Randomization, Surprise, and Adversarial Forecasters

Roberto Corrao[*] Drew Fudenberg[†] Florian Mudekereza[§]
David Levine[‡]

August 6, 2026

Abstract

An *adversarial forecaster representation* sums an expected utility function and a measure of surprise that depends on an adversary's forecast. These representations are concave and satisfy a smoothness condition, and any concave utility function that satisfies the smoothness condition has an adversarial forecaster representation. We provide several tractable classes of adversarial forecaster preferences. Concavity typically leads the agent to randomize. We characterize the support size of optimally chosen lotteries, and how it depends on preference for surprise.

JEL codes: D81, D91

Keywords: Surprise; random choice; adversarial forecaster

[*]Department of Economics, Stanford University, rcorrao@mit.edu

[†]Department of Economics, MIT, drew.fudenberg@gmail.com

[‡]Department of Economics, RHUL and WUStL,david@dklevine.com

[§]Department of Economics, MIT

We are grateful to Attila Ambrus, Oğuzhan Çelebi, Simone Cerreia-Vioglio, Philip Dybvig, Glenn Ellison, Joel Flynn, Mira Frick, Duarte Gonçalves, Ryota Ijima, Giacomo Lanzani, Fabio Maccheroni, Massimo Marinacci, Moritz Meyer-ter-Vehn, Stephen Morris, Pietro Ortoleva, Philipp Sadowski, Chris Shannon, Lorenzo Stanca, Tomasz Strzalecki, Jonathan Weinstein, and Alexander Wolitzky for helpful comments and conversations, Sivakorn Sanguanmoo for excellent research assistance. We thank the Gordon B. Pye Dissertation Fellowship, NSF grant SES 1951056 and SES-2417162. MIUR (PRIN 20103S5RN3). and the Leverhulme Trust for financial support.

1 Introduction

The idea that stochastic choices observed in data may come from a deliberate desire to randomize was first advanced by Machina (1985) and is empirically supported by e.g. Agranov and Ortoleva (2017). As expected utility does not allow a preference for randomization, we propose the notion of *continuous local expected utility*, a departure from expected utility that allows preference for randomization. Continuous local expected utility requires that expected utility is approximately correct for comparing lotteries that are close, and that small changes in the lottery do not change these approximations much. By formulating this condition in terms of supporting hyperplanes we guarantee that utility is concave in probabilities, so our representation captures a preference for surprise.

Although *continuous local expected utility* has the properties we desire, it is not easy to work with. This leads us to introduce the more tractable *adversarial forecaster* model, where the agent enjoys being "surprised," and the surprisingness of an outcome is measured by the minimized error of a fictitious adversary who tries to forecast the outcome in advance. We show that this model is equivalent to continuous local expected utility, and that a lottery is optimal for an adversarial forecaster utility if and only if it maximizes the local utility evaluated at that lottery. This alternative way of describing continuous local expected utility lets us bring our intuitions to bear: it is easier to evaluate what people would consider surprising under particular circumstances than the abstract question of how local utility might be expected to vary with the lottery they choose. It is also a powerful mathematical tool that enables us to construct classes of preferences with various properties, such as a preference for continuous densities or preferences that satisfy stochastic dominance properties.

We develop and apply a broad class of continuous local expected utility preference where the forecast error has a finite-dimensional parameterization. With the *generalized methods of moments* (GMM) preferences, if the forecast error is a function of m parameters and there are k moment restrictions, there is an optimal lottery with support of no more than $m + k + 1$ points. The bounded support size of GMM representations is not a necessary consequence of continuous local utility. In particular, optimal lotteries can have "thick" (i.e. uncountable) support in adversarial forecaster preferences that arise when the agent trades off the interests of different potential selves. Moreover, when the selves' preferences are more diverse, more outcomes are included in the support of the optimal lottery.[1]

We also study the monotonicity properties of adversarial forecaster preferences with respect to stochastic orders. 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 respects the stochastic order. We 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 a preference for surprise may lead an agent with a risk-averse expected utility component to have preferences that are overall risk loving.

Finally, we apply adversarial forecaster preferences to the timing and acquisition of information. Extending Ely, Frankel, and Kamenica (2015), we first show that a preference for surprise shapes when information is revealed: agents who place more weight on suspense at the final resolution prefer to delay learning, whereas those who place more weight on earlier suspense prefer early revelation, with intermediate preferences allowing partial revelation. We then combine adversarial forecaster preferences with the rational-inattention model of Matějka and McKay (2015). Information can improve decisions while making realized payoffs easier to forecast. This tradeoff can generate deliberate mistakes even when information is free and can make accuracy nonmonotone in the stakes: information costs drive inattention at low stakes, whereas a preference for surprise reduces accuracy at high stakes. In some environments, a preference for surprise can also create demand for costly information that has no instrumental value.

Related Work Our paper is related to three distinct classes of risk preference models. It is closest to other 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 Ely, Frankel, and Kamenica (2015), where the adversary is left implicit. The adversarial forecaster representation imposes more differentiability properties than those models because of its assumption that for each lottery the forecaster has a unique choice that minimizes expected forecast error. These differentiability properties and the concavity of the representation let us characterize optimal lotteries via

[1] In the one-dimensional case, monotone preferences of this sort correspond to *ordinally independent* preferences (Green and Jullien, 1988) with concave utility.

first-order conditions. When the possible outcomes are an interval of real numbers, Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella (2019) introduce a weakening of expected utility that allows optimal choices to be strictly mixed; adversarial forecaster preferences satisfy their axioms if the local utilities are strictly increasing. The adversarial forecaster model 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).

The preferences studied in Quiggin (1982), Green and Jullien (1988), and Galichon and Henry (2012) all have adversarial forecaster representations provided that a supermodularity condition holds. The preferences induced by temporal risk in Machina (1984) are similar to adversarial forecaster preferences, but have a convex representation and so do not generate a preference for randomization.

Our analysis of monotonicity is related to the work on stochastic orders and preferences over lotteries in e.g. Cerreia-Vioglio (2009), Cerreia-Vioglio, Maccheroni, and Marinacci (2017), and Sarver (2018). Unlike the previous results, we do not assume Fréchet differentiability or finite-dimensional outcomes, and characterize monotonicity with respect to stochastic orders given a representation rather than constructing one.[2]

2 The General Model

We study utility functions that are concave and approximately linear, a modest departure from the linearity of expected utility theory. This section defines the relevant notion of continuous local expected utility and introduces an alternative formulation, adversarial forecaster utility, which decomposes utility into an expected utility component and a preference for being surprised. Our main result is that these two formulations are equivalent, so optimal lotteries can be characterized in terms of the relative preferences for surprise. This gives us a powerful tool for constructing examples and classes of examples and analyzing their properties.

[2]See Online Appendix II.D for a more detailed discussion of these and other related results.

2.1 Setup and Definitions

We analyze preferences over lotteries that are represented by a continuous but not necessarily linear utility function V, where the lotteries $F \in \mathcal{F}$ are Borel probability measures over a compact metric space X of outcomes, endowed with the weak topology on measures. We say that a continuous function $w : X \to \mathbb{R}$ is a *local expected utility* (EU) of V at F if it is a supporting hyperplane, that is $\int w(x)d\tilde{F}(x) \geqslant V(\tilde{F})$ for every $\tilde{F} \in \mathcal{F}$ and $\int w(x)dF(x) = V(F)$.

Note that if V has a local EU w at F and $\int w(x)dF(x) \geqslant \int w(x)d\tilde{F}(x)$, then $V(F) \geqslant V(\tilde{F})$, so that for the class of lotteries that have lower local EU than F, V ranks F versus alternative lotteries *exactly* according to their local EU. This is an important difference with the notion of local utility in (Machina, 1982), where the local EU only *approximately* ranks lotteries the same way as the original utility V. Below we compare our notion to Machina's in more detail.

Note also that if V has a local expected utility at each F then V must be weakly concave, so the induced preferences over lotteries are convex: If the decision maker is indifferent between lotteries F and $\tilde{F}$, they weakly prefer any mixture of the two, i.e they have a preference for randomization.[3] This is a trait for which there is overwhelming evidence (Prelec, 1990), but it is not consistent with expected utility, rank-dependent utility, or cumulative prospect theory.

Expected utility preferences have the same local expected utility at each lottery. We weaken this to require that V has a local expected utility at every lottery F and that the local expected utility varies continuously with the lottery.

Definition 1. V has *continuous local expected utility* if there is a continuous function $w : X \times \mathcal{F} \to \mathbb{R}$ such that $w(x, F)$ is a local expected utility of V at every $F \in \mathcal{F}$.

As we show in Online Appendix IV, V has continuous local expected utility if and only if it is concave and Gâteaux differentiable with continuous Gâteaux derivative. This is a weaker form of differentiability than in Machina (1982), who considers Fréchet derivatives.[4] At the same time, concavity rules out many cases that would fall within Machina's model. For example, the betweenness preferences of Dekel (1986)

[3]This becomes a strict preference for surprise when V is strictly concave, as in some special cases we analyze in Sections 5. See for example Cerreia-Vioglio (2009) for a general analysis of convex preferences under risk.

[4]Observe that to define Fréchet differentiability one needs to endow the set of probability measures with a metric, a property not needed for the Gâteaux derivative.

admit a utility representation that is at the same time quasi-concave and quasi-convex, that is, if the decision maker is indifferent between two lotteries, then they are also indifferent among any mixture of those lotteries. Due to concavity, our decision maker prefers any mixture of the two indifferent lotteries, typically strictly so. As we show below, one way to interpret this is as a preference for surprise: Mixing between two indifferent lotteries enhances the unpredictability, hence the enjoyment, of entertaining the mixture lottery. Continuous local expected utility also rules out preferences with a strictly convex utility representation, such as the convex quadratic utility of Chew, Epstein, and Segal (1991), and even those convex preferences (i.e., quasiconcave utilities) that need not admit a Gâteaux differentiable representation, such as cautious expected utility in Cerreia-Vioglio, Dillenberger, and Ortoleva (2015).[5]

The continuous local utility of V at F is a valid Gâteaux derivative for V, which lets us compute the continuous local utility whenever it exists. Let δ_x denote the Dirac measure on x.

Proposition 1. *If V has continuous local expected utility, then it is concave, and the continuous local expected utility is*[6]

$$w(x,F) = V(F) + \left.\frac{dV((1-\lambda)F + \lambda\delta_x)}{d\lambda}\right|_{\lambda=0}.$$

We now introduce a representation where the agent prefers lotteries whose outcomes are difficult to predict, in the sense that even the best prediction has a large expected error. To formalize this, we use the device of a fictitious adversarial forecaster who picks a forecast over outcomes to minimize the expected forecast error given the agent's chosen lottery.

Definition 2. A *forecast* is an element y of a compact metric space Y. A continuous function $\sigma : X \times Y \to \mathbb{R}$ is a *forecast error* if it is non-negative, $\hat{y}(F) := \operatorname{argmin}_{y\in Y} \int \sigma(x,y)dF(x)$ is a singleton for all $F \in \mathcal{F}$, and $\hat{y}(x) := \hat{y}(\delta_x)$ satisfies $\sigma(x,\hat{y}(x)) = 0$ for all $x \in X$.

[5] For example, when their maxmin representation is with respect to a finite set of utilities, the representation in Cerreia-Vioglio, Dillenberger, and Ortoleva (2015) is not Gâteaux differentiable.

[6] The result says there is a unique way to specify a continuous local expected utility function. This does not imply that there is a unique local expected utility at each point; generally, there will be a continuum of local expected utilities at boundary points. With the topology of weak convergence, all points are on the boundary of $\Delta(X)$ when X is infinite.

This definition allows for quite general forecast spaces. Perhaps the simplest case is where $X \subseteq \mathbb{R}$ and Y is the convex hull of X as in Example 1, so that a forecast corresponds to an expected value of x. We also consider the cases when the forecast is on both the mean and variance of x. We also allow fairly general forecast errors; in Example 1 we use the familiar squared error. We normalize the error to be non-negative, and assume that for any lottery F there is a unique forecast $\hat{y}(F)$ that minimizes the expectation of $\sigma(x, y)$ with respect to F.[7] We interpret σ as the loss function of the adversarial forecaster, and as with the typical loss functions in statistics (e.g., Huber, 2011) we require there is a unique optimal forecast for each lottery. Moreover, since it is easy to forecast the outcome of a lottery that assigns probability 1 to a single outcome, we require that the unique minimizing forecast $\hat{y}(x)$ given a degenerate distribution that assigns probability 1 to x has forecast error 0, i.e. $\sigma(x, \hat{y}(\delta_x)) = 0$. We call $\sigma(x, \hat{y}(F))$ the *surprise* of the decision maker at outcome x in lottery F.

The adversarial forecaster tries to produce good forecasts by minimizing the expected forecast error. That is, the forecaster knows F and chooses y to minimize $\int \sigma(x, y)dF(x)$. We refer to the minimum value $\Sigma(F) = \min_{y \in Y} \int \sigma(x, y)dF(x)$ as the *suspense* of lottery F; it is also the expected surprise of the agent at lottery F.[8] The minimum isolates the residual unpredictability of F: forecast error that could be eliminated by a better forecast reflects a poor forecast rather than suspense in the lottery. Let $C(X)$ denote the set of continuous real-valued functions over X.

Definition 3. A function $V : \mathcal{F} \to \mathbb{R}$ is an *adversarial forecaster utility* if

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

for some $v \in C(X)$, forecast space Y, and forecast error function σ.

This representation can be interpreted as follows: The agent has a preference over outcomes described by the baseline utility function v, and a preference for surprise captured by the forecast error σ. Given a forecast of the adversary, the agent's total utility is the sum of their expected baseline utility and the lottery's suspense. Equation 1 implies that V is continuous and concave, and that $V(\delta_x) = v(x)$. When two

[7] One sufficient condition is that $Y \subseteq \mathbb{R}^k$ is convex and $\sigma(x, y)$ is strictly convex in y for every x.

[8] As in Ely, Frankel, and Kamenica (2015), surprise is a function of realized outcomes and suspense is a measure of uncertainty for the outcome computed before its realization.

lotteries induce the same suspense, their ranking is determined by expected baseline utility. Note that adversarial forecaster preferences do not need to respect first-order stochastic dominance: As in the next example, the agent might prefer a risky (and hence exciting) option to a sure thing that stochastically dominates it.

Example 1. In a sports match, the outcome is $x = 1$ if the preferred team wins and $x = -1$ if it loses. Here lotteries can be represented by the probability $p \in [0, 1]$ that the preferred team wins. Fix $\gamma \geqslant 0$. For $\gamma > 0$, assume that $v(x) = x$ and $\sigma(x, y) = \gamma(x - y)^2$, where the forecast space is $Y = [-1, 1]$; the adversary's unique optimal forecast is $y = 2p - 1$, the expected value of the lottery chosen. When $\gamma = 0$, take Y to be a singleton and $\sigma \equiv 0$. With this, the resulting suspense is $4\gamma p(1 - p)$, which is γ times the variance of the chosen lottery, and the agent's overall utility over lotteries is represented by $V(p) = 2p - 1 + 4\gamma p(1 - p)$. For $\gamma > 0$, simple algebra gives that the optimal lottery is $p(\gamma) := \min \{1/2 + 1/(4\gamma), 1\}$; when $\gamma = 0$, the optimum is $p(0) = 1$. As illustrated in Figure 1, if $\gamma > 1/2$ and the agent can choose any $p \in [0, 1]$, the best lottery $p(\gamma)$ is such that the preferred team might lose, while if $\gamma \leqslant 1/2$ the best lottery assigns probability one to the preferred team winning the match. $\triangle$

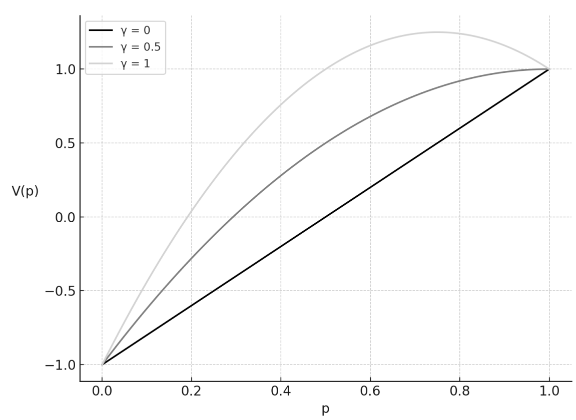

Figure 1: $V(p) = 2p - 1 + 4\gamma p(1 - p)$

2.2 Equivalence of the Two Representations

We now show that the two representations developed above are in fact equivalent.

Theorem 1. *A utility function has continuous local expected utility if and only if it is an adversarial forecaster utility for some forecast space and forecast error function.*

The formal proofs of all results are in the appendix except where otherwise noted. It is easy to see that if V is an adversarial forecaster representation, then $w(\cdot, F) = v + \sigma(\cdot, \hat{y}(F))$ is a local expected utility of V, and the continuity of σ implies that w is continuous. Conversely, given a representation V with continuous local expected utility w, we can set $v(x) = V(\delta_x)$, $Y = \{w(\cdot, F)\}_{F \in \mathcal{F}}$, and $\sigma(x, \hat{y}(F)) = w(x, F) - v(x)$. Because w is continuous, Y is compact, σ is continuous, and σ attains its minimum value of 0 at degenerate lotteries. Finally, by Proposition 1, $w(\cdot, F)$ is the unique affine function tangent to V at each F, so σ satisfies the uniqueness property.

Given V, the baseline utility $v(x) = V(\delta_x)$, the suspense $\Sigma(F) = V(F) - \int v(x)dF(x)$, and the local utility $w(\cdot, F)$ are pinned down, although different forecast spaces and forecast errors (Y, σ) can represent the same preference for surprise. Theorem 1 shows that, among the preferences that exhibit a taste for randomization, those with a continuous local utility are exactly those that capture a taste for surprise in the sense of our adversarial forecaster model.

Relation with support functions Let $\mathcal{U} = \{v + \sigma(\cdot, y) : y \in Y\}$. The adversarial forecaster representation can be written as $V(F) = \min_{u \in \mathcal{U}} \int u(x)dF(x)$. Thus, each forecast generates an expected utility function, and the forecaster selects the one that gives the lowest expected value at the chosen lottery. The uniqueness of the optimal forecast implies that every lottery F selects a unique element $v + \sigma(\cdot, \hat{y}(F))$ of $\mathcal{U}$, which is the continuous local expected utility of V at F. In convex-analysis notation, $V(F) = -h_{-\mathcal{U}}(F)$, where $h_{-\mathcal{U}}$ is the support function of $-\mathcal{U}$. When X is finite, the same minimum is obtained over $co(\mathcal{U})$, and the minimizing utility is unique for every lottery.[9] This lower-envelope interpretation provides the geometric basis for the characterization of optimal lotteries in the next section.

3 Implications for Choice

This section illustrates the implications of adversarial forecaster utility with two applications. The first gives a characterization of optimal choices; and the second relates

[9]When X is finite, so that $\mathcal{F}$ is finite-dimensional, the concavity and continuity of V are equivalent to a generalized version of adversarial forecaster utility where the minimum in equation 1 is replaced by an infimum and the uniqueness property is not necessarily satisfied. Corrao, Fudenberg, and Levine (2024) shows that this infimum cannot, in general, be strengthened to a minimum even when V also satisfies *best outcome independence* (cf. Maccheroni (2002)).

the model to models of stochastic choice.

3.1 Optimal lotteries

Our analysis below makes extensive use of the following result, which extends the usual first-order condition for maximization to our infinite-dimensional setting. It can be thought of as a "fixed-point" characterization of optimal lotteries, because it shows that an optimal lottery maximizes the local expected utility $v(x) + \sigma(x, \hat{y}(F^*))$ which depends on the chosen lottery F^*.

Proposition 2. *If V is an adversarial forecaster utility, then for any convex and compact set of feasible lotteries $\overline{\mathcal{F}} \subseteq \mathcal{F}$,*

$$F^* \in \operatorname*{argmax}_{F \in \overline{\mathcal{F}}} V(F) \iff F^* \in \operatorname*{argmax}_{F \in \overline{\mathcal{F}}} \int v(x) + \sigma(x, \hat{y}(F^*)) dF(x). \tag{2}$$

Maximizing local expected utility is a sufficient condition for maximizing V, whether or not the local utility is continuous. The proof of necessity relies on the fact that V has continuous local expected utility.[10] The proposition says that if F^* is optimum, it is also optimal with respect to the expected utility function $w(x, F^*) = v(x) + \sigma(x, \hat{y}(F^*))$. To see how this works, consider Example 1. Here it is easy to see that the two degenerate lotteries δ_{-1} and δ_1 do not satisfy this condition when $\gamma > 1/2$. Instead, each optimal lottery p must assign strictly positive probability to both outcomes and, by Proposition 2, the local expected utility at p is the same for both outcomes. Some simple algebra shows that the only lottery satisfying this indifference condition is $p(\gamma) = 1/2 + 1/(4\gamma)$.

Proposition 2 can be directly used to link the shape of the forecast error function (or equivalently the local utility) to whether the decision maker randomizes, and if so, how.

Corollary 1. *Suppose that X is a compact and convex subset of a Euclidean space and V has continuous local expected utility w.*

1. *If $w(x, F)$ is strictly quasiconcave in x for every $F \in \mathcal{F}$, then any optimal lottery F^* over $\mathcal{F}$ is degenerate.*

[10] For example, if $X = Y = [-1, 1]$ and $V(F) = \min_{y \in [-1,1]} \int_{-1}^{1} (2y - 1) x dF(x)$, then $F^* = \delta_0$ is optimal over $\mathcal{F}$ for V. However, $w(x, y) = (2y - 1)x$ is a local expected utility for V at F^* for every $y \in [-1, 1]$, yet F^* is strictly suboptimal for all of these local utility functions except for the one corresponding to $y = 1/2$.

