Adaptive Preferences: An Evolutionary Model of Non-Expected Utility and Ambiguity Aversion*

Philipp Sadowski† Todd Sarver‡

Current Draft:
December 18, 2023

Abstract

We enrich an evolutionary model with common and idiosyncratic uncertainty as in Robson (1996) by allowing for hidden actions (or phenotypic flexibility). In contexts where common uncertainty is ambiguous and idiosyncratic uncertainty is risky, the model generates both ambiguity aversion and non-expected-utility preferences for risk, thereby providing a link between the two types of behavior. While the general evolutionarily optimal objective function does not have an obvious similarity to functional forms studied in the literature, our main results show that it can be recast in a form that is immediately relatable to some common models of ambiguity aversion and non-expected utility. The special cases we consider include ones that embed rank-dependent utility or divergence preferences within a model of ambiguity aversion.

Keywords: Evolution of preferences, ambiguity, phenotypic flexibility
JEL Classification: D81, D83, D84

*First Version: December 8, 2019. This research was supported by the National Science Foundation through Grant SES-1851664. We thank the editor and two referees for comments that significantly improved the paper. We also thank Jiayun Dong, Andrew Ellis, José Heleno Faro, Yoram Halevy, Cosmin Ilut, Frank Johannes, Bart Lipman, George Mailath, Massimo Marinacci, Sujoy Mukerji, Wolfgang Pesendorfer, Arthur Robson, Larry Samuelson, and numerous seminar participants for helpful comments and discussions.

†Duke University, Department of Economics, 213 Social Sciences/Box 90097, Durham, NC 27708. Email: p.sadowski@duke.edu.

‡Duke University, Department of Economics, 213 Social Sciences/Box 90097, Durham, NC 27708. Email: todd.sarver@duke.edu.

1 Introduction

This paper provides an evolutionary perspective on choice under uncertainty, based on the notion that natural selection not only can influence physical traits, but can also shape choice behavior. Using this approach, we develop a foundation for a non-expected-utility and ambiguity-averse model of choice. Systematic violations of expected utility are common, but at least when risk is objective, they appear to be at odds with evolutionary optimality. A central contribution of this paper is to expand the scope of the evolutionary approach by allowing individuals to simultaneously make multiple decisions, some of which are observable and others which are hidden from the modeler. When chosen optimally, as evolution will require, the presence of such hidden actions will generate preferences that appear to violate expected utility from the perspective of the analyst. We show that the resulting class of evolutionarily optimal preferences, which we call *adaptive preferences*, includes rank-dependent expected utility in the context of risk, and variants of the smooth model, variational preferences, and multiple prior preferences in the contexts of both risk and ambiguity. Importantly, while ambiguity-averse preferences are typically assumed to reduce to expected utility when facing objective risk, our model excludes this benchmark version of many of the ambiguity models it nests and instead links different uncertainty attitudes to violations of expected utility. Thus, our evolutionary approach can help to address both the question of model selection and the potential link between Allais and Ellsberg behavior.

The starting point of our analysis is an observation which dates back to a seminal paper by Robson (1996): Evolutionary optimality generates a preference for idiosyncratic uncertainty over common uncertainty, and ambiguity is closely associated with common uncertainty in many instances. Hence, natural selection favors ambiguity aversion. The intuition for why evolution can generate aversion to common uncertainty is actually quite simple. To illustrate, suppose there are two actions between which all individuals must choose in every period. For both actions, individual growth (meaning net expected number of offspring) will be either 2 or 4, each with probability $\frac{1}{2}$. The only difference is that one action bears common uncertainty, where realized per-period growth is perfectly correlated across individuals, while the other bears idiosyncratic uncertainty, where realized growth is independent across individuals. By the law of large numbers, the per-period growth of a (large subpopulation with a common) genotype who consistently chooses the idiosyncratic uncertainty will be approximately $\frac{1}{2}(2+4) = 3$. In contrast, a genotype who chooses the common uncertainty will grow by either 2 or 4, each in approximately half of the periods. Heuristically, this leads to a long-run average growth over two periods of $2 \times 4 = 8$, which is less than $3 \times 3 = 9$. This example illustrates the detrimental effect of correlation on growth: The genotype who chooses the idiosyncratic uncertainty will have a higher long-run growth rate, which implies it will almost surely dominate in the long run (Lemma 1).[1] We discuss and justify the close connection

[1]The existence and exact form of this aversion to common uncertainty depend on both the frequency of reproduction (Robatto and Szentes (2017)) and timing of reproduction within the life cycle of organisms

between ambiguity and correlated uncertainty in detail in Section 1.1.

The main innovation of our paper is the incorporation of adaptation via hidden actions. The incorporation of hidden actions is motivated not only by economic settings—where data sets often capture only a subset of the many decision being make by individuals—but also by biological settings—where hidden actions might take the form of unobservable aspects of physical adaptation of organisms. In an economic context, data sets could contain information such as occupation choice, investments, or even vaccination decisions, while omitting information about other complementary decisions such as housing choice, other investments or insurance, or social distancing measures, respectively. In a biological context, hidden actions could take the form of rapid and reversible physical adaptation, known as *phenotypic flexibility*, which has recently gained increased attention in evolutionary biology.[2]

It is well known that hidden actions can lead to revealed preferences over *observed* choices that violate expected utility, even if the individual's actual joint preferences over *all* choices satisfy expected utility.[3] In particular, since different hidden actions may be optimal for different observable actions, individuals may be averse to probabilistic mixtures over observable outcomes (see Sarver (2018)).

