DV(F)(\tilde{F} - \overline{F})$ the *directional derivative* of V at F in direction $\tilde{F} - \overline{F}$.

Section 5

Recall that $\hat{Y}(F) = \{y \in Y : \forall \tilde{y} \in Y, (F, y)\overline{\precsim}(F, \tilde{y})\}$.

Proof of Proposition 3. Assume that $\gtrsim$ has an adversarial expected utility representation. Given the assumptions on X, Y, and $\overline{\gtrsim}$, by Theorem 1 in Grandmont [1972] there is a jointly continuous $u : X \times Y \to \mathbb{R}$ such that $G\overline{\gtrsim}\tilde{G}$ if and only if $\int u(x, y)dG(x, y) \geqslant \int u(x, y)d\tilde{G}(x, y)$. This immediately gives the representation $V(F) \equiv \min_{y\in Y} \int u(x, y)dF(x)$. Conversely, assume that $\gtrsim$ is represented by a function V defined as in Equation 7. Then $\gtrsim$ has an adversarial expected utility representation with $\overline{\gtrsim}$ given by the EU preference induced by u over $\mathcal{G}$.

Proof of Theorem 4.

We will prove that: $(i) \implies (ii)$, $(ii) \implies (iii)$, and $(iii) \implies (i)$.

(i) implies (ii) Define $\mathcal{W}_{v,\sigma} = cl\left(\{v + \sigma(\cdot, F)\}_{F\in\mathcal{F}}\right)$, where cl denotes the closure operation, and $M = \max_{F\in\mathcal{F}} |V(F)|$. For every $F \in \mathcal{F}$, we have $\max_{x\in X} |v(x) + \sigma(x, F)| \leqslant M$. Therefore, we have $\max_{x\in X} |w(x)| \leqslant M$ for all $w \in \mathcal{W}_{v,\sigma}$. Next, because X is compact, v is uniformly continuous and $\{\sigma(\cdot, F)\}_{F\in\mathcal{F}}$ is uniformly equicontinuous ,so there is a continuous function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\omega(0) = 0$ and$|v(x) + \sigma(x, F) - v(x') - \sigma(x', F)| \leqslant \omega(d(x, x'))$ for every $x, x' \in X$ and $F \in \mathcal{F}$, Thus $\mathcal{W}_{v,\sigma}$ is equicontinuous and, by the Arzaela-Ascoli theorem, a compact metric space. This implies that $V(F) = \min_{w\in\mathcal{W}_{v,\sigma}} \int w(x)dF(x)$, and, by setting $Y = \mathcal{W}_{v,\sigma}$ and $u(x, y) = y(x)$, that V admits a representation as in Equation 7. Finally, Proposition 3 implies (ii).

(ii) implies (iii). Define $\mathcal{W} = \{u(\cdot, y)\}_{y\in Y}$. We will show that $\mathcal{W}$ is uniformly bounded, equicontinuous, and closed. By the Arzaela-Ascoli theorem this will in turn imply that $\mathcal{W}$ is compact. Since u is continuous on the compact set $X \times Y$ it is bounded, hence the family $\mathcal{W}$ is uniformly bounded. By the Heine-Cantor theorem it is also uniformly continuous, so the family $\mathcal{W} = \{u(\cdot, y)\}_{y\in Y}$ is equicontinuous. Next, consider a sequence $w^n \in \mathcal{W}$ converging to w. For all $n \in \mathbb{N}$, there exists $y^n \in Y$ such that $w^n = u(\cdot, y^n)$, and since Y is compact, this sequence has an accumulation point y. Given that u is jointly continuous, we have $w^n = u(\cdot, y^n) \to u(\cdot, y)$, so $w = u(\cdot, y) \in \mathcal{W}$. This shows that $\mathcal{W}$ is closed. Finally, fix $F \in \mathcal{F}$ and $y \in \hat{Y}(F)$. By definition we have $\int u(x, y)dF(x) = V(F)$. Moreover, for all $\tilde{F} \in \mathcal{F}$, we have $\int u(x, y)d\tilde{F}(x) \geqslant \min_{y'\in Y} \int u(x, y')d\tilde{F}(x) = V(\tilde{F})$. This proves that $u(\cdot, y) \in \mathcal{W} \cap \mathcal{W}_V(F) \neq \varnothing$. Given that F was arbitrary, this shows that V has a continuous local expected utility.

(iii) implies (i). By assumption $\gtrsim$ has a representation V with compact local expected utility. Let w denote the corresponding equicontinuous selection from $\mathcal{W}_V$. By Theorem 6, there exist $v \in C(X)$ and a pseudo surprise function σ such that V can be written as in Equation 12. In particular, by point 2 of Theorem 6, σ satisfies 12 if and only if $\sigma(x,F) = B(\delta_x, F)$ for some Bregman divergence of V. Let B_w be the Bregman divergence induced by w and let $\sigma(x,F) = B_w(\delta_x, F) = w(x,F) - v(x)$, hence $\{\sigma(\cdot, F)\}_{F\in\mathcal{F}}$ is equicontinuous. This implies that $\gtrsim$ has a weakly adversarial forecaster representation.

Now we prove the second part of the statement. We first show that if $\gtrsim$ has a representation V with a continuous local expected utility, then it has an adversarial expected utility representation that satisfies uniqueness. Let $w_V : \mathcal{F} \to C(X)$ be a continuous selection from $\mathcal{W}_{\tilde{V}}(F)$, and define $Y = \{w_V(\cdot, F)\}_{F\in\mathcal{F}} \subseteq C(X)$. Since $X, \mathcal{F}$ are compact and w_V is continuous, it follows that Y is closed, bounded, and equicontinuous, so it is compact. For all $y = w_V(\cdot, F)$ and $x \in X$, define $u(x,y) = w_V(x,F)$ and observe that it is jointly continuous. By the properties of the continuous selection w_V, for all $F \in \mathcal{F}$ and for all $\tilde{y} \in Y$, we have

$$\min_{y\in Y} \int u(x,y)dF(x) = V(F) = \int w_V(x,F)dF(x) \leqslant \int u(x,\tilde{y})dF(x).$$

It remains to show that $\int u(x,y)dF(x)$ has a unique minimum over y. Suppose that for some F there is a $\tilde{F} \neq F$ such that $V(F) = \int w_V(x,\tilde{F})dF(x)$. For every $\lambda \in (0,1)$, define $F_\lambda = \lambda\tilde{F} + (1-\lambda)F$. Then because V is concave and the w_V are local expected utility functions, for all $\lambda \in [0,1]$

$$\begin{aligned}\lambda V(\tilde{F}) + (1-\lambda)V(F) \leqslant V(F_\lambda) &\leqslant \lambda \int w_V(x,\tilde{F})d\tilde{F}(x) + (1-\lambda)\int w_V(x,\tilde{F})dF(x) \\ &= \lambda V(\tilde{F}) + (1-\lambda)V(F),\end{aligned}$$

so that

$$V(F_\lambda) = \int w_V(x,\tilde{F})dF_\lambda(x) \tag{13}$$

Next, fix $\mu \in (0,1)$. By the properties of w_V, we have $V(\tilde{F}) \leqslant \int w_V(x,F_\mu)d\tilde{F}(x)$.

Moreover,

$$\begin{aligned}\lambda V(\tilde{F})+(1-\mu)V(F)=V(F_\mu)&=\int w_V(x,F_\mu)dF_\mu(x)\\&=\mu\int w_V(x,F_\mu)d\tilde{F}(x)+(1-\mu)\int w_V(x,F_\mu)dF(x)\end{aligned}$$

so that, by rearranging the terms,

$$V(\tilde{F})=\int w_V(x,F_\mu)d\tilde{F}(x)+\frac{(1-\mu)}{\mu}\left(\int w_V(x,F_\mu)dF(x)-V(F)\right)\geqslant\int w_V(x,F_\mu)d\tilde{F}(x)$$

where the last inequality follows because $\mu\in(0,1)$ and $\int w_V(x,F_\mu)dF(x)\geqslant V(F)$. With this, we have

$$V(\tilde{F})=\int w_V(x,F_\mu)d\tilde{F}(x). \tag{14}$$

Next, fix $\tilde{x}\in X$. Since $\mu>0$, it follows that there exists $\lambda\in(0,\mu)$ such that $F_\mu+\lambda(\delta_{\tilde{x}}-\tilde{F})\in\mathcal{F}$. Therefore,

$$\begin{aligned}w_V(\tilde{x},F_\mu)-V(\tilde{F})=w_V(\tilde{x},F_\mu)-\int w_V(x,F_\mu)d\tilde{F}(x)&=\lim_{\lambda\downarrow 0}\frac{V(F_\mu+\lambda(\delta_{\tilde{x}}-\tilde{F}))-V(F_\mu)}{\lambda}\\&\leqslant\lim_{\lambda\downarrow 0}\frac{\int w_V(x,\tilde{F})d\left(F_\mu+\lambda(\delta_{\tilde{x}}-\tilde{F})\right)(x)-V(F_\mu)}{\lambda}\\&=\int w_V(x,\tilde{F})d\left(\delta_{\tilde{x}}-\tilde{F}\right)(x)=w_V(\tilde{x},\tilde{F})-V(\tilde{F}),\end{aligned}$$

where the first equality follows by (14), the second equality by Lemma 2, the inequality by the properties of w_V, the third equality by (13), and the last equality by the properties of w_V again. This implies that $w_V(\tilde{x},F_\mu)\leqslant w_V(\tilde{x},\tilde{F})$. Similarly,

$$\begin{aligned}w_V(\tilde{x},\tilde{F})-V(\tilde{F})=w_V(\tilde{x},\tilde{F})-\int w_V(x,\tilde{F})d\tilde{F}(x)&=\lim_{\lambda\downarrow 0}\frac{V(\tilde{F}+\lambda(\delta_{\tilde{x}}-\tilde{F}))-V(\tilde{F})}{\lambda}\\&\leqslant\lim_{\lambda\downarrow 0}\frac{\int w_V(x,F_\mu)d\left(\tilde{F}+\lambda(\delta_{\tilde{x}}-\tilde{F})\right)(x)-V(\tilde{F})}{\lambda}\\&=\int w_V(x,F_\mu)d\left(\delta_{\tilde{x}}-\tilde{F}\right)(x)=w_V(\tilde{x},F_\mu)-V(\tilde{F})\end{aligned}$$

where the first equality follows by the properties of w_V, the second equality follows by Lemma 2, the inequality by the properties of w_V, and the third and the last equality by (14). This implies that $w_V(\tilde{x}, \tilde{F}) \leqslant w_V(\tilde{x}, F_\mu)$ and we conclude that $w_V(\tilde{x}, F_\mu) = w_V(\tilde{x}, \tilde{F})$. Since this is true for all $\mu > 0$ and w_V is continuous it holds also in the limit: $w_V(\tilde{x}, F) = w_V(\tilde{x}, \tilde{F})$. Given that $\tilde{x}$ was arbitrary, the minimizer is unique, which proves that (iii) implies (i).

We next show that if $\gtrsim$ has an adversarial expected utility representation that satisfies uniqueness, then it has an adversarial forecaster representation. Let Y and u denote the adversarial expected utility representation of $\gtrsim$. For all $F \in \mathcal{F}$, let $\hat{y}(F) \in Y$ denote the unique minimizer of $\int u(x, \tilde{y}) dF(x)$. Define $v(x) = \min_{y \in Y} u(x, y)$, $\sigma(x, F) = u(x, y(F)) - v(x)$, and $V(F) = \int v(x) dF(x) + \int \sigma(x, F) dF(x)$. Observe that, by construction, we have $V(F) = \min_{y \in Y} \int u(x, y) dF(x)$, hence V represents $\gtrsim$. Finally, fix $F, \tilde{F} \in \mathcal{F}$ and observe that

$$
\begin{aligned}
\int \sigma(x, F) dF(x) &= \int u(x, y(F)) dF(x) - \int v(x) dF(x) \\
&\leqslant \int u(x, y(\tilde{F})) dF(x) - \int v(x) dF(x) = \int \sigma(x, \tilde{F}) dF(x)
\end{aligned}
$$

showing that σ is a surprise function. To see that if $\gtrsim$ has an adversarial forecaster representation then it can be represented by a function V with continuous local expected utility, let v and σ correspond to the adversarial forecaster representation of $\gtrsim$. The map $w_V : \mathcal{F} \to C(X)$ given by $w_V(x, F) = v(x) + \sigma(x, F)$ is a continuous local utility of $V(F) = \min_{\tilde{F} \in \mathcal{F}} \int w_V(x, \tilde{F}) dF(x)$. ■

Section 2

Proof of Theorem 1. Immediate from the proof of Theorem 4 above. ■

Proposition 1 follows from the following three lemmas. The first two are standard and we relegate their proof to Online Appendix II.