2. *If $w(x, F)$ is strictly quasiconvex in x for every $F \in \mathcal{F}$, then any optimal lottery F^* over $\mathcal{F}$ is supported on the extreme points ext(X) of X.*

This result follows from the fact that F^* is optimal if and only if $\text{supp}(F^*) \subseteq \text{argmax}_{x \in X} w(x, F^*)$. Therefore, when each $w(x, F)$ is strictly quasiconcave, each candidate optimal lottery must be supported on the single maximizer of the local utility at that lottery. Similarly, when $w(x, F)$ is strictly quasiconvex, each candidate optimal lottery must be supported on the extreme points of X. We know that risk averse individuals have a preference for degenerate lotteries, and risk loving individuals for extremal points. This generalizes to quasi-concavity provided the local expected utility functions have that property. In the strictly quasi-convex case, the solution can be degenerate, as in the next example.

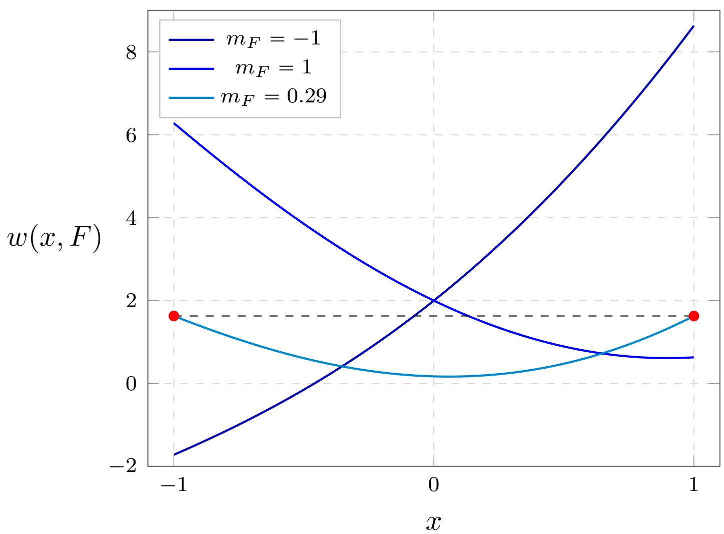

Figure 2: $w(x, F) = (1 - \exp(-x)) + 2(x - m_F)^2$

Example 2. Here we extend the sport-match preferences of Example 1 by allowing risk-averse (CARA) baseline preferences. We set $X = Y = [-1, 1]$, $v(x) = (1 - \exp(-\lambda x))/\lambda$ with $\lambda > 0$, and $\sigma(x, y) = \gamma(x - y)^2$ with $2\gamma > \lambda \exp(\lambda)$. The local utility at any lottery F is $w(x, F) = v(x) + \gamma(x - m_F)^2$, and $w''(x, F) = 2\gamma - \lambda \exp(-\lambda x) > 0$ for every $x \in [-1, 1]$. From Corollary 1, a lottery F^* is optimal only if it is supported on -1 or 1. For any lottery F on $[-1, 1]$, let $q_F(t) := \inf\{x \in [-1, 1] : F(x) \geqslant t\}$ for $t \in (0, 1]$ and $q_F(0) := \inf \text{supp}(F)$, and call q_F its quantile function. The unique solution to this maximization has quantile function

$$q^*(t) = \begin{cases} -1 & \text{if} \quad t \leqslant \frac{1-\overline{q}^*}{2} \\ 1 & \text{if} \quad t > \frac{1-\overline{q}^*}{2}, \end{cases} \tag{3}$$

where $\overline{q}^* = \min\{r(\lambda)/(\lambda\gamma), 1\}$ is the mean of the optimal lottery and $r(\lambda) = (\exp(\lambda) - \exp(-\lambda))/4$. This corresponds to the binary lottery on $\{-1, 1\}$ that assigns probability $(1 + \overline{q}^*)/2$ to 1. The variance of this lottery is $1 - (\overline{q}^*)^2$; when $\overline{q}^* < 1$, it is $1 - (r(\lambda)/(\lambda\gamma))^2$, which is decreasing in λ and increasing in γ: agents with lower baseline risk aversion and more taste for surprise are willing to sacrifice more expected value for higher variance. For comparison, Figure 2 plots the local utilities for $\lambda = 1$ and $\gamma = 2$; at this calibration $r(\lambda)/(\lambda\gamma) = 0.29$, and the corresponding local utility is indifferent between -1 and 1. $\triangle$

Example 3. In the previous example, the optimal lottery was either degenerate or binary. This next example shows how a different specification of the surprise function leads to optimal lotteries with thick (i.e., non-finite) support. As above, consider the adversarial forecaster utility function $V(F) = \mathbb{E}_F[-(1-\gamma)x^2] + \gamma \max_{T\in\Delta(F,U)} \mathrm{Cov}_T(\theta, x)$, where $\Delta(F, U)$ is the set of joint distributions over $X \times \Theta$ with marginals F and U, $\mathrm{Cov}_T(\theta, x)$ is the covariance between x and θ under T, and U is the uniform distribution on $[-1, 1]$ with quantile $q_U(t) = 2t - 1$.

For any fixed F, covariance is maximized by pairing larger values of θ with larger values of x: if T is uniform on $[0, 1]$, the maximizing joint distribution is induced by $(\theta, x) = (q_U(T), q_F(T))$. It follows that $\max_{T\in\Delta(F,U)} Cov_T(\theta, x) = \int_0^1 q_U(t)q_F(t)dt = \frac{1}{2}\mathbb{E}|X - X'|$, where X and X' are independent with distribution F.[11] Consequently, $V(F) = \int_0^1 \left[\gamma(2t-1)q_F(t) - (1-\gamma)q_F(t)^2\right] dt$.

The pointwise maximizer of the preceding integrand is nondecreasing in t, and

[11] To verify the first equality, set $A = (\theta + 1)/2$ and $B = (x + 1)/2$. For every coupling, $\mathbb{E}[AB] = \int_0^1\int_0^1 \Pr(A > s, B > z)ds\,dz \leqslant \int_0^1\int_0^1 \min\{\Pr(A > s), \Pr(B > z)\}ds\,dz$. Equality holds when A and B are generated by the same uniform random variable, because their upper-tail events are then nested. Since their means are fixed, this coupling also maximizes covariance. For the second equality, if $X = q_F(T)$ and $X' = q_F(S)$ for independent uniform T, S, then $\mathbb{E}|X - X'| = 2\int_0^1\int_0^t \big(q_F(t) - q_F(s)\big)dsdt = 2\int_0^1(2t-1)q_F(t)dt$. Finally, the covariance term has an adversarial forecaster representation. Take $Y = \mathcal{F}$ and $\sigma(x, G) = \gamma\int_{-1}^1 \left(G(z) - \mathbf{1}_{\{x\leqslant z\}}\right)^2 dz$. Equivalently, $\sigma(x, G) = \gamma\left[\int|x - z|dG(z) - \frac{1}{2}\int\int|z - z'|dG(z)dG(z')\right]$, so σ is continuous. Moreover, $\int\sigma(x, G)dF(x) - \int\sigma(x, F)dF(x) = \gamma\int_{-1}^1(G(z) - F(z))^2dz$. Thus, $G = F$ is the unique minimizing forecast, $\sigma(x, \delta_x) = 0$, and the minimized expected forecast error is $\gamma\mathbb{E}|X - X'|/2$.

therefore is a valid quantile function. Thus, the quantile function $q_{F*}(t)$ of the optimal lottery solves

$$q_{F*}(t) \in \operatorname*{argmax}_{x \in [-1,1]} \varphi(q_U(t), x) = \operatorname*{argmax}_{x \in [-1,1]} \left\{\gamma(2t-1)x - (1-\gamma)x^2\right\} \tag{4}$$

for all $t \in [0, 1]$. The unique solution of (4) is

$$q_{F*}(t) = \max\left\{-1, \min\left\{1, \frac{\gamma}{1-\gamma}(t - 1/2)\right\}\right\},$$

which induces an optimal distribution that depends on γ and has thick (i.e. uncountable support) for all $\gamma \in (0, 1)$.[12] $\triangle$

3.2 Stochastic Choice

To relate the adversarial forecaster representation to stochastic choice, suppose that the agent has a finite set $\tilde{\mathcal{F}} \subset \mathcal{F}$ of feasible lotteries, and can implement any randomization over these lotteries, so they can choose any lottery over outcomes given by $F_\lambda = \sum_{\tilde{F} \in \tilde{\mathcal{F}}} \tilde{F}\lambda(\tilde{F})$ for $\lambda \in \Delta(\tilde{\mathcal{F}})$. We assume that the agent reduces compound lotteries and so only cares about the final lottery F_λ. The concavity of the adversarial forecaster representation then implies that each compound lottery F_λ is weakly preferred to at least one lottery in the support of λ. This preference is strict when V is strictly concave and λ assigns positive probability to at least two distinct lotteries.[13] Machina (1985) pointed out via examples that intrinsic preferences for randomization, as modeled by convex preferences, generate stochastic optimal choices without the shocks to utility or information. In a setting with real-valued outcomes, Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella (2019) characterizes deliberately stochastic choices generated by preferences that are convex and monotone with respect to first-order stochastic dominance (FOSD). In particular, it shows that a necessary condition for deliberately stochastic choice is a violation of the regularity property that enlarging the choice set cannot increase the probability of pre-existing alternatives.[14]

[12] When γ is less than 2/3 the optimum has no mass points, as $\gamma \to 1$ the probability assigned to mass points goes to 1.

[13] See Proposition 4 for a class of strictly concave adversarial forecaster representations.

[14] Recall that convex preferences correspond to value functions that are quasi-concave, and that a stochastic choice function $P : \mathcal{X} \to \Delta(X)$, where $\mathcal{X} \subseteq 2^X$ is the collection of feasible menus, satisfies *regularity* if $P(x|\overline{X}) \leqslant P(x|\overline{X}')$ for all $x \in \overline{X}' \subseteq \overline{X}$.

The next example shows how a preference for surprise can also lead to violations of regularity with real-valued outcomes.

Example 4. Suppose that $X = Y = [-1, 1]$, that the agent's baseline utility is $v(x) = x$, and that the agent's preference for surprise is $\sigma(x, y) = (x - y)^2$. As in Example 2, the continuous local utility of V is $w(x, F) = v(x) + (x - m_F)^2$, where $m_F = \int_{-1}^{1} \tilde{x} dF(\tilde{x})$. Observe that the agent's ranking of two lotteries with the same expected value $\overline{x}$ is the same as that of an expected utility agent with utility function $w(x) = v(x) + (x - \overline{x})^2$, which is less risk averse than v. Moreover, the stochastic choice rule induced by these preferences need not satisfy regularity. For example, when the set of feasible lotteries is $\Delta\big(\{-1, 0\}\big)$, the unique optimal choice is δ_0, so there is no suspense. In contrast, when the feasible lotteries are $\Delta\big(\{-1, 0, 1\}\big)$, the optimal lottery is $1/4\delta_{-1} + 3/4\delta_1$: the agent tolerates the risk of the bad outcome -1 when it can be accompanied by a larger chance of outcome 1.[15]

$\triangle$

Some classes of adversarial forecaster preferences do satisfy regularity. This is true for example of the weak APU of Fudenberg, Iijima, and Strzalecki (2015). The weak APU representation is defined only for finite sets X; it is given by $V(F) = \sum_{x \in X} \left(u(x)F(x) - c(F(x))\right)$ where $F(x)$ denotes the probability mass function of F and the cost function $c : [0, 1] \to \mathbb{R} \cup \{\infty\}$ is strictly convex and continuously differentiable on $(0, 1)$. To have continuous local expected utility we also need to assume that the derivative c' is bounded and then the local expected utility at F^* is $w(x, F^*) = u(x) - c'(F^*(x))$.[16]

4 Risk aversion and prudence

To see that preference for surprise can alter the agent's risk preference, consider a parametric adversarial asymmetric forecaster utility with $X = Y = [0, 1]$ as in

[15] Note that any binary lottery with $p\delta_1 + (1 - p)\delta_{-1}$ is strictly preferred to a point mass at 0 provided that $p > (3 - \sqrt{5})/4$. Thus the example satisfies the sufficient condition for the failure of regularity in Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella (2019) Theorem 2.

[16] The stronger version of APU requires $\lim_{q \to 0} c'(q) = -\infty$ which is not consistent with continuous local expected utility.

Example 7 with loss function $\rho(z)$, and a baseline utility function $v(x)$:

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

Theorem 2 in Appendix A shows there are optimal lotteries in $\mathcal{F}$ that are supported on at most two points. Moreover, because the local expected utility of the agent is $w(x, F) = v(x) + \rho(x - \hat{y}(F))$ with second derivative $w''(x, F) = v''(x) + \rho''(x - \hat{y}(F))$, Proposition 10 implies that V preserves the MPS order when v is not too concave. This implies that there is an optimal distribution of the form $p^*\delta_1 + (1-p^*)\delta_0$ for some $p^* \in [0, 1]$, and Proposition 2 can be used to explicitly compute p^*.

Suppose in particular that $\rho(z) = \lambda(\exp(z) - z - 1)$ for some $\lambda > 0$. The local expected utility is $w(x, F) = v(x) + \lambda [\exp(x - \hat{y}(F)) - (x - \hat{y}(F)) - 1]$, where $\hat{y}(F) = \log\left(\int_0^1 \exp(x)dF(x)\right)$, that is, the (normalized) cumulant generating function evaluated at 1. With this loss function the agent prefers a positive surprise $x > \hat{y}(F)$ to a negative surprise $x < \hat{y}(F)$ of the same absolute value. The second derivative of the local expected utility at an arbitrary lottery F is $w''(x, F) = v''(x) + \lambda \exp(x - \hat{y}(F))$, so the surprise term shifts the agent's local risk attitudes toward risk loving, with a larger effect at outcomes above $\hat{y}(F)$. Similarly, preference for surprise can alter the agent's higher-order risk preference. The n-th order derivative of each local utility is $w^{(n)}(x, F) = v^{(n)}(x) + \lambda \exp(x - \hat{y}(F))$, so for λ high enough, $w^{(n)} > 0$. From Proposition 10, this implies that higher enjoyment for surprise induces preferences 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, that is 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 whenever λ is high.[17]

As an example, suppose $v(x) = 1 - \exp(-ax)/a$ for $a > 0$. If there is no preference for surprise, the agent has standard CARA EU preferences. Let the forecaster's loss function be $\rho(z) = \lambda(\exp(z) - z - 1)$ for some $\lambda > 0$. If there is no preference for surprise ($\lambda = 0$), the agent is mixed risk averse, as are most of the risk-averse subjects in Deck and Schlesinger (2014). As λ increases, the sign of the even derivatives of the local expected utilities switches from negative to positive, while the sign of the odd

[17] 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 $\tilde{v}^{(n)} = \max_{x \in X} |v^{(n)}(x)|$.

derivatives remains positive, so the agent shifts from mixed risk averse to mixed risk loving when $a \leqslant 1$, whereas no finite λ does so when $a > 1$.[18]

4.1 Relative risk aversion and adversarial forecasters

In this section, we assume that X is a compact interval of real numbers and compare the risk attitudes of an adversarial forecaster utility $V(F)$ and the baseline expected utility $v(x) = V(\delta_x)$ that comes from ignoring the suspense term of V. We first recall the notion of relative risk attitudes introduced in Chew, Karni, and Safra (1987) for non-EU preferences.

Definition 4. For all $v \in C(X)$ and $F, \tilde{F} \in \mathcal{F}$, we say that F is a *simple compensated spread* of $\tilde{F}$ with respect to v if $\int v(x)dF(x) = \int v(x)d\tilde{F}(x)$ and there exists $x_0 \in X$ such that $F(x) \geqslant \tilde{F}(x)$ for all $x < x_0$ and $F(x) \leqslant \tilde{F}(x)$ for all $x \geqslant x_0$. A continuous utility V is *more risk loving* than a continuous expected utility v if $V(F) \geqslant V(\tilde{F})$ whenever F is a *simple compensated spread* of $\tilde{F}$ with respect to v.

In other words, F is a *simple compensated spread* of $\tilde{F}$ with respect to v if an expected-utility agent with utility v is indifferent between these two lotteries and F increases in the upper tail and decreases in the lower tail with respect to $\tilde{F}$. We can use Definition 4 to compare the risk attitudes of an adversarial forecaster utility with those of an agent with the same baseline utility v but with no preference for surprise. This isolates the role of the forecast error σ in the agent's attraction for risk.

Corollary 2. *Fix an adversarial forecaster utility V with representation v and σ, and suppose the baseline utility v is strictly increasing. Then V is more risk loving than v if and only if for all $F \in \mathcal{F}$ there exists a continuous and convex function $\phi_F : v(X) \to \mathbb{R}$ such that*

$$\sigma(x, \hat{y}(F)) = \phi_F(v(x)). \tag{5}$$

This result follows from combining Proposition 3 in Cerreia-Vioglio, Maccheroni, and Marinacci (2017) and our Proposition 10. The former implies that V is more risk loving than v if and only if V preserves $\gtrsim_{\mathcal{W}_v}$ where $\mathcal{W}_v$ is the set of functions $\phi(v(x))$ where $\phi(t)$ is continuous and convex. By Proposition 10 this is equivalent to $v + \sigma(\cdot, \hat{y}(F)) \in \mathcal{W}_v$ for all $F \in \mathcal{F}$, which is equivalent to 5. A sufficient condition for

[18]For even n, $w^{(n)}(x, F) = -a^{n-1}\exp(-ax) + \lambda \exp(x - \hat{y}(F))$. If $a \leqslant 1$, $\lambda > ea$ makes this expression positive for every $n \geqslant 2$, $x \in [0, 1]$, and $F \in \mathcal{F}$. If $a > 1$, then $w^{(n)}(0, F) \leqslant -a^{n-1} + \lambda$, which is negative for all sufficiently large even n.

5 is that the baseline utility v is strictly increasing and concave and that the surprise function is increasing and convex in x.[19] The next example applies Corollary 2 to GMM preferences.

Example 5. Consider the GMM preferences $V(F) = \int v(x)dF(x) + \lambda \min_{y \in Y} \int (v(x) - y)^2 dF(x)$, with $Y \equiv v(X)$ and $\lambda \geqslant 0$. Here the adversarial forecaster tries to predict the realized utility of the agent, so $\sigma(x, \hat{y}(F)) = \lambda \left(v(x) - \int v(\tilde{x})dF(\tilde{x})\right)^2$, which satisfies (5). When $\lambda = 0$, take the forecast space to be a singleton and set $\sigma \equiv 0$. Thus, for every $\lambda > 0$ the adversarial forecaster utility V is relatively more risk-loving than the baseline expected utility v. △

4.2 Repeated choices and correlation aversion

In experimental settings, participants appear to be averse to correlation between risks across time periods (see, e.g., Andersen et al., 2018). Here we show that allowing the adversarial forecaster to make second-period forecasts after observing the first-period realizations is a special case of the one-period adversarial forecaster model and that the induced preferences can exhibit correlation aversion.

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 action $y_0 \in Y_0$ with no additional information about F, and then takes $y_1 \in Y_1$ after observing the realization of x_0, where 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. Moreover, assume that, for every F, the adversary has a unique optimal strategy. The agent knows that the adversary picks $y_1 \in Y_1$ conditional on the realization of x_0, and their induced preferences over lotteries on $X_0 \times X_1$ take into account the adversary's conditional best response.

Example 6. Let $X_0 = \{0, 1\}$, $X_1 = [0, 1]$, and $v(x_0, x_1) = v_0(x_0) + v_1(x_1)$. The adversary chooses a strategy $y = (y_0, y_1(0), y_1(1)) \in Y = [0, 1]^3$, where y_0 forecasts x_0 and $y_1(i)$ forecasts x_1 after observing $x_0 = i$. Fix $\eta > 0$ and define the forecast error by $\sigma_\eta((x_0, x_1), y) = (x_0 - y_0)^2 + (x_1 - y_1(x_0))^2 + \eta(y_1(1) - y_1(0))^2$. The last term is a small cost that uniquely determines the forecast following a zero-probability

[19] To see this, observe that v^{-1} is strictly increasing and convex when v is strictly increasing and concave. If in addition σ is increasing and convex in x, we can rewrite $\sigma(x, y) = \sigma(v^{-1}(v(x)), y)$, so 5 is satisfied by the continuous and convex function $\phi_F(t) = \sigma(v^{-1}(t), \hat{y}(F))$.

history, so σ_η is a forecast error.[20] The resulting adversarial forecaster utility is $V_\eta(F) = \int[v_0(x_0) + v_1(x_1)]dF(x_0, x_1) + \Sigma_\eta(F)$, where, writing $p_i = F_0(i)$,

$$\Sigma_\eta(F) = \mathrm{Var}_F(x_0) + \mathrm{Var}_F(x_1) - \frac{\mathrm{Cov}_F(x_0, x_1)^2}{p_0 p_1 + \eta}.$$

Let $F_0 \otimes F_1$ denote the joint lottery that has the same marginal distributions as F but makes x_0 and x_1 independent. Since the baseline utility is additively separable, replacing F by $F_0 \otimes F_1$ leaves expected baseline utility and both marginal variances unchanged, while setting covariance equal to zero. Thus, $V_\eta(F_0 \otimes F_1) \geqslant V_\eta(F)$, with strict inequality whenever $\mathrm{Cov}_F(x_0, x_1) \neq 0$. Therefore, when all joint lotteries are feasible, every optimal lottery has zero covariance. If both realizations of x_0 occur with positive probability, this is equivalent to $\int \tilde{x}_1 dF_1(\tilde{x}_1 \mid 1) = \int \tilde{x}_1 dF_1(\tilde{x}_1 \mid 0)$, so the adversary's optimal second-period forecast is independent of x_0.[21] $\triangle$

4.3 Stochastic Orders

Online Appendix II.D gives a general characterization of when adversarial forecaster preferences preserve an arbitrary stochastic order. It shows this property reduces to whether the local utility induced by the forecaster's optimal prediction respects the order at every lottery. This provides a unified test for when a preference for surprise preserves or overturns familiar behavioral rankings.