1.1 Ambiguity as Common Uncertainty

In many examples and applications of ambiguity, the unknown probabilities concern common factors that affect all individuals in the population. For example, in one of the earliest applications of ambiguity to economics, Dow and Werlang (1992) and Epstein and Wang (1994) examined the implications of ambiguity about asset returns. Returns to financial assets are obviously common to all individuals who invest in them. Similarly, in applications to macroeconomics, ambiguity typically concerns aggregate variables, such as factor productivity (Ilut and Schneider (2014), Bianchi, Ilut, and Schneider (2018)). Other examples of uncertainty about aggregate variables that can affect individual outcomes and where probabilities are poorly understood could include the timing of new technological breakthroughs, natural disasters such as earthquakes or tsunamis, or climate change and its implications.

One reason common uncertainty in the examples mentioned so far may be subject to greater ambiguity than idiosyncratic uncertainty is that idiosyncratic random variables can be studied using cross-sectional data, whereas aggregate variables by definition cannot. Greater abundance of data may lead to a better understanding. Nonetheless, there could be common uncertainty for which the probabilities are well understood by individuals, and our results would be equally relevant in those settings.

(Robson and Samuelson (2019)). We discuss these considerations further in Section 7.1.

[2]We discuss the potential relevance of our results for evolutionary biology in more detail in Section 7.2.

[3]Prior studies of the impact of physical commitments on risk preferences include Grossman and Laroque (1990), Gabaix and Laibson (2001), and Chetty and Szeidl (2007, 2016). Unobservable commitments in particular are explored in Kreps and Porteus (1979), Machina (1984), and Ergin and Sarver (2015).

In addition to ambiguity taking the form of common uncertainty about aggregate variables, there is also a fundamental and systematic link between common uncertainty and any instance of ambiguity involving *model uncertainty*—ambiguity about the true data generating process. Even if the risks faced by each individual are well understood and idiosyncratic conditional on some common underlying model parameter, if that parameter is unknown and ambiguous then all individuals share in the resulting common uncertainty.[4] For a simple illustration, consider a medical treatment. If the efficacy (success rate) of the treatment for a population with a given set of characteristics is known, then whether it is successful for one individual is independent of whether it succeeds for another. However, if the treatment has undergone limited testing, then its success rate may be unknown and would itself be a source of common uncertainty for all individuals. In fact, most instances of ambiguity can be cast as common uncertainty about idiosyncratic probabilities.

Of course, we should be careful to point out that the correlation mechanism at play in this paper may not be the only driver of ambiguity aversion. We would not go so far as to claim that every instance of ambiguity corresponds to common uncertainty; nor would we suggest that every instance of common uncertainty involves ambiguous beliefs. Nonetheless, the main thrust of the preceding discussion is that there are indeed many situations in which ambiguity is tightly linked to common uncertainty, and our results speak specifically to these instances of ambiguity. In other cases where ambiguity is not connected to common uncertainty, we remain agnostic about whether ambiguity aversion is driven by heuristics developed by genotypes from the case of common uncertainty or whether some other source of ambiguity aversion is at play.

1.2 Outline

The remainder of the paper is structured as follows. Section 2 formally sets up our model. Section 3 establishes that adaptive preferences are evolutionarily optimal.

Section 4 sets the stage by illustrating, via an example, how the model in Robson (1996), which is a version of the smooth model of ambiguity aversion and a special case of our model without hidden actions, predicts Ellsberg behavior. In Section 5, we explore an alternative special case involving hidden actions but with no common uncertainty, and we show that the evolutionarily optimal preferences in this case correspond to the optimal risk attitude preferences studied by Sarver (2018). In particular, we show that our model nests rank-dependent expected utility (RDU) and divergence preferences as special cases.

Section 6 analyzes the general case of our model when hidden actions and common uncertainty are simultaneously at play. Our main result is a dual formulation of our representation that greatly simplifies the comparison to existing models. We use this result to show that

[4] This interpretation is closely connected to the macroeconomic literature on robustness to model uncertainty (Hansen and Sargent (2001, 2008)), and is discussed in the evolutionary context in Robson (1996).

versions of several prominent representations, including variational preferences, multiple priors expected utility, and rank-dependent utility, can be embedded in our general model. Importantly, these special cases provide a link between Ellsberg- and Allais-type behaviors, as we illustrate with an example.

In Section 7, we discuss some of the simplifying assumptions that are commonly made in economic applications of the evolutionary approach and the robustness of our results to relaxing them. We also describe the biological evidence of phenotypic flexibility, which provides an alternative interpretation and motivation for the hidden actions in our model. This connection suggests that our model may have relevance not just in economic contexts, but also in the framework of evolutionary biology. Finally, in the Supplementary Appendix, we provide proofs for some supporting results that are omitted from the main paper.

2 Evolutionary Setting

The basic idea behind the evolutionary approach is that a large population of individuals is initially made up of subpopulations with different genotypes, where a genotype specifies the physical traits as well as the programmed behavior (choices) of an organism. These choices lead to a possibly uncertain outcome, and this outcome together with the physical traits of the organism determine its evolutionary fitness, that is, its number of offspring and its own survival. The offspring inherit the parent's genotype. In the next period the offspring and the parent (if still alive) will face a choice of their own, and so on. In this way, the number of individuals who share a particular genotype may shrink or grow over time, relative to the whole population. A genotype is evolutionarily optimal among those initially present if the relative size of its subpopulation does not vanish over time.[5]

2.1 Uncertainty

Common components of uncertainty are modeled via a state space Ω. The realization of $\omega \in \Omega$ is common to all individuals in the population. In addition, given ω, idiosyncratic uncertainty is captured via a state space S, where each individual in the population receives