Lemma 3. $\sigma(x, F)$ *defined by a methods of moments forecast is surprise function.*

Lemma 4. *Let* $H(x, \tilde{x}) = \int h(x, s) h(\tilde{x}, s) d\mu(s)$. *Then*

$$
V(F) = \int H(x, x) dF(x) - \int \int H(x, \tilde{x}) dF(x) dF(\tilde{x})
$$

with directional derivatives for relevant directions $(\delta_z - F)$ *given by*

$$
DV(F)(\delta_z - F) =
$$

$$H(z,z) - \int H(x,x)dF(x) - 2\left[\left[\int H(z,x)dF(x)\right) - \int H(x,\tilde{x})dF(x)dF(\tilde{x})\right].$$

When $F \mapsto h(F,\cdot)$ is one-to-one we have an additional property:

Lemma 5. *If $F \mapsto h(F,\cdot)$ is one-to-one and μ assigns positive probability to open sets of S then $V(F)$ is strictly concave.*

Proof. From Lemma 4 it suffices to prove that the positive semi-definite quadratic form $\int\int H(x,\tilde{x})dM(x)dM(\tilde{x})$ is positive definite on the linear subspace of signed measures where $\int dM(x) = 0$. Recall that $H(x,\tilde{x}) = \int h(x,s)h(\tilde{x},s)d\mu(s)$ and suppose that for some $\hat{s}$ we have $\int h(x,\hat{s})dM(x) \neq 0$. Since h is continuous there is an open set $\tilde{S} \subseteq S$ such that $\hat{s} \in \tilde{S}$ and $\int h(x,s)dM(x) \neq 0$ for all $s \in \tilde{S}$. Since μ assigns positive probability to open sets of S this implies that

$$\int\int H(x,\tilde{x})dM(x)dM(\tilde{x}) = \int\left[\left(\int h(x,s)dM(x)\right)\int h(\tilde{x},s)dM(\tilde{x})\right]\mu(s)ds > 0.$$

Hence it suffices for $V(F)$ to be strictly convex that $\int h(x,s)dM(x) \neq 0$ for any signed measure M with $\int dM(x) = 0$. Using the Jordan decomposition we may write $M = \lambda(F - \tilde{F})$ where $F, \tilde{F}$ are probability measures and $\lambda > 0$ if $M \neq 0$. Hence $\int h(x,s)dM(x) = 0$ for $M \neq 0$ if and only if for all s we have

$$h(F,s) = \int h(x,s)dF(x) = \int h(x,s)d\tilde{F}(x) = h_{\tilde{F}}(s).$$

Since $h \to h(F,\cdot)$ is one-to-one this implies $F = \tilde{F}$ and $M = 0$. ■

Appendix II: Section 4

Appendix II.A: General characterization

As in Section 4, we fix an adversarial EU representation (Y,u). First, we consider arbitrary convex and compact subsets $\overline{\mathcal{F}} \subseteq \mathcal{F}$ of feasible lotteries. Let $\mathcal{H}$ denote the set of probability measures over Y, and $ext(\overline{\mathcal{F}})$ the set of extreme points of $\overline{\mathcal{F}}$. By Choquet's theorem, for all $F \in \overline{\mathcal{F}}$, there exists $\lambda \in \Delta\left(ext\left(\overline{\mathcal{F}}\right)\right)$ such that $F = \int \tilde{F}d\lambda(\tilde{F})$.

For every $\overline{\mathcal{F}}$, define $V^*\left(\overline{\mathcal{F}}\right) = \max_{F \in \overline{\mathcal{F}}} V(F)$. By Sion's minmax theorem,

$$V^*\left(\overline{\mathcal{F}}\right) = \max_{F \in \overline{\mathcal{F}}} \min_{y \in Y} \int u(x,y)\, dF(x) = \min_{H \in \mathcal{H}} \max_{F \in ext\left(\overline{\mathcal{F}}\right)} \int \int u(x,y)\, dF(x)\, dH(y).$$

Next we characterize the optimal lotteries given an arbitrary feasibility set $\overline{\mathcal{F}}$. Let $\Lambda_F \subseteq \Delta\left(ext\left(\overline{\mathcal{F}}\right)\right)$ be the set of probability measures over extreme points that satisfy $F = \int \tilde{F} d\lambda\left(\tilde{F}\right)$ for F.

Theorem 7. *Fix $\hat{H} \in \arg\min_{H \in \mathcal{H}} \max_{F \in ext\left(\overline{\mathcal{F}}\right)} \int\int u(x,y)\, dF(x)\, dH()$. Then $\hat{F} \in \arg\max_{F \in \overline{\mathcal{F}}} V(F)$ if and only if for all $\tilde{F} \in ext\left(\overline{\mathcal{F}}\right)$, $V(\hat{F}) \geqslant \int\int u(x,y)\, d\tilde{F}(x)\, d\hat{H}(y)$, and, for all $\tilde{F} \in \bigcup_{\lambda \in \Lambda_{\hat{F}}} \operatorname{supp} \lambda$, $V(\hat{F}) = \int\int u(x,y)\, d\tilde{F}(x)\, d\hat{H}(y)$.*

Proof. Fix $\hat{H}$ as in the statement. Then fix $\hat{F} \in \arg\max_{F \in \overline{\mathcal{F}}} V(F)$, $\tilde{F} \in ext\left(\overline{\mathcal{F}}\right)$, and observe that

$$\begin{aligned} &\int \int u(x,y)\, d\tilde{F}(x)\, d\hat{H}(y) \leqslant \max_{F \in ext\left(\overline{\mathcal{F}}\right)} \int \int u(x,y)\, dF(x)\, d\hat{H}(y) \\ &= \min_{H \in \mathcal{H}} \max_{F \in ext\left(\overline{\mathcal{F}}\right)} \int \int u(x,y)\, dF(x)\, dH(y) = V^*\left(\overline{\mathcal{F}}\right) = V\left(\hat{F}\right), \end{aligned}$$

yielding the first part of the desired condition. Next, observe that

$$\begin{aligned} V^*\left(\overline{\mathcal{F}}\right) &= \max_{F \in ext\left(\overline{\mathcal{F}}\right)} \int \int u(x,y)\, dF(x)\, d\hat{H}(y) \\ &\geqslant \int \int u(x,y)\, d\hat{F}(x)\, d\hat{H}(y) \geqslant \min_{H \in \mathcal{H}} \int \int u(x,y)\, d\hat{F}(x)\, dH(y) = V^*\left(\overline{\mathcal{F}}\right), \end{aligned}$$

By combining the first two chains of inequalities, we have

$$\int \int u(x,y)\, d\hat{F}(x)\, d\hat{H}(y) \geqslant \int \int u(x,y)\, d\tilde{F}(x)\, d\hat{H}(y) \qquad \forall \tilde{F} \in ext\left(\overline{\mathcal{F}}\right). \tag{15}$$

Next, fix $\lambda \in \Lambda_{\hat{F}}$, $F^* \in \operatorname{supp} \lambda$, and assume toward a contradiction that

$$V\left(\hat{F}\right) > \int \int u(x,y)\, dF^*(x)\, d\hat{H}(y).$$

It follows that $\int \left(\int u(x,y)\, d\tilde{F}(x)\right) d\hat{H}(y)\, d\lambda\left(\tilde{F}\right) = \int u(x,y)\, d\hat{F}(x)\, d\hat{H}(y)$

$\geqslant V\left(\hat{F}\right) > \int\int u(x,y)\,dF^{*}(x)\,d\hat{H}(y)$, so there exists $F^{\star} \in \operatorname{supp}\lambda$ and $\varepsilon > 0$ such that

$$\int\int u(x,y)\,dF^{\star}(x)\,d\hat{H}(y) > \int\int u(x,y)\,d\tilde{F}(x)\,d\hat{H}(y)$$

for all $\tilde{F} \in \operatorname{supp}\lambda \cap B_{\varepsilon}(F^{*})$, where $B_{\varepsilon}(F^{*}) \subseteq \mathcal{F}$ is the ball of radius ε (in the Kantorovich-Rubinstein metric metric) centered at F^{*}.

Next, define the probability measure $\lambda^{\star} = \lambda(B_{\varepsilon}(F^{*}))\delta_{F^{\star}} + (1-\lambda(B_{\varepsilon}(F^{*})))\,\lambda(\cdot|B_{\varepsilon}(F^{*})^{c})$ and the lottery $F_{\lambda^{\star}} = \int \tilde{F} d\lambda^{\star}(\tilde{F})$. Then

$$\int\int u(x,y)\,dF_{\lambda^{\star}}(x)\,d\hat{H}(y) = \int\left(\int u(x,y)\,d\tilde{F}(x)\right)d\hat{H}(y)\,d\lambda^{\star}\left(\tilde{F}\right)$$
$$= \lambda(B_{\varepsilon}(F^{*}))\int u(x,y)\,dF^{\star}(x) + (1-\lambda(B_{\varepsilon}(F^{*})))\int\left(\int u(x,y)\,d\tilde{F}(x)\right)d\hat{H}(y)\,d\lambda\left(\tilde{F}|B_{\varepsilon}(F^{*})^{c}\right)$$
$$> \lambda(B_{\varepsilon}(F^{*}))\int\left(\int u(x,y)\,d\tilde{F}(x)\right)d\hat{H}(y)\,d\lambda\left(\tilde{F}|B_{\varepsilon}(F^{*})\right)$$
$$+ (1-\lambda(B_{\varepsilon}(F^{*})))\int\left(\int u(x,y)\,d\tilde{F}(x)\right)d\hat{H}(y)\,d\lambda\left(\tilde{F}|B_{\varepsilon}(F^{*})^{c}\right)$$
$$= \int\left(\int u(x,y)\,d\tilde{F}(x)\right)d\hat{H}(y)\,d\lambda\left(\tilde{F}\right) = \int\int u(x,y)\,d\hat{F}(x)\,d\hat{H}(y)$$

which contradicts equation (15).
Conversely, fix $\tilde{F} \in ext\left(\overline{\mathcal{F}}\right)$ and observe that the implication follows by

$$\begin{aligned} V\left(\hat{F}\right) &\geqslant \max_{\tilde{F}\in ext\left(\overline{\mathcal{F}}\right)} \int\int u(x,y)\,d\tilde{F}(x)\,d\hat{H}(y) \\ &= \min_{H\in\mathcal{H}} \max_{F\in ext\left(\overline{\mathcal{F}}\right)} \int\int u(x,y)\,dF(x)\,dH(y) = V^{*}\left(\hat{\mathcal{F}}\right) \geqslant V\left(\hat{F}\right). \qquad \blacksquare \end{aligned}$$

Note that when $\overline{\mathcal{F}} = \Delta\left(\bar{X}\right)$ for some closed subset $\bar{X}$, the extreme points $ext\left(\overline{\mathcal{F}}\right) = \bar{X}$ are simply point masses over the set of feasible outcomes. In this case, Theorem 7 implies that F is optimal if and only if $V(F) \geqslant \int u(x,y)d\hat{H}(y)$ for all $x \in \overline{X}$, with equality for $x \in \operatorname{supp} F$.

Appendix II.B: Section 4

For the rest of this section, we fix a closed subset $\overline{X} \subseteq X$ and a finite collection of functions $\Gamma = \{g_1, ..., g_k\} \subset C(\overline{X})$. As in the main text, we consider $\mathcal{F}_{\Gamma}(\overline{X}) \subseteq \mathcal{F}$ defined in in equation 4. By Theorem 2.1 in Winkler [1988], $\tilde{F} \in ext\left(\mathcal{F}_{\Gamma}(\overline{X})\right)$ if and only if $\tilde{F} \in \mathcal{F}_\Gamma(\overline{X})$ and $\tilde{F} = \sum_{i=1}^{p} \alpha_i \delta_{x_i}$ for some $p \leqslant k+1$, $\alpha \in \Delta(\{1, ..., p\})$, and $\{x_1, ..., x_p\} \subseteq \bar{X}$ such that the vectors $\{(g_1(x_i), ..., g_k(x_i), 1)\}_{i=1}^{p}$ are linearly independent. For every finite subset of extreme points $\mathcal{E} \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$, define

$$\overline{X}_{\mathcal{E}} = \bigcup_{\tilde{F} \in \mathcal{E}} \operatorname{supp} \tilde{F} \subseteq \overline{X},$$

which is finite from Winkler's theorem. We identify $co(\mathcal{E})$ with the subset of $\mathcal{F}_\Gamma(\overline{X})$ composed of all convex linear combinations of extreme points in $\mathcal{E}$.