5 The Bounds of Optimal Randomization

This section analyzes the extent of optimal randomization in a class of adversarial forecaster models called *generalized method of moments*, which is based on the idea that the forecaster makes a prediction by targeting certain moments of the lottery. These are in a sense the simplest examples of non-linear preferences with continuous local expected utility, because they are always quadratic. We first show that when the forecaster only cares about finitely many moments, the extent of optimal randomization is bounded by the number of moments. We then show that as the number

[20] The expected loss is strictly convex in y for every F. At a degenerate lottery on (x_0, x_1), its unique minimum is attained at $y_0 = x_0$ and $y_1(0) = y_1(1) = x_1$, and the minimum loss is zero.

[21] For $p_0 p_1 > 0$, if $\mu_i = \int \tilde{x}_1 dF_1(\tilde{x}_1 \mid i)$, direct minimization gives $\hat{y}_1(1) - \hat{y}_1(0) = p_0 p_1(\mu_1 - \mu_0)/(p_0 p_1 + \eta)$.

of moments grows to infinity, the extent of randomization can increase to the point where optimal lotteries randomize over the entire space of outcomes.

5.1 Generalized Method of Moments Preferences

Suppose X is a closed bounded subset of a Euclidean space, and let S be any finite set. Given any continuous 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 define the forecast space $Y = \prod_{s \in S} h(\mathcal{F}, s) \subseteq \mathbb{R}^S$, a compact set, and call it the set of *generalized moments*: these correspond to functions of the outcomes that are indexed by s. We now suppose that the adversary's goal is to match the collection of moments of F given by $h(x, s)$.

Definition 5. The loss function σ is based on the *generalized method of moments* (GMM)[22] if there is a finite probability space (S, μ) and a continuous function $h : X \times S \to \mathbb{R}$ such that

$$\sigma(x, y) = \sum_{s \in S} (h(x, s) - y(s))^2 \mu(s). \tag{6}$$

Proposition 4 shows that when μ has full support, any loss function σ based on the generalized methods of moments is a forecast error, and moreover the associated suspense is quadratic. If $X \subseteq \mathbb{R}$ and $S = \{s_1, ..., s_m\}$ is a finite set of non-negative integers, taking $h(x, s_j) = x^{s_j}$ for every $s_j \in S$ yields the standard method of moments.[23] The simplest case is $X \subseteq \mathbb{R}$ and $S = \{1\}$, as in Examples 1 and 4.

5.2 Moment Restrictions and Bounds on Optimal Supports

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. We focus on the extent of optimal randomization, that is, the size of the supports of optimal distributions.

To define the support restrictions formally, fix a closed subset $\overline{X} \subseteq X$ and a finite collection of k continuous functions $\Gamma = \{g_1, ..., g_k\} \subseteq C(X)$ together with the

[22]In econometrics, the generalized method of moments means minimizing a quadratic loss function on the data under the constraint that a number of generalized moment restrictions are satisfied.

[23]See for example Chapter 18 in Greene (2003).

feasibility set

$$\mathcal{F}_\Gamma = \left\{ F \in \Delta\left(\overline{X}\right) : \forall g_i \in \Gamma, \int g_i\left(x\right) dF\left(x\right) \leqslant 0 \right\}, \tag{7}$$

which we assume is non-empty. For example, if x is money, then $\int x dF(x) = 0$ is the constraint that the agent must choose a fair lottery.[24] When the constraint set Γ is empty, the agent can pick any lottery with support $\overline{X}$.

When an expected-utility agent maximizes over $\mathcal{F}_\Gamma$, there are optimal lotteries that are extreme points of the set $\mathcal{F}_\Gamma$, and all the extreme points of this set are supported on at most $k+1$ points of $\overline{X}$. We now generalize this idea to the class of GMM preferences and show that the upper bound on the support of an optimal lottery depends on the number of moments defining the adversary's loss function as well as the number of moment restrictions.

Proposition 3. *When the agent has GMM utility with m moments and Γ contains k moment restrictions, there is an optimal lottery that puts positive probability on at most $m+k+1$ points.*

The proof is relatively simple, so we present it here. First, given the forecast error in equation 6, for every F the optimal forecast is $\hat{y}(F) = (h(F,s))_{s\in S}$ and define $\overline{Y} = \hat{y}(\mathcal{F}_\Gamma)$. Then the optimization problem becomes

$$\begin{aligned}\max_{F\in\mathcal{F}_\Gamma} V(F) &= \max_{F\in\mathcal{F}_\Gamma} \int \left\{ v(x) + \sum_{s\in S} \left(h(x,s) - h(F,s)\right)^2 \mu(s) \right\} dF(x) \\ &= \max_{\overline{y}\in\overline{Y}} \max_{F\in\mathcal{F}_\Gamma : \hat{y}(F)=\overline{y}} \int \left\{ v(x) + \sum_{s\in S} \left(h(x,s) - \overline{y}(s)\right)^2 \mu(s) \right\} dF(x).\end{aligned}$$

Now fix an optimal solution $\overline{y}^*$ of the outer maximization problem.[25] F^* solves the

[24] Equality constraints can be incorporated in (7) by considering both $g_i(x)$ and $-g_i(x)$.

[25] An optimizer $\overline{y}^* \in \overline{Y}$ exists because the function

$$R(\overline{y}) = \max_{F\in\mathcal{F}_\Gamma : \hat{y}(F)=\overline{y}} \int \left\{ v(x) + \sum_{s\in S} \left(h(x,s) - \overline{y}(s)\right)^2 \mu(s) \right\} dF(x)$$

is upper semicontinuous by Berge Maximum theorem.

original problem and is consistent with $\overline{y}^*$ if and only if it solves

$$\max_{F \in \mathcal{F}_\Gamma : \hat{y}(F) = \overline{y}^*} \int \left\{ v(x) + \sum_{s \in S} \left(h(x,s) - \overline{y}^*(s)\right)^2 \mu(s) \right\} dF(x) \tag{8}$$

which is linear in F: The agent behaves as if they were maximizing expected utility over all lotteries that have the optimal values of the relevant moments. Because the objective in (8) is linear in F, there is a solution in the set of extreme points of $\{F \in \mathcal{F}_\Gamma : h(F, \cdot) = \overline{y}^*\}$. This set is defined by the k inequality moment restrictions in Γ and the m equality restrictions $h(F, s) = \overline{y}^*(s)$, $s \in S$. Corollary 6 therefore implies that each of its extreme points is supported on at most $k + m + 1$ points of $\overline{X}$, as claimed.

5.3 Infinitely many moments and unbounded randomization

So far we have analyzed the minimal support of optimal lotteries under the assumption that the parameter space Y is finite dimensional. When Y is infinite dimensional, every optimal distribution can have "thick" (i.e. non-finite) support. We will show this for a class of GMM preferences with infinitely many relevant moments.

We extend GMM utilities by considering a compact probability space (S, μ) endowed with its Borel sigma algebra and a continuous function $h : X \times S \to \mathbb{R}$. As before, the forecast space is the compact set $Y = \{h(F, \cdot) \in C(S) : F \in \mathcal{F}\}$,[26] and the forecast error is

$$\sigma(x, y) = \int \left(h(x,s) - y(s)\right)^2 d\mu(s). \tag{9}$$

We now show that σ based on the generalized methods of moments is a forecast error, and moreover, the associated suspense is quadratic.

Proposition 4. *If μ has full support, then any loss function σ based on the generalized methods of (infinite) moments is a forecast error, 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 Σ and V are strictly concave

[26] The Arzelà–Ascoli theorem implies that Y is compact: Y is closed because $\mathcal{F}$ is compact, it is uniformly bounded because $\mathcal{F} \times S$ is compact, and it is equicontinuous because h is continuous.

When $X \subseteq \mathbb{R}$, the generalized moments $h(x,s) = \exp(sx)$, $-s_0 \leqslant s \leqslant s_0$, with $s_0 > 0$ and μ of full support on $[-s_0, s_0]$, correspond to the moment generating function and so induce a one-to-one mapping. Proposition 4 implies that the GMM preference $V(F)$ induced by this class is strictly concave, thereby exhibiting a strict preference for randomization. For example, Theorem 2 in Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella (2019) implies that the stochastic choice induced by these preferences is in general non-degenerate. Moreover, strict concavity of V is inconsistent with EU. Chew, Epstein, and Segal (1991) show that quadratic utilities satisfy mixture symmetry, a weakening of both independence and betweenness that is more consistent with some experimental findings such as Hong and Waller (1986). Proposition 3 in Dillenberger (2010) shows that preferences represented by quadratic utilities satisfy negative certainty independence (NCI) only if they are expected utility preferences.[27] Therefore, when V is induced by a GMM forecast error and its continuous local utility is not constant, the corresponding preference does not satisfy NCI. This is intuitive, because GMM suspense depends on the moments of the entire lottery: mixing both alternatives with the same third lottery changes the forecaster's best prediction and hence changes their suspense by different amounts, so the original ranking can reverse.

When a GMM utility has infinitely many moments, we call H its *kernel*. Next, we provide sufficient conditions for an infinite GMM utility to induce a unique optimal lottery that has full support over the outcome space. For simplicity, we consider the one-dimensional case and do not impose exogenous moment restrictions on the feasible lotteries.

Proposition 5. *Assume that* $X = [0,1]$, $\Gamma = \varnothing$, *and* $v \equiv 0$. *If the kernel of the GMM representation* $H(x,\tilde{x}) = \int h(x,s)h(\tilde{x},s)d\mu(s) = G(x-\tilde{x})$ *is strictly positive definite on zero-mass signed measures,*[28] *and* $H(0,\tilde{x})$ *is non-negative and strictly decreasing when positive, and strictly convex in* $\tilde{x}$, *then there is a unique optimal lottery, and it has full support over* X.

The restriction $v \equiv 0$ isolates the pure value of surprise. With a nonzero baseline utility, strict positive definiteness of the kernel still yields uniqueness, but full support

[27] NCI says that if the agent prefers a lottery to a certain outcome, this ranking is not reversed by mixing each option with a third lottery.

[28] That is, $\iint H(x,\tilde{x})dM(x)dM(\tilde{x}) > 0$ for every nonzero finite signed Borel measure M on X such that $M(X) = 0$.

need not hold: if v puts sufficiently large weight on some outcomes relative to the suspense term, the unique optimum may have smaller support. For the hypotheses of the result to be satisfied, the GMM adversary must have a sufficiently large set of forecasts, as in Example 11 in Online Appendix III. The proof uses Proposition 4 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 which, together with the assumptions on H, imply that there cannot be an open set in X to which F assigns probability zero.

We close this section with a corollary of Proposition 5; its proof is in Online Appendix I.

Corollary 3. *Maintain the assumptions of Proposition 5, and let F denote the unique fully supported solution. There exists a sequence of GMM representations V^n with $|S^n| \in \mathbb{N}$, and a sequence of lotteries F^n such that each F^n is optimal for V^n, is supported on at most $|S^n| + 1$ points, and $F^n \to F$ 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 every outcome is in the support of the optimal lottery.

5.4 Parametric Adversarial Forecaster and Randomization

For GMM preferences, the forecast space is the set of generalized moments, $\prod_{s \in S} h(\mathcal{F}, s)$. Because S is finite, Y is a subset of a Euclidean space, so $\hat{y}(F) = (h(F, s))_{s \in S}$ can be interpreted as a finite-dimensional parameter that represents the best forecast for F. Parametric adversarial forecaster representations generalize these properties and let us relax the symmetric loss function of the GMM case.

Definition 6. A forecast error σ is *parametric* if $Y \subseteq \mathbb{R}^m$ for some finite integer m, and σ is continuously differentiable in y. A function $V : \mathcal{F} \to \mathbb{R}$ is a *parametric adversarial forecaster utility* if it has an adversarial forecaster representation with a parametric forecast error.

This definition is tailored for utility functions with an explicit adversarial forecaster representation (v, σ). However, the proof of Theorem 1 constructs a forecast error σ starting from a continuous local expected utility w of V. It is then straightforward to provide conditions on w that imply V is parametric.[29]

Example 7. This example relaxes the GMM representation by allowing the forecaster to have different preferences regarding positive and negative surprises. Let $X = [0,1]$, set $Y = X$, and 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 $z > 0$. The utility function

$$V(F) = \int_0^1 v(x)dF(x) + \min_{y \in Y} \int_0^1 \rho(x-y)dF(x), \tag{10}$$

arises from the parametric adversarial forecaster representation with forecast error $\sigma(x,y) = \rho(x-y)$. Here $\hat{y}(F)$ is the unique minimizer in (10), and the suspense function is $\Sigma(F) = \int \rho(x - \hat{y}(F))dF(x)$ which can be interpreted as an index of the dispersion of F, without requiring symmetry. As we show in Section 4, this can lead to more "prudent" preferences than those induced by $\rho(z) = z^2$. $\triangle$

Example 8. Proposition 7 in Fudenberg, Iijima, and Strzalecki (2015) shows that V has an APU representation $V(F) = \sum_{x \in X}[u(x)F(x) - c(F(x))]$ if and only if it has an AVU representation, that is,

$$V(F) = \sum_{x \in X} u(x)F(x) + \min_{y \in \mathbb{R}^X} \sum_{x \in X} \left[y(x) + \sum_{\tilde{x} \in X} \phi(y(\tilde{x})) \right] F(x), \tag{11}$$

where $\phi(z) := c^*(-z)$ and c^* is the convex conjugate of the original cost function c. If c' is bounded, let I be the compact interval given by the closure of $-c'((0,1))$; all minimizing coordinates in (11) lie in I. For $y \in I^X$, set $\ell(x,y) = y(x) + \sum_{\tilde{x} \in X} \phi(y(\tilde{x}))$ and let $a = \min_{y \in I^X} \ell(x,y)$, which is independent of x. Then, $v(x) = u(x) + a$ and $\sigma(x,y) = \ell(x,y) - a$ give an adversarial forecaster representation: σ is non-negative, its minimum at every degenerate lottery is zero, and strict convexity of c implies that the minimizing forecast is unique. Since a is constant, the baseline utility v differs from u only by an additive constant.

[29] It is sufficient that $w(x,F) = \tilde{w}(x, P(F))$ for some continuous functions $P : \mathcal{F} \to Y$ and $\tilde{w} : X \times Y \to \mathbb{R}$, where $Y = P(\mathcal{F})$ has the structure of a compact m-dimensional C^1 manifold with boundary, $\tilde{w}$ is continuously differentiable in y, and $y \mapsto \tilde{w}(\cdot, y)$ is one-to-one on Y. The reason $y \mapsto \tilde{w}(\cdot, y)$ being one-to-one is sufficient is that if two parameters minimize expected forecast error, the uniqueness argument in Theorem 1 would imply that they generate the same local utility; one-to-one parametrization then implies that the parameters coincide.

The AVU representation in Equation 11 is an example of a parametric adversarial forecaster utility where the parameter space Y has dimension $m = |X|$. In the spirit of the nested logit model, we generalize the APU representation by considering uncertain taste shocks $y \in \mathbb{R}^X$ that are the same across certain classes of outcomes in X, reducing the dimensionality of the parameter space.[30] Fix a partition $\mathcal{P} = \{E_1, \ldots, E_m\}$ of X, and let Y be a compact convex m-dimensional C^1 manifold with boundary contained in the set of vectors in I^X that are measurable with respect to this partition. Writing $y(x) = r_j$ for $x \in E_j$, the adversary minimizes $\sum_{j=1}^m [F(E_j)r_j + |E_j|\phi(r_j)]$. Since ϕ is strictly convex on I, this objective is strictly convex on Y, so its minimizer is unique. For each $x \in X$, let $a(x) = \min_{y \in Y} \ell(x, y)$, define $v(x) = u(x) + a(x)$ and $\sigma(x, y) = \ell(x, y) - a(x)$. Then, $V(F)$ as in (11), with $\mathbb{R}^X$ replaced by Y, has an adversarial forecaster representation whose parameter space has dimension m. △

Theorem 2 in Appendix A extends the support bounds of Proposition 3 to parametric adversarial preferences that are not GMM, as in Example 7. For the asymmetric losses in Example 7, we show there is an optimal lottery supported on no more than $2(k + 1)$ points given the k moment restrictions in Γ. In our generalization of AVU in Example 8, the number of parameters coincides with the number of cells of the partition describing the uncertainty shock. This can be smaller than the cardinality of $|\overline{X}|$, so Theorem 2 yields a meaningful bound on the support of optimal lotteries. When the optimum is unique and X is finite, our result gives a testable prediction on the support of stochastic choices induced by AVU preferences with perfectly correlated shocks.[31]

6 Applications to Information

We now apply adversarial forecaster preferences to two information problems. We first study when uncertainty is resolved and then how much information an agent chooses to acquire.

[30]Nested logit divides items into groups, with correlated value shocks within each group. Here we consider the case where items within the same group have perfectly correlated shocks.

[31]Online Appendix II.B provides an extension to the case of infinite X.

6.1 The timing of information

We now study the implications of adversarial forecaster preferences for the timing of information acquisition. We consider an agent choosing among two-stage lotteries that represent distributions over both states and intermediate information. We show that the "preference for surprise" in Ely, Frankel, and Kamenica (2015) (EFK) has an adversarial forecaster representation.[32] EFK assumed that the agent does not directly care about the outcome itself; our approach makes it easy to model agents that directly care about the outcome.

Let $\Omega = \{0, 1\}$ be a binary state space. The outcomes $x = (p, \omega)$ are elements of $X = \Delta(\Omega) \times \Omega$. The agent chooses an element F of the set $\overline{\mathcal{F}}$ of lotteries satisfying $\Pr_F(\omega = 1 \mid p) = p$ almost surely; we write $dF(p)$ for the marginal distribution of p and let $p_F = \Pr_F(\omega = 1) = \int p dF(p)$. The lottery resolves over two time periods: In period 1, the agent learns their interim belief $p \in \Delta(\Omega)$, and in period 2, $\omega \in \Omega$ realizes. Following EFK, we assume that the agent has preference for suspense in both periods, and assume that the preference for first-period suspense is $V_1(F) = g(E(F))$ for $E(F) = \int_0^1 \frac{1}{2}||p - p_F||^2 dF(p) = \int_0^1 p^2 dF(p) - p_F^2$ and some function $g : \mathbb{R} \to \mathbb{R}$ that is twice continuously differentiable, strictly increasing, and concave, with $g(0) = 0$. The resulting utility function V_1 has continuous local utility, so it is an adversarial forecaster representation by Theorem 1. The suspense in period 2 given interim belief p is $\sum_{\omega \in \Omega} \frac{1}{2}||\delta_\omega - p||^2 p(\omega)$, and the expected period-2 suspense is

$$V_2(F) = \int g\left(\sum_{\omega \in \Omega} \frac{1}{2}||\delta_\omega - p||^2 p(\omega)\right) dF(p) = \int_0^1 g(p - p^2) dF(p).$$

Finally, we generalize EFK so that the agent gets direct utility equal to $\tilde{v} \in \mathbb{R}$ when the realized state is $\omega = 1$ and direct utility 0 when $\omega = 0$; the case $\tilde{v} = 0$ yields the preferences in Ely, Frankel, and Kamenica (2015).[33]

The overall utility of the agent is $V_\beta(F) = p_F \tilde{v} + (1 - \beta)V_1(F) + \beta V_2(F)$, where $\beta \in [0, 1]$ captures the relative importance of suspense across periods. The discussion above shows V_β has continuous local expected utility, so by Theorem 1 it admits an

[32] Here we assume there are only two states, but it is true for any finite state space.

[33] In Ely, Frankel, and Kamenica (2015), p_F is fixed, so all the feasible two-stage lotteries induce the same prior belief over Ω, and flow utility at each period depends on the expected surprise for the next period given the current belief. Here p_F may vary because the choice set includes different binary risks as well as their information structures.

adversarial forecaster representation. The local utilities of V_β are:

$$w_\beta(p,\omega,F) = \omega\tilde{v} + (1-\beta)\left[g\left(D_2(F)\right) + g'\left(D_2(F)\right)\left((p-p_F)^2 - D_2(F)\right)\right] + \beta g(p-p^2),$$

where $D_2(F) = \int \tilde{p}^2 dF(\tilde{p}) - p_F^2$. As we show in Proposition 9 in Online Appendix II.C, when β is near 1 (i.e., the agent mostly cares about second-period surprise), the optimum is to reveal no information in the first period so the set of interim beliefs is a singleton, and when β is near 0, so the agent mostly cares about first-period surprise, the state is fully revealed then. Finally, for intermediate values of β the optimum can have three different interim beliefs (see Online Appendix III).

6.2 Information acquisition and rational inattention

This section applies adversarial forecaster preferences to rational inattention. The combination of adversarial forecaster preferences and rational inattention is natural when the realized payoff has both instrumental and experiential value. For example, a trader deciding how much research to conduct before taking a speculative position, or a sports bettor deciding how carefully to analyze a match, may value not only the expected return but also the suspense surrounding the eventual gain or loss. Information then improves expected performance but also makes the realized payoff easier to forecast. The agent may therefore remain inattentive not only because information is costly, but also because information reduces suspense. Moreover, a preference for surprise can generate costly attention even when information has no instrumental value.

The main result in binary-decision problems is that information costs and preference for surprise govern opposite ends of the stakes. At low stakes, a positive information cost makes the action nearly independent of the state because the instrumental return to state-contingent behavior is small; we refer to this as low-stakes inattention. At high stakes, increasing the stakes reduces accuracy because the value of payoff suspense grows quadratically while the expected-payoff benefit of accuracy grows only linearly. This high-stakes decline persists when information is free. Thus, information costs drive low-stakes inattention, while the preference for surprise drives the decline in accuracy at high stakes.