[5]For expositional clarity, we consistently interpret fitness as the number of offspring plus possible survival of the organism, so that evolutionary selection happens over many generations. We note that our model and results do not rely on the naturalistic interpretation of the evolutionary process. In particular, evolutionary selection can be faster if it is not based on procreation, but rather on imitation or reallocation. For example, the reallocation of resources based on fitness is the idea behind selection-based theories following Friedman (1953) that are central in the evolutionary economics literature (see Nelson and Winter (1982)), and versions of the replicator dynamics have been established based on reinforcement learning (Börgers and Sarin (1997)) or imitation (Schlag (1998)) at the individual level. While correlation was not emphasized in these studies, we suspect that similar non-biological interpretations can be developed for the setting with correlated uncertainty studied here.

an independent draw of the state $s \in S$. The entire payoff-relevant state space is then $\Omega \times S$. We assume that Ω and S are Polish spaces, that is, complete and separable metrizable spaces. We endow the spaces Ω and S with their Borel σ-algebras $\mathcal{B}_\Omega$ and $\mathcal{B}_S$, respectively, and we endow the product of these spaces with the product σ-algebra $\mathcal{B}_\Omega \otimes \mathcal{B}_S$.

Given any measurable space $(Y, \mathcal{Y})$, let $\triangle(Y)$ denote the set of countably additive probability measures on Y, and let $\triangle_s(Y)$ denote the set of all simple probability measures on Y (i.e., measures with finite support). The state is drawn each period according to a measure $\mu \in \triangle(\Omega \times S)$. The marginal distribution of μ on Ω assigns probability $\mu(E)$ to any measurable event $E \in \mathcal{B}_\Omega$. As noted, there is a common draw of the ω dimension of the state for all individuals in the population according to this marginal distribution. However, conditional on ω, the s dimension of the state is drawn independently for each individual according to the conditional probability distribution $\mu(s|\omega)$ on S.[6]

2.2 Consumption and Fitness

Let Z denote a nonempty set of outcomes. Both the ω and s dimensions of the state space are potentially relevant for the outcome of an action. Formally, let $\mathcal{F}$ denote the set of simple acts, that is, the set of all measurable and finite-valued functions $f : \Omega \times S \to Z$. An *evolutionary fitness function* $\psi : Z \to \mathbb{R}$ specifies the (net expected) individual reproductive growth associated with each outcome.[7] Given an act $f \in \mathcal{F}$, the individual growth in state (ω, s) is then $\psi(f(\omega, s))$. For example, for a population of individuals, aggregate fitness of zero indicates extinction, fitness of one indicates that the birth rate is equal to the death rate and hence there is no change in the size of the population, and fitness of 1.5 indicates a 50% growth in the population. Aggregate fitness can obviously never be negative. Whether or not individual fitness functions take negative values is not important for the evolutionary optimality of adaptive preferences. However, in order to derive exact dual characterizations of some special cases of our model, it will be technically useful to allow some outcomes to generate negative individual fitness, which could be interpreted as an externality that eliminates other individuals.

Individuals face the task of choosing acts in each period before learning the realization of the state (ω, s). Each genotype determines preferences that are used for this choice. In addition to the observable choice of act f, we assume that individuals might simultaneously take hidden actions, that is, actions that are unobservable to the modeler. Incomplete data

[6]More precisely, since S may be an infinite set, the conditional probability distribution given ω assigns probability $\mu(E|\omega)$ to an event $E \in \mathcal{B}_S$. Note that since S is a Polish space, the existence of a regular conditional probability distribution is ensured by Proposition 10.2.8 in Dudley (2002).

[7]Realized net individual growth, which includes survival and offspring, must be an integer, but since reproduction and survival may be uncertain given the outcome $z \in Z$, expected individual growth may take non-integer values. As the main results of Section 3 show, evolutionary fitness of a genotype with a large population depends only on the expected reproductive growth $\psi(z)$ its individuals attain from each outcome z.

of this sort is pervasive in economic analysis, as data sets often contain only a snapshot of one dimension of the full spectrum of decisions being made by individuals. We model hidden actions in a simple and tractable reduced form by allowing individuals to select a fitness function ψ from some feasible set Ψ in each period.[8] As we discuss in Section 7.2, our use of multiple fitness functions can also be interpreted in terms of phenotypic flexibility in the context of evolutionary biology.

We aim to uncover various preferences that can be nested within our evolutionary model, thereby illustrating the structure imposed by our model on choice under uncertainty. At the outset, we therefore impose only minimal technical restrictions on the set of fitness functions: We assume Ψ is nonempty and that $\sup_{\psi \in \Psi} \psi(z)$ is finite for every $z \in Z$. Of course, additional structure and restrictions on the set Ψ may be appropriate depending on the application, as the availability of various hidden actions and their impact on fitness will naturally depend on the choice context, and such restrictions will serve to refine the exact preferences under uncertainty generated by our model. In the context of idiosyncratic risk in Section 5, we provide examples of easy to interpret sets Ψ that give rise to some established functional forms from the literature on non-expected-utility. Utilizing our main result (Theorem 2), those examples are extended to the general setup with common uncertainty (ambiguity) in Section 6.

2.3 Growth Rates

In a given time period, the aggregate growth rate of a genotype will be determined by the common preferences each individual in its subpopulation is programmed to use when choosing (deterministically or possibly randomly) an act f and a fitness function ψ. We assume each decision problem is faced repeatedly, leading to a stochastic sequence of aggregate growth rates for each genotype. Our analysis of natural selection and evolutionary optimality will center around the comparison of long-run growth rates of different genotypes (with different programmed preferences), which we state in log terms.