Theorem 8. *Fix a finite set $\mathcal{E} \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$, and suppose that Y has the structure of an m-dimensional manifold with boundary, that u is continuously differentiable in y, and that Y and u satisfy the uniqueness property. We have:*

1. *For an open dense full measure set of $w \in \mathcal{W} \subseteq \mathbb{R}^{\overline{X}_{\mathcal{E}}}$, every lottery F that solves $\max_{\tilde{F} \in co(\mathcal{E})} \min_{y \in Y} \int (u(x,y) + w(x)) d\tilde{F}(x)$ has finite support on no more than $(k+1)(m+1)$ points of $\overline{X}_{\mathcal{E}}$.*

2. *There exists a lottery F that solves $\max_{\tilde{F} \in co(\mathcal{E})} \min_{y \in Y} \int u(x,y) d\tilde{F}(x)$ and has finite support on no more than $(k+1)(m+1)$ points of $\overline{X}_{\mathcal{E}}$.*

Proof. Let $|\mathcal{E}| = n$ and $|\overline{X}_{\mathcal{E}}| = r \leqslant n(k+1)$. Because $|\operatorname{supp} \tilde{F}| \leqslant k+1$ for every $\tilde{F} \in ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$, both statements are trivial if $(m+1) \geqslant n$. For $(m+1) < n$, for every $w \in \mathbb{R}^{\overline{X}_{\mathcal{E}}}$, define $u_w(x,y) = u(x,y) + w(x)$ and $V_w(F) = \min_{y \in Y} \int u_w(x,y) dF(x)$, and fix $H_w \in \arg\min_{H \in \mathcal{H}} \max_{F \in \mathcal{E}} \int\int u_w(x,y)\, dF(x)\, dH(y)$. For every $w \in \mathbb{R}^{\overline{X}_{\mathcal{E}}}$, the uniqueness property implies that $H_w = \hat{y}(F_w) \in Y$ for some $F_w \in \operatorname{argmax}_{F \in co(\mathcal{E})} V_w(F)$, and the expectation of each w with respect to each $F \in co(\mathcal{E})$ is well defined since $\operatorname{supp} F \subseteq \overline{X}_{\mathcal{E}}$ by construction.

We first prove point 1. Fix an arbitrary subset of $m+2$ extreme points $\overline{\mathcal{E}} = \left\{\tilde{F}_1, ..., \tilde{F}_{m+2}\right\} \subseteq \mathcal{E}$ and consider the map $U_{\overline{\mathcal{E}}} : Y \times \mathbb{R} \times \mathbb{R}^{\overline{X}_{\mathcal{E}}} \to \mathbb{R}^{m+2}$ defined by

$$U_{\overline{\mathcal{E}}}(y, v, w)_\ell = u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell) \qquad \forall \ell \in \{1, ..., m+2\}$$

where, for every $y \in Y$, $u(\tilde{F}_\ell, y) = \int u(x,y) d\tilde{F}_\ell(x)$ and $w(\tilde{F}_\ell) = \int w(x) d\tilde{F}_\ell(x)$. For every $(y, v) \in Y \times \mathbb{R}$, the derivative of $U_{\overline{\mathcal{E}}}$ with respect to $w \in \mathbb{R}^{\overline{X}_{\mathcal{E}}}$ is a $(m+2) \times r$ matrix whose ℓ-th row coincides with the probability vector $\tilde{F}_\ell$, and because the $\left\{\tilde{F}_1, ..., \tilde{F}_{m+2}\right\}$ are extreme points of $\mathcal{F}_\Gamma(\overline{X})$, this matrix has full rank, so the total derivative of $U_{\overline{\mathcal{E}}}$ has full rank as well. Hence by the parametric transverality theorem[35], for an open dense full measure subset of $\mathbb{R}^{\overline{X}_{\mathcal{E}}}$, denoted $\mathcal{W}(\overline{\mathcal{E}})$, the manifold $(y, v) \mapsto u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell)$ intersects zero transversally. Since $\dim(Y \times \mathbb{R}) < m+2$,there is no (y, v) that solve $u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell) = 0$ for all $\ell \leqslant m+2$. And since $\mathcal{E}$ has finitely many subsets $\overline{\mathcal{E}}$ of $m+2$ extreme points, the intersection $\mathcal{W} = \bigcap_{\overline{\mathcal{E}}} \mathcal{W}(\overline{\mathcal{E}})$ is open dense and of full measure since it is the finite intersection of full measure sets. Thus, for $w \in \mathcal{W}$ and for all $y \in Y$ and $v \in \mathbb{R}$, $u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell) = 0$ for at most $m + 1$ extreme points in $\mathcal{E}$.

Next, fix $w \in \mathcal{W}$, $F^* \in \operatorname{argmax}_{F \in co(\mathcal{E})} V_w$, and $\lambda \in \Lambda_{F*}$. By Theorem 7, for all $\tilde{F} \in \operatorname{supp} \lambda \subseteq \mathcal{E}$, we have $u(\tilde{F}, H_w) - V_w(F^*) + w(\tilde{F}) = 0$. By the previous part of the proof and Theorem 7, we then have $|\operatorname{supp} \lambda| \leqslant m+1$. Therefore, F_w is the linear combination of up to $m+1$ extreme points in $\mathcal{E}$. Each $\tilde{F} \in \mathcal{E}$ is supported on up to $k+1$ points of $\overline{X}_{\mathcal{E}}$, so F_w is supported on up to $(m+1)(k+1)$ points of $\overline{X}_{\mathcal{E}}$.

Now we prove point 2. Because $\mathcal{W}$ is dense in $\mathbb{R}^{\overline{X}_{\mathcal{E}}}$, there exists a sequence $w^n \in \mathcal{W}$ such that $w^n(x) \to 0$ for all $x \in \overline{X}_{\mathcal{E}}$, and a sequence of corresponding optimal lotteries F^n with support of no more than $(m+1)(k+1)$ points of $\overline{X}_{\mathcal{E}}$. Choose a convergent subsequence of $F^n \to F$, and observe that lotteries with no more than $(m+1)(k+1)$ points of support cannot converge weakly to a lottery with larger support. Finally, because V_w is continuous with respect to w, the Berge Maximum Theorem implies that F solves $\max_{F \in co(\mathcal{E})} V_0(F)$, concluding the proof. ∎

Lemma 6. *Suppose that for every finite set $\mathcal{E} \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$ there exists a lottery $F_{\mathcal{E}}$ that solves $\max_{F \in co(\mathcal{E})} V(F)$ and has finite support on no more than $(m+1)(k+1)$ points of $\overline{X}$. Then there exists a lottery F^* that solves $\max_{F \in \mathcal{F}_\Gamma(\overline{X})} V(F)$ and that has finite support on no more than $(m+1)(k+1)$ points of $\overline{X}$.*

Proof of Theorem 2. By Theorem 8 and Lemma 6, there exists a solution F^* that is supported on no more than $(k+1)(m+1)$ points of $\overline{X}$. ∎

[35]See e.g. Guillemin and Pollack [2010].

Proof of Theorem 3. Stationarity implies that $H(x, x)$ is constant, so the directional derivatives from Lemma 4 simplify to

$$DV(F)(\delta_z - F) = -2\left[\left[\int H(z, x)dF(x)\right) - \int H(x, \tilde{x})dF(x)dF(\tilde{x})\right].$$

Since $V(F)$ is continuous and concave on a compact set the maximum exists, and is characterized by the condition that no directional derivative is positive, which is

$$\int H(z, x)dF(x) \geqslant \int H(x, \tilde{x})dF(x)dF(\tilde{x}) \text{ for all } z \in X. \tag{16}$$

This implies the complementary slackness condition: if there exists $z \in A$ such that z satisfies (16) with strict inequality, then $F(A) = 0$.[36]

Next we show that for any $0 < a \leqslant 1$ and interval $A = [0, a)$ there is $z \in A$ such that $\int H(z, x)dF(x)) = \int H(x, \tilde{x})dF(x)dF(\tilde{x})$. By continuity this implies $\int H(0, x)dF(x)) = \int H(x, \tilde{x})dF(x)dF(\tilde{x})$ and by symmetry $\int H(1, x)dF(x)) = \int H(x, \tilde{x})dF(x)dF(\tilde{x})$. Suppose instead that for all $z \in A$ we have $\int H(z, x)dF(x) > \int H(x, \tilde{x})dF(x)dF(\tilde{x})$, and take $a \in X$ to be the supremum of the set $\left\{x' \in X : \int H(x', x)dF(x) > \int H(x, \tilde{x})dF(x)dF(\tilde{x})\right\}$, so that $\int H(a, x)dF(x) = \int H(x, \tilde{x})dF(x)dF(\tilde{x})$. By complementary slackness we have $F(A) = 0$. Positive definiteness, that is $\int H(x, \tilde{x})dF(x)dF(\tilde{x}) > 0$, implies that for some non-trivial interval $x \in [a, b]$ we have $H(a, x) > 0$. Since $H(0, \tilde{x})$ is decreasing and $H(a, a) = \max_{\tilde{x}} H(a, \tilde{x})$ it follows that $H(a, x) > H(0, x)$. Hence $\int H(x, \tilde{x})dF(x)dF(\tilde{x}) = \int H(a, x)dF(x) > \int H(0, x)dF(x))$ violating the first order condition at $z = 0$.

Finally, suppose there is a non-trival open interval $A = (a, b)$ such that $F(A) = 0$. We may assume w.l.o.g. that $\int H(a, x)dF(x) = \int H(x, \tilde{x})dF(x)dF(\tilde{x})$, $\int H(b, x)dF(x)) = \int H(x, \tilde{x})dF(x)dF(\tilde{x})$. Then for $x \notin A$ by strict convexity either $(1/2)(H(a, x) + H(b, x)) > H((a + b)/2, x)$ or both the left-hand side and the right-hand side are equal to zero. The latter cannot be true for a positive measure set of $x \notin A$, so $\int H(x, \tilde{x})dF(x)dF(\tilde{x}) = (1/2)\left(\int H(a, x)dF(x) + \int H(b, x)dF(x)\right) > \int H((a+b)/2, x)dF(x))$ violating the first order condition at $(a + b)/2$. ∎

[36] Assume there is such z with $F(A) > 0$. By continuity of H, there is an open set $\tilde{A} \subseteq A$ such that $z \in \tilde{A}$, $F(\tilde{A}) > 0$, and every $x \in \tilde{A}$ satisfies (16) with strict inequality. Then $\int H(x, \tilde{x})dF(x)dF(\tilde{x}) \int_{\tilde{A}} \int_X H(x, \tilde{x})dF(\tilde{x})dF(x) + \int_{\tilde{A}^c} \int_X H(x, \tilde{x})dF(\tilde{x})dF(x) > F(\tilde{A}) \int H(x, \tilde{x})dF(x)dF(\tilde{x}) + (1 - F(\tilde{A})) \int H(x, \tilde{x})dF(x)dF(\tilde{x}) = \int H(x, \tilde{x})dF(x)dF(\tilde{x})$, a contradiction.

Proof of Corollary 1. By Theorem 15.10 in Aliprantis and Border [2006], there exists a sequence of finitely supported $\mu^n \in \Delta(S)$ such that $\mu^n \to \mu$. The GMM adversarial forecaster representation V^n induced by h together with each μ^n satisfies all the properties of Theorem 2 by defining $Y^n = \prod_{s \in \operatorname{supp} \mu^n} h(X, s) \subseteq \mathbb{R}^{m^n}$, where $m_n = |\operatorname{supp} \mu^n|$. Therefore Theorem 2 implies that for every $n \in \mathbb{N}$, there exists a solution F^n of the problem $\max_{F \in \Delta(\overline{X})} V^n(F)$ that is supported on up to $m_n + 1$ points of $\overline{X}$. Because the constraint set $\Delta(\overline{X})$ is compact and V is continuous, the Berge maximum theorem implies that all the accumulation points of the sequence F^n are solutions of the problem $\max_{F \in \Delta(\overline{X})} V(F)$, where V is the GMM adversarial forecaster representation induced by h and μ. Given that Theorem 3 established that this problem has a unique full-support solution F, it follows that F is the unique accumulation point of F^n. Because $\overline{X}$ is compact, the sequence $\operatorname{supp} F^n$ converges to some set $\hat{X} \subseteq \overline{X}$ in the Haussdorf sense. By Box 1.13 in Santambrogio [2015], $F^n \to F$ implies that $\operatorname{supp} F \subseteq \hat{X}$, and, given that $\operatorname{supp} F = X$, it follows that $\operatorname{supp} F^n \to \overline{X}$. ∎

Appendix III: Section 6

Recall that $\hat{Y}(F) = \operatorname{argmin}_{y \in Y} \int u(x, y) dF(x)$, and let $\mathcal{H}(\hat{Y}(F)) \subseteq \mathcal{H}$ denote the probability measures over $\hat{Y}(F)$.