6.2.1 Preliminaries

As in Matějka and McKay (2015)'s discrete-choice formulation of the rational-inattention model introduced by Sims (2003), information acquisition and action choice are summarized by a state-contingent rule. Let Ω and A be finite sets of states and actions, let $\mu \in \Delta(\Omega)$ have full support, and let $x(a, \omega) \in \mathbb{R}$ be the payoff from action a in state ω. Set $X = \{x(a, \omega) : (a, \omega) \in A \times \Omega\}$. The agent chooses $\pi \in \Delta(A)^{\Omega}$, where $p_\pi(a) = \sum_\omega \mu(\omega)\pi(a \mid \omega)$ is the unconditional probability of action a, and the induced payoff lottery is denoted F_π. The rule π costs $\kappa I(\pi)$, where $\kappa \geqslant 0$ is the unit cost of information and $I(\pi) = \sum_{\omega,a} \mu(\omega)\pi(a \mid \omega)\log(\pi(a \mid \omega)/p_\pi(a))$ is mutual information. The timing is as follows: First the agent chooses π, then the forecaster then observes π and forecasts the payoff without observing the state, information, or action, and finally π is implemented and the payoff is realized.

Suppose the forecaster chooses a forecast in the convex hull of X and minimizes squared error, and let $\gamma \geqslant 0$ measure the agent's preference for surprise. The optimal forecast is the expected payoff $m_\pi = \sum_{\omega,a} \mu(\omega)\pi(a \mid \omega)x(a, \omega)$,[34] so the agent solves

$$\max_\pi \left\{ \sum_{\omega,a} \mu(\omega)\pi(a \mid \omega)x(a, \omega) + \gamma \mathrm{Var}_\pi(x) - \kappa I(\pi) \right\}, \tag{12}$$

where the first two terms in (12) are the adversarial forecaster utility of F_π.

6.2.2 Free information and deliberate mistakes

We begin with free information, so $\kappa = 0$. This isolates the implications of adversarial forecaster preferences before introducing any information-processing friction. For each state ω, let $x_\omega^- = \min_{a \in A} x(a, \omega)$, $x_\omega^+ = \max_{a \in A} x(a, \omega)$, and their midpoint is $c_\omega = (x_\omega^- + x_\omega^+)/2$. The next result characterizes optimal mistakes.

Proposition 6. *Suppose $\kappa = 0$ and $\gamma > 0$. Then,*

1. *All optimal rules induce the same expected payoff.*

[34] Any information structure followed by an action rule induces a state-contingent choice rule $\pi(a \mid \omega)$. The same rule can be implemented by revealing only the action to be taken, which leaves the induced payoff lottery F_π unchanged and cannot increase the Shannon information cost. Since the adversarial forecaster utility depends only on F_π, it is therefore without loss to optimize directly over π, with cost $\kappa I(\pi)$.

2. *For every optimal rule π^* and state ω, only actions yielding x_ω^+ or x_ω^- can be chosen. In particular, π^* chooses*

   (a) *only x_ω^+ if $c_\omega > m_{\pi^*} - 1/(2\gamma)$*

   (b) *only x_ω^- if $c_\omega < m_{\pi^*} - 1/(2\gamma)$*

   (c) *to mix between x_ω^+ and x_ω^- only if $c_\omega = m_{\pi^*} - 1/(2\gamma)$.*

3. *If the midpoints c_ω are distinct, then an optimal rule randomizes over different payoffs in at most one state.*

The threshold in Proposition 6 is endogenous because it depends on the expected payoff generated by the optimal rule itself. In each state, the agent chooses either a best action or a worst action; an intermediate mistake is never optimal because it sacrifices expected payoff without creating as much suspense as an extreme mistake. States are ordered by their payoff midpoints. In states with a high midpoint, the gain from taking the best action dominates. In states with a low midpoint, taking the worst action moves the realized payoff sufficiently far below the forecast to compensate for its instrumental cost. Generically, the agent therefore makes deterministic extreme mistakes in a cutoff set of states and mixes over payoff levels in at most one state.

Corollary 4. *If $0 < \gamma_1 < \gamma_2$, then every pair of rules (π_1^*, π_2^*) that is optimal at these parameters satisfies $m_{\pi_2^*} \leqslant m_{\pi_1^*}$ and $Var_{\pi_2^*}(x) \geqslant Var_{\pi_1^*}(x)$.*

This result gives a global comparative static. A stronger preference for surprise weakly increases the variance of the chosen payoff and weakly lowers its expectation. To determine where deliberate mistakes begin, suppose that $x_\omega^- < x_\omega^+$ for at least one state, let $\overline{x}^+ = \sum_\omega \mu(\omega) x_\omega^+$, and let $c_{\min} = \min_{\omega : x_\omega^- < x_\omega^+} c_\omega$. The rule that always chooses a payoff-maximizing action is optimal if and only if $\frac{1}{2\gamma} \geqslant \overline{x}^+ - c_{\min}$. If $\overline{x}^+ > c_{\min}$, this is equivalent to $\gamma \leqslant [2(\overline{x}^+ - c_{\min})]^{-1}$, and deliberate mistakes first occur in states with the lowest midpoint once γ exceeds this cutoff. If $\overline{x}^+ \leqslant c_{\min}$, the payoff-maximizing rule is optimal for every $\gamma > 0$. If $x_\omega^- = x_\omega^+$ for every state, the payoff distribution is independent of the decision rule, so a payoff-maximizing rule is optimal for every γ.

The next example shows that a preference for surprise can also create a demand for information that has no instrumental value.

Example 9. Let $\Omega = A = \{0, 1\}$, let the prior be uniform, and set $x(a, \omega) = a + \omega$. Action 1 strictly dominates action 0 in both states, so an expected-utility agent always chooses action 1 and acquires no information. When information is free, the unique optimal rule satisfies $\pi^*(1 \mid 1) = 1$ and

$$\pi^*(1 \mid 0) = \begin{cases} 1 & \text{if } \gamma \leqslant 1/2, \\ 1/\gamma - 1 & \text{if } 1/2 < \gamma < 1, \\ 0 & \text{if } \gamma \geqslant 1. \end{cases} \tag{13}$$

Notably, if $\gamma > 1$, every solution of (12) has $I(\pi) > 0$ for every finite $\kappa \geqslant 0$. △

When $\gamma > 1$ and information is free, the agent deliberately takes the inferior action in the low state but takes the superior action in the high state. This sacrifices one unit of payoff when conditions are unfavorable, but it makes the difference between unfavorable and favorable realizations twice as large. The eventual payoff is therefore more suspenseful. With costly information, some state dependence remains optimal no matter how large the finite information cost is. Starting from a rule that ignores the state, the agent can make a very small state-contingent adjustment: choose the superior action slightly more often in the high state and slightly less often in the low state. The resulting increase in payoff dispersion is proportional to the size of the adjustment, while the associated information cost is comparatively negligible when the adjustment is sufficiently small. Some costly attention is therefore worthwhile even though it does not improve expected payoff.

6.2.3 A binary decision problem

Consider the canonical binary problem in which $\Omega = A = \{-1, 1\}$, the prior is uniform, and the payoff is $b > 0$ if the action matches the state and 0 otherwise. For example, this can represent a trader choosing between two positions or a bettor choosing between two contestants. Let $r_\pi = \mathbb{P}_\pi(a = \omega)$ denote accuracy. For any given accuracy, a symmetric decision rule minimizes the information cost, so problem (12) reduces to

$$\max_{r \in [1/2, 1]} \left\{ br + \gamma b^2 r(1 - r) - \kappa \left(\log 2 - H(r) \right) \right\}, \tag{14}$$

where $H(r) = -r \log r - (1 - r) \log(1 - r)$ is binary entropy. The first term rewards accuracy, the second rewards unpredictability of the realized payoff, and the third is the least attention cost required to attain accuracy r.

We first characterize the costly-information problem. For $\kappa > 0$, let $I^* = \log 2 - H(r^*)$ denote the information used at the optimum. The next result separates the effects of the preference for surprise and the information cost on optimal accuracy.

Proposition 7. *Suppose $\kappa > 0$. Optimal accuracy $r^* \in (1/2, 1)$ is unique and satisfies*

$$b + \gamma b^2(1 - 2r^*) = \kappa \log\left(\frac{r^*}{1 - r^*}\right). \tag{15}$$

Both r^* and I^* are strictly decreasing in γ and κ. Moreover, $r^* \to 1/2$ and $I^* \to 0$ as $\gamma \to \infty$.

The left-hand side of (15) is the local payoff advantage of a correct action over an incorrect action. Its first term is the instrumental gain from being correct. Since $r^* > 1/2$, its second term is negative: a higher probability of success makes the realized payoff more predictable and therefore reduces suspense. The right-hand side is the marginal information cost required to support accuracy r^*. Equivalently, dividing (15) by b gives $1 - \gamma b(2r^* - 1) = \frac{\kappa}{b} \log(r^*/(1 - r^*))$. Thus, γb measures the importance of surprise relative to expected payoff, whereas κ/b measures the importance of information costs relative to the stakes. Increasing either γ or κ lowers accuracy in this binary environment, but for different reasons: γ reduces the local benefit of a correct action, while κ raises the cost of making the action depend on the state.

Under standard rational inattention, $\gamma = 0$, so (15) gives the solution $r^* = \exp(b/\kappa)/(1+\exp(b/\kappa))$. Thus, the standard comparative static is that both accuracy and attention increase strictly with the stakes: as $b \downarrow 0$, accuracy converges to $1/2$ and mutual information converges to zero, while as $b \to \infty$, accuracy converges to one and mutual information converges to $\log 2$. The next result shows that a preference for surprise reverses this comparative static once the stakes are sufficiently high.

Corollary 5. *Suppose $\gamma > 0$. Then,*

1. *If $\kappa = 0$, optimal accuracy is*

$$r^* = \begin{cases} 1 & \text{if } \gamma b \leqslant 1, \\ \dfrac{1}{2} + \dfrac{1}{2\gamma b} & \text{if } \gamma b > 1. \end{cases} \tag{16}$$

Thus, the agent deliberately makes mistakes if and only if $\gamma b > 1$, and optimal accuracy is strictly decreasing in b on this region.

2. *If $\kappa > 0$, then r^* and I^* are strictly increasing in b up to a unique stake and strictly decreasing thereafter. At the turning point, $2\gamma b(2r^* - 1) = 1$. Moreover,*

$$\lim_{b \downarrow 0} \frac{r^* - 1/2}{b} = \frac{1}{4\kappa} \qquad \textit{and} \qquad \lim_{b \to \infty} b\left(r^* - \frac{1}{2}\right) = \frac{1}{2\gamma}. \tag{17}$$

3. *Write $r^*(b, \kappa)$ to display its dependence on stakes and information cost. Then,*

$$\lim_{b \downarrow 0} \lim_{\kappa \downarrow 0} r^*(b, \kappa) = 1 \qquad \textit{whereas} \qquad \lim_{\kappa \downarrow 0} \lim_{b \downarrow 0} r^*(b, \kappa) = \frac{1}{2}. \tag{18}$$

When information is free ($\kappa = 0$), imperfect accuracy need not reflect inattention: the agent can learn the state and then deliberately randomize between the correct and incorrect actions. The expected-payoff benefit of accuracy is proportional to b, while the suspense generated by an unpredictable success or failure is proportional to b^2.[35] Full accuracy is therefore optimal at low stakes, but sufficiently high stakes make deliberate errors attractive.

When $\kappa > 0$ is fixed and b is small, the instrumental benefit of state-contingent behavior is too small to justify its information cost. Consequently, $I^* \to 0$ and $r^* \to 1/2$ as $b \downarrow 0$. This low-stakes inattention effect disappears when information is free, as shown by the first iterated limit in (18). At high stakes, by contrast, the quadratic surprise term dominates the linear expected-payoff term. Accuracy and mutual information then fall as the stakes increase. This effect is not caused by costly information: the free-information solution in (16) is also strictly decreasing in b when $\gamma b > 1$, and the second limit in (17) coincides with its high-stakes behavior. Thus, information costs drive low-stakes inattention, while the preference for surprise drives the decline in accuracy at high stakes.

[35] Since the suspense term is quadratic in payoffs, the numerical values of γ and κ depend on the payoff unit. This raises the question of whether the comparative static in b depends on whether payoffs are measured, for example, in dollars or cents. If all payoffs are multiplied by a constant $c > 0$, the same preferences are represented by replacing γ with γ/c and κ with $c\kappa$; the objective in (12) is then multiplied by c, so the agent's choices are unchanged. For the comparative static in b, the payoff unit and the parameters γ and κ remain fixed while b—the payoff gain from choosing the correct action rather than the incorrect action—varies.

7 Conclusion

Adversarial forecaster preferences arise naturally in many settings. They allow the interpretation of random choice as a preference for surprise, and also allow sharp characterizations of the optimal "amount" (i.e., support size) of randomization and of various monotonicity properties. Ongoing work considers a more general "adversarial expected utility representation" that inherits many of the optimality and monotonicity properties of the adversarial forecaster representation, but does not require continuous local utility. This lets us consider cases where the adversary has only finitely many actions or where the loss function has kinks, as it does with an absolute-deviation forecast error. This more general representation can also be applied to settings where the agent first chooses a distribution of qualities or outcomes and then chooses an allocation rule or an information-revelation policy.

Appendix A: Sections 2, 3, and 5

Here we prove the main results in Sections 2, 3, and 5. The proofs of the ancillary results that are first stated in this section are in Online Appendix I.A. and the proofs of Section 6.2 are in Online Appendix I.C.

Lemma 1. *If V has continuous local expected utility $w(x,F)$, then for all $F, \tilde{F} \in \mathcal{F}$:*

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

Proof. (1) This is immediate, as by definition $\int w(x,F)dF(x) = V(F) \leqslant \int w(x,\tilde{F})dF(x)$ for all $F, \tilde{F} \in \mathcal{F}$.
(2) 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, $\int w(x,F)d\overline{F}(x) \geqslant V(\overline{F})$ so

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

Similarly, since $w(x,\overline{F})$ is a local utility function at $\overline{F}$, $\int w(x,\overline{F})dF(x) \geqslant V(F)$, so

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

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

as $\lambda \to 0$, since $w(x,\overline{F})$ is continuous in $\overline{F}$. Putting these together yields

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

which yields the statement. ∎

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

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

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

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

where the second equality follows by Lemma 1. ∎

We can now prove Proposition 1 and Theorem 1.

Proof of Proposition 1. Assume that V has continuous local expected utility $w(x,F)$. As argued in the main text, V is concave. Lemma 2 implies that $D(F, (\delta_x - F)) = w(x,F) - \int w(x,F) dF(x) = w(x,F) - V(F)$, where the second equality follows from the properties of $w(x,F)$. This implies that $D(F, (\delta_x - F))$ is well-defined and continuous and that $w(x,F) = V(F) + D(F, (\delta_x - F))$ as desired. ∎

Proof of Theorem 1. (If). Let v and σ correspond to the adversarial forecaster representation of V. Since $(F, y) \mapsto \int \sigma(x,y) dF(x)$ is continuous and Y is compact, every limit point of $\hat{y}(F_n)$ for $F_n \to F$ minimizes $\int \sigma(x,y) dF(x)$; uniqueness therefore implies $\hat{y}(F_n) \to \hat{y}(F)$. The map $w : X \times \mathcal{F} \to \mathbb{R}$ given by $w(x, F) = v(x) + \sigma(x, \hat{y}(F))$ is a continuous local utility of $V(F) = \min_{\tilde{F} \in \mathcal{F}} \int w(x, \tilde{F}) dF(x)$.

(Only if). Let $w(x, F)$ denote the continuous local expected utility of V, and define $Y = \{w(\cdot, F)\}_{F \in \mathcal{F}} \subseteq C(X)$, endowed with the sup norm. Since $\mathcal{F}$ is compact and $F \mapsto w(\cdot, F)$ is continuous, Y is compact. For all $y \in Y$, define $u(x, y) = y(x)$; this function is continuous on $X \times Y$. For all $F \in \mathcal{F}$ and all $\tilde{y} \in Y$, $V(F) = \int w(x, F) dF(x) \leqslant \int u(x, \tilde{y}) dF(x)$, where the equality follows from the tangency of $w(\cdot, F)$ at F and the inequality from the supporting-hyperplane property. Thus, $V(F) = \min_{y \in Y} \int u(x, y) dF(x)$.

We will now show that the minimum is unique. Fix $F \in \mathcal{F}$, and suppose that $\tilde{y} \in Y$ satisfies $V(F) = \int u(x, \tilde{y}) dF(x)$. Choose $\tilde{F} \in \mathcal{F}$ such that $\tilde{y} = w(\cdot, \tilde{F})$. For every $\lambda \in [0, 1]$, define $F_\lambda = \lambda \tilde{F} + (1 - \lambda) F$. Then,

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

where the first inequality follows from concavity of V, the second inequality from the supporting-hyperplane property of $w(\cdot, \tilde{F})$, and the last equality from the tangency of $w(\cdot, \tilde{F})$ at $\tilde{F}$ and from the hypothesis that $\tilde{y}$ minimizes at F. Therefore, for all $\lambda \in [0, 1]$,

$$V(F_\lambda) = \lambda V(\tilde{F}) + (1 - \lambda) V(F) = \int w(x, \tilde{F}) dF_\lambda(x). \tag{19}$$

Fix $\mu \in (0, 1)$. Since $w(\cdot, F_\mu)$ is a local expected utility of V at F_μ,

$$V(F_\mu) = \int w(x, F_\mu) dF_\mu(x) = (1 - \mu) \int w(x, F_\mu) dF(x) + \mu \int w(x, F_\mu) d\tilde{F}(x).$$

Moreover, the supporting-hyperplane property gives

$$\int w(x, F_\mu) dF(x) \geqslant V(F) \quad \text{and} \quad \int w(x, F_\mu) d\tilde{F}(x) \geqslant V(\tilde{F}).$$

Combining these inequalities with (19) evaluated at $\lambda = \mu$, and $\mu \in (0,1)$, implies

$$V(\tilde{F}) = \int w(x, F_\mu) d\tilde{F}(x). \tag{20}$$

Fix $\tilde{x} \in X$. Since $F_\mu + \lambda(\delta_{\tilde{x}} - \tilde{F}) = (1-\mu)F + (\mu - \lambda)\tilde{F} + \lambda\delta_{\tilde{x}} \in \mathcal{F}$ for all $\lambda \in [0, \mu]$, Lemma 2 and (20) give $\lim_{\lambda \downarrow 0} \frac{V(F_\mu + \lambda(\delta_{\tilde{x}} - \tilde{F})) - V(F_\mu)}{\lambda} = w(\tilde{x}, F_\mu) - V(\tilde{F})$. Using the supporting-hyperplane property of $w(\cdot, \tilde{F})$ and (19),

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

Thus, $w(\tilde{x}, F_\mu) \leqslant w(\tilde{x}, \tilde{F})$. Conversely, Lemma 2, the supporting-hyperplane property of $w(\cdot, F_\mu)$, and (20) imply

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

Therefore, $w(\tilde{x}, F_\mu) = w(\tilde{x}, \tilde{F})$. Since this holds for all $\mu \in (0,1)$, continuity of w implies $w(\tilde{x}, F) = w(\tilde{x}, \tilde{F})$. As $\tilde{x}$ was arbitrary, $\tilde{y} = w(\cdot, \tilde{F}) = w(\cdot, F)$, proving that $\int u(x,y) dF(x)$ has a unique minimizer over Y.

Lastly, define $v(x) = V(\delta_x)$ and $\sigma(x,y) = u(x,y) - v(x)$. Then, $v \in C(X)$, σ is continuous, and $\sigma \geqslant 0$ because $v(x) = V(\delta_x) = \min_{y \in Y} u(x,y)$. Moreover, the minimizer of $\int \sigma(x,y) dF(x)$ is the same unique minimizer as that of $\int u(x,y) dF(x)$, and $\sigma(x, \hat{y}(\delta_x)) = 0$ for all $x \in X$. Thus, σ is a forecast error and

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

so V has an adversarial forecaster representation. ■

Proof of Proposition 2. For every $G \in \mathcal{F}$, Theorem 1 implies that $x \mapsto v(x) +$

$\sigma(x, \hat{y}(G))$ is a continuous local expected utility of V at G.