Definition 1. Suppose the aggregate growth rate of a genotype is given by $(\lambda_t)_{t \in \mathbb{N}}$, where λ_t is the random variable that describes the aggregate growth rate in period t of the entire subpopulation of individuals with that genotype. We say that α is the *(log) long-run growth rate* of the genotype if $\frac{1}{T} \sum_{t=1}^{T} \ln(\lambda_t) \to \alpha$ almost surely as $T \to \infty$.

For an arbitrary sequence $(\lambda_t)_{t \in \mathbb{N}}$ of random variables, the long-run growth rate may not exist, since the series above may not converge. However, we will see in the next section that in our model, the long-run growth rate exists for any act f and fitness function ψ.

[8]This reduced form derives immediately from a more explicit model of hidden actions, where individuals take a hidden action $y \in Y$ and have a single fixed fitness function $\hat{\psi}(z, y)$ for outcome/action pairs. The resulting set of fitness functions in our model would then be $\Psi = \{\hat{\psi}(\cdot, y) : y \in Y\}$.

To establish that the long-run growth rate is the appropriate statistic for comparison in our evolutionary model, the next lemma demonstrates how it relates to long-run dominance of a particular genotype over others. First, note that throughout the paper, we follow the standard convention of assuming that the number of agents of each genotype is (infinitely) large, which we formally model by treating the set of individuals of each genotype i as a continuum with measure $N^i(t)$ at time period t.[9] Thus, if the sequence of aggregate growth rates of genotype i is $(\lambda_t^i)_{t\in\mathbb{N}}$ and the initial measure of this genotype is $N^i(0)$, then the measure of its subpopulation at time $T \in \mathbb{N}$ is

$$N^i(T) = N^i(0) \prod_{t=1}^{T} \lambda_t^i.$$

Lemma 1. *Consider two genotypes $i = A, B$, where genotype i has a sequence of stochastic aggregate growth rates $(\lambda_t^i)_{t\in\mathbb{N}}$ yielding long-run growth rate α^i, that is, $\frac{1}{T}\sum_{t=1}^{T}\ln(\lambda_t^i) \xrightarrow{a.s.} \alpha^i$. If $\alpha^A > \alpha^B$, then regardless of the initial measures $N^A(0) > 0$ and $N^B(0) > 0$ of their respective subpopulations at time $t = 0$, we have $N^A(t)/N^B(t) \to \infty$ almost surely as $t \to \infty$.*

Note that Lemma 1 does *not* imply that a higher long-run growth rate yields higher expected population size as t grows large, as indeed it is possible to have the expected value of $N^B(t)$ exceed that of $N^A(t)$ for all t. Nonetheless, the lemma implies that the event where $N^B(t)$ exceeds $N^A(t)$ vanishes (has probability zero) in the limit as $t \to \infty$.

Evolutionary theory aims to explain which genotypes can be observed in the long run. Lemma 1 clarifies why maximizing long-run growth, rather than the expected population size, is evolutionarily optimal. If in the present moment organisms have already been evolving for t periods, then the relative population sizes of different genotypes that we observe today are a snapshot of the evolutionary process in period t. Assuming this process has been underway for some time (t is large), the probability is very high that the dominant genotype observed today is precisely the one with the highest long-run growth rate.

3 Evolutionarily Optimal Choice

We begin our analysis by deriving the long-run growth rates associated with (possibly random) choices of action and fitness function. Since evolutionary optimality requires maximizing long-run growth, the optimal value function over random choice follows immediately.

Definition 2. A *random choice* $\pi \in \mathcal{R}(\mathcal{F}, \Psi) \equiv \triangle_s(\mathcal{F} \times \Psi)$ is a simple probability measure over the space of acts and feasible fitness functions.

[9]Using results from the theory of branching processes, it can be shown that our results involving continuum populations are the correct limiting approximations for large but finite populations.

The random choice π assigns probability $\pi(f, \psi)$ to a pair (f, ψ). Therefore, for a given state realization (ω, s), the random choice π achieves an expected fitness of

$$\mathbb{E}_\pi\big[\psi(f(\omega, s))\big] = \int_{\mathcal{F}\times\Psi} \psi(f(\omega, s))\, d\pi(f, \psi).$$

We adopt the convention that the domain of the natural logarithm includes nonpositive numbers and its range is the extended reals by setting $\ln(x) = -\infty$ for all $x \leq 0$.

Theorem 1 (Long-Run Growth)**.** *Suppose Ψ and μ are fixed, and consider a genotype with an (infinitely) large subpopulation of individuals. The long-run growth rate of the genotype from choosing the random choice $\pi \in \mathcal{R}(\mathcal{F}, \Psi)$ in every period is*

$$\Lambda(\pi) = \int_\Omega \ln\bigg(\int_S \mathbb{E}_\pi\big[\psi(f(\omega, s))\big]\, d\mu(s|\omega)\bigg) d\mu(\omega). \tag{1}$$

The concavity of the logarithm implies that Λ is more adversely affected by common uncertainty about ω than by idiosyncratic uncertainty about s. Also, since Λ expresses the long-run average growth rate in log terms, $\Lambda(\pi) = -\infty$ corresponds to extinction and $\Lambda(\pi) = 0$ corresponds to constant population size. At the heart of the proof of Theorem 1 is the same logic that is behind the seminal result of Robson (1996), who considered the special case of no adaptation ($\Psi = \{\psi\}$) and deterministic choice.