Proof of Theorem 5. We only prove the equivalence between (i) and (ii) since the other implications are explained in the main text. (i) implies (ii). As a preliminary step we show that, for every $F \in \mathcal{F}$ and for every $G, \hat{G} \in \mathcal{F}$ such that $G \gtrsim_{\mathcal{W}} \hat{G}$, there exists $y \in \hat{Y}(F)$ such that $\int u(x, y) d\hat{G}(x) \leqslant \int u(x, y) dG(x)$. Observe that $\lambda G + (1 - \lambda)F \gtrsim_{\mathcal{W}} \lambda\hat{G} + (1 - \lambda)F$, for all $\lambda \in (0, 1]$. By hypothesis, this implies that $V(\lambda G + (1 - \lambda)F) \geqslant V(\lambda\hat{G} + (1 - \lambda)F)$, for all $\lambda \in (0, 1]$. Next, consider a sequence $\lambda_n \to 0$. For every $n \in \mathbb{N}$, fix two any $\hat{y}_n \in \hat{Y}(\lambda_n\hat{G} + (1 - \lambda_n)F)$, and $y_n \in \hat{Y}(\lambda_n G + (1 - \lambda_n)F)$. Observe that, for every $n \in \mathbb{N}$, we have

$$\begin{aligned}
\int u(x, \hat{y}_n) d(\lambda_n\hat{G} + (1 - \lambda_n)F)(x) &= V(\lambda_n\hat{G} + (1 - \lambda_n)F) \\
&\leqslant V(\lambda_n G + (1 - \lambda_n)F) = \int u(x, y_n) d(\lambda_n G + (1 - \lambda_n)F)(x) \\
&\leqslant \int u(x, \hat{y}_n) d(\lambda_n G + (1 - \lambda_n)F)(x)
\end{aligned}$$

where the last inequality follows since $y_n \in \hat{Y}(\lambda_n G + (1-\lambda_n)F)$. This implies that

$$\lambda_n \int u(x,\hat{y}_n)d\hat{G}(x)+(1-\lambda_n)\int u(x,\hat{y}_n)dF(x) \leqslant \lambda_n \int u(x,\hat{y}_n)dG(x)+(1-\lambda_n)\int u(x,\hat{y}_n)dF(x),$$

which in turn gives $\int u(x,\hat{y}_n)d\hat{G}(x) \leqslant \int u(x,\hat{y}_n)dG(x)$. Take a subsequence $\hat{y}_n$ converging to y. By Lemma 1 $y \in \hat{Y}(F)$ and $\int u(x,y)d\hat{G}(x) \leqslant \int u(x,y)dG(x)$ as desired.

Next, fix $F \in \mathcal{F}$ and define the subset of the signed measures on X in the weak topology $\mathcal{M} = \left\{G - \hat{G} : G, \hat{G} \in \mathcal{F}, G \gtrsim_{\mathcal{W}} \hat{G}\right\}$; for every $M \in \mathcal{M}$, there exists $y \in \hat{Y}(F)$ such that $\int u(x,y)dM(x) \geqslant 0$. Let $\mathcal{U}(F)$ denote the convex hull of $\left\{u(\cdot,y) : y \in \hat{Y}(F)\right\}$. Since $\hat{Y}(F)$ is compact so is $\mathcal{U}(F)$, so $\max_{w\in\mathcal{U}(F)} \int w(x)dM(x)$ exists, and is nonnegative for all $M \in \mathcal{M}$. Thus $\inf_{M\in\mathcal{M}} \max_{w\in\mathcal{U}(F)} \int w(x)dM(x) \geqslant 0$. Now we show that $\mathcal{M}$ is convex and compact. Fix $M, M' \in \mathcal{M}$ and $\lambda \in [0,1]$, and probability measures $G, G', \hat{G}, \hat{G}'$ such that $G \gtrsim_{\mathcal{W}} \hat{G}$, $G' \gtrsim_{\mathcal{W}} \hat{G}'$, such that $M = G - \hat{G}$ and $M' = G' - \hat{G}'$. From the definition of $\gtrsim_{\mathcal{W}}$, $\lambda G + (1-\lambda)G' \gtrsim_{\mathcal{W}} \lambda\hat{G} + (1-\lambda)\hat{G}'$, so $\lambda M + (1-\lambda)M' \in \mathcal{M}$. Moreover, the subset in $\mathcal{F} \times \mathcal{F}$ of points $G, \hat{G}$ such that $G \gtrsim_{\mathcal{W}} \hat{G}$ is closed so it is compact. As subtraction is continuous, $\mathcal{M}$ is the continuous image of a compact set, so it is also compact. Given that $\mathcal{U}(F)$ and $\mathcal{M}$ are compact and convex, and the objective function is bilinear and continuous in each argument separately, the Sion minimax Theorem implies that $\max_{w\in\mathcal{U}(F)} \min_{M\in\mathcal{M}} \int w(x)dM(x) \geqslant 0$.

Letting $v \in \mathcal{U}(F)$ be a solution, we see that $G \gtrsim_{\mathcal{W}} \hat{G}$ implies $\int v(x)d\left(G - \hat{G}\right)(x) \geqslant 0$, that is $\gtrsim_v$ preserves $\gtrsim_{\mathcal{W}}$. Hence, because v continuous, Theorem 2 in Castagnoli and Maccheroni [1999] implies that $v \in \langle\mathcal{W}\rangle$.

(ii) implies (i). Consder $F, G \in \mathcal{F}$ such that $F \gtrsim_{\mathcal{W}} G$, and a probability distribution H over $\hat{Y}(F)$ such that $\int u(x,y)dH(y) \in \langle\mathcal{W}\rangle$. Then $y \in \hat{Y}(F)$ implies $V(G) \leqslant \int u(x,y)dG(x)$ and $\int u(x,y)dF(x) = V(F)$. By Fubini's theorem this implies $V(G) \leqslant \int\int u(x,y)dH(y)dG(x)$ and $\int\int u(x,y)dH(y)dF(x) = V(F)$. Since $\int u(x,y)dH(y) \in \langle\mathcal{W}\rangle$ and $G \preceq_{\mathcal{W}} F$, it follows that

$$V(G) \leqslant \int\int u(x,y)dH(y)dG(x) \leqslant \int\int u(x,y)dH(y)dF(x) = V(F). \qquad \blacksquare$$

References

Aliprantis, C. D. and K. C. Border (2006). *Infinite dimensional analysis.* Springer.

Andersen, S. et al. (2018). “Multiattribute utility theory, intertemporal utility, and correlation aversion.” *International Economic Review*, 59, 537–555.

Battigalli, P., R. Corrao, and M. Dufwenberg (2019). “Incorporating belief-dependent motivation in games.” *Journal of Economic Behavior & Organization*, 167, 185–218.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2018). “Disclosure and choice.” *The Review of Economic Studies*, 85, 1471–1501.

Bergemann, D., T. Heumann, and S. Morris (2022). “Screening with Persuasion.”

Caplin, A. and J. Leahy (2001). “Psychological expected utility theory and anticipatory feelings.” *The Quarterly Journal of Economics*, 116, 55–79.

Castagnoli, E. and F. Maccheroni (1999). *Generalized stochastic dominance and unanimous preferences*. Pitagora.

Cerreia-Vioglio, S. (2009). “Maxmin expected utility on a subjective state space: Convex preferences under risk.”

Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva (2015). “Cautious expected utility and the certainty effect.” *Econometrica*, 83, 693–728.

— (2020). “An explicit representation for disappointment aversion and other betweenness preferences.” *Theoretical Economics*, 15, 1509–1546.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2019). “Deliberately stochastic.” *American Economic Review*, 109, 2425–45.

Cerreia-Vioglio, S., F. Maccheroni, and M. Marinacci (2017). “Stochastic dominance analysis without the independence axiom.” *Management Science*, 63, 1097–1109.

Chatterjee, K. and R. V. Krishna (2011). “A nonsmooth approach to nonexpected utility theory under risk.” *Mathematical Social Sciences*, 62, 166–175.

Chew, S. H., L. G. Epstein, and U. Segal (1991). “Mixture symmetry and quadratic utility.” *Econometrica: Journal of the Econometric Society*, 139–163.

Corrao, R., D. Fudenberg, and D. Levine (2022). “On Concave Functions over Lotteries.”

Deck, C. and H. Schlesinger (2014). “Consistency of higher order risk preferences.” *Econometrica*, 82, 1913–1943.

Dworczak, P. and A. Kolotilin (2022). “The persuasion duality.” *Available at SSRN 3474376*.

Eeckhoudt, L. and H. Schlesinger (2006). “Putting risk in its proper place.” *American Economic Review*, 96, 280–289.

Ely, J., A. Frankel, and E. Kamenica (2015). “Suspense and surprise.” *Journal of Political Economy*, 123, 215–260.

Frankel, A. and E. Kamenica (2019). "Quantifying information and uncertainty." *American Economic Review*, 109, 3650–80.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015). "Stochastic choice and revealed perturbed utility." *Econometrica*, 83, 2371–2409.

Fudenberg, D. and D. K. Levine (2006). "A dual-self model of impulse control." *American economic review*, 96, 1449–1476.

Grandmont, J.-M. (1972). "Continuity properties of a von Neumann-Morgenstern utility." *Journal of economic theory*, 4, 45–57.

Greene, W. H. (2003). *Econometric analysis.* Pearson Education India.

Guillemin, V. and A. Pollack (2010). *Differential topology.* Vol. 370. American Mathematical Soc.

Gul, F. and W. Pesendorfer (2001). "Temptation and self-control." *Econometrica*, 69, 1403–1435.

Kamenica, E. and M. Gentzkow (2011). "Bayesian persuasion." *American Economic Review*, 101, 2590–2615.

Levine, D. K., S. Modica, and J. Sun (2021). "Twin Peaks: Expressive Voting and Soccer Hooliganism."

Loertscher, S. and E. V. Muir (2022). "Monopoly pricing, optimal randomization, and resale." *Journal of Political Economy*, 130, 566–635.

Loseto, M. and A. Lucia (2021). "Deliberate Randomization under Risk." *Available at SSRN 3811073.*

Maccheroni, F. (2002). "Maxmin under risk." *Economic Theory*, 19, 823–831.

Machina, M. J. (1982). """ Expected Utility" Analysis without the Independence Axiom." *Econometrica: Journal of the Econometric Society*, 277–323.

— (1984). "Temporal risk and the nature of induced preferences." *Journal of economic theory*, 33, 199–231.

Menezes, C., C. Geiss, and J. Tressler (1980). "Increasing downside risk." *The American Economic Review*, 70, 921–932.

Mussa, M. and S. Rosen (1978). "Monopoly and product quality." *Journal of Economic theory*, 18, 301–317.

Sarver, T. (2018). "Dynamic Mixture-Averse Preferences." *Econometrica*, 86, 1347–1382.

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

Stanca, L. (2021). "Recursive Preferences, Correlation Aversion, and the Temporal Resolution of Uncertainty."

Winkler, G. (1988). "Extreme points of moment sets." *Mathematics of Operations Research*, 13, 581–587.

Online Appendix I: Sections 3 and 7

This appendix gives proofs of the secondary results stated in the main text. We begin by stating and proving a result that implies Proposition 2, and then spell out the details for the linear case $g(d) = d$ that were sketched in the main text.

For every $F \in \mathcal{F}$, define $\xi_{\beta,F} : [0,1] \to \mathbb{R}$ as $\xi_{\beta,F}(\tilde{p}) = (1-\beta)g'\left(D_2(F)\right)\tilde{p}^2 + \beta g(\tilde{p} - \tilde{p}^2)$ and let $cav(\xi_{\beta,F})$ denote its concavification.

Proposition 6. *For every $\beta \in [0,1]$, there exists an optimal distribution F^*_Δ supported on no more than three beliefs and if (p^*_F, F^*_Δ) is optimal then p^*_F solves*

$$\max_{p\in[0,1]} \left\{p\tilde{v} - (1-\beta)g'\left(D_2(F^*)\right)p^2 + cav(\xi_{F^*})(p)\right\}$$

Moreover, there exist $\underline{\beta}, \overline{\beta} \in (0,1)$ with $\underline{\beta} \leqslant \overline{\beta}$ such that

1. *When $\beta \geqslant \overline{\beta}$, no disclosure is uniquely optimal and p^*_F is an optimal distribution over states if and only if it solves* $\max_{p\in[0,1]}\{p\tilde{v} + \beta g(p-p^2)\}$.
2. *When $\beta \leqslant \underline{\beta}$, full disclosure is uniquely optimal and p^*_F is an optimal distribution over states if and only if it solves* $\max_{p\in[0,1]}\{p\tilde{v} + (1-\beta)g'(p-p^2)(p-p^2)\}$.

Proof of Proposition 6. We first show that $F^* \in \operatorname{argmax}_{F\in\overline{\mathcal{F}}} V_\beta(F)$ if and only if $F^* \in \operatorname{argmax}_{F\in\mathcal{F}(x_0)} \int w_\beta(x, F^*)dF(x)$. Assume that $F^* \in \operatorname{argmax}_{F\in\overline{\mathcal{F}}} V_\beta(F)$. Define $\mathcal{W}_\beta = \{w_\beta(\cdot, F)\}_{F\in\mathcal{F}} \subset C(X)$ and observe that it is compact since w_β is jointly continuous. Because $w_\beta(\cdot, F^*) \in \operatorname{argmin}_{w\in\mathcal{W}_\beta} \max_{F\in\overline{\mathcal{F}}} \int w(x)dF(x)$,

$$\begin{aligned}
\max_{F\in\overline{\mathcal{F}}} \min_{w\in\mathcal{W}_\beta} \int w(x)dF(x) &= \max_{F\in\overline{\mathcal{F}}} V(F) = V(F^*) = \min_{w\in\mathcal{W}_\beta} \int w(x)dF^*(x) \\
&\leqslant \int w_\beta(x, F^*)dF^*(x) \leqslant \max_{F\in\overline{\mathcal{F}}} \int w_\beta(x, F^*)dF(x) \\
&= \min_{w\in\mathcal{W}_\beta} \max_{F\in\overline{\mathcal{F}}} \int w(x)dF(x) = \max_{F\in\overline{\mathcal{F}}} \min_{w\in\mathcal{W}_\beta} \int w(x)dF(x),
\end{aligned}$$

where the last equality follows from Sion minamx theorem given that $\overline{\mathcal{F}}$ is compact and convex. This shows that $F^* \in \operatorname{argmax}_{F\in\overline{\mathcal{F}}} \int w_\beta(x, F^*)dF(x)$.