(If). For all $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} \int v(x) + \sigma(x, \hat{y}(F^*))dF(x)$ and $F \in \overline{\mathcal{F}}$, $V(F^*) = \int v(x) + \sigma(x, \hat{y}(F^*))dF^*(x) \geqslant \int v(x) + \sigma(x, \hat{y}(F^*))dF(x) \geqslant V(F)$, where the first equality and last inequality follow from the definition of continuous local utility and the fact that $v(x)+\sigma(x, \hat{y}(F))$ is a continuous local utility of V, and the first inequality follows by assumption. This implies that $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} V(F)$.[36]

(Only if). Fix $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} V(F)$. We show that F^* maximizes the local expected utility $v + \sigma(\cdot, \hat{y}(F^*))$ over $\overline{\mathcal{F}}$. Fix any $\hat{F} \in \overline{\mathcal{F}}$. Since $\overline{\mathcal{F}}$ is convex, $(1-\lambda)F^* + \lambda\hat{F} \in \overline{\mathcal{F}}$ for every $\lambda \in [0, 1]$. Hence, by the optimality of F^* for V over $\overline{\mathcal{F}}$,

$$\frac{V((1-\lambda)F^* + \lambda\hat{F}) - V(F^*)}{\lambda} \leqslant 0$$

for every $\lambda \in (0, 1]$. Taking $\lambda \downarrow 0$ and applying Lemma 2 at F^* in the direction $\hat{F} - F^*$ gives $\int \big(v(x) + \sigma(x, \hat{y}(F^*))\big)d\hat{F}(x) - \int \big(v(x) + \sigma(x, \hat{y}(F^*))\big)dF^*(x) \leqslant 0$, where we used Theorem 1 to identify the continuous local expected utility at F^* with $v + \sigma(\cdot, \hat{y}(F^*))$. Since $\hat{F}$ was arbitrary, $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} \int \big(v(x) + \sigma(x, \hat{y}(F^*))\big)dF(x)$. ∎

Proof of Corollary 1. By Proposition 2, F^* maximizes $V(F)$ over $\mathcal{F}$ if and only if $F^* \in \operatorname{argmax}_{F \in \mathcal{F}} \int v(x)+\sigma(x, \hat{y}(F^*))dF(x)$, that is, if and only if $x \in \operatorname{argmax}_{\tilde{x} \in X} v(\tilde{x}) + \sigma(\tilde{x}, \hat{y}(F^*))$ for all $x \in \operatorname{supp}(F^*)$. Assume that $w(x, F)$ is strictly quasiconcave in x for all $F \in \mathcal{F}$ and assume by contradiction that $x, x' \in \operatorname{supp}(F^*)$ with $x \neq x'$. The set $\operatorname{argmax}_{\tilde{x} \in X} v(\tilde{x}) + \sigma(\tilde{x}, \hat{y}(F^*))$ must be a singleton and therefore x and x' cannot be both optimal, contradicting the optimality of F^*. Next assume that $w(x, F)$ is strictly quasiconvex in x for all $F \in \mathcal{F}$ and suppose, toward a contradiction, that $x \in \operatorname{supp}(F^*)\backslash \operatorname{ext}(X)$. Then, there exist distinct $x_1, x_2 \in X$ and $\alpha \in (0, 1)$ such that $x = \alpha x_1 + (1-\alpha)x_2$. Strict quasiconvexity gives $w(x, F^*) < \max\{w(x_1, F^*), w(x_2, F^*)\}$, which contradicts optimality of F^*: $x \in \operatorname{argmax}_{\tilde{x} \in X} w(\tilde{x}, F^*)$. ∎

Proof of Proposition 10. In Proposition 12 in Online Appendix IV, we show that V is Gâteaux differentiable with derivative given by the local utility $w(x, F)$ as in

[36] See Proposition 8 in Online Appendix II.A for an alternative proof that can also be applied to the more general adversarial expected utility model.

Proposition 1. Theorem 1 then implies that the local utility is $w(x, F) = v(x) + \sigma(x, \hat{y}(F))$ for every $F \in \mathcal{F}$. With this, exactly the same argument of Proposition 1 in Cerreia-Vioglio, Maccheroni, and Marinacci (2017) yields the desired result. ∎

Proof of Proposition 4. The result follows from the following three lemmas. The first two are standard and are proved in Online Appendix I.A. Recall that here we allow the set S to be any compact metric space.

Lemma 3. *If μ has full support, then the function $\sigma(x, y)$ defined in equation 9, with $Y = \{h(F, \cdot) \in C(S) : F \in \mathcal{F}\}$, is a forecast error.*

Given $F, \tilde{F} \in \mathcal{F}$, the direction $\tilde{F} - F$ is *relevant* at F if for some $\lambda > 0$ the signed measure $F + \lambda(\tilde{F} - F) \geqslant 0$ is an ordinary measure.

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

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

The directional derivatives $D\Sigma(F, \delta_z - F)$ for directions $(\delta_z - F)$ at F are

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

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 has full support, then $\Sigma(F)$ and $V(F)$ are strictly concave.*

Proof. From Lemma 4, $\Sigma(F)$ is the sum of an affine function of F and the negative of the quadratic form $Q(F) = \int\int H(x, \tilde{x})dF(x)dF(\tilde{x})$. It is therefore enough 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 M such that $\int dM(x) = 0$. For such an M, $\int\int H(x, \tilde{x})dM(x)dM(\tilde{x}) = \int \left(\int h(x, s)dM(x)\right)^2 d\mu(s) \geqslant 0$. Suppose $M \neq 0$ and $\int dM(x) = 0$. By the Jordan decomposition, $M = \lambda(F - \tilde{F})$ for some $F, \tilde{F} \in \mathcal{F}$ and some $\lambda > 0$. If $\int h(x, s)dM(x) = 0$ for every $s \in S$, then $h(F, \cdot) = h(\tilde{F}, \cdot)$. Since $F \mapsto h(F, \cdot)$ is one-to-one, this implies $F = \tilde{F}$, hence $M = 0$, a contradiction. Thus there exists $\hat{s} \in S$ such that $\int h(x, \hat{s})dM(x) \neq 0$. Since $s \mapsto \int h(x, s)dM(x)$ 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 has full support, $\mu(\tilde{S}) > 0$, and therefore $\int\int H(x,\tilde{x})dM(x)dM(\tilde{x}) = \int \left(\int h(x,s)dM(x)\right)^2 d\mu(s) > 0$. Thus, the quadratic form is positive definite on the subspace of signed measures with total mass zero. Therefore, Σ is strictly concave. Since $V(F) = \int v(x)dF(x) + \Sigma(F)$ and the first term is affine in F, V is strictly concave as well. ■

Now we extend the support bounds of Proposition 3 to parametric adversarial preferences. Consider a utility V with parametric forecast error σ and an arbitrary compact and convex set $\overline{\mathcal{F}} \subseteq \mathcal{F}$ of feasible lotteries. Define $\overline{Y} = \hat{y}(\overline{\mathcal{F}})$. The same steps as for GMM show that $\max_{F\in\overline{\mathcal{F}}} V(F) = \max_{\overline{y}\in\overline{Y}} \max_{F\in\overline{\mathcal{F}}:\hat{y}(F)=\overline{y}} \int v(x)+\sigma(x,\overline{y})dF(x)$.[37] We can fix an optimal solution $\overline{y}^*$ of the outer maximization problem and maximize $\int v(x) + \sigma(x,\overline{y}^*)dF(x)$ over the lotteries $\overline{\mathcal{F}}$ that satisfy $\hat{y}(F) = \overline{y}^*$.

Theorem 2. *Let V be a parametric adversarial forecaster utility with representation (Y, v, σ), and suppose that Y has the structure of a compact m-dimensional C^1 manifold with boundary. Fix a closed set $\overline{X} \subseteq X$, $\{g_1, ..., g_k\} \subseteq C(X)$, and let $\overline{\mathcal{F}} = \mathcal{F}_\Gamma(\overline{X})$. Then there is an optimal lottery for V over $\overline{\mathcal{F}}$ that assigns positive probability to at most $(k+1)(m+1)$ points of $\overline{X}$.*

The proof of Theorem 2 uses the following two results. Let $\mathcal{H}$ denote the set of probability measures over Y. Let $ext(\overline{\mathcal{F}})$ denote the extreme points of any convex and compact $\overline{\mathcal{F}} \subseteq \mathcal{F}$. Let $\Lambda_F \subseteq \Delta\left(ext\left(\overline{\mathcal{F}}\right)\right)$ be the probability measures over extreme points that satisfy $F = \int \tilde{F} d\lambda(\tilde{F})$.[38]

Lemma 6. *Fix $\hat{H} \in \arg\min_{H\in\mathcal{H}} \max_{F\in ext(\overline{\mathcal{F}})} \int\int u(x,y)\, dF(x)\, dH(y)$. 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 $\tilde{F} \in \bigcup_{\lambda\in\Lambda_{\hat{F}}} \operatorname{supp}\lambda$, $V(\hat{F}) = \int\int u(x,y)\, d\tilde{F}(x)\, d\hat{H}(y)$.*

Now fix a closed subset $\overline{X} \subseteq X$ and a finite collection of functions $\Gamma = \{g_1, ..., g_k\} \subset C(\overline{X})$, and consider $\mathcal{F}_\Gamma(\overline{X}) \subseteq \mathcal{F}$. By Theorem 4 in Online Appendix I.B (cf. Theorem 2.1 in Winkler (1988)), every $\tilde{F} \in ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$ is supported on at most $k+1$ points of $\overline{X}$. 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 by Theorem 4. Recall that $\hat{Y}(F) \equiv \operatorname{argmin}_{y\in Y} \int u(x,y)dF(x)$.

[37] As in the GMM case, a maximizer exists because the function $R(\overline{y}) = \max_{F\in\overline{\mathcal{F}}:\hat{y}(F)=\overline{y}} \int v(x) + \sigma(x,\overline{y})dF(x)$ is upper semicontinuous and $\overline{Y}$ is compact because the function $\hat{y}$ is continuous.

[38] This set is non-empty from Choquet's theorem.

Theorem 3. *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 $\hat{Y}(F)$ is a singleton for all $F \in \mathcal{F}$. Then:*

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 $P = co(\mathcal{E})$, let $N = (m+1)(k+1)$, and let $|\mathcal{E}| = n$. Since every $\tilde{F} \in \mathcal{E}$ is an extreme point of $\mathcal{F}_\Gamma(\overline{X})$, Theorem 4 implies that $|\operatorname{supp} \tilde{F}| \leqslant k+1$. Therefore, if $n \leqslant m+1$, every lottery in P is supported on no more than $n(k+1) \leqslant N$ points of $\overline{X}_\mathcal{E}$. In this case both conclusions follow by taking $\mathcal{W} = \mathbb{R}^{\overline{X}_\mathcal{E}}$. We therefore assume throughout the rest of the proof that $n \geqslant m+2$.

For every $w \in \mathbb{R}^{\overline{X}_\mathcal{E}}$, define $u_w(x,y) = u(x,y) + w(x)$ and $V_w(F) = \min_{y \in Y} \int u_w(x,y) dF(x)$. Since w does not depend on y, the set of minimizers of $\int u_w(x,y) dF(x)$ over $y \in Y$ is the same as the set of minimizers of $\int u(x,y) dF(x)$ over $y \in Y$. Thus, the uniqueness assumption on $\hat{Y}(F)$ also holds for the perturbed utility V_w.

Proof of 1. Let $\mathcal{A}$ be the collection of all subsets $\overline{\mathcal{E}} = \{\tilde{F}_1, ..., \tilde{F}_{m+2}\} \subseteq \mathcal{E}$ whose elements are affinely independent as probability vectors on the finite set $\overline{X}_\mathcal{E}$. Fix $\overline{\mathcal{E}} \in \mathcal{A}$. Define $\Phi_{\overline{\mathcal{E}}} : Y \to \mathbb{R}^{m+1}$ by $\Phi_{\overline{\mathcal{E}}}(y)_\ell = u(\tilde{F}_\ell, y) - u(\tilde{F}_{m+2}, y)$ $\forall \ell \in \{1, ..., m+1\}$, where $u(\tilde{F}, y) = \int u(x,y) d\tilde{F}(x)$. Also define the linear map $L_{\overline{\mathcal{E}}} : \mathbb{R}^{\overline{X}_\mathcal{E}} \to \mathbb{R}^{m+1}$ by $L_{\overline{\mathcal{E}}}(w)_\ell = w(\tilde{F}_\ell) - w(\tilde{F}_{m+2})$ $\forall \ell \in \{1, ..., m+1\}$, where $w(\tilde{F}) = \int w(x) d\tilde{F}(x)$.

We first show that $L_{\overline{\mathcal{E}}}$ is onto. Suppose $a_1, ..., a_{m+1} \in \mathbb{R}$ satisfy $\sum_{\ell=1}^{m+1} a_\ell \big(\tilde{F}_\ell - \tilde{F}_{m+2}\big) = 0$. Then, $\sum_{\ell=1}^{m+1} a_\ell \tilde{F}_\ell - \left(\sum_{\ell=1}^{m+1} a_\ell\right) \tilde{F}_{m+2} = 0$, and the coefficients in this linear combination sum to zero. This is an affine dependence among $\tilde{F}_1, ..., \tilde{F}_{m+2}$. Since these lotteries are affinely independent, $a_\ell = 0$ for every ℓ. Therefore, the signed measures $\{\tilde{F}_\ell - \tilde{F}_{m+2}\}_{\ell=1}^{m+1}$ are linearly independent, so the linear functionals defining $L_{\overline{\mathcal{E}}}$ are linearly independent. Thus, $L_{\overline{\mathcal{E}}}$ has rank $m+1$ and is onto.

A common tie among the $m+2$ lotteries $\tilde{F}_1, ..., \tilde{F}_{m+2}$ for some $y \in Y$ occurs if and only if $u(\tilde{F}_\ell, y) + w(\tilde{F}_\ell) = u(\tilde{F}_{m+2}, y) + w(\tilde{F}_{m+2})$ $\forall \ell \in \{1, ..., m+1\}$, which is equivalent to $L_{\overline{\mathcal{E}}}(w) = -\Phi_{\overline{\mathcal{E}}}(y)$ for some $y \in Y$. Since u is continuously differentiable in y and Y is an m-dimensional compact manifold with boundary, the set $\Phi_{\overline{\mathcal{E}}}(Y) \subseteq \mathbb{R}^{m+1}$ is compact and has Lebesgue measure zero. Since $L_{\overline{\mathcal{E}}}$ is onto, the set $B_{\overline{\mathcal{E}}} = L_{\overline{\mathcal{E}}}^{-1}(-\Phi_{\overline{\mathcal{E}}}(Y))$ is closed and has Lebesgue measure zero in $\mathbb{R}^{\overline{X}_{\mathcal{E}}}$. Therefore, $\mathcal{W}_{\overline{\mathcal{E}}} = \mathbb{R}^{\overline{X}_{\mathcal{E}}} \backslash B_{\overline{\mathcal{E}}}$ is open, dense, and has full measure. If $\mathcal{A}$ is empty, set $\mathcal{W} = \mathbb{R}^{\overline{X}_{\mathcal{E}}}$; otherwise set $\mathcal{W} = \bigcap_{\overline{\mathcal{E}} \in \mathcal{A}} \mathcal{W}_{\overline{\mathcal{E}}}$. Since $\mathcal{A}$ is finite, $\mathcal{W}$ is open, dense, and has full measure.

Fix $w \in \mathcal{W}$ and let $F \in P$ solve $\max_{\tilde{F} \in P} V_w(\tilde{F})$. Let y_F denote the unique element of $\hat{Y}(F)$. By Proposition 2, applied to the perturbed utility V_w, the lottery F maximizes the linear functional $\tilde{F} \mapsto \int u_w(x, y_F) d\tilde{F}(x)$ over $\tilde{F} \in P$. Hence, F belongs to the exposed face $G_F = \operatorname{argmax}_{\tilde{F} \in P} \int u_w(x, y_F) d\tilde{F}(x)$. We claim that $\dim G_F \leqslant m$. Suppose not. Then, G_F contains $m+2$ affinely independent extreme points $\tilde{F}_1, ..., \tilde{F}_{m+2}$ of P. Since $P = co(\mathcal{E})$ and every element of $\mathcal{E}$ is an extreme point of $\mathcal{F}_\Gamma(\overline{X})$, the extreme points of P are elements of $\mathcal{E}$. Therefore, $\{\tilde{F}_1, ..., \tilde{F}_{m+2}\} \in \mathcal{A}$. Since all these lotteries belong to G_F, they all attain the same value under the linear functional $\tilde{F} \mapsto \int u_w(x, y_F) d\tilde{F}(x)$. Thus, $u(\tilde{F}_\ell, y_F) + w(\tilde{F}_\ell) = u(\tilde{F}_{m+2}, y_F) + w(\tilde{F}_{m+2})$ $\forall \ell \in \{1, ..., m+1\}$. This contradicts $w \in \mathcal{W}_{\{\tilde{F}_1, ..., \tilde{F}_{m+2}\}}$. As a result, $\dim G_F \leqslant m$.

By Carathéodory's theorem applied in the affine hull of the face G_F, the lottery $F \in G_F$ can be written as a convex combination of at most $\dim G_F + 1 \leqslant m+1$ extreme points of G_F. Each such extreme point is an element of $\mathcal{E}$ and is supported on at most $k+1$ points of $\overline{X}_{\mathcal{E}}$. Therefore, F is supported on at most $(m+1)(k+1)$ points of $\overline{X}_{\mathcal{E}}$. Since F was an arbitrary solution of the perturbed problem, the first statement follows.

Proof of 2. Since $\mathcal{W}$ is dense in $\mathbb{R}^{\overline{X}_{\mathcal{E}}}$, there exists a sequence $w^n \in \mathcal{W}$ such that $\|w^n\|_\infty \to 0$. For each n, let $F^n \in \operatorname{argmax}_{\tilde{F} \in P} V_{w^n}(\tilde{F})$. By the first part of the proof, each F^n is supported on at most N points of $\overline{X}_{\mathcal{E}}$. Since P is compact, there is a subsequence, still denoted by F^n, that converges to some $F \in P$. Because $\overline{X}_{\mathcal{E}}$ is finite, the set of probability measures on $\overline{X}_{\mathcal{E}}$ with support of cardinality at most N is closed: indeed, it is the finite union, over all subsets $A \subseteq \overline{X}_{\mathcal{E}}$ with $|A| \leqslant N$, of the closed faces $\{\mu : \mu(\overline{X}_{\mathcal{E}} \backslash A) = 0\}$. Hence F is supported on at most N points.

It remains to show that F solves the unperturbed problem. For every $G \in P$ and every n, $V_{w^n}(F^n) \geqslant V_{w^n}(G)$. Since $V_{w^n}(H) = V_0(H) + \int w^n(x) dH(x)$ $\forall H \in P$, we obtain $V_0(F^n) \geqslant V_0(G) - 2\|w^n\|_\infty$ $\forall G \in P$. Taking limits along the convergent subsequence and using the continuity of V_0 gives $V_0(F) \geqslant V_0(G)$ $\forall G \in P$. Therefore, F solves $\max_{\tilde{F} \in co(\mathcal{E})} \min_{y \in Y} \int u(x, y) d\tilde{F}(x)$ and is supported on at most $(m+1)(k+1)$

points of $\overline{X}_{\mathcal{E}}$. ∎

Lemma 7. *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. Fix a parametric adversarial forecaster representation (Y, v, σ), and define $u = v + \sigma$. By hypothesis, the adversarial expected utility representation (Y, u) is such that Y has the structure of an m-dimensional manifold with boundary, u is continuously differentiable in y, and Y and u satisfy the uniqueness property. By Theorem 3 and Lemma 7, there exists a solution F^* that is supported on no more than $(k+1)(m+1)$ points of $\overline{X}$. ∎

Proof of Proposition 5. Since $v \equiv 0$ and $\Gamma = \varnothing$, the objective is the suspense term $V(F) = \Sigma(F) = \int H(x,x)dF(x) - \int\int H(x,\tilde{x})dF(x)dF(\tilde{x})$. Since $H(x,\tilde{x}) = G(x-\tilde{x})$, we have $H(x,x) = G(0)$ for every $x \in X$. Thus, $V(F) = G(0) - Q(F)$, $\quad Q(F) := \int\int H(x,\tilde{x})dF(x)dF(\tilde{x})$. The set $\Delta(X)$ is compact in the weak topology and V is continuous, so an optimal lottery exists.

We first show uniqueness. Let $F \neq F'$ and $\lambda \in (0,1)$. Set $M = F - F'$. Then $M(X) = 0$ and $M \neq 0$. By strict positive definiteness of H on signed measures with total mass zero, $\int\int H(x,\tilde{x})dM(x)dM(\tilde{x}) > 0$. A direct expansion gives

$$Q(\lambda F + (1-\lambda)F') = \lambda Q(F) + (1-\lambda)Q(F') - \lambda(1-\lambda)\int\int H(x,\tilde{x})dM(x)dM(\tilde{x}).$$

Therefore, Q is strictly convex on $\Delta(X)$, and so $V = G(0) - Q$ is strictly concave. Thus, the optimal lottery is unique. Denote it by F^*.

We show that $\operatorname{supp} F^* = X$. For notational simplicity, write $\phi(z) := \int H(z,x)dF^*(x)$, and $q := \int\int H(x,\tilde{x})dF^*(x)dF^*(\tilde{x})$. For every $z \in X$ and every $\lambda \in [0,1]$, the lottery $(1-\lambda)F^* + \lambda\delta_z$ is feasible. Since F^* is optimal, the right derivative of $V((1-\lambda)F^* + \lambda\delta_z)$ at $\lambda = 0$ is non-positive. Using $H(x,x) = G(0)$ and differentiating the quadratic term yields $\frac{d}{d\lambda}V((1-\lambda)F^* + \lambda\delta_z)\big|_{\lambda=0} = -2(\phi(z) - q) \leqslant 0$. Thus,

$$\phi(z) \geqslant q \qquad \text{for every } z \in X. \tag{21}$$

Moreover, $\int \phi(z)dF^*(z) = q$. Since $\phi - q$ is continuous and non-negative by (21), then

$$\phi(z) = q \qquad \text{for every } z \in \operatorname{supp} F^*. \tag{22}$$

Indeed, if $\phi(z) > q$ at some point of $\operatorname{supp} F^*$, then by continuity $\phi - q$ would be strictly positive on a neighborhood with positive F^*-probability, contradicting $\int(\phi - q)dF^* = 0$. Let $f(t) := H(0,t)$ for $t \in [0,1]$. Since $H(x,\tilde{x}) = \int h(x,s)h(\tilde{x},s)d\mu(s)$ is symmetric and $H(x,\tilde{x}) = G(x - \tilde{x})$, we have $H(x,\tilde{x}) = f(|x - \tilde{x}|)$for all $x, \tilde{x} \in [0,1]$. By assumption, f is non-negative, strictly decreasing whenever it is positive, and strictly convex.