Proof. Recall that, conditional on ω, the s dimension of the state is independently distributed for each individual in the population. Randomization in choice is also idiosyncratic. Therefore, by the law of large numbers, conditional on the realized ω_t at time t, the aggregate growth rate of a large population of individuals with random choice π is approximately

$$\lambda_t(\omega_t) = \int_S \mathbb{E}_\pi\big[\psi(f(\omega_t, s))\big]\, d\mu(s|\omega_t).$$

Since we consider infinite subpopulations in our model, we can treat this approximation as exact.[10] Taking the product over a sequence of realized common components $\omega_1, \ldots, \omega_T$ and raising to the power $1/T$ gives the realized annualized growth rate over this sequence of periods:

$$\prod_{t=1}^{T}\bigg(\int_S \mathbb{E}_\pi\big[\psi(f(\omega_t, s))\big]\, d\mu(s|\omega_t)\bigg)^{1/T}.$$

[10]Note that an approximate (limiting) version of Theorem 1 also holds for finite populations, provided the initial population size is sufficiently large. Using the theory of branching processes (Athreya and Ney (1972, Chapter 5)), it can be shown that the average growth rate of a finite population converges to $\Lambda(\pi)$ conditional on non-extinction. Moreover, it can be shown that when $\Lambda(\pi) > 0$, the probability of extinction converges to zero as the initial population becomes large.

Taking the logarithm of this expression and then the limit as $T \to \infty$, we have

$$\frac{1}{T}\sum_{t=1}^{T} \ln\left(\int_S \mathbb{E}_\pi\big[\psi(f(\omega_t, s))\big]\, d\mu(s|\omega_t)\right) \\ \to \int_\Omega \ln\left(\int_S \mathbb{E}_\pi\big[\psi(f(\omega, s))\big]\, d\mu(s|\omega)\right) d\mu(\omega) \quad \text{a.s.,}$$

by the law of large numbers. ■

The long-run growth rate of the population is optimized if individuals choose π to maximize Equation (1). However, since only the random choice of act is observed while the choice of fitness function corresponds to some unobservable action, it will be useful to decompose π into its (observable) marginal distribution over acts and (unobservable) conditional distribution over fitness functions given the act.

Definition 3. A *(random) action* $\rho \in \mathcal{R}(\mathcal{F}) \equiv \triangle_s(\mathcal{F})$ is a simple probability measure over acts, where $\rho(f)$ denotes the probability assigned to f. A *(random) adaptation plan* is a function $\tau \in \mathcal{R}(\Psi|\mathcal{F}) \equiv (\triangle_s(\Psi))^{\mathcal{F}}$ mapping from the space of acts to the set of simple probability measures over the feasible fitness functions, where $\tau(\psi|f)$ is the probability assigned to fitness function ψ following the observable choice of act f.

The random choice π can equivalently be expressed as a pair ρ and τ. Formally, let $\tau \otimes \rho$ denote the measure with marginal distribution ρ on $\mathcal{F}$ and conditional distribution $\tau(\cdot|f)$ on Ψ. Then, the expectation of $\psi(f(\omega, s))$ with respect to this measure is

$$\mathbb{E}_{\tau\otimes\rho}\big[\psi(f(\omega, s))\big] = \int_{\mathcal{F}} \int_{\Psi} \psi(f(\omega, s))\, d\tau(\psi|f)\, d\rho(f).$$

Given a random action ρ and random adaptation plan τ, the corresponding joint choice of both action and adaptation is given by $\pi = \tau \otimes \rho$. Therefore, the highest possible long-run growth rate associated with an action ρ (and subsequent optimal choice of adaptation plan) is

$$\begin{aligned} V(\rho) &= \sup_{\tau \in \mathcal{R}(\Psi|\mathcal{F})} \Lambda(\tau \otimes \rho) \\ &= \sup_{\tau \in \mathcal{R}(\Psi|\mathcal{F})} \int_\Omega \ln\left(\int_S \mathbb{E}_{\tau\otimes\rho}\big[\psi(f(\omega, s))\big]\, d\mu(s|\omega)\right) d\mu(\omega). \end{aligned} \tag{2}$$

Robson (1996) considered the special case with a single fitness function ψ, without random choice, in which case the long-run growth rate associated with the deterministic choice of act f reduces to[11]

$$V(f) = \int_{\Omega} \ln\left(\int_S \psi(f(\omega, s))\, d\mu(s|\omega)\right) d\mu(\omega). \tag{3}$$

By Lemma 1, the evolutionarily optimal genotype is the one that maximizes the long-run growth rate; that is, it chooses among actions to maximize Equation (2). We refer to the preferences over random actions represented by this function V as *adaptive preferences*. As is usual in random choice contexts, we do not directly observe these preferences, only the implied random choice rule. Formally, a *decision problem* A specifies a nonempty and finite set of available acts. The resulting set of feasible action choices is

$$\mathcal{R}(A) \equiv \{\rho \in \mathcal{R}(\mathcal{F}) : \mathrm{supp}(\rho) \subset A\}.$$

Corollary 1 (Evolutionarily Optimal Choice)**.** *Suppose Ψ and μ are fixed. Then, for every infinitely repeated decision problem A, the genotype that chooses a random action in* $\mathrm{argmax}_{\rho \in \mathcal{R}(A)} V(\rho)$ *achieves a weakly higher long-run growth rate than all others.*

The adaptive preferences represented by Equation (2) specify the optimal response to correlated and uncorrelated uncertainty, but do not concern ambiguity per se. However, as laid out in Section 1.1, in many examples and applications of ambiguity, the unknown probability concerns a common factor that affects all individuals in the population. Thus, the evolutionary mechanism described in Theorem 1 may capture one important source of ambiguity aversion. In particular, the Robson (1996) representation in Equation (3) is a special case of the issue-preference model studied by Nau (2006) and Ergin and Gul (2009), and it is a special case of the smooth model of Klibanoff, Marinacci, and Mukerji (2005) when restricted to acts f that depend only on s. We discuss this special case in detail in Section 4.