Conversely, let F^* be such that $F^* \in \operatorname{argmax}_{F\in\overline{\mathcal{F}}} \int w_\beta(x, F^*)dF(x)$ and fix $\hat{F} \in \overline{\mathcal{F}}$. Since V has a continuous local expected utility, $0 \geqslant \int w_\beta(x, F^*)d(\hat{F} - F^*)(x) \geqslant$

$V(\hat{F})-V(F^*)$, where the first inequality follows by assumption. Thus $V(F^*) \geqslant V(\hat{F})$ and, since $\hat{F}$ was arbitrarily chosen, the claim follows.

We now prove the first part of the statement. Let $\beta \in [0,1]$, fix an arbitrary optimal distribution F^* with marginals (p_F^*, F_Δ^*), and denote $q^* = \int p^2 dF_\Delta^*(p)$. Define the set

$$\Delta(p_F^*, q^*) = \left\{F_\Delta \in \Delta(S) : \int p^2 dF(p) = p_F^*, \int p^2 dF(p) = q^*\right\}.$$

Consider the maximization problem:

$$\max_{F_\Delta \in \Delta(p_F^*, q^*)} \int g(p-p^2) dF_\Delta(p). \tag{17}$$

If F_Δ is feasible for Problem 2, it yields a weakly higher utility than F_Δ^* because F_Δ has the same second moment as F_Δ^* and the latter is feasible for Problem 17, so any solution F_Δ of Problem 17 is also a solution of Problem 2. Finally, observe that $\Delta(p_F^*, q^*)$ is a moment set with $k = 2$ moment conditions. The objective function of Problem 17 is linear in F_Δ, so it follows from Theorem 2.1. in Winkler [1988] that there is solution of Problem 17, and hence of Problem 2, that is supported on no more than three points of $\Delta(S)$, concluding the proof of the first statement.

Next, assume that (p_F^*, F_Δ^*) is optimal, that is, there exists an optimal $F^* \in \overline{\mathcal{F}}$ whose marginals are given by (p_F^*, F_Δ^*). By the initial claim and equation 3, we have that (p_F^*, F_Δ^*) solve

$$\max_{p \in \overline{\Delta}, F_\Delta \in \Delta(S): \int \tilde{p} dF(\tilde{p}) = p} \left\{p\tilde{v} + (1-\beta) g'\left(D_2(F^*)\right) \int (\tilde{p}^2 - p^2) dF_\Delta(\tilde{p}) + \beta \int g(\tilde{p} - \tilde{p}^2) dF_\Delta(p)\right\}$$

$$= \max_{p \in \overline{\Delta}} \left\{p\tilde{v} - (1-\beta) g'\left(D_2(F^*)\right) p^2 + \max_{F_\Delta : \int \tilde{p} dF(\tilde{p}) = p} \left[\int (1-\beta) g'\left(D_2(F^*)\right) \tilde{p}^2 + \beta g(\tilde{p} - \tilde{p}^2) dF_\Delta(\tilde{p})\right]\right\} \tag{18}$$

$$= \max_{p \in \overline{\Delta}} \left\{p\tilde{v} - (1-\beta) g'\left(D_2(F^*)\right) p^2 + cav(\xi_{\beta, F^*})(p)\right\}$$

Given the assumptions on g and given that $\overline{\Delta}$ is compact, there exist $\underline{\beta}, \overline{\beta} \in (0,1)$ with $\underline{\beta} \leqslant \overline{\beta}$ such that ξ_{β,F^*} is strictly concave over $\overline{\Delta}$ for all $\beta \geqslant \overline{\beta}$ and ξ_{β,F^*} is strictly convex over $\overline{\Delta}$ for all $\beta \leqslant \underline{\beta}$. We now prove points 1 and 2.

1. When $\beta \geqslant \overline{\beta}$, ξ_{β,F^*} is strictly concave so that $cav(\xi_{\beta,F^*}) = \xi_{\beta,F^*}$. By Corollary 2 in Kamenica and Gentzkow [2011], the inner maximization problem in equation 18 is uniquely solved by $F_\Delta = \delta_p$, that is, no disclosure is uniquely optimal. This

implies that $F^*_\Delta = \delta_{p^*_F}$. Next, we have that $p\tilde{v} - (1-\beta)g'\left(D_2(F^*)\right)p^2 + \xi_{\beta,F^*}(p) = p\tilde{v} + \beta g(p - p^2)$. Given that the optimal (p^*_F, F^*_Δ) are arbitrary, the statement follows.

2. When $\beta \leqslant \underline{\beta}$, ξ_{β,F^*} is strictly convex. By Corollary 2 in Kamenica and Gentzkow [2011], the inner maximization problem in equation 18 is uniquely solved by $F_\Delta = (1-p)\delta_0 + p\delta_1$, that is, full disclosure is uniquely optimal, and $cav(\xi_{\beta,F^*})(\tilde{p}) = (1-\beta)g'\left(D_2(F^*)\right)\tilde{p}$. This implies that $F^*_\Delta = (1-p^*_F)\delta_0 + p^*_F\delta_1$. Next, we have that $p\tilde{v} - (1-\beta)g'\left(D_2(F^*)\right)p^2 + cav(\xi_{\beta,F^*})(p) = p\tilde{v} + (1-\beta)g'(D_2(F^*))(p-p^2)$. Given that the optimal (p^*_F, F^*_Δ) are arbitrary, the statement follows. ■

The linear case Consider the setting of Section 3 with an arbitrary finite state space S and $X = S \times \Delta(S)$. As before, the broadcaster chooses a joint distribution $F \in \mathcal{F}$ over states and conditional beliefs of the watcher, where the feasible joint distributions are those such that the marginal over states is the feasible set $\overline{\Delta} \subseteq \Delta(S)$ and the conditional distribution over states given the belief p is equal to p itself.

The preferences of the watcher over joint distributions of states and beliefs have an adversarial forecaster representation, where preferences over states are given by utility function $v \in \mathbb{R}^S$, and the surprise given the realization $x = (s,p)$ and the forecast $\hat{F}$ of the adversary is

$$\sigma_\beta((s,p),\hat{F}) = (1-\beta)\sigma_0(p,\hat{F}_\Delta) + \beta\sigma_1(s,\hat{F}(\cdot|p)).$$

Here $\hat{F}_\Delta$ and $\hat{F}(\cdot|p)$ are respectively the marginal distribution over $\Delta(S)$ and the conditional distribution over S given p, while σ_0 and σ_1 are surprise functions for the outcome spaces $X_0 = \Delta(S)$ and $X_1 = S$ respectively, and $\beta \in [0,1]$ a parameter capturing the relative importance of interim and ex post surprise.

Clearly, σ_β satisfies all the properties of Definition 1. Indeed,

$$\sigma_\beta((s,p),\delta_{(s,p)}) = (1-\beta)\sigma_0(p,\delta_p) + \beta\sigma_1(s,\delta_s) = 0$$

because σ_0 and σ_1 are surprise functions, and for every $\hat{F} \in \mathcal{F}$,

$$\begin{aligned}\int \sigma_\beta(x, F) dF(x) &= (1-\beta) \int \sigma_0(p, F_\Delta) dF_\Delta(p) + \beta \int\int \sigma_1(s, F(\cdot|p)) dF(s|p) dF_\Delta(p) \\ &\leqslant (1-\beta) \int \sigma_0(p, \hat{F}_\Delta) dF_\Delta(p) + \beta \int\int \sigma_1(s, \hat{F}(\cdot|p)) dF(s|p) dF_\Delta(p) \\ &= \int \sigma_\beta(x, \hat{F}) dF(x)\end{aligned}$$

With this, the preferences of the watcher over joint lotteries F are given by $V_\beta(F) = \int v(s) dF(s,p) + \min_{\hat{F} \in \mathcal{F}} \int \sigma_\beta((s,p), \hat{F}) dF(s,p)$. The broadcaster solves $\max_{F \in \overline{\mathcal{F}}} V(F)$.

Next, consider the binary state case $S = \{0,1\}$, $\Delta(S) = [0,1]$, with $\overline{\Delta} = [0,1]$ and the surprise functions: $\sigma_0(p, \hat{F}_\Delta) = \frac{1}{2}(p - \int \tilde{p} d\hat{F}_\Delta(\tilde{p}))^2$ and $\sigma_1(s, \hat{p}) = (s - \hat{p})^2$. Also assume that the watcher gets utility $\tilde{v} \in \mathbb{R}$ when the state is equal to $s = 1$. For every feasible lottery $F \in \overline{\mathcal{F}}$ let $p_F \in [0,1]$ denote induced probability that $s = 1$ and let F_Δ the marginal over $\Delta(S)$. The definition of $\mathcal{F}$ implies that $p_F = \int p dF_\Delta(p)$. The total payoff of the watcher simplifies to

$$\begin{aligned}V_\beta(F) &= \tilde{v} p_F + (1-\beta) \int (p - p_F)^2 dF_\Delta(p) + \beta \int p(1-p) dF_\Delta(p) \\ &= p_F(\tilde{v} + \beta) - p_F^2(1-\beta) + \int (1-2\beta) p^2 dF_\Delta(p),\end{aligned}$$

which is W's payoff in Section 3 when g is linear $g(d) = d$. Therefore, the maximization problem of the broadcaster simplifies to

$$\max_{F \in \overline{\mathcal{F}}} V_\beta(F) = \max_{p \in [0,1]} \left\{ p(\tilde{v} + \beta) - p^2(1-\beta) + \max_{F_\Delta \in \Delta[0,1] : \int \tilde{p} dF_\Delta(\tilde{p}) = p} \int (1-2\beta) \tilde{p}^2 dF_\Delta(\tilde{p}) \right\}$$

When $\beta < 1/2$, the integrand in the inner maximization is strictly convex, so full disclosure is uniquely optimal. When $\beta > 1/2$, the integrand in the inner maximization is strictly concave, so that no disclosure is uniquely optimal. When $\beta = 1/2$, then the corresponding term disappears, and the watcher is indifferent over all the information structures. And simple computations show that $p_F^* = \max\left\{0, \min\left\{1, \frac{\tilde{v} + \max\{\beta, 1-\beta\}}{2\max\{\beta, 1-\beta\}}\right\}\right\}$ solves the outer maximization problem. $\triangle$

Next, we prove the results in Section 7. We endow $\mathcal{F}$ with the Kantorovich-

Rubinstein norm

$$d_1(F, \tilde{G}) = \sup \left\{ \int f(x) d(F-G)(x) : f \in Lip_1(X) \right\}$$

where $\mathrm{Lip}_1(X)$ is the set of 1-Lipshitz continuous functions over X.

Proof of Proposition 4. By Theorem 4 in Dworczak and Kolotilin [2022], it follows that there exists a compact set $\mathcal{W} \subseteq \mathcal{W}(S^*)$ such that, for every $F \in \mathcal{F}$,

$$V_S(F) = \min_{w \in \mathcal{W}(S^*)} \int w(x) dF(x) = \min_{w \in \mathcal{W}} \int w(x) dF(x).$$

This yields immediately the first part of the statement. Next, assume that R is strictly concave in a and continuously differentiable. Define

$$\mathcal{W}(S, R_a) = \left\{ \max_{a \in A} \{S(\cdot, a) + q(a) R_a(\cdot, a)\} \in C(X) : q \in B(A) \right\}$$

It follows from Lemma 1 in Kolotilin, Corrao, and Wolitzky [2022] (and its proof) that there exists a compact set $\tilde{\mathcal{W}} \subseteq \mathcal{W}(S, R_a)$ such that:

$$V_S(F) = \min_{w \in \mathcal{W}(S, R_a)} \int w(x) dF(x) = \min_{w \in \mathcal{W}} \int w(x) dF(x).$$

This proves the second part of the statement. ■

Proof of Proposition 5. Because X and Θ are compact subsets of Euclidean spaces, and π is continuous (hence uniformly continuous), Proposition 1.11 in Santambrogio [2015] implies that for every $F \in \mathcal{F}$, we have that

$$V_{\pi,Q}(F) = \max_{J \in \Delta(\Theta \times X): marg_\Theta(J) = Q, marg_X(J) = F} \int \pi(\theta, x) dJ(\theta, x) \tag{19}$$

$$= \min_{w \in \mathcal{W}_{\pi,Q}} \int w(x) dF(x) \tag{20}$$

Inspection of the proof of Proposition 1.11 shows rgar we can restrict the minimization in 19 to a compact subset $\mathcal{W} \subset \mathcal{W}_{\pi,Q}$. Under the same assumptions, Proposition 7.17 Santambrogio [2015] implies that $V_{\pi,Q}$ is concave, superdifferentiable, and its superdifferential coincides with the minimizers of the minimization problem in 19. In turn, Theorem 4 yields the first part of the statement.