We first rule out a gap at the left endpoint. Suppose, toward a contradiction, that $\inf \operatorname{supp} F^* = a > 0$. Then $a \in \operatorname{supp} F^*$. Since $f(0) > 0$, by continuity there is $\varepsilon > 0$ such that $f(t) > 0$ for all $t \in [0,\varepsilon]$. Because $a \in \operatorname{supp} F^*$, the set $[a, a + \varepsilon/2] \cap X$ has positive F^*-probability. For every $x \in \operatorname{supp} F^* \subseteq [a,1]$, $H(a,x) = f(x - a) \geqslant f(x) = H(0,x)$, and the inequality is strict for every $x \in [a, a + \varepsilon/2] \cap X$, because then $0 \leqslant x - a < x$ and $f(x - a) > 0$. Therefore, $\phi(a) = \int H(a,x)dF^*(x) > \int H(0,x)dF^*(x) = \phi(0)$. By (22), $\phi(a) = q$, so $\phi(0) < q$, contradicting (21) at $z = 0$. As a result, $\inf \operatorname{supp} F^* = 0$.

Next, we rule out a gap at the right endpoint. Suppose, toward a contradiction, that $\sup \operatorname{supp} F^* = b < 1$. Then $b \in \operatorname{supp} F^*$. By the same argument as above, there is $\varepsilon > 0$ such that $[b - \varepsilon/2, b] \cap X$ has positive F^*-probability and $f(t) > 0$ for all $t \in [0,\varepsilon]$. For every $x \in \operatorname{supp} F^* \subseteq [0,b]$, $H(b,x) = f(b - x) \geqslant f(1 - x) = H(1,x)$, and the inequality is strict for every $x \in [b - \varepsilon/2, b] \cap X$, so $\phi(b) = \int H(b,x)dF^*(x) > \int H(1,x)dF^*(x) = \phi(1)$. By (22), $\phi(b) = q$, so $\phi(1) < q$, contradicting (21) at $z = 1$. As a result, $\sup \operatorname{supp} F^* = 1$.

Finally, we rule out an interior gap. Suppose, toward a contradiction, that there exist $0 \leqslant a < b \leqslant 1$ such that $a, b \in \operatorname{supp} F^*$, $(a,b) \cap \operatorname{supp} F^* = \varnothing$. Let $m = (a + b)/2$. Then, $\operatorname{supp} F^* \subseteq [0,a] \cup [b,1]$. For every $x \in [0,a]$, strict convexity of f gives $H(m,x) = f(m - x) < \frac{1}{2}f(a - x) + \frac{1}{2}f(b - x) = \frac{1}{2}H(a,x) + \frac{1}{2}H(b,x)$. For every $x \in [b,1]$, strict convexity of f gives

$$H(m,x) = f(x - m) < \frac{1}{2}f(x - a) + \frac{1}{2}f(x - b) = \frac{1}{2}H(a,x) + \frac{1}{2}H(b,x).$$

Thus, the strict inequality $H(m,x) < \frac{1}{2}H(a,x) + \frac{1}{2}H(b,x)$ holds for every $x \in \operatorname{supp} F^*$.

Integrating with respect to F^* yields $\phi(m) < \frac{1}{2}\phi(a) + \frac{1}{2}\phi(b) = q$, where the equality uses (22) and the fact that $a, b \in \operatorname{supp} F^*$. This contradicts (21) at $z = m$.

We have shown that the support has no left-end gap, no right-end gap, and no interior gap. Therefore, $\operatorname{supp} F^* = [0, 1] = X$. ■

References

Agranov, M. and P. Ortoleva (2017). "Stochastic choice and preferences for randomization." *Journal of Political Economy*, 125, 40–68.

Aliprantis, C. D. and K. C. Border (2006). *Infinite Dimensional Analysis*. Springer.

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.

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*, 139–163.

Corrao, R., D. Fudenberg, and D. K. Levine (2024). "On concave functions over lotteries." *Journal of Mathematical Economics*, 110, 102936.

Dekel, E. (1986). "An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom." *Journal of Economic theory*, 40, 304–318.

Dillenberger, D. (2010). "Preferences for one-shot resolution of uncertainty and Allais-type behavior." *Econometrica*, 78, 1973–2004.

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

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.

Galichon, A. and M. Henry (2012). "Dual theory of choice with multivariate risks." *Journal of Economic Theory*, 147, 1501–1516.

Green, J. R. and B. Jullien (1988). "Ordinal independence in nonlinear utility theory." *Journal of Risk and Uncertainty*, 1, 355–387.

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

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

Hong, C. S. and W. S. Waller (1986). "Empirical tests of weighted utility theory." *Journal of Mathematical Psychology*, 30, 55–72.

Loseto, M. and A. Lucia (2021). "Deliberate randomization under risk." *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*, 277–323.

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

— (1985). "Stochastic choice functions generated from deterministic preferences over lotteries." *The economic journal*, 95, 575–594.

Matějka, F. and A. McKay (2015). "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review*, 105, 272–98.

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

Prelec, D. (1990). "A "Pseudo-endowment" effect, and its implications for some recent nonexpected utility models." *Journal of Risk and Uncertainty*, 3, 247–259.

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

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

Sarver, T. (2018). "Dynamic mixture-averse preferences." *Econometrica*, 86, 1347–1382.

Sims, C. A. (2003). "Implications of rational inattention." *Journal of monetary Economics*, 50, 665–690.

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

Online Appendix I: Ancillary results

This appendix gives proofs of the ancillary results stated in the main appendix, states some useful results from optimal transportation theory, and proofs of Section 6.2.

Online Appendix I.A: Ancillary results for Appendix A

Proof of Lemma 3. Non-negativity is immediate. Continuity of σ follows from the continuity of h, compactness of $X \times S$, and the sup-norm topology on $C(S)$. For every $F \in \mathcal{F}$, define $h_F(s) = \int h(x,s)dF(x)$. By definition, $h_F \in Y$. For any $y \in Y$,

$$\int \sigma(x,y)dF(x) = \int \int (h(x,s) - y(s))^2 \, d\mu(s)dF(x).$$

Expanding the square around $h_F(s)$ gives

$$\int \sigma(x,y)dF(x) - \int \sigma(x,h_F)dF(x) = \int (y(s) - h_F(s))^2 \, d\mu(s) \geqslant 0.$$

Thus, h_F is a minimizer. If equality holds, then $y(s) = h_F(s)$ for μ-almost every s. Since y and h_F are continuous and μ has full support, this implies $y = h_F$ on all of S. Hence the minimizer is unique. Finally, for every $x \in X$, $\hat{y}(\delta_x) = h(x,\cdot)$ and $\sigma(x,\hat{y}(\delta_x)) = \int (h(x,s) - h(x,s))^2 \, d\mu(s) = 0$. Therefore, σ is a forecast error. ■

Proof of Lemma 4. By definition, $\Sigma(F) = \int\int (h(x,s) - h(F,s))^2 \, d\mu(s)dF(x)$. Expanding the square gives $\Sigma(F) = \int\int h(x,s)^2 d\mu(s)dF(x) - \int h(F,s)^2 d\mu(s)$. Since $H(x,x) = \int h(x,s)^2 d\mu(s)$ and

$$\int h(F,s)^2 d\mu(s) = \int\int\int h(x,s)h(\tilde{x},s)d\mu(s)dF(x)dF(\tilde{x}) = \int\int H(x,\tilde{x})dF(x)dF(\tilde{x}),$$

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

We extend Σ to the space of signed measures by

$$\Sigma(F+M) = \int H(x,x)d(F+M)(x) - \int\int H(x,\tilde{x})d(F+M)(x)d(F+M)(\tilde{x}).$$

Expanding around F yields

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

Therefore, for the direction $M = \delta_z - F$,

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

Thus,

$$D\Sigma(F,\delta_z-F) = H(z,z)-2\int H(z,\tilde{x})dF(\tilde{x})-\int H(x,x)dF(x)+2\int\int H(x,\tilde{x})dF(x)dF(\tilde{x}),$$

which is equivalently

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

$\blacksquare$

Proof of Lemma 6. 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

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

Combining the first two chains of inequalities yields

$$\int\int u(x,y)\, d\hat{F}(x)\, d\hat{H}(y) \geqslant \int\int u(x,y)\, d\tilde{F}(x)\, d\hat{H}(y) \qquad \forall\tilde{F} \in ext\left(\overline{\mathcal{F}}\right). \tag{23}$$

Now 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) 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

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

which contradicts equation (23).

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

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

Before proving Lemma 7, we state and prove an intermediate result.

Lemma 8. *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 combination of finitely many points in $ext(\mathcal{F}_\Gamma(\overline{X}))$.*

Proof. Let $\mathcal{F}_e = ext(\mathcal{F}_\Gamma(\overline{X}))$ and endow it with the relative topology inherited from $\mathcal{F}_\Gamma(\overline{X})$. Since $\mathcal{F}_\Gamma(\overline{X})$ is compact and metrizable, Choquet's theorem implies that for every $F \in \mathcal{F}_\Gamma(\overline{X})$ there exists $\lambda \in \Delta(\mathcal{F}_e)$ such that

$$\int \varphi(x)dF(x) = \int_{\mathcal{F}_e} \left(\int \varphi(x)d\tilde{F}(x) \right) d\lambda(\tilde{F}) \qquad \forall \varphi \in C(\overline{X}).$$

By Theorem 15.10 in Aliprantis and Border (2006), there exists a sequence $\lambda^n \in \Delta_0(\mathcal{F}_e)$ such that $\lambda^n \to \lambda$. Define $F^n = \int_{\mathcal{F}_e} \tilde{F} d\lambda^n(\tilde{F})$. Each F^n is a finite convex combination of points in $ext(\mathcal{F}_\Gamma(\overline{X}))$. Moreover, for every $\varphi \in C(\overline{X})$, the map $\tilde{F} \mapsto \int \varphi(x)d\tilde{F}(x)$ is continuous, so

$$\int \varphi(x)dF^n(x) = \int_{\mathcal{F}_e} \left(\int \varphi(x)d\tilde{F}(x) \right) d\lambda^n(\tilde{F}) \to \int_{\mathcal{F}_e} \left(\int \varphi(x)d\tilde{F}(x) \right) d\lambda(\tilde{F}) = \int \varphi(x)dF(x).$$

Thus, $F^n \to F$ weakly. ■

Proof of Lemma 7. Let $N = (m+1)(k+1)$. Since $\mathcal{F}_\Gamma(\overline{X})$ is compact and V is continuous, there exists $\hat{F} \in \operatorname{argmax}_{F \in \mathcal{F}_\Gamma(\overline{X})} V(F)$. By Lemma 8, there exists a sequence $\hat{F}^n \to \hat{F}$ such that $\hat{F}^n \in co(\mathcal{E}^n)$ for some finite set $\mathcal{E}^n \subseteq ext(\mathcal{F}_\Gamma(\overline{X}))$. By Theorem 3, for every n there exists $F^n \in \operatorname{argmax}_{F \in co(\mathcal{E}^n)} V(F)$ such that F^n is supported on at most N points of $\overline{X}$. Since $\hat{F}^n \in co(\mathcal{E}^n)$, we have $V(F^n) \geqslant V(\hat{F}^n)$ for every n.

By compactness, a subsequence of F^n converges weakly to some $F^* \in \mathcal{F}_\Gamma(\overline{X})$. The set of probability measures on $\overline{X}$ with support of cardinality at most N is weakly closed: indeed, writing $F^n = \sum_{i=1}^N \alpha_i^n \delta_{x_i^n}$, compactness of $\Delta(\{1, \ldots, N\}) \times \overline{X}^N$ lets us pass to a further subsequence with $\alpha_i^n \to \alpha_i$ and $x_i^n \to x_i$ for every i, so the limit is $\sum_{i=1}^N \alpha_i \delta_{x_i}$. Thus, F^* is supported on at most N points.

Finally, by continuity of V, $V(F^*) = \lim_n V(F^n) \geqslant \lim_n V(\hat{F}^n) = V(\hat{F})$. Since $\hat{F}$ is optimal over $\mathcal{F}_\Gamma(\overline{X})$, F^* is also optimal, as desired. ■

Proof of Corollary 3. By Theorem 15.10 in Aliprantis and Border (2006), there exists a sequence of finitely supported $\mu^n \in \Delta(S)$ such that $\mu^n \to \mu$. For every n, let $m^n = |\operatorname{supp} \mu^n|$. The utility V^n is a GMM utility with m^n moments and no exogenous moment restrictions, so Proposition 3 implies that there exists a solution F^n of $\max_{F\in\Delta(X)} V^n(F)$ supported on at most m^n+1 points of X.

We next show that $V^n \to V$ uniformly on $\Delta(X)$, where V is the GMM adversarial forecaster representation induced by (h,μ). For every n, define $H^n(x,\tilde{x}) = \int h(x,s)h(\tilde{x},s)d\mu^n(s)$ and $H(x,\tilde{x}) = \int h(x,s)h(\tilde{x},s)d\mu(s)$. We claim that $H^n \to H$ uniformly on $X \times X$. To see this, let $\mathcal{G} = \{s \mapsto h(x,s)h(\tilde{x},s) : (x,\tilde{x}) \in X \times X\} \subseteq C(S)$. Since h is continuous and $X \times X$ is compact, $\mathcal{G}$ is compact in the sup-norm topology of $C(S)$. Fix $\varepsilon > 0$ and choose $g_1,\ldots,g_J \in \mathcal{G}$ such that every $g \in \mathcal{G}$ is within ε of some g_j in the sup norm. Since $\mu^n \to \mu$ weakly, for all large enough n,

$$\max_{j\leqslant J}\left|\int g_j(s)d\mu^n(s) - \int g_j(s)d\mu(s)\right| < \varepsilon.$$

Therefore, for all large enough n, $\sup_{g\in\mathcal{G}}\left|\int g(s)d\mu^n(s) - \int g(s)d\mu(s)\right| < 3\varepsilon$, which proves that $H^n \to H$ uniformly on $X \times X$.

For every $F \in \Delta(X)$, $V^n(F) = \int H^n(x,x)dF(x) - \int\int H^n(x,\tilde{x})dF(x)dF(\tilde{x})$, and the analogous formula holds for V. Thus,

$$\sup_{F\in\Delta(X)} |V^n(F) - V(F)| \leqslant 2 \sup_{(x,\tilde{x})\in X\times X} |H^n(x,\tilde{x}) - H(x,\tilde{x})|,$$

so $V^n \to V$ uniformly on $\Delta(X)$.

We now prove that $F^n \to F$ weakly. Since $\Delta(X)$ is compact, every subsequence of F^n has a further subsequence, still denoted by F^n, that converges weakly to some $\hat{F} \in \Delta(X)$. Let $\eta_n = \sup_{G\in\Delta(X)} |V^n(G) - V(G)|$. Then $\eta_n \to 0$. Since F^n maximizes V^n, for every $G \in \Delta(X)$, $V(F^n) \geqslant V^n(F^n) - \eta_n \geqslant V^n(G) - \eta_n \geqslant V(G) - 2\eta_n$. Taking limits along the convergent subsequence and using the continuity of V gives $V(\hat{F}) \geqslant V(G)$ for every $G \in \Delta(X)$. Thus, $\hat{F}$ solves $\max_{G\in\Delta(X)} V(G)$. Proposition 5 established that this problem has a unique full-support solution F, so $\hat{F} = F$. Since every weakly convergent subsequence of F^n has the same limit F, the full sequence satisfies $F^n \to F$ weakly.

It remains to prove the convergence of the supports. Let $K^n = \operatorname{supp} F^n$. Since X is compact, every subsequence of K^n has a further subsequence that converges in the Hausdorff topology to some compact set $\hat{X} \subseteq X$. Along such a subsequence, we also have $F^n \to F$ weakly. By Box 1.13 in Santambrogio (2015), this implies that $\operatorname{supp} F \subseteq \hat{X}$. Since Proposition 5 gives $\operatorname{supp} F = X$ and $\hat{X} \subseteq X$, it follows that

$\hat{X} = X$. Therefore, every Hausdorff accumulation point of $\operatorname{supp} F^n$ is equal to X, so $\operatorname{supp} F^n \to X = \operatorname{supp} F$ in the Hausdorff topology. ■

Online Appendix I.B: Theorems cited in the main appendix

Theorem 4 (Theorem 2.1 in Winkler (1988)). *Let $(X, \mathcal{X})$ be a measurable space and let $\overline{\mathcal{F}} \subseteq \mathcal{F}$ be a simplex of probability measures whose extreme points are Dirac measures. Fix measurable functions $g_1, ..., g_n$ over X and real number $c_1, ..., c_n$. Consider the set*

$$\mathcal{H} = \left\{ F \in \overline{\mathcal{F}} : \forall i \in \{1, ..., n\}, g_i \text{ is } F\text{-integrable and } \int g_i(x)dF(x) \leqslant c_i \right\}.$$

Then $\mathcal{H}$ is convex and each of its extreme points is supported on up to $n+1$ points of X. The same support bound holds if the inequalities defining $\mathcal{H}$ are replaced by equalities.

The following corollary records the mixed equality and inequality case needed in Proposition 3.

Corollary 6. *Fix continuous functions $g_1, \ldots, g_k, h_1, \ldots, h_m \in C(\overline{X})$ and constants $c_1, \ldots, c_k, d_1, \ldots, d_m$. Every extreme point of*

$$\mathcal{H} = \left\{ F \in \Delta(\overline{X}) : \int g_i(x)dF(x) \leqslant c_i \text{ for all } i, \quad \int h_j(x)dF(x) = d_j \text{ for all } j \right\}$$

is supported on at most $k+m+1$ points of $\overline{X}$.

Proof. Fix $F \in ext(\mathcal{H})$, and let $I(F) = \left\{ i \in \{1, \ldots, k\} : \int g_i(x)dF(x) = c_i \right\}$. Define

$$\mathcal{H}_{I(F)} = \left\{ G \in \Delta(\overline{X}) : \int g_i(x)dG(x) = c_i \text{ for all } i \in I(F), \quad \int h_j(x)dG(x) = d_j \text{ for all } j \right\}.$$

We claim that $F \in ext(\mathcal{H}_{I(F)})$. Otherwise, there are distinct $F^+, F^- \in \mathcal{H}_{I(F)}$ such that $F = (F^+ + F^-)/2$. For $\varepsilon \in (0,1)$, set $F_\varepsilon^\pm = (1-\varepsilon)F + \varepsilon F^\pm$. These lotteries preserve all the equalities defining $\mathcal{H}_{I(F)}$. For every $i \notin I(F)$, $\int g_i(x)dF(x) < c_i$ and $\int g_i(x)dF_\varepsilon^\pm(x) \to \int g_i(x)dF(x)$ as $\varepsilon \downarrow 0$. Since there are finitely many inactive inequalities, for all sufficiently small $\varepsilon > 0$ both F_ε^+ and F_ε^- belong to $\mathcal{H}$. They are distinct and satisfy $F = (F_\varepsilon^+ + F_\varepsilon^-)/2$, contradicting $F \in ext(\mathcal{H})$.

Thus, $F \in ext(\mathcal{H}_{I(F)})$. If $|I(F)| + m = 0$, then F is an extreme point of $\Delta(\overline{X})$ and hence is a Dirac measure. Otherwise, the equality case of Theorem 4, applied to the $|I(F)| + m$ equality moment conditions defining $\mathcal{H}_{I(F)}$, implies that $|\operatorname{supp} F| \leqslant |I(F)| + m + 1 \leqslant k + m + 1$. ■

Online Appendix I.C: Proofs of Section 6.2

We first establish two technical lemmas. The first gives an optimality condition for the free-information problem.

Lemma 9. *Suppose $\kappa = 0$ and $\gamma > 0$. A rule π solves* (12) *if and only if, for every $\omega \in \Omega$, it assigns positive probability only to actions that maximize $w(x(a,\omega), F_\pi) = x(a,\omega) + \gamma(x(a,\omega) - m_\pi)^2$.*

Proof. For a rule π, write $U_\gamma(\pi) = m_\pi + \gamma \mathrm{Var}_\pi(x)$. For any rules π and $\widetilde{\pi}$, direct expansion gives

$$U_\gamma(\widetilde{\pi}) - U_\gamma(\pi) = \sum_{\omega,a} \mu(\omega)\big(\widetilde{\pi}(a \mid \omega) - \pi(a \mid \omega)\big) w(x(a,\omega), F_\pi) - \gamma(m_{\widetilde{\pi}} - m_\pi)^2. \quad (24)$$

Suppose first that the support condition in the statement holds. For every $\widetilde{\pi}$, statewise optimality gives $\sum_{\omega,a} \mu(\omega)\widetilde{\pi}(a \mid \omega) w(x(a,\omega), F_\pi) \leqslant \sum_{\omega,a} \mu(\omega)\pi(a \mid \omega) w(x(a,\omega), F_\pi)$. (24) then implies $U_\gamma(\widetilde{\pi}) \leqslant U_\gamma(\pi)$, so π is optimal.

Conversely, suppose that π is optimal but that, for some state ω, it assigns positive probability to an action a that does not maximize $w(x(b,\omega), F_\pi)$ over $b \in A$. Choose a maximizing action a' and, for sufficiently small $\varepsilon > 0$, obtain π^ε by moving probability ε from a to a' in state ω. The first term on the right-hand side of (24) is $\varepsilon\mu(\omega)\big[w(x(a',\omega), F_\pi) - w(x(a,\omega), F_\pi)\big] > 0$, whereas the absolute value of the final term is bounded by a constant times ε^2. Thus, $U_\gamma(\pi^\varepsilon) > U_\gamma(\pi)$ for all sufficiently small $\varepsilon > 0$, contradicting optimality. This proves the equivalence. ■

Lemma 10. *In the binary problem, no optimal rule has accuracy below* $1/2$. *Among all rules with accuracy $r \in [1/2, 1]$, the minimum mutual information is* $\log 2 - H(r)$, *and it is attained by the symmetric rule with $\pi(1 \mid 1) = \pi(-1 \mid -1) = r$.*

Proof. If a rule π has accuracy $r < 1/2$, define $\tilde{\pi}(a \mid \omega) = \pi(-a \mid \omega)$. The rule $\tilde{\pi}$ has accuracy $1 - r > 1/2$, the same mutual information because relabeling actions is a bijection, and the same payoff variance because $r(1-r) = (1-r)r$. Its expected payoff is larger by $b(1-2r) > 0$, so the original rule cannot be optimal.