4 Ambiguity Aversion

To set the stage, this section illustrates via an example the special case of our model previously analyzed by Robson (1996): a single fitness function ($\Psi = \{\psi\}$), which allows the supremum over Ψ to be dropped from the representation in Equation (2).

Example 1 (Ellsberg)**.** Consider an Ellsberg urn with one black ball and two balls that could each be either red or yellow. Each individual independently draws one ball from the urn, which we model using the state space $S = \{b, r, y\}$ for independent risk. The individual may be offered the following bets on colors of the ball drawn:

[11] The survey by Robson and Samuelson (2011) summarizes these results, as well as some recent developments in the literature on the evolution of preferences.

	b	r	y
B	1	0	0
R	0	1	0
BY	1	0	1
RY	0	1	1

In this table, B denotes the act that pays \$1 if the ball drawn is black and \$0 otherwise, BY indicates the act that pays \$1 if the ball is either black or yellow, and so on. The typical preference pattern documented by Ellsberg (1961) is $B \succ R$ and $BY \prec RY$, in violation of Savage's sure-thing principle.

To understand such preferences within the evolutionary model described above, note that although the draw of the ball is independent across individuals, the composition of the urn itself may be common for all individuals. In this case, we can model the possible urn compositions using the set $\Omega = \{\omega_1, \omega_2, \omega_3\}$, where $\omega_1 = (b, r, r)$, $\omega_2 = (b, r, y)$, and $\omega_3 = (b, y, y)$. Even if individuals form subjective probability assessments on the possible urn compositions, this correlated uncertainty is treated differently than uncorrelated uncertainty. For ease of illustration, suppose that μ assigns equal weight to each urn composition and that there is a single fitness function ψ that takes values $\psi(0) = 0$ and $\psi(1) = 1$. When acts only depend on s, the long run growth rate for deterministic choice in Equation (3) is a special case of the smooth model (Klibanoff, Marinacci, and Mukerji (2005)) with a concave transformation function, so these evolutionarily optimal preferences exhibit Ellsberg behavior:

$$V(B) = \ln\left[\frac{1}{3}\right] > \frac{1}{3}\ln\left[\frac{2}{3}\right] + \frac{1}{3}\ln\left[\frac{1}{3}\right] + \frac{1}{3}\ln[0] = V(R),$$

and

$$V(BY) = \frac{1}{3}\ln\left[\frac{1}{3}\right] + \frac{1}{3}\ln\left[\frac{2}{3}\right] + \frac{1}{3}\ln[1] < \ln\left[\frac{2}{3}\right] = V(RY).$$

This example is also useful for illustrating the role of random choice within our model, and why restricting to deterministic choice of acts is *not* without loss of generality. When choosing between B and R (or between BY and RY), it is easy to see that strict randomization in choice is not optimal. However, for other decision problems, randomization can have a strict benefit within our evolutionary model. Consider an act Y that bets on yellow, paying \$1 if the ball is yellow and \$0 otherwise. Suppose the individual is asked to choose between R and Y. Notice that $V(R) = V(Y)$ (both of which are strictly less than $V(B)$). However, if the random action ρ sets $\rho(R) = \alpha$ and $\rho(Y) = 1 - \alpha$, then it is easy to verify that $V(\rho)$ is maximized at $\alpha = 1/2$, giving a value of $V(\rho) = \ln(1/3) = V(B)$. In this case, the individual strictly prefers randomization to the deterministic choice of either Y or R. ▲

In Example 1, the crucial assumption for generating ambiguity aversion is that the composition of the urn is common across all individuals. In contrast, if a different urn is composed for each individual and if there is no correlation in how these urns are constructed, then correlation aversion alone would not produce ambiguity aversion—a different mechanism would be required to generate Ellsberg behavior. This example is therefore useful for illustrating both the scope and the limitations of the evolutionary model: Adaptive preferences generate ambiguity aversion anytime there is uncertainty about the model itself or some other factor that is common to all individuals in the population, which we contend is the case in the vast majority of examples and applications of ambiguity.[12] As noted earlier, in cases involving idiosyncratic ambiguity, we do not take a stand on whether ambiguity aversion is driven by heuristics developed by the genotypes from the case of common uncertainty or if it arises from some other source.[13]

In line with the interpretation of ambiguity as model uncertainty, we favor a statistical interpretation of the smooth model used in this section, where each $\omega \in \Omega$ is a candidate for the true model (the law of nature governing the distribution of $s \in S$) and the marginal distribution of μ on Ω is a prior over the candidate models.[14] For simplicity, we treat μ as constant over time. In that case, evolutionary optimality requires that individuals' preferences (eventually) assign the correct weights, so that μ becomes objective—it accurately reflects the data generating process. However, our evolutionary approach can easily be extended to allow μ_t to change with time t. For a simple example, suppose there is an index set K and a set of possible distributions $\mu^k \in \triangle(\Omega \times S)$, where $k \in K$ is redrawn periodically after finitely many periods. Then, in each period t, it is again evolutionarily optimal for individuals to maximize the growth rate in Equation (2), this time using their "best guess" of the distribution μ_t given all information available at time t. This information evolves as follows: One $\omega \in \Omega$ is commonly drawn each period, so that in between draws of k the marginal of μ^k on Ω is gradually revealed. At the same time, with a large number of individuals who each independently draw a state $s \in S$ each period, the conditional $\mu^k(\cdot|\omega)$ on S can be fully revealed in one period. In other words, in this situation ambiguity will only linger in the case of common uncertainty, in line with the discussion in Section 1.1.