Next, assume that Q is absolutely continuous and π is continuously differentiable. By Propositions 7.17 and 7.18 in Santambrogio [2015], for every $F \in \mathcal{F}$, there exists a unique (deterministic) maximizer in 19 and the minimization problem in 19 admits a unique minimizer w_F (up to an additive constant) which coincides with the Gateaux derivative of $V_{\pi,Q}$. Next, continuity of the map $F \mapsto W_F$ follows by Theorem 1.52 in Santambrogio. This shows that $V_{\pi,Q}$ is continuously superdifferentiable, hence, by Theorem 4, that it admits an adversarial forecaster representation. Finally, given that π is strictly supermodular and continuously differentiable, it follows from Theorem 2.9 and Remark 2.13 in Santambrogio [2015], that $\xi_F = F^{-1} \circ Q$ is the unique (deterministic) solution of the allocation problem and that $\int_{\xi_F(0)}^{x} \pi_x\left((\xi_F)^{-1}(z), z\right) dz$ is the unique (up to a constant) solution of the minimization problem in 19. ■

Online Appendix II: Ancillary results

This appendix gives proofs of the ancillary results stated in the main appendix.

Proof of Theorem 6. It is immediate that under (ii), condition (iii) for V is obtained by setting $Y = \{\sigma(\cdot, F)\}_{F \in \mathcal{F}}$ and $u(x, y) = y(x)$. It is also immediate that (iii) implies (i) since, for all $F \in \mathcal{F}$ and $y \in \hat{Y}(F)$, we have that $u(x, y)$ is a local expected utility of V at F. We next prove that (i) implies (ii). Given that V has a local expected utility, it follows that $\mathcal{W}_V(F) \neq \varnothing$ for all $F \in \mathcal{F}$. Fix $w_F \in \mathcal{W}_V^*(F)$ for all $F \in \mathcal{F}$ and let B denote the corresponding Bregman divergence as defined in Definition 10. Observe that for every F we have

$$\begin{aligned}\int B(\delta_x, F) dF(x) &= V(F) - \int V(\delta_x) dF(x) - \int w_F(x) dF(x) + \int w_F(x) dF(x)\\ &= V(F) - \int V(\delta_x) dF(x),\end{aligned}$$

so $V(F) = \int V(\delta_x) dF(x) + \int B(\delta_x, F) dF(x)$. Now define $v(x) = V(\delta_x)$ and $\sigma(x, F) = B(\delta_x, F)$ for all x and F. Given that V is continuous, it follows that v is continuous.

Next, we show that σ is a pseudo surprise function. First, observe that, for every F,

$$\sigma(x,F) = V(F) - v(x) - \int w_F(\tilde{x})dF(\tilde{x}) + w_F(x)$$

is continuous in x since v and w_F are continuous. Second, $\sigma(x,\delta_x) = B(\delta_x,\delta_x) = 0$ for every x. Finally, fix $F, \tilde{F} \in \mathcal{F}$ and observe that

$$\begin{aligned}\int \sigma(x,\tilde{F})dF(x) &= V(\tilde{F}) - \int v(x)dF(x) - \int w_{\tilde{F}}(x)d\tilde{F}(x) + \int w_{\tilde{F}}(x)dF(x) \\ &\geqslant V(F) - \int v(x)dF(x) = \int \sigma(x,F)dF(x),\end{aligned}$$

where the inequality follows since $w_{\tilde{F}} \in \mathcal{W}_V^*(\tilde{F})$. This shows that σ is a pseudo surprise function. Thus $V(F) = \int v(x)dF(x) + \min_{\hat{F}\in\mathcal{F}} \int \sigma(x,\hat{F})dF(x)$, as desired. (ii) implies (i).

Next, we prove point 1. Assume that there exist $\hat{v} \neq v$ that satisfy equation 12 for V, possibly with respect to a different pseudo surprise functions σ and $\hat{\sigma}$. Then we have $v(x) = V(\delta_x) = \hat{v}(x) + \min_{\hat{F}\in\mathcal{F}} \hat{\sigma}(x,\hat{F}) = \hat{v}(x) + \hat{\sigma}(x,\delta_x) = \hat{v}(x)$, a contradiction.

We finally prove point 2. First let $\sigma(x,F) = B(\delta_x,F)$ for some Bregman divergence of V. It follows from the proof of (i) implies (ii) that σ satisfies 12 for V. Conversely, assume that a pseudo surprise function σ satisfies 12 for V. Fix F, and for every F and x, define $w_F(x) = v(x) + \sigma(x,F)$. Given that σ is a pseudo surprise function, we have

$$V(F) = \int v(x)dF(x) + \int \sigma(x,F)dF(x) = \int w_F(x)dF(x).$$

Next, fix $\tilde{F} \in \mathcal{F}$ and observe that

$$V(\tilde{F}) \leqslant \int v(x)d\tilde{F}(x) + \int \sigma(x,F)d\tilde{F}(x) = \int w_F(x)d\tilde{F}(x).$$

This proves that $w_F \in \mathcal{W}_V(F)$. Given that F was arbitrarily chosen, it follows that w_F is a local expected utility for V. Consider the corresponding Bregman divergence

B and observe that, for every $\tilde{F} \in \mathcal{F}$,

$$\begin{aligned} B(\tilde{F}, F) &= V(F) - V(\tilde{F}) - \int (v(x) - \sigma(x, F))\, d(F - \tilde{F})(x) \qquad \forall F \in \mathcal{F} \\ &= \int \Big(\sigma(x, F) - \sigma(x, \tilde{F})\Big)\, d\tilde{F}(x) \end{aligned}$$

where the second equality follows from equation 12. With this, we have $B(\delta_x, F) = \sigma(x, F)$ for every x. Given that F was arbitrarily chosen, the implication follows. ■

Proof of Lemma 1. Write

$$\int w^n(x) dF^n(x) - \int w(x) dF(x) = \int (w^n(x) - w(x))\, dF^n(x) + \int w(x) d(F^n(x) - F(x)).$$

For the second term $\int w(x) d(F^n(x) - F(x)) \to 0$ by the definition of weak convergence. Analyzing the first term

$$\int (w^n(x) - w(x))\, dF^n(x) \leqslant \sup |w^n(x) - w(x)| \int dF^n(x) = \sup |w^n(x) - w(x)| \to 0.$$ [37]

Finally, we wish to show that if $F^n \to F$ and w^n are local expected utility functions for F^n with $w^n \to w$ then w is a local expected utility function for F. Suppose we are given $\int w^n(x) d\tilde{F}(x) \geqslant V(\tilde{F})$ and $\int w^n(x) dF^n(x) = V(F^n)$. We have $\int w(x) d\tilde{F}(x) \geqslant V(\tilde{F})$ by the definition of weak convergence. It remains to show that $\int w(x) dF(x) = V(F)$. As $V(F)$ is continuous so it suffices to show that $\int w^n(x) dF^n(x) = \int w(x) dF(x)$. This follows directly from the first result. ■

Proof of Lemma 2. Choose $\mu > 0$ as in the statement and observe that

$$\begin{aligned} \lim_{\lambda \downarrow 0} \frac{V(F + \lambda(\tilde{F} - \overline{F})) - V(F)}{\lambda} &= \frac{1}{\mu} \lim_{\lambda \downarrow 0} \frac{V((1 - \lambda/\mu)F + (\lambda/\mu)(F + \mu(\tilde{F} - \overline{F})) - V(F)}{\lambda/\mu} \\ &= \int w(x, F) d\tilde{F}(x) - \int w(x, F) d\overline{F}(x). \end{aligned}$$ ■

[37] This highlights an important difference between positive and signed measures. In the case of a signed measure it is not true that $\int (w^n(x) - w(x))\, dF^n(x) \leqslant \sup |w^n(x) - w(x)| \int dF^n(x)$ and in fact the lemma is false for signed measures on infinite dimensional spaces.

Proof of Lemma 3. We must show that σ is non-negative, weakly continuous, that $\sigma(x,x) = 0$ and that $\int\sigma(x,F)dF(x) \leqslant \int\sigma(x,G)dF(x)$. Non-negativity is obvious. Since $h(x,s)$ is continous in x we have $F^n \to F$ implies that $h_{F^n}(s)$ converges pointwise to $h_n(s)$. Hence $\left(h(x,s) - \int h(\tilde{x},s)dF^n(\tilde{x})\right)^2$ converges pointwise to $\left(h(x,s) - \int h(\tilde{x},s)dF(\tilde{x})\right)^2$. Given that h is square-integrable over (S,μ), the dominated convergence theorem implies that

$$\int\left(h(x,s) - \int h(\tilde{x},s)dF^n(\tilde{x})\right)^2 d\mu(s) \to \int\left(h(x,s) - \int h(\tilde{x},s)dF(\tilde{x})\right)^2 d\mu(s).$$

For the last property, $\sigma(x,x) = \int\left(h(x,s) - h(x,s)\right)^2 d\mu(s) = 0$, and so

$$\int\sigma(x,G)dF(x) = \int\int\left(h(x,s) - h_G(s)\right)^2 d\mu(s)dF(x) = \int\left(\int\left(h(x,s) - h_G(s)\right)^2 dF(x)\right) d\mu(s).$$

Since mean square error is minimized by the mean for each s,

$$h(F,s) = \int h(x,s)dF(x) \in \arg\min_{H\in\mathbb{R}} \int\left(h(x,s) - H\right)^2 dF(x)$$

implying that $\int\sigma(x,F)dF(x) \leqslant \int\sigma(x,G)dF(x)$. ∎

Proof of Lemma 4. By definition $V(F) = \int\int\left(h(x,s) - h(F,s)\right)^2 d\mu(s)dF(x)$, and simple manipulations show this is equal to

$$\int H(x,x)dF(x) - \int\int\int\left[h(x,s)h(\tilde{x},s)d\mu(s)\right]dF(x)dF(\tilde{x}).$$

We next extend V to the space of signed measures by

$$V(F{+}M) = \int H(x,x)d\left(F(x) + M(x)\right) - \int\int H(x,\tilde{x})d\left(F(x) + M(x)\right)d\left(F(\tilde{x}) + M(\tilde{x})\right)$$

and observe that the cross term is

$$-2\int\left(\int H(x,\tilde{x})dF(\tilde{x})\right)dM(x) = -2\int\int h(x,s)h(\tilde{x},s)d\mu(s)dF(\tilde{x})dM(x)$$

so that

$$V(F+M) = V(F)+\int \left[H(x,x) - 2\int h(x,s)h(\tilde{x},s)d\mu(s)dF(\tilde{x}) \right] dM(x) - \int\int H(x,\tilde{x})dM(x)dM(\tilde{x}).$$

This enables us to compute the directional derivatives. The directional derivative in the direction $M = \delta_z - F$ is given as

$$\begin{aligned} DV(F)(\delta_z - F) &= \int \left[\int h^2(x,s)d\mu(s) - 2\int h(x,s)h(\tilde{x},s)d\mu(s)dF(\tilde{x}) \right] (d\delta_z - dF(x)) \\ &= \int h^2(z,s)d\mu(s) - 2\int h(z,s)h(\tilde{x},s)d\mu(s)dF(\tilde{x}) \\ &- \int h^2(x,s)dF(x)d\mu(s) + 2\int h(x,s)h(\tilde{x},s)d\mu(s)dF(\tilde{x})dF(x). \end{aligned}$$ ■

Before proving Lemma 6, we state and prove an intermediate result.