Now fix a rule with accuracy $r \geqslant 1/2$, let the random variables A and Ω denote its action and state, and let $E = \mathbf{1}_{A \neq \Omega}$ be its error indicator. Since the prior is uniform, $H(\Omega) = \log 2$. Conditional on A, the variables E and Ω determine one another: $E = 0$ if and only if $\Omega = A$, while $E = 1$ if and only if $\Omega = -A$. Therefore, $H(\Omega \mid A) = H(E \mid A) \leqslant H(E) = H(r)$, where the last equality uses the symmetry of binary entropy. It follows that $I(\pi) = I(A; \Omega) = H(\Omega) - H(\Omega \mid A) \geqslant \log 2 - H(r)$. Under the symmetric rule, the unconditional distribution of A is uniform and the error probability conditional on either action is $1 - r$. Thus, E is independent of A, so $H(E \mid A) = H(E)$ and the inequality holds with equality. ∎

Proof of Proposition 6. For every rule π, write $s_\pi = \sum_{\omega,a} \mu(\omega)\pi(a \mid \omega)x(a,\omega)^2$, so its utility is $m_\pi + \gamma(s_\pi - m_\pi^2)$. Let π and $\tilde{\pi}$ be optimal, and set $\overline{\pi} = (\pi + \tilde{\pi})/2$. Since $m_{\overline{\pi}} = (m_\pi + m_{\tilde{\pi}})/2$ and $s_{\overline{\pi}} = (s_\pi + s_{\tilde{\pi}})/2$, we have $U_\gamma(\overline{\pi}) = \frac{U_\gamma(\pi) + U_\gamma(\tilde{\pi})}{2} + \frac{\gamma}{4}(m_\pi - m_{\tilde{\pi}})^2$. If the two means differed, $\overline{\pi}$ would be strictly better than both optimal rules. Thus, all optimal rules induce the same expected payoff.

Fix an optimal rule π^*. By Lemma 9, every action in the support of $\pi^*(\cdot \mid \omega)$ maximizes $x + \gamma(x - m_{\pi^*})^2$ over the payoffs available in state ω. This function is strictly convex in x. Therefore, no payoff strictly between x_ω^- and x_ω^+ can maximize it: if $x = \lambda x_\omega^- + (1-\lambda)x_\omega^+$ for some $\lambda \in (0,1)$, strict convexity gives

$$x + \gamma(x - m_{\pi^*})^2 < \lambda\left[x_\omega^- + \gamma(x_\omega^- - m_{\pi^*})^2\right] + (1-\lambda)\left[x_\omega^+ + \gamma(x_\omega^+ - m_{\pi^*})^2\right],$$

which is no greater than the larger endpoint value.

The difference between the local payoffs at the two endpoints is

$$\begin{aligned} x_\omega^+ + \gamma(x_\omega^+ - m_{\pi^*})^2 - x_\omega^- - \gamma(x_\omega^- - m_{\pi^*})^2 &= (x_\omega^+ - x_\omega^-)\left[1 + \gamma(x_\omega^+ + x_\omega^- - 2m_{\pi^*})\right] \\ &= 2\gamma(x_\omega^+ - x_\omega^-)\left(c_\omega - m_{\pi^*} + \frac{1}{2\gamma}\right). \end{aligned}$$

Its sign yields the threshold rule. If the c_ω are distinct, equality can hold in at most one state, so different payoff levels can both receive positive probability in at most one state. ∎

Proof of Corollary 4. Fix $0 < \gamma_1 < \gamma_2$ and let π_i^* be optimal at γ_i. Write $m_i = m_{\pi_i^*}$ and $v_i = \mathrm{Var}_{\pi_i^*}(x)$. Optimality gives $m_1 + \gamma_1 v_1 \geqslant m_2 + \gamma_1 v_2$ and $m_2 + \gamma_2 v_2 \geqslant m_1 + \gamma_2 v_1$. Adding these inequalities yields $(\gamma_2 - \gamma_1)(v_2 - v_1) \geqslant 0$, so $v_2 \geqslant v_1$. Substituting this into the first inequality gives $m_1 - m_2 \geqslant \gamma_1(v_2 - v_1) \geqslant 0$, so $m_2 \leqslant m_1$. ∎

Proof of Example 9. Let $q_i = \pi(1 \mid i)$ for $i \in \{0, 1\}$. When $\kappa = 0$, expected payoff is $(q_0 + 1 + q_1)/2$ and the second moment of payoff is $(q_0 + 1 + 3q_1)/2$. Holding $q_0 + q_1$ fixed, increasing q_1 and decreasing q_0 by the same amount leaves the mean unchanged and strictly raises the second moment. Therefore, an optimum with $q_0 > 0$ must have $q_1 = 1$. If $q_0 = 0$ and $q_1 < 1$, the derivative of utility with respect to q_1 is $1/2 + \gamma(3/2 - m_\pi) > 0$, so this rule is not optimal either. Thus, every optimum has $q_1 = 1$. With $q = q_0$, the derivative of free-information utility is $(1 - \gamma(1 + q))/2$. It follows that $q = 1$ if $\gamma \leqslant 1/2$, $q = 1/\gamma - 1$ if $1/2 < \gamma < 1$, and $q = 0$ if $\gamma \geqslant 1$, proving (13).

Now suppose $\gamma > 1$ and $\kappa < \infty$. Assume, toward a contradiction, that an optimal rule has $I(\pi) = 0$. Then, the action is independent of the state. Let q be the probability of action 1. Expected payoff is $1/2 + q$, and payoff variance is $1/4 + q(1-q)$. Among uninformative rules, the objective is therefore $N(q) = \frac{1}{2} + q + \gamma\big(\frac{1}{4} + q(1-q)\big)$. Since N is strictly concave and $\gamma > 1$, its unique maximizer is $q^\circ = (1 + \gamma)/(2\gamma) \in (0, 1)$. Thus, any globally optimal uninformative rule must use this marginal action probability.

For sufficiently small $d > 0$, consider the rule with $\pi^d(1 \mid 0) = q^\circ - d$ and $\pi^d(1 \mid 1) = q^\circ + d$. Its marginal probability of action 1, and hence its expected payoff and the variance of the action, are unchanged. Since $\mathrm{Cov}_{\pi^d}(a, \omega) = d/2$, its payoff variance increases by d. Its mutual information is $I(\pi^d) = H(q^\circ) - \frac{1}{2}H(q^\circ - d) - \frac{1}{2}H(q^\circ + d)$. Because $q^\circ \in (0, 1)$, choose $\overline{d} > 0$ such that $[q^\circ - \overline{d}, q^\circ + \overline{d}] \subset (0, 1)$. The second derivative $H''(q) = -1/[q(1-q)]$ is bounded in absolute value on this interval; let M be such a bound. Taylor's theorem applied separately at $q^\circ - d$ and $q^\circ + d$ implies that, for every $d \in (0, \overline{d})$, $0 \leqslant I(\pi^d) = H(q^\circ) - \frac{1}{2}H(q^\circ - d) - \frac{1}{2}H(q^\circ + d) \leqslant \frac{M}{2}d^2$. The change in the objective is therefore at least $\gamma d - \kappa M d^2/2$, which is strictly positive for all sufficiently small $d > 0$ because $\gamma > 0$ and $\kappa < \infty$. This contradicts optimality of the uninformative rule. Thus, every optimum has strictly positive mutual information. ■

Proof of Proposition 7. By Lemma 10, the binary problem is equivalent to (14). Let $J(r)$ denote its objective. For $r \in (1/2, 1)$, $J'(r) = b + \gamma b^2(1-2r) - \kappa \log\left(\frac{r}{1-r}\right)$ and $J''(r) = -2\gamma b^2 - \kappa(1/r + 1/(1-r)) < 0$. Thus, J is strictly concave. Moreover, $J'(1/2) = b > 0$, while $J'(r) \to -\infty$ as $r \uparrow 1$. Therefore, there is a unique maximizer $r^* \in (1/2, 1)$, characterized by (15). Define $G(r, \gamma, \kappa) = b + \gamma b^2(1-2r) - \kappa \log(r/(1-r))$. At the optimum, $G(r^*, \gamma, \kappa) = 0$, while

$$G_r = -2\gamma b^2 - \kappa\left(\frac{1}{r} + \frac{1}{1-r}\right) < 0, \quad G_\gamma = b^2(1-2r^*) < 0, \quad G_\kappa = -\log\left(\frac{r^*}{1-r^*}\right) < 0.$$

The implicit function theorem gives $\partial r^*/\partial\gamma = -G_\gamma/G_r < 0$ and $\partial r^*/\partial\kappa = -G_\kappa/G_r < 0$. Since $d(\log 2 - H(r))/dr = \log(r/(1-r)) > 0$ for $r > 1/2$, I^* is strictly decreasing in both parameters as well.

Finally, suppose that r^* does not converge to $1/2$ as $\gamma \to \infty$. Then there are $\varepsilon > 0$ and a sequence $\gamma_n \to \infty$ such that $r^*(\gamma_n) \geqslant 1/2 + \varepsilon$. (15) then implies $b - 2\varepsilon\gamma_n b^2 \geqslant \kappa \log\left(\frac{1/2+\varepsilon}{1/2-\varepsilon}\right)$, which is impossible because the left-hand side converges to $-\infty$ and the right-hand side is finite. Thus, $r^* \to 1/2$, and continuity of binary entropy gives $I^* \to 0$. ■

Proof of Corollary 5. Suppose first that $\kappa = 0$. By Lemma 10, the objective is $J(r) = br + \gamma b^2 r(1-r)$. Its derivative is $J'(r) = b + \gamma b^2(1-2r)$. If $\gamma b \leqslant 1$, then $J'(1) = b(1-\gamma b) \geqslant 0$, and since J' is weakly decreasing, J is increasing on $[1/2, 1]$ and is uniquely maximized at $r = 1$. If $\gamma b > 1$, the unique solution of $J'(r) = 0$ is $r = 1/2 + 1/(2\gamma b) \in (1/2, 1)$; strict concavity makes it the unique optimum. Differentiating this expression with respect to b gives $-1/(2\gamma b^2) < 0$.

Now suppose $\kappa > 0$, and set $z^* = 2r^* - 1 \in (0, 1)$. (15) is equivalent to

$$\gamma b^2 z^* + 2\kappa \operatorname{arctanh}(z^*) = b. \tag{25}$$

For every $b > 0$, the left-hand side is continuous and strictly increasing in z^* from zero to infinity, so (25) uniquely determines z^*. Implicit differentiation gives

$$\frac{dz^*}{db} = \frac{1 - 2\gamma b z^*}{\gamma b^2 + 2\kappa/(1-(z^*)^2)}. \tag{26}$$

The denominator is strictly positive. A stationary point therefore satisfies $2\gamma b z^* = 1$. Substituting this equality into (25) yields $8\gamma\kappa z^* \operatorname{arctanh}(z^*) = 1$. The function $z \mapsto z \operatorname{arctanh}(z)$ is strictly increasing from zero to infinity on $(0, 1)$, because its derivative is $\operatorname{arctanh}(z) + z/(1 - z^2) > 0$. Hence, there is exactly one stationary point.

As $b \downarrow 0$, (25) implies $2\kappa \operatorname{arctanh}(z^*) \leqslant b$, so $z^* \to 0$. Dividing (25) by b gives

$$\gamma b z^* + 2\kappa \frac{\operatorname{arctanh}(z^*)}{z^*} \frac{z^*}{b} = 1.$$

Since $\operatorname{arctanh}(z)/z \to 1$ as $z \to 0$, it follows that $z^*/b \to 1/(2\kappa)$. In particular, the numerator in (26) is positive for all sufficiently small b.

As $b \to \infty$, (25) implies $z^* \leqslant 1/(\gamma b)$, so $z^* \to 0$. Dividing (25) by b gives $\gamma b z^* + \frac{2\kappa}{b} \operatorname{arctanh}(z^*) = 1$. The second term converges to zero, and therefore $\gamma b z^* \to 1$. Thus, the numerator in (26) converges to -1 and is negative for all sufficiently large b. Since the derivative is continuous and has exactly one zero, z^*, and therefore r^*, is strictly increasing before the unique stationary point and strictly decreasing afterward. Since $I^* = \log 2 - H((1 + z^*)/2)$ is strictly increasing in z^*, it has the same monotonicity.

Because $r^* - 1/2 = z^*/2$, the limits $z^*/b \to 1/(2\kappa)$ and $\gamma b z^* \to 1$ give (17). The turning-point condition follows from $2\gamma b z^* = 1$.

It remains to prove (18). Fix $b > 0$. As $\kappa \downarrow 0$, the objective in (14) converges uniformly on $[1/2, 1]$ to the free-information objective, because $0 \leqslant \kappa\,(\log 2 - H(r)) \leqslant \kappa \log 2$ for every $r \in [1/2, 1]$. Since the free-information optimizer is unique, the maximizer $r^*(b, \kappa)$ converges to the value in (16). For all sufficiently small b, that value is one, and hence

$$\lim_{b \downarrow 0} \lim_{\kappa \downarrow 0} r^*(b, \kappa) = 1.$$

Conversely, the low-stakes limit proved above gives $\lim_{b \downarrow 0} r^*(b, \kappa) = 1/2$ for every fixed $\kappa > 0$. Therefore,

$$\lim_{\kappa \downarrow 0} \lim_{b \downarrow 0} r^*(b, \kappa) = \frac{1}{2}.$$

This proves (18). ∎

Online Appendix II: Optimization

This appendix collects additional optimization results for the adversarial forecaster and adversarial expected utility representations that are of independent interest.

Online Appendix II.A: Optimal lotteries in the adversarial EU model

Here we provide two alternative characterizations of optimal lotteries under the adversarial expected utility model.

Proposition 8. *Let V be an adversarial expected utility representation (Y, u) and let $\overline{\mathcal{F}} \subseteq \mathcal{F}$ be a convex and compact set. The following are equivalent:*

(i) $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} V(F)$

(ii) There exists $H \in \mathcal{H}(\hat{Y}(F^*))$ such that $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} \int\int u(x, y) dH(y) dF(x)$.

(iii) For all $F \in \overline{\mathcal{F}}$, there exists $y \in \hat{Y}(F^*)$ such that $\int u(x, y) dF^*(x) \geqslant \int u(x, y) dF(x)$.

The equivalence between (i) and (iii) is similar to Proposition 1 in Loseto and Lucia (2021), with the important difference that they consider quasiconcave representations and restrict to a finite set of utilities (which corresponds to a finite Y in our notation).

Proof. For every $F \in \overline{\mathcal{F}}$ and $H \in \mathcal{H}$, write $L(F, H) = \int\int u(x, y) dH(y) dF(x)$. Also, let $\mathcal{W} = \{u(\cdot, y) : y \in Y\} \subseteq C(X)$ and let $co(\mathcal{W})$ denote its closed convex hull. Since Y is compact and u is continuous, $\mathcal{W}$ is compact. For every F,

$$V(F) = \min_{y \in Y} \int u(x, y) dF(x) = \min_{w \in co(\mathcal{W})} \int w(x) dF(x),$$

because the minimum of a continuous linear functional over $co(\mathcal{W})$ is the same as its minimum over $\mathcal{W}$.

We first prove that (ii) implies (i). If $H \in \mathcal{H}(\hat{Y}(F^*))$ and F^* maximizes $L(\cdot, H)$ over $\overline{\mathcal{F}}$, then for every $\tilde{F} \in \overline{\mathcal{F}}$, $V(F^*) = L(F^*, H) \geqslant L(\tilde{F}, H) \geqslant V(\tilde{F})$, so $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} V(F)$.

Next we prove that (i) implies (ii). Suppose $F^* \in \operatorname{argmax}_{F\in\overline{\mathcal{F}}} V(F)$. By Sion's minmax theorem,

$$\max_{F\in\overline{\mathcal{F}}} \min_{w\in co(\mathcal{W})} \int w(x)dF(x) = \min_{w\in co(\mathcal{W})} \max_{F\in\overline{\mathcal{F}}} \int w(x)dF(x).$$

Choose $w^* \in co(\mathcal{W})$ solving the minimization problem on the right-hand side. Then

$$V(F^*) \leqslant \int w^*(x)dF^*(x) \leqslant \max_{F\in\overline{\mathcal{F}}} \int w^*(x)dF(x) = \max_{F\in\overline{\mathcal{F}}} V(F) = V(F^*).$$

Thus all inequalities are equalities, so F^* maximizes $F \mapsto \int w^*(x)dF(x)$ over $\overline{\mathcal{F}}$, and $\int w^*(x)dF^*(x) = \min_{w\in co(\mathcal{W})} \int w(x)dF^*(x) = V(F^*)$. Since $w^* \in co(\mathcal{W})$, there exists $H \in \mathcal{H}$ such that $w^*(x) = \int u(x,y)dH(y)$. Moreover, $\int\int u(x,y)dH(y)dF^*(x) = V(F^*) = \min_{y\in Y} \int u(x,y)dF^*(x)$. Because $\int u(x,y)dF^*(x) \geqslant V(F^*)$ for every $y \in Y$, the last equality implies that H is supported on $\hat{Y}(F^*)$. Hence $H \in \mathcal{H}(\hat{Y}(F^*))$, and F^* maximizes $L(\cdot, H)$ over $\overline{\mathcal{F}}$, proving (ii).

It remains to prove the equivalence between (ii) and (iii). If (ii) holds, then for every $F \in \overline{\mathcal{F}}$, $\int_{\hat{Y}(F^*)} \left(\int u(x,y)dF^*(x) - \int u(x,y)dF(x)\right) dH(y) \geqslant 0$. Therefore at least one $y \in \hat{Y}(F^*)$ satisfies $\int u(x,y)dF^*(x) \geqslant \int u(x,y)dF(x)$, which gives (iii).

Conversely, suppose (iii) holds. Define

$$G(F,y) = \int u(x,y)dF^*(x) - \int u(x,y)dF(x) \qquad (F,y) \in \overline{\mathcal{F}} \times \hat{Y}(F^*).$$

Condition (iii) says that $\max_{y\in\hat{Y}(F^*)} G(F,y) \geqslant 0$ for all $F \in \overline{\mathcal{F}}$. Since G is continuous, affine in F, and its integral is affine in probability measures $H \in \mathcal{H}(\hat{Y}(F^*))$, Sion's minmax theorem gives

$$0 \leqslant \min_{F\in\overline{\mathcal{F}}} \max_{y\in\hat{Y}(F^*)} G(F,y) = \max_{H\in\mathcal{H}(\hat{Y}(F^*))} \min_{F\in\overline{\mathcal{F}}} \int_{\hat{Y}(F^*)} G(F,y)dH(y).$$

Thus, there exists $H \in \mathcal{H}(\hat{Y}(F^*))$ such that $\int\int u(x,y)dH(y)dF^*(x) \geqslant \int\int u(x,y)dH(y)dF(x)$ $\forall F \in \overline{\mathcal{F}}$, which is exactly (ii). ∎

Online Appendix II.B: Robust solutions

This section shows that the finite-support property of Theorem 2 holds generically for "robustly optimal" lotteries for a parametric adversarial forecaster V over $\overline{\mathcal{F}}$. Without loss of generality, assume $\overline{X} = cl\left(\bigcup_{F\in\mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F\right)$. Let $\overline{\mathcal{E}} = cl\left(ext\left(\mathcal{F}_\Gamma(\overline{X})\right)\right)$, and fix an increasing sequence of finite sets $\mathcal{E}^n \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$ such that $cl\left(\bigcup_{n\in\mathbb{N}} \mathcal{E}^n\right) = \overline{\mathcal{E}}$. Then $cl\left(\bigcup_{n\in\mathbb{N}} co(\mathcal{E}^n)\right) = \mathcal{F}_\Gamma(\overline{X})$ and $cl\left(\bigcup_{n\in\mathbb{N}} \overline{X}_{\mathcal{E}^n}\right) = \overline{X}$.[39] For every $F \in \mathcal{F}_\Gamma(\overline{X})$, we say that a sequence $F^n \in co(\mathcal{E}^n)$ is a *finitely approximating sequence of F along $(\mathcal{E}^n)$* if $F^n \to F$ weakly.

Definition 7. Fix $w \in C(\overline{X})$ and a lottery F that solves

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

We say that F is a *robust solution at w along $(\mathcal{E}^n)$* if there exists a finitely approximating sequence $F^n \in co(\mathcal{E}^n)$ of F such that, for every $n \in \mathbb{N}$,

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

In words, an optimal lottery F is robust along $(\mathcal{E}^n)$ if it can be approximated by lotteries that are generated by the finite sets $\mathcal{E}^n$ and that are optimal within the corresponding finite-dimensional feasible sets.

Theorem 5. *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. There exists a residual subset $\overline{\mathcal{W}} \subseteq C(\overline{X})$ such that, for every $w \in \overline{\mathcal{W}}$, every robust solution at w along $(\mathcal{E}^n)$ has finite support on no more than $(k+1)(m+1)$ points of $\overline{X}$.*

The proof uses 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*

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

[39] The first equality follows from Choquet's theorem and the density of $\bigcup_n \mathcal{E}^n$ in the closure of the extreme points. For the second, every neighborhood of a point in the support of a feasible lottery has positive probability under some extreme point in a Choquet representation; weak approximation of that extreme point by elements of $\bigcup_n \mathcal{E}^n$ and the Portmanteau theorem imply that the neighborhood meets $\overline{X}_{\mathcal{E}^n}$ for some n.

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

Proof. Let $R : C(\overline{X}) \to \mathbb{R}^{\hat{X}}$ denote the restriction map $R(w) = w_{|\hat{X}}$. Since R is continuous and $\hat{\mathcal{W}}$ is open, $\overline{\mathcal{W}} = R^{-1}(\hat{\mathcal{W}})$ is open.