[12] Halevy and Feltkamp (2005) suggested another mechanism by which correlation can generate ambiguity aversion: Risk aversion alone implies that an individual who makes repeated bets on an urn would rather draw from a risky than an ambiguous urn. In our evolutionary context, instead, the maximization of long-run growth generates an aversion to correlation in the contemporaneous draws of different individuals.

[13] If one is not convinced that the Ellsberg urn is a perfect fit for our model, the objects in the example can be recast in terms of other examples discussed in the introduction. For instance, the acts B, R, Y could represent different medical treatments for a condition and the idiosyncratic states b, r, y could represent the events in which each treatment is successful for an individual, with B being a better understood treatment than R and with the efficacy of the combined treatment RY being better understood than that of BY.

[14] See Klibanoff, Marinacci, and Mukerji (2005) or Marinacci (2015) for a discussion of this interpretation. An alternative interpretation is that the marginal of μ on Ω is a preference parameter that captures subjective plausibility of different first-order probabilistic beliefs $\mu(\cdot|\omega)$ on S.

5 Non-Expected Utility

The adaptive model also accommodates violations of expected utility in the context of pure (uncorrelated) risk. In this section, we consider two such special cases of our model: rank-dependent utility in Section 5.1 and a class of divergence preferences involving pessimistic distortions of an objective probability distribution in Section 5.2.

In contrast to Section 4, we now permit non-degenerate adaptation (non-singleton Ψ); however, to focus solely on risk preferences for the moment, we will restrict attention in this section to the special case of our model without common uncertainty ($\Omega = \{\omega\}$). Since a central contribution of this paper is to illustrate the *joint* restrictions on ambiguity attitudes and non-expected-utility preferences imposed by our evolutionary environment, Section 6 will characterize a set of equivalent representations for the general model with both non-trivial common uncertainty and non-trivial adaptation. The main result of this paper, Theorem 2 in Section 6, will enable us to embed the examples from this section within a model that incorporates ambiguity aversion.

Since this section focuses on the special case without common uncertainty, we will suppress the Ω dimension from the state space and focus on acts defined on the state space S. In this case, there is no strict benefit to randomization, so it is without loss of generality to restrict attention to deterministic action choices f and adaptation choices ψ.[15] Therefore, Equation (2) becomes

$$\begin{aligned} V(f) &= \sup_{\psi\in\Psi} \ln\left(\int_S \psi(f(s))\, d\mu(s)\right) \\ &= \ln\left(\sup_{\psi\in\Psi} \int_S \psi(f(s))\, d\mu(s)\right). \end{aligned} \tag{4}$$

Note that in this case the logarithm can also be dropped by taking a monotone transformation, but we will retain it for consistency in expressing growth rates in log terms and for ease of comparing the formulas in this section to later results.

In order to accommodate certain special cases, it will be technically convenient to permit the fitness functions ψ to take the value $-\infty$, so we henceforth assume that Ψ is a nonempty set of functions $\psi : Z \to [-\infty, \infty)$.[16] We have assumed throughout that the set Ψ is pointwise bounded above. In anticipation of our results in Section 6, we will also focus attention on sets that are closed, as formalized in the next assumption. We will see that the special cases

[15] Formally, after dropping the expectation over Ω from Equation (2), we have

$$V(\rho) = \sup_{\tau\in\mathcal{R}(\Psi|\mathcal{F})} \ln\left(\int_S \mathbb{E}_{\tau\otimes\rho}\big[\psi(f(s))\big]\, d\mu(s)\right).$$

Since the expression inside the logarithm is linear in both τ and ρ, it is maximized by a deterministic action choice and adaptation plan.

[16] Recall that we take $\ln(x) = -\infty$ for all $x \le 0$.

of rank-dependent utility and divergence preferences considered in the following subsections will be characterized by sets Ψ that satisfy this assumption.

Assumption 1. Suppose Ψ is a nonempty set of functions $\psi : Z \to [-\infty, \infty)$ that is pointwise bounded above (that is, $\sup_{\psi \in \Psi} \psi(z) < \infty$ for every $z \in Z$) and closed in the topology of pointwise convergence (on the extended reals).

5.1 Rank-Dependent Utility

Although the connection is nontrivial, the following result shows that rank-dependent utility with a pessimistic probability distortion function can be expressed as a special case of our model.

Proposition 1 (Rank-Dependent Utility)**.** *Suppose $Z \subset \mathbb{R}$. Fix $\mu \in \triangle(S)$, and fix any bounded nondecreasing function $u : Z \to \mathbb{R}$ and any function $\varphi : [0,1] \to [0,1]$ that is nondecreasing, concave, and onto. Then, there exists a set Ψ of functions $\psi : Z \to \mathbb{R}$ satisfying Assumption 1 such that for any simple act $f : S \to Z$,*[17]

$$\sup_{\psi \in \Psi} \int_S \psi(f(s))\, d\mu(s) = \int_Z u(z)\, d(\varphi \circ F_{f,\mu})(z)$$

where

$$F_{f,\mu}(z) = \int_S \mathbf{1}[f(s) \le z]\, d\mu(s)$$

denotes the cumulative distribution function of f given μ. In particular, the value function V in Equation (4) can be equivalently expressed as

$$V(f) = \ln\left(\int_Z u(z)\, d(\varphi \circ F_{f,\mu})(z)\right).$$

Since we identify idiosyncratic uncertainty over S with pure risk, the distribution of outcomes $F_{f,\mu}$ amounts to an unambiguous risky prospect. Thus, given the appropriate set of fitness functions Ψ, adaptive preferences are equivalent to rank-dependent utility where individuals overweight the probability assigned to worse outcomes.[18] The following example from Ben-Tal and Teboulle (2007) and Sarver (2018, supplementary material) illustrates more concretely one instance of this duality.