Lemma 7. *For every $F \in \mathcal{F}_\Gamma(\overline{X})$, there exists a sequence $F^n \to F$ such that each F^n is the convex linear combination of finitely many points in $ext(\mathcal{F}_\Gamma(\overline{X}))$.*

Proof. Define $\mathcal{F}_e = ext(\mathcal{F}_\Gamma(\overline{X}))$ and endow it with the relative topology inherited from the weak topology of $\mathcal{F}$. This makes $\mathcal{F}_e$ metrizable. Next, by the Choquet Theorem, $\mathcal{F}_\Gamma(\overline{X})$ can be embedded in the set $\Delta(\mathcal{F}_e)$ of Borel probability measures over $\mathcal{F}_e$. By Theorem 15.10 in Aliprantis and Border [2006], we have that the subset $\Delta_0(\mathcal{F}_e)$ of finitely supported probability measures over $\mathcal{F}_e$ is dense in $\Delta(\mathcal{F}_e)$. In turn this implies the statement. ■

Proof of Lemma 6. Let $\hat{F}$ solve $\max_{F\in\mathcal{F}_\Gamma(\overline{X})} V(F)$. By Lemma 7, there exists a sequence $\hat{F}^n \to \hat{F}$ such that, for every $n \in \mathbb{N}$, we have $\hat{F}^n \in co(\mathcal{E}^n)$ for some finite set $\mathcal{E}^n \subseteq ext(\mathcal{F}_\Gamma(\overline{X}))$. By Theorem 8, for every $n \in \mathbb{N}$, there exists a lottery $F^n \in co(\mathcal{E}^n)$ that is supported on no more that $(k+1)(m+1)$ points of $\overline{X}$ and such that $V(F^n) \geqslant V(\hat{F}^n)$. Given that $\mathcal{F}_\Gamma(\overline{X})$ is compact, there exists a subsequence of F^n that converges to some lottery $F^* \in \mathcal{F}_\Gamma(\overline{X})$. Since each F^n has support on at most $(k+1)(m+1)$ points, the same is true for F^*. And since V is continuous $V(F^n) \to V(F^*)$ and $V(\hat{F}_n) \to V(\hat{F})$ hence $V(F^*) \geqslant V(\hat{F})$, F^* is optimal. ■

Online Appendix III: Optimization

Online Appendix III.A: finite Y

This section states and proves additional results on the optimization problem of Section 4. Fix an arbitrary compact and convex set $\overline{\mathcal{F}} \subseteq \mathcal{F}$ of feasible lotteries. We start with a simple lemma that establishes the existence of a saddle pair (F^*, y^*).

Lemma 8. *There exists $F^* \in \overline{\mathcal{F}}$ and $y^* \in Y$ such that*

$$\int u(x, y^*) dF^*(x) = V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F) \tag{21}$$

Proof. Because $\overline{\mathcal{F}}$ is compact and V is continuous in the weak topology, it follows that there exists $F^* \in \overline{\mathcal{F}}$ such that $V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F)$. And because Y is compact and u is continuous in y, there exists $y^* \in Y$ such that $\int u(x, y^*) dF^*(x) = V(F^*)$, yielding the statement. ■

For every (F^*, y^*) as in Lemma 8, define the set

$$\overline{\mathcal{F}}(F^*, y^*) = \left\{ F \in \overline{\mathcal{F}} : \forall y \in Y \backslash \{y^*\}, \int u(x, y) dF(x) \geqslant \int u(x, y) dF^*(x) \right\} \tag{22}$$

Observe that $\overline{\mathcal{F}}(F^*, y^*)$ is nonempty, since it contains F^*, and convex since it is defined by (possibly infinitely many) linear inequalities. In addition, $\overline{\mathcal{F}}(F^*, y^*)$ is the intersection of closed sets since $u(., y)$ is a continuous function for all $y \in Y \backslash \{y^*\}$, so it too is closed.

Lemma 9. *Fix (F^*, y^*) as in Lemma 8 and a nonempty, closed, and convex set $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$. The set* $\operatorname{argmax}_{F \in K} \int u(x, y^*) dF(x)$ *is nonempty, convex, and closed.*

Proof. Given that K is nonempty, convex, and closed, hence compact, and the map $F \mapsto \int u(x, y^*) dF(x)$ is linear and continuous, the statement immediately follows. ■

We next state and prove a general, yet simple, result about the existence of maximizers of Problem 21 that are extreme points of convex, closed sets $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$.

Lemma 10. *Fix (F^*, y^*) as in Lemma 8 and a nonempty, closed, and convex set $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$ such that $F^* \in K$. We have*

$$\operatorname*{argmax}_{F \in K} \int u(x, y^*) dF(x) \subseteq \operatorname*{argmax}_{F \in \overline{\mathcal{F}}} V(F). \tag{23}$$

In particular, there exists $F_0 \in ext(K)$ such that $V(F_0) = V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F)$.

Proof. Fix $F^* \in \mathcal{F}$ and $y^* \in Y$ as in Lemma 8 and a nonempty, closed, and convex set $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$. Let $\hat{F} \in \operatorname{argmax}_{F \in K} \int u(x, y^*) dF(x)$. We need to show that $V(\hat{F}) = V(F^*)$. Observe that

$$\int u(x, y) d\hat{F}(x) \geqslant \int u(x, y) dF^*(x) \qquad \forall y \in Y \backslash \{y^*\} \tag{24}$$

since $\hat{F} \in K \subseteq \overline{\mathcal{F}}(F^*, y^*)$. Moreover, we have

$$\int u(x, y^*) d\hat{F}(x) \geqslant \int u(x, y^*) dF^*(x) \tag{25}$$

since $\hat{F} \in \operatorname{argmax}_{F \in K} \int u(x, y^*) dF(x)$ and $F^* \in K$. Then for all $y \in Y$, we have that

$$\int u(x, y) d\hat{F}(x) \geqslant \int u(x, y) dF^*(x) \geqslant V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F) \tag{26}$$

and in particular that $V(\hat{F}) \geqslant \max_{F \in \overline{\mathcal{F}}} V(F)$. Given that $\hat{F} \in \overline{\mathcal{F}}$, we must have $V(\hat{F}) = V(F^*)$, so $\hat{F} \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} V(F)$. This proves the first part of the theorem. The second part immediately follows from the Bauer maximum principle since the map $F \mapsto \int u(x, y^*) dF(x)$ is linear over the convex set K. ∎

Lemma 10 is not very insightful per se since the set $\overline{\mathcal{F}}(F^*, y^*)$ depends on the particular choice of (F^*, y^*). However, whenever we can find a set K as in the statement of Lemma 10 whose extreme points satisfy interesting properties, the theorem lets us conclude that there is an optimizer of the original problem with those properties. We next apply this strategy to optimization problems with additional structure on $\overline{\mathcal{F}}$ and on Y by relying on known characterizations of extreme points of sets of probability measures. For completeness, we report here the original results mentioned.

Theorem 9 (Proposition 2.1 in Winkler [1988])**.** *Fix a convex and closed set $\overline{\mathcal{F}} \subset \mathcal{F}$, an affine function $\Lambda : \overline{\mathcal{F}} \to \mathbb{R}^{n-1}$, and a convex set $C \subset \Lambda(\overline{\mathcal{F}})$. The set $\Lambda^{-1}(C)$ is convex and every extreme point of $\Lambda^{-1}(C)$ is a convex combination of at most n extreme points of $\overline{\mathcal{F}}$.*

We can combine this result with Lemma 10 to obtain the following result.

Theorem 10. *Suppose that Y is finite with m elements. There exists a solution $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} V(F)$ that is a convex combination of at most m extreme points of $\overline{\mathcal{F}}$.*

Proof. Fix (F^*, y^*) as in Lemma 8. Observe that $|Y \backslash \{y^*\}| = m - 1$ by assumption. For simplicity we write $Y \backslash \{y^*\} = \{y_1, ..., y_m - 1\}$. Define the map $\Lambda : \overline{\mathcal{F}} \to \mathbb{R}^{m-1}$ as

$$\Lambda(F)_i = \int u(x, y_i) dF(x) \qquad \forall i \in \{1, ..., m-1\} \tag{27}$$

Also define the convex set

$$C \equiv \Lambda(\overline{\mathcal{F}}(F^*, y^*)) \subseteq \Lambda(\overline{\mathcal{F}}) \tag{28}$$

It is easy to see that $\Lambda^{-1}(C) = \overline{\mathcal{F}}(F^*, y^*)$. Therefore, by Theorem 9 it follows that every extreme point of $\overline{\mathcal{F}}(F^*, y^*)$ is a convex combination of at most n extreme points of $\overline{\mathcal{F}}$. Finally, the statement follows by a direct application of Theorem 10. ∎

We close this section with a result that improves on the bound of the support of an optimal lottery for the optimization problem under moment constraints analyzed in Section 4, provided that Y is finite.

Theorem 11. *Suppose that Y is finite with m elements. For every closed $\overline{X} \subseteq X$, there exists an optimal lottery F^* for problem 5 that has finite support on no more than $k + m$ points of $\overline{X}$.*

Proof of Theorem 11. Let $\overline{\mathcal{F}} = \mathcal{F}_\Gamma(\overline{X})$ for some closed $\overline{X} \subseteq X$, and fix (F^*, y^*) as in Lemma 8. The set $\overline{\mathcal{F}}(F^*, y^*)$ is defined by $k + m - 1$ moment restrictions: k moments restrictions from Γ and $m - 1$ from the definition of $\overline{\mathcal{F}}(F^*, y^*)$. By Lemma 10 there exists $F^* \in \operatorname{ext}(\overline{\mathcal{F}}(F^*, y^*))$ such that $V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F)$. By Winkler's Theorem the each $\tilde{F} \in \overline{\mathcal{F}}(F^*, y^*)$ is supported on up to $k+m$ points of $\overline{X}$ as desired. ∎

Online Appendix III.B: Robust solutions

This section shows that the finite-support property of Theorem 2 generically holds for all solutions of the optimization problem in 5 that are "robust" in the following sense. For every $F \in \mathcal{F}_\Gamma(\overline{X})$, we call a sequence as in Lemma 7 a *finitely approximating sequence of* F.

Definition 12. Fix $w \in C(\overline{X})$ and a lottery F that solves

$$\max_{F \in \mathcal{F}_\Gamma(\overline{X})} \min_{y \in Y} \int u(x,y) + w(x) dF(x)$$

We say that F is a *robust solution at* w if

$$F^n \in \operatorname*{argmax}_{\tilde{F} \in co(\mathcal{E}^n)} \left\{ \min_{y \in Y} \int u(x,y) + w(x) dF(x) \right\}$$

for some approximating sequence $F^n \in co(\mathcal{E}^n)$ of F, with $\mathcal{E}^n$ being any finite set of extreme points generating F^n.

In words, an optimal lottery F is robust if it can be approximated by a sequence of lotteries that are generated by finitely many extreme points and that are optimal within the set of lotteries generated by the same extreme points.

Theorem 12. *Suppose that Y is an m-dimensional manifold with boundary, that u is continuously differentiable in y, and that Y and u satisfy the uniqueness property. For an open dense set of $w \in \overline{\mathcal{W}} \subseteq C(\overline{X})$, every robust solution at w has finite support on no more than $(k+1)(m+1)$ points of $\overline{X}$.*

The proof will use the following lemma.

Lemma 11. *Fix a finite set $\hat{X} \subseteq \overline{X}$ and an open dense subset $\hat{\mathcal{W}}$ of $\mathbb{R}^{\hat{X}}$. The set*

$$\overline{\mathcal{W}} = \left\{ w \in C(\overline{X}) : w_{|\hat{X}} \in \hat{\mathcal{W}} \right\}$$

is open and dense in $C(\overline{X})$, where $w_{|\hat{X}}$ denotes the restriction of w on $\hat{X}$.

Proof. Because $\hat{\mathcal{W}}$ is open, so is $\overline{\mathcal{W}}$. Fix $w \in C(\overline{X})$. Given that $w_{|\hat{X}} \in \mathbb{R}^{\hat{X}}$, there exists a sequence $\hat{w}^n \in \hat{\mathcal{W}}$ such that $\hat{w}^n \to w_{|\hat{X}}$. Next, fix $n \in \mathbb{N}$ large enough so that

$B_{1/n}(\hat{x}) \cap B_{1/n}(\hat{x}') = \varnothing$ for all $\hat{x}, \hat{x}' \in \hat{X}$.[38] By Urysohn's Lemma (see Lemma 2.46 in Aliprantis and Border [2006]), for every $\hat{x} \in \hat{X}$, there exists a continuous function $v^n_{\hat{x}}$ such that $v^n_{\hat{x}}(x) = 0$ for all $x \in \overline{X} \backslash B_{1/n}(\hat{x})$ and $v^n_{\hat{x}}(\hat{x}) = 1$. Now define the continuous function

$$w^n(x) = w(x)(1 - \max_{\hat{x} \in \hat{X}} v^n_{\hat{x}}(x)) + \sum_{\hat{x} \in \hat{X}} \hat{w}^n(x) v^n_{\hat{x}}(x).$$

Clearly, we have $w^n \in \overline{\mathcal{W}}$. Because $\hat{X}$ is finite and $\overline{X}$ is compact, $w^n \to w$ as desired. ∎

We are now ready to prove Theorem 12.

Proof of Theorem 12. Without loss of generality, we assume that $\overline{X} = \bigcup_{F \in \mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F$.[39] Define $\overline{\mathcal{E}} = cl\left(ext\left(\mathcal{F}_\Gamma(\overline{X})\right)\right)$ and consider an increasing sequence of finite sets of extreme points $\mathcal{E}^n \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$ such that $\mathcal{E}^n \uparrow \overline{\mathcal{E}}$. Observe that, by construction, we have $\overline{X}_{\mathcal{E}^n} \uparrow \overline{X}$.[40] For every $n \in \mathbb{N}$, let $\hat{\mathcal{W}}^n$ the open dense subset of $\mathbb{R}^{\overline{X}_{\mathcal{E}^n}}$ that satisfies the property of point 2 in Theorem 8. By Lemma 11 the set

$$\overline{\mathcal{W}}^n = \left\{ w \in C(\overline{X}) : w_{|\overline{X}_{\mathcal{E}^n}} \in \hat{\mathcal{W}}^n \right\}$$

is an open dense subset of $C(\overline{X})$. By the Baire category theorem (see Theorem 3.46 in Aliprantis and Border [2006]), it follows that the set $\overline{\mathcal{W}} = \bigcap_{n \in \mathbb{N}} \overline{\mathcal{W}}^n$ is dense in $C(\overline{X})$.