To prove density, fix $w \in C(\overline{X})$ and $\varepsilon > 0$. Since $\hat{\mathcal{W}}$ is dense in $\mathbb{R}^{\hat{X}}$, choose $\hat{w} \in \hat{\mathcal{W}}$ such that $\max_{\hat{x}\in\hat{X}} |\hat{w}(\hat{x}) - w(\hat{x})| < \varepsilon/3$. By uniform continuity of w and finiteness of $\hat{X}$, choose $\eta > 0$ such that the balls $B_\eta(\hat{x})$ are pairwise disjoint and $d(x, \hat{x}) < \eta \Rightarrow |w(x) - w(\hat{x})| < \varepsilon/3$. 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 $v_{\hat{x}} : \overline{X} \to [0, 1]$ such that $v_{\hat{x}}(\hat{x}) = 1$ and $v_{\hat{x}}(x) = 0$ for all $x \in \overline{X} \backslash B_\eta(\hat{x})$. Define

$$\tilde{w}(x) = w(x)\big(1 - \max_{\hat{x}\in\hat{X}} v_{\hat{x}}(x)\big) + \sum_{\hat{x}\in\hat{X}} \hat{w}(\hat{x})v_{\hat{x}}(x).$$

Then, $\tilde{w}_{|\hat{X}} = \hat{w} \in \hat{\mathcal{W}}$, so $\tilde{w} \in \overline{\mathcal{W}}$. Moreover, $\tilde{w}(x) = w(x)$ outside $\cup_{\hat{x}\in\hat{X}} B_\eta(\hat{x})$, while for $x \in B_\eta(\hat{x})$, $|\tilde{w}(x) - w(x)| \leqslant |\hat{w}(\hat{x}) - w(x)| \leqslant |\hat{w}(\hat{x}) - w(\hat{x})| + |w(\hat{x}) - w(x)| < 2\varepsilon/3$. Thus, $\|\tilde{w} - w\|_\infty < \varepsilon$, proving that $\overline{\mathcal{W}}$ is dense in $C(\overline{X})$. ■

Proof of Theorem 5. For every $n \in \mathbb{N}$, let $\hat{\mathcal{W}}^n$ be the open dense subset of $\mathbb{R}^{\overline{X}_{\mathcal{E}^n}}$ given by point 1 of Theorem 3. By Lemma 11, the set $\overline{\mathcal{W}}^n = \big\{w \in C(\overline{X}) : w_{|\overline{X}_{\mathcal{E}^n}} \in \hat{\mathcal{W}}^n\big\}$ is open and dense in $C(\overline{X})$. Therefore, $\overline{\mathcal{W}} = \bigcap_{n\in\mathbb{N}} \overline{\mathcal{W}}^n$ is a residual subset of $C(\overline{X})$ by the Baire category theorem (see Theorem 3.46 in Aliprantis and Border (2006)). Fix $w \in \overline{\mathcal{W}}$ and let F^* be a robust solution at w along $(\mathcal{E}^n)$. By definition, there exists a sequence $F^n \to F^*$ such that, for every n,

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

Since every lottery in $co(\mathcal{E}^n)$ is supported on $\overline{X}_{\mathcal{E}^n}$ and $w_{|\overline{X}_{\mathcal{E}^n}} \in \hat{\mathcal{W}}^n$, point 1 of Theorem 3 implies that each F^n is supported on no more than $N = (k + 1)(m + 1)$ points of $\overline{X}_{\mathcal{E}^n}$.

It remains only to pass the support bound to the weak limit. The set of probability measures on $\overline{X}$ with support of cardinality at most N is weakly closed. Indeed, if $G^j \to G$ weakly and $G^j = \sum_{i=1}^{N} \alpha_i^j \delta_{x_i^j}$, allowing zero weights, then compactness of

$\Delta(\{1,\ldots,N\}) \times \overline{X}^N$ lets us pass to a subsequence with $\alpha_i^j \to \alpha_i$ and $x_i^j \to x_i$ for every i. Thus, for every $\varphi \in C(\overline{X})$,

$$\int \varphi dG = \lim_j \int \varphi dG^j = \lim_j \sum_{i=1}^N \alpha_i^j \varphi(x_i^j) = \sum_{i=1}^N \alpha_i \varphi(x_i),$$

so $G = \sum_{i=1}^N \alpha_i \delta_{x_i}$. Applying this observation to $F^n \to F^*$ shows that F^* is supported on no more than $N = (k+1)(m+1)$ points. Since $w \in \overline{\mathcal{W}}$ and the robust solution F^* were arbitrary, the result follows. ■

Online Appendix II.C: Formal result from Section 6.1

For the following result, assume g' is bounded away from zero on $[0, 1/4]$. Let $\underline{g} = \min_{z\in[0,1/4]} g'(z) > 0$ and $\overline{g} = \max_{z\in[0,1/4]} g'(z)$, and define $\underline{\beta} = \underline{g}/(2(\underline{g} + \overline{g}))$ and $\overline{\beta} = 1 - \underline{\beta}$. In particular, $0 < \underline{\beta} < \overline{\beta} < 1$.

Proposition 9. *For every $\beta \in [0, 1]$, there exists an optimal distribution F^* whose marginal over interim beliefs is supported on no more than three points. Moreover:*

1. *When $\beta \geqslant \overline{\beta}$, every optimal distribution F^* satisfies $F^*_\Delta = \delta_{p^*_F}$ (so the intermediate stage reveals no information), and p^*_F is optimal if and only if $p^*_F \in \text{argmax}_{p\in[0,1]}\{p\tilde{v} + \beta g(p - p^2)\}$.*
2. *When $\beta \leqslant \underline{\beta}$, every optimal distribution F^* satisfies $F^*_\Delta = (1-p^*_F)\delta_0 + p^*_F\delta_1$ (the state is fully revealed), and p^*_F is optimal if and only if $p^*_F \in \text{argmax}_{p\in[0,1]}\{p\tilde{v} + (1-\beta)g'(p^*_F - (p^*_F)^2)(p - p^2)\}$.*

Proof of Proposition 9. Existence of an optimal distribution follows because $\overline{\mathcal{F}}$ is compact in the weak topology and V_β is continuous.

To show that there is an optimal distribution whose marginal over interim beliefs is supported on no more than three points, let $\beta \in [0, 1]$, fix an arbitrary optimal distribution F^* with marginals (p^*_F, F^*_Δ), and let $q^* = \int_0^1 p^2 dF^*_\Delta(p)$. Define $\Delta(p^*_F, q^*) = \{F_\Delta \in \Delta([0, 1]) : \int_0^1 p dF_\Delta(p) = p^*_F, \int_0^1 p^2 dF_\Delta(p) = q^*\}$, and consider the problem of maximizing $\int_0^1 g(p - p^2)dF_\Delta(p)$ over $F_\Delta \in \Delta(p^*_F, q^*)$. Every such marginal is implemented by a feasible joint distribution by setting the conditional probability of $\omega = 1$ given p equal to p. Since F^*_Δ is feasible for this problem, any solution has the same prior and second moment as F^*_Δ and weakly increases the only term of V_β that is not fixed on $\Delta(p^*_F, q^*)$; hence it also solves the original problem. The objective is linear in F_Δ, and $\Delta(p^*_F, q^*)$ is a moment set with two moment conditions, so Theorem 2.1 in Winkler (1988) implies that this problem has a solution supported on no more than three points.

We now establish the inequalities used to prove points 1 and 2. Fix any feasible distribution F, and let $m = p_F = \int_0^1 p dF_\Delta(p)$, $q = \int_0^1 p^2 dF_\Delta(p)$, $a = m - q = \int_0^1 p(1-p) dF_\Delta(p)$, and $b = q - m^2 = D_2(F)$. Then $a, b \geqslant 0$ and $a + b = m - m^2 \leqslant 1/4$. By Jensen's inequality, $\int_0^1 g(p-p^2) dF_\Delta(p) \leqslant g(a)$. Moreover, the bounds $\underline{g} \leqslant g'(z) \leqslant \overline{g}$ on $[0, 1/4]$ imply that $\underline{g}t \leqslant g(s+t) - g(s) \leqslant \overline{g}t$ whenever $s, t \geqslant 0$ and $s + t \leqslant 1/4$, and, because $g(0) = 0$, that $g(t) \leqslant \overline{g}t$ for every $t \in [0, 1/4]$.

For point 1, suppose that $\beta \geqslant \overline{\beta}$, and let F^0 be the feasible no-information distribution whose marginal over interim beliefs is δ_m. Since F^0 and F have the same prior, their baseline expected utilities coincide. The first-period suspense of F^0 is zero and its second-period suspense is $m - m^2 = a + b$. Therefore, the preceding inequalities imply that $V_\beta(F^0) - V_\beta(F) \geqslant \beta[g(a+b) - g(a)] - (1-\beta)g(b) \geqslant [\beta\underline{g} - (1-\beta)\overline{g}]b$. By the definition of $\overline{\beta}$, the coefficient in brackets is at least $\underline{g}/2 > 0$. Thus, F^0 is strictly preferred to F whenever $b > 0$. Since b is the variance of the interim belief, $b = 0$ if and only if $F_\Delta = \delta_m$. Therefore, every optimal distribution F^* satisfies $F^*_\Delta = \delta_{p^*_F}$. The utility of a no-information distribution with prior p is $p\tilde{v} + \beta g(p - p^2)$, so p^*_F is optimal if and only if $p^*_F \in \operatorname{argmax}_{p \in [0,1]}\{p\tilde{v} + \beta g(p - p^2)\}$. This proves point 1.

For point 2, suppose that $\beta \leqslant \underline{\beta}$, and let F^1 be the feasible full-information distribution whose marginal over interim beliefs is $(1-m)\delta_0 + m\delta_1$. Again, F^1 and F have the same prior. The first-period suspense of F^1 is $m - m^2 = a + b$ and its second-period suspense is zero. Therefore,

$$V_\beta(F^1) - V_\beta(F) \geqslant (1-\beta)[g(a+b) - g(b)] - \beta g(a) \geqslant [(1-\beta)\underline{g} - \beta\overline{g}]a.$$

By the definition of $\underline{\beta}$, the coefficient in brackets is at least $\underline{g}/2 > 0$. Hence, F^1 is strictly preferred to F whenever $a > 0$. Since $p(1-p)$ is nonnegative on $[0, 1]$ and vanishes only at 0 and 1, $a = 0$ if and only if F_Δ is supported on $\{0, 1\}$. Its mean then uniquely determines it as $(1-m)\delta_0 + m\delta_1$. Thus, every optimal distribution F^* satisfies $F^*_\Delta = (1 - p^*_F)\delta_0 + p^*_F\delta_1$.

Under full information, the utility associated with prior p is $H(p) = p\tilde{v} + (1 -$

$\beta)g(p-p^2)$. Thus, p_F^* is optimal if and only if it maximizes H over $[0,1]$. For a candidate prior p_F^*, let $Q_{p_F^*}(p) = p\tilde{v} + (1-\beta)g'(p_F^* - (p_F^*)^2)(p-p^2)$. Both H and $Q_{p_F^*}$ are differentiable and concave, and $H'(p_F^*) = Q'_{p_F^*}(p_F^*)$. Since a point x maximizes a differentiable concave function f on $[0,1]$ if and only if $f'(x)(p-x) \leqslant 0$ for every $p \in [0,1]$, it follows that p_F^* maximizes H if and only if it maximizes $Q_{p_F^*}$. This proves point 2. ∎

Online Appendix II.D: Monotonicity and behavior

Preferences that preserve stochastic orders capture the idea that individuals prefer lotteries that are better according to the stochastic order. When $x \in \mathbb{R}$ represents monetary outcomes, the class of increasing functions generates the first-order stochastic dominance relation, and a preference that preserves this order is monotone increasing with respect to the realized wealth. Similarly, a preference that preserves the stochastic order generated by concave functions will exhibit risk aversion. Here we give necessary and sufficient conditions for the adversarial forecaster model to preserve a stochastic order, and give applications to absolute and relative risk aversion of various orders. Section 4.2 applies these results to aversion to correlation in risks across time periods.

7.1 Stochastic orders and monotonicity

We start with the definition of the stochastic order induced by a set of continuous real-valued functions.

Definition 8. Fix a set $\mathcal{W} \subseteq C(X)$.

(i) The stochastic order $\gtrsim_{\mathcal{W}}$ is defined as:

$$F \gtrsim_{\mathcal{W}} \tilde{F} \iff \int w(x)dF(x) \geqslant \int w(x)d\tilde{F}(x) \quad \forall w \in \mathcal{W}. \tag{27}$$

(ii) A utility V *preserves* $\gtrsim_{\mathcal{W}}$ if for all $F, \tilde{F} \in \mathcal{F}$, $F \gtrsim_{\mathcal{W}} \tilde{F}$ implies $V(F) \geqslant V(\tilde{F})$.

Let $\langle \mathcal{W} \rangle$ denote the smallest closed convex cone containing $\mathcal{W}$ and all constant functions. Because the adversarial forecaster utility has a max-min representation with local utility at F given by $v + \sigma(\cdot, \hat{y}(F))$ and this coincides with the Gâteaux derivative of V at F, we can apply Theorem 1 in Cerreia-Vioglio, Maccheroni, and Marinacci (2017) to obtain the following characterization.[40]

Proposition 10. *Let V be an adversarial forecaster representation with baseline utility function v and surprise function σ, and fix a set $\mathcal{W} \subseteq C(X)$. Then V preserves $\succsim_{\mathcal{W}}$ if and only if $v + \sigma(\cdot, \hat{y}(F)) \in \langle \mathcal{W} \rangle$ for all $F \in \mathcal{F}$.*

This finding shows that, when X is a subset of the real line, adversarial forecaster preferences are compatible with FOSD and SOSD provided that, for every $F \in \mathcal{F}$, the function $v+\sigma(\cdot, \hat{y}(F))$ is increasing (for FOSD) or increasing and concave (for SOSD). FOSD requires that the enjoyment of surprise never compensates for a lower realized payoff, while SOSD additionally requires that it never overturns aversion to mean-preserving spreads. This result can also be used to check whether the adversarial forecaster representation favors mean-preserving spreads; in this case it is enough to check V preserves the convex order. For example, in Example 4, when $v'' \geqslant -2$, the local utility is convex in x for all forecasts F. Thus Proposition 10 implies that the agent weakly prefers any mean-preserving spread $\tilde{F}$ of F to F.

Online Appendix III: Additional examples

This section presents two examples. The first solves a three-point timing-of-information problem, and the second gives an infinite-moment GMM utility with a unique optimal lottery that has full support.

Example 10. Consider the problem of maximizing $V(F) = \frac{1}{2}V_1(F) + \frac{1}{2}V_2(F)$ over distributions on $[0,1]$ with no more than three points in their support, where $V_1(F) = \sqrt{\mathrm{Var}_F(x)}$ and $V_2(F) = \int \sqrt{x - x^2}dF(x)$. Writing $z = x - 1/2$ and $a = \mathbb{E}_F[z^2]$, Jensen's inequality and $\mathrm{Var}_F(x) \leqslant a$ imply $V(F) \leqslant \frac{1}{2}(\sqrt{a} + \sqrt{1/4 - a}) \leqslant 1/(2\sqrt{2})$. Equality holds if and only if $\mathbb{E}_F[z] = 0$, $z^2 = 1/8$ almost surely, and the two possible values of z have equal probability. Thus, the unique optimal lottery is $F^* = \frac{1}{2}\delta_{\frac{1}{2}-\frac{1}{2\sqrt{2}}} + \frac{1}{2}\delta_{\frac{1}{2}+\frac{1}{2\sqrt{2}}}$, and its value is $V(F^*) = 1/(2\sqrt{2}) \approx 0.354$. △

Example 11 (Wiener Process Example)**.** We interpret $x \in [0,1]$ as time. Let W_x be a standard Wiener process and consider the stationary Ornstein–Uhlenbeck process h_x

[40]Theorem 1 and Lemma 1 in Cerreia-Vioglio, Maccheroni, and Marinacci (2017) are stated under the assumption that $X \subseteq \mathbb{R}$. However, an inspection of their proof shows that the same results hold for any compact metric space X. Therefore, we omit the proof of Proposition 10.

defined by $dh_x = -h_x dx + \sqrt{2} dW_x$, where h_0 is standard normal and independent of the future increments of W. Its solution is $h_x = \exp(-x)h_0 + \sqrt{2}\int_0^x \exp(-(x-r))dW_r$, so $\mathbb{E}[h_x] = 0$, $\mathbb{E}[h_x^2] = 1$, and $H(x,\tilde{x}) = \mathbb{E}[h_x h_{\tilde{x}}] = \exp(-|x-\tilde{x}|)$. Therefore, $H(x,\tilde{x}) = G(x-\tilde{x})$ for $G(z) = \exp(-|z|)$. As a covariance kernel of a non-degenerate Gaussian process, H is strictly positive definite, and it admits the continuous factorization required for an infinite-moment GMM representation.[41] Moreover, $H(0,\tilde{x}) = \exp(-\tilde{x})$ is positive, strictly decreasing, and strictly convex on $[0,1]$. Thus, this kernel satisfies the assumptions of Proposition 5, so the associated GMM utility has a unique optimal lottery with full support. $\triangle$

Online Appendix IV: Adversarial forecasters, local utilities, and Gâteaux 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 continuous in (x,F).

Proposition 11. *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. The first implication follows from $V(F) = \int w(x,F)dF(x) \geqslant \int w(x,F)d\tilde{F}(x) \geqslant V(\tilde{F})$. To prove the second, let $V(\tilde{F}) \geqslant V(F)$ and observe that $\int w(x,F)d\tilde{F}(x) \geqslant$

[41] A factorization follows from the Fourier expansion of $\exp(-|z|)$ on $[-\pi,\pi]$: its cosine coefficients are strictly positive and summable. Splitting each cosine term into sine and cosine features and indexing these features by two sequences converging to a common limit point yields a compact probability space (S,μ) and a continuous function $\bar{h} : [0,1] \times S \to \mathbb{R}$ such that $H(x,\tilde{x}) = \int_S \bar{h}(x,s)\bar{h}(\tilde{x},s)d\mu(s)$. Strict positive definiteness on signed measures follows from $\exp(-|x-\tilde{x}|) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{\exp(it(x-\tilde{x}))}{1+t^2}dt$. Indeed, for every finite signed measure M on $[0,1]$, $\int\int \exp(-|x-\tilde{x}|)dM(x)dM(\tilde{x}) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{|\widehat{M}(t)|^2}{1+t^2}dt$, $\widehat{M}(t) = \int \exp(itx)dM(x)$. If this expression is zero, continuity gives $\widehat{M}(t) = 0$ for every t. The Stone-Weierstrass theorem then implies $M = 0$, because the linear span of $\{\exp(itx) : t \in \mathbb{R}\}$ is dense in $C([0,1])$.

$V(\tilde{F}) \geqslant V(F) = \int w(x, F)dF(x)$, implying that $\tilde{F} \gtrsim_F F$ 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 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 assuming V is *Fréchet differentiable* over $\mathcal{F}$, a strong notion of differentiability.[42]

Our notion of local expected utility is neither weaker nor stronger than Fréchet differentiability. If V has continuous local expected utility, then it is concave, which is not implied by Fréchet differentiability. Conversely, Example 12 shows that continuous local expected utility does not imply Fréchet differentiability.

Now we discuss the relationship between continuous local expected utility and the weaker notion of *Gâteaux differentiability*, which has been used to extend Machina's notion of local utility to functions that are not necessarily Fréchet differentiable.

In particular, Chew, Karni, and Safra (1987) develops a theory of local utilities for rank-dependent preferences and Chew and Nishimura (1992) extends it to a broader class. Recall that V is Gâteaux differentiable[43] 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 Gâteaux derivative of V at F we can define the directional derivative operator $DV(F)(\tilde{F} - \overline{F}) = \int w(x, F)d\tilde{F}(x) - \int w(x, F)d\overline{F}(x)$. We can restate Lemma 1 with the language of Gâteaux derivatives just introduced.

Proposition 12 (Lemma 1 in Appendix A)**.** *If V has continuous local expected utility $w(x, F)$, then V is Gâteaux differentiable and $w(\cdot, F)$ is the Gâteaux derivative of V at F, for all F.*

[42] The notion of Fréchet differentiability depends on the norm used. Here, following Machina, we use the L_1-norm.

[43] Here we follow Huber (2011) and subsequent authors and modify the definition of the Gâteaux derivative to only consider directions that lie within the set of probability measures.

Corollary 7. *V has continuous local expected utility if and only if it is concave and Gâteaux differentiable with continuous Gâteaux derivative.*

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

Example 12. 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 Fréchet differentiable. Integration by parts gives $V(F) = \int_0^1 [1 - g(F(x))]dx$, which is strictly concave in F. For $F_\lambda = (1-\lambda)F + \lambda\tilde{F}$, differentiation under the integral sign gives $\lim_{\lambda\downarrow 0}[V(F_\lambda) - V(F)]/\lambda = -\int_0^1 g'(F(z))(\tilde{F}(z) - F(z))dz$. If $w(x,F) = \int_0^x g'(F(z))dz$, integration by parts yields $-\int_0^1 g'(F(z))(\tilde{F}(z) - F(z))dz = \int w(x,F)d(\tilde{F} - F)(x)$, so $w(\cdot,F)$ is a Gâteaux derivative of V at F. Moreover, if F_n converges weakly to F, then $F_n(z) \to F(z)$ at almost every z, and dominated convergence implies $\sup_{x\in[0,1]} |w(x,F_n) - w(x,F)| \to 0$. Since every $w(\cdot,F)$ is $\|g'\|_\infty$-Lipschitz, w is continuous in (x,F). Therefore, by Corollary 7, V has continuous local expected utility and, by Theorem 1, it admits an adversarial forecaster representation. $\triangle$