Next, fix $w \in \overline{\mathcal{W}}$ and a robust optimal lottery F^* for

$$\max_{F \in \mathcal{F}_\Gamma(\overline{X})} \min_{y \in Y} \int u(x, y) + w(x) dF(x)$$

It follows that F^* is the weak limit of a sequences of solutions F^n of the problem

$$\max_{F \in co(\mathcal{E}^n)} \min_{y \in Y} \int u(x, y) + w(x) dF(x)$$

In particular, given that, for every $n \in \mathbb{N}$, we have $w_{|\overline{X}_{\mathcal{E}^n}} \in \hat{\mathcal{W}}^n$, it follows from The-

[38] Here, $B_{1/n}(\hat{x})$ is the open ball centered at $\hat{x}$ and of radius $1/n$.

[39] Assume not, then we could just consider lotteries over the closed set $\overline{X}' = cl\left(\bigcup_{F \in \mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F\right)$.

[40] This follows from the fact that $\overline{X} = \bigcup_{F \in \mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F$ by assumption. See also footnote 39.

orem 8 that F^n is supported on up to $(k+1)(m+1)$ points of $\overline{X}_{\mathcal{E}^n}$. Given that $F^n \to F^*$, it follows that F is supported on up to $(k+1)(m+1)$ points of $\overline{X}$. Given that F^* and w were arbitrarily chosen, the result follows. ∎

Online Appendix IV: additional examples

Example 6 (Example 3 continued)**.** Consider the setting of Example 3 with an adversarial forecaster that minimizes the absolute value of the error: $V(F) = \int_0^1 v(x)dF(x) + \min_{c\in[0,1]} \int |c-x|dF(x)$. For every $F \in \mathcal{F}$ define

$$F^{-1}(t) := \inf\{x \in X : F(x) \geqslant t\}$$

for every $t \in [0,1]$. The quantile correspondence is defined as $\mathcal{Q}^t(F) := [F^{-1}(t), F^{-1}(t^+)]$ for every $t \in [0,1]$ and $F \in \mathcal{F}$, and $\operatorname{argmin}_{c\in[0,1]} \int |c-x|dF(x) = \mathcal{Q}^{0.5}(F)$. Moreover, even if the correspondence $\mathcal{Q}^{0.5}$ has a closed graph, it is not possible to select a median $q^{0.5}(F) \in \mathcal{Q}^{0.5}$ for every F so that the map $F \mapsto q^{0.5}(F)$ is continuous (see for example Huber [2011]). Now assume by contradiction that V has a continuous local expected utility, that is, there exists a jointly continuous function $w(x,F)$ such that $V(F) = \int w(x,F)dF(x)$. This implies that there exists a selection $q^{0.5} \in \mathcal{Q}^{0.5}$ such that $\int |q^{0.5}(F) - x|dF(x) = \int w(x,F) - v(x)dF(x)$ for every F, yielding a contradiction with the discontinuity of $q^{0.5}$. △

Example 7 (Weiner Process Example)**.** In this example, we interpret $x \in [0,1]$ as time. While it is natural to think of $h(\cdot, s)$ as a random function of s with distribution induced by F, there is a dual interpretation in which we think of $h(x, \cdot)$ as a random function of x (a random field) with distribution induced by μ. In this interpretation, the $H(x,\tilde{x})$ are the second (non-central) moments of that random variable between different points $x, \tilde{x}$ in the random field. If, for example, $X = [0,1]$, then this random field is a stochastic process, and $H(x,\tilde{x})$ the second moments of the process h between times $x, \tilde{x}$. It is well known that continuous time Markov process are equivalent to stochastic differential equations and that an underlying measure space S and measure μ can be found for each such process. Specifically, consider the process generated by the stochastic differential equation $dh = -h + dW$ where W is the standard Weiner process on (S,μ) and the initial condition $h(0,s)$ has a standard normal distribution.

Then the distribution of the difference between $h(x, \cdot)$and $h(\tilde{x}, \cdot)$ depends only on the time difference $\tilde{x} - x$, and in particular $H(x, \tilde{x}) = \int h(x, s)h(\tilde{x}, s)d\mu(s) = G(x - \tilde{x})$. In this case $H(0, \tilde{x}) = e^{-\tilde{x}}$, which is non-negative, strictly decreasing and strictly convex. $\triangle$

Online Appendix V: Adversarial forecasters, local utilities, and Gateaux derivatives

In this section, we discuss the relationship between our notion of local utility and the one in Machina [1982]. This is closely related to the differentiability properties of a function V with a continuous local expected utility, which we also discuss.

Fix a continuous functional $V : \mathcal{F} \to \mathbb{R}$. Recall that V has a local expected utility if, for every $F \in \mathcal{F}$ there exists $w(\cdot, F) \in C(X)$ such that $V(F) = \int w(x, F)dF(x)$ and $V(\tilde{F}) \leqslant \int w(x, F)d\tilde{F}(x)$ for all $\tilde{F} \in \mathcal{F}$, We say that this local expected utility is continuous if w is jointly continuous in (x, F).

Proposition 7. *Let $\succsim$ admit a representation V with a local expected utility w and, for every $F \in \mathcal{F}$, let $\succsim_F$ denote the expected utility preference induced by $w(\cdot, F)$. Then $F \succsim_F \tilde{F}$ (resp. $F \succ_F \tilde{F}$) implies that $F \succsim \tilde{F}$ (resp. $F \succ \tilde{F}$).*

Proof. To see the first implication, observe that $V(F) = \int w(x, F)dF(x) \geqslant \int w(x, F)d\tilde{F}(x) \geqslant V(\tilde{F})$. We prove the second implication by contraposition . Let $V(\tilde{F}) \geqslant V(F)$ and observe that $\int w(x, F)d\tilde{F}(x) \geqslant V(\tilde{F}) \geqslant V(F) = \int w(x, F)dF(x)$, implying that $\tilde{F} \succsim_F F$ as desired. ■

Machina [1982] introduced the concept of local utilities for a preference over lotteries with $X \subseteq \mathbb{R}$. For ease of comparison we make assume here that $X = [0, 1]$ for the rest of this section. Machina [1982] says that V has a local utility if, for every $F \in \mathcal{F}$, there exists a function $m(\cdot, F) \in C(X)$ such that

$$V(\tilde{F}) - V(F) = \int m(x, F)d(\tilde{F} - F)(x) + o(||\tilde{F} - F||),$$

where $||\cdot||$ is the L_1-norm. This is equivalent to assume that V is *Frechet differentiable* over $\mathcal{F}$, a strong notion of differentiability.

Our notion of local expected utility is neither weaker nor stronger. If V has a continuous local expected utility, then it is concave, which is not implied by Machina's representation. However, having a continuous local expected utility does not imply Frechet differentiability, as we argue in Example 8 below. We first highlight the relationship between continuous local expected utility and the weaker notion of *Gateaux differentiability*. Recall that V is *Gateaux differentiable*[41] at F if there is a $w(\cdot, F) \in C(X)$ such that at F if there is a $w(\cdot, F) \in C(X)$ such that

$$\int w(x, F)d\tilde{F}(x) - \int w(x, F)dF(x) = \lim_{\lambda \downarrow 0} \frac{V((1-\lambda)F + \lambda\tilde{F}) - V(F)}{\lambda}.$$

If $w(\cdot, F)$ is the Gateaux derivative of V at F we can define the directional derivative operator $DV(F)(\tilde{F} - \overline{F}) = \int w(x, F)d\tilde{F}(x) - \int w(x, F)d\overline{F}(x)$ and we say that the direction $\tilde{F} - \overline{F}$ is *relevant* at F if for some $\lambda > 0$ the signed measure $F + \lambda(\tilde{F} - \overline{F}) \geqslant 0$ is in fact an ordinary measure.

Gateaux differentiability has been used to extend Machina's notion of local utility to functions that are not necessarily Frechet differentiable. In particular, Chew, Karni, and Safra [1987] develop a theory of local utilities for rank-dependent preferences and Chew and Nishimura [1992] extend it to a broader class. Next, we show that if V has a continuous local expected utility, then it is Gateaux differentiable.

Proposition 8. *If V has a continuous local expected utility $w(x, F)$, then V is Gateaux differentiable and, for every $F \in \mathcal{F}$, $w(\cdot, F)$ is the Gateaux derivative of V at F.*

Proof. Fix F and $\tilde{F}$, and for $0 < \lambda \leqslant 1$ and $\overline{F} = (1-\lambda)F + \lambda\tilde{F}$ define

$$\Delta(\lambda) = \frac{V(\overline{F}) - V(F)}{\lambda}.$$

Since $w(x, F)$ is a local expected utility function at F we have $\int w(x, F)d\overline{F}(x) - V(F) \geqslant V(\overline{F}) - V(F)$ so

$$\Delta(\lambda) = \frac{V(\overline{F}) - V(F)}{\lambda} \leqslant \frac{\int w(x, F)d\overline{F}(x) - V(F)}{\lambda} = \int w(x, F)d\tilde{F}(x) - \int w(x, F)dF(x).$$

[41] Here we Huber [2011] and subsequent authors and adapt the standard definition of the Gateaux derivative to only consider directions that lie within the set of probability measures.

On the other hand since $w(x,\overline{F})$ is a local utility function at $\overline{F}$ we have $\int w(x,\overline{F})dF(x) - V(\overline{F}) \geqslant V(F) - V(\overline{F})$ so

$$\Delta(\lambda) = \frac{V(\overline{F}) - V(F)}{\lambda} \geqslant \frac{V(\overline{F}) - \int w(x,\overline{F})dF(x)}{\lambda}$$

$$= \frac{\int w(x,\overline{F})\left(d\overline{F}(x) - dF(x)\right)}{\lambda} = \int w(x,\overline{F})d\tilde{F}(x) - \int w(x,\overline{F})dF(x)$$

$$\to \int w(x,F)d\tilde{F}(x) - \int w(x,F)dF(x)$$

since $w(x,\overline{F})$ is continuous in $\overline{F}$. Putting these together we have

$$\int w(x,F)d\tilde{F}(x) - \int w(x,F)dF(x) \leqslant \lim_{\lambda\downarrow 0}\Delta(\lambda) \leqslant \int w(x,F)d\tilde{F}(x) - \int w(x,F)dF(x)$$

which shows that $w(\cdot,F)$ is the Gateaux derivative of V at F. ■

As an immediate corollary we have:

Corollary 4. *A function V has a continuous local expected utility if and only if it is concave and Gateaux differentiable with a jointly continuous Gateaux derivative $w(x,F)$.*

We conclude by providing an example of an important class of preferences that have a continuous local expected utility but not a local utility in Machina's sense.

Example 8. Consider a function V with a *Yaari's dual representation*, that is, $V(F) = \int x d(g(F))(x)$ for some continuous, strictly increasing, and onto function $g : [0,1] \to [0,1]$. In addition, assume that g is strictly convex and continuously differentiable, for example $g(t) = t^2$. By Lemma 2 in Chew, Karni, and Safra [1987], V is not Frechet differentiable, hence it does not have a Machina's local utility, but we can rewrite the dual representation as $V(F) = \int_0^1 1 - g(F(x))dx$, so V is strictly concave in F. Moreover, by Corollary 1 in Chew, Karni, and Safra [1987], V is Gateaux differentiable with Gateaux derivative $w(x,F) = \int_0^x g'(F(z))dz$, which is jointly continuous in (x,F). Therefore, by Corollary 4, V has a continuous local expected utility and, by Theorem 1, it admits an adversarial forecaster representation. △

References

Aliprantis, C. D. and K. C. Border (2006). *Infinite dimensional analysis*. Springer.

Chew, h. S. and N. Nishimura (1992). "Differentiability, comparative statics, and non-expected utility preferences." *Journal of Economic Theory*, 56, 294–312.

Chew, S. H., E. Karni, and Z. Safra (1987). "Risk aversion in the theory of expected utility with rank dependent probabilities." *Journal of Economic theory*, 42, 370–381.

Dworczak, P. and A. Kolotilin (2022). "The persuasion duality." *Available at SSRN 3474376*.

Huber, P. J. (2011). "Robust statistics." *International encyclopedia of statistical science*. Springer, 1248–1251.

Kolotilin, A., R. Corrao, and A. Wolitzky (2022). "Persuasion as matching." *Available at SSRN 4140246*.

Santambrogio, F. (2015). "Optimal transport for applied mathematicians." *Birkäuser, NY*, 55, 94.

Winkler, G. (1988). "Extreme points of moment sets." *Mathematics of Operations Research*, 13, 581–587.