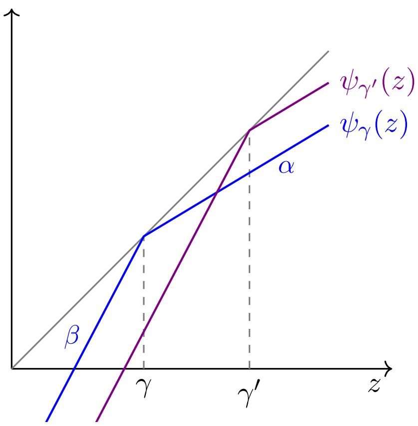
[17] This dual formula is similar to several existing results in the literature. See, for example, Wakker (1994), Chatterjee and Krishna (2011), or the Supplementary Material of Sarver (2018).

[18] In this paper, we focus on exploring the scope of adaptive preferences by identifying special cases that can be nested. Sarver (2018) considers a similar representation to Equation (4). He shows that his model does not overlap with other prominent non-expected-utility models (disappointment aversion, betweenness preferences, cautious expected utility) except in the case of expected utility. These insights are easily extended to our model and help delineate the boundary of the set of preferences that it nests.

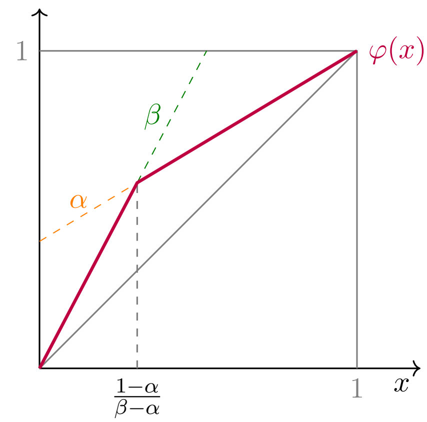
Example 2 (RDU Fitness Functions)**.** Suppose $Z \subset \mathbb{R}$, and fix some $0 \leq \alpha < 1 < \beta$. Consider a parametric class of fitness functions where for each $\gamma \in \mathbb{R}$, we define $\psi_\gamma : \mathbb{R} \to \mathbb{R}$ by

$$\psi_\gamma(z) = \begin{cases} \gamma + \beta(z - \gamma) & \text{if } z < \gamma \\ \gamma + \alpha(z - \gamma) & \text{if } z \geq \gamma, \end{cases}$$

The fitness function ψ_γ can be interpreted as a piecewise-linear gain-loss function with a target consumption level γ. It is concave (since $\alpha < \beta$), takes value γ at $z = \gamma$, and takes values strictly below z for $z \neq \gamma$. Figure 1a illustrates. ▲



(a) Gain-loss functions ψ_γ



(b) Probability distortion function φ.

Figure 1: Illustration of Example 2 and Claim 1.

Claim 1. *Define ψ_γ as in Example 2. Then, for any simple act $f : S \to \mathbb{R}$,*

$$\max_{\gamma \in \mathbb{R}} \int_S \psi_\gamma(f(s))\, d\mu(s) = \int_Z z\, d(\varphi \circ F_{f,\mu})(z)$$

where

$$\varphi(x) = \begin{cases} \beta x & \text{if } x < \frac{1-\alpha}{\beta-\alpha} \\ \alpha x + (1-\alpha) & \text{if } x \geq \frac{1-\alpha}{\beta-\alpha}. \end{cases}$$

Figure 1b illustrates the probability distortion function φ from Claim 1. Sarver (2018, supplementary material) proves this result for the special case of $\alpha = 1 - \theta$ and $\beta = 1 + \theta$ for $\theta \in [0, 1]$. The general proof follows from similar arguments and we therefore omit it. For intuition, if there is a $\gamma \in \mathbb{R}$ such that $F_{f,\mu}(\gamma) = \frac{1-\alpha}{\beta-\alpha}$, then the optimal fitness function for the act f is ψ_γ.[19] From this, it is easy to show that outcomes to the left of γ are weighted by β times their probability and outcomes to the right are weighted by α, consistent with the distortion function φ.

[19]Since f is a simple act and $F_{f,\mu}$ is therefore not continuous, there may be no such γ. In general, any γ that satisfies $F_{f,\mu}(\gamma-) \leq \frac{1-\alpha}{\beta-\alpha} \leq F_{f,\mu}(\gamma)$ will be optimal, where $F_{f,\mu}(\gamma-)$ is the left limit of $F_{f,\mu}$ at γ.

5.2 Divergence Preferences

Definition 4. Fix a convex and lower semicontinuous function $\phi : \mathbb{R}_+ \to [0, \infty]$ such that $\phi(1) = 0$ and there exists some $\alpha < 1 < \beta$ such that ϕ is finite on the interval $[\alpha, \beta]$, and let p and q be probability measures on a given state space. The *ϕ–divergence* of p with respect to q is given by

$$D_\phi(p \,\|\, q) = \begin{cases} \displaystyle\int \phi\left(\frac{dp}{dq}\right) dq & \text{if } p \ll q, \\ \infty & \text{otherwise.} \end{cases} \tag{5}$$

The notation $p \ll q$ indicates that p is *absolutely continuous* with respect to q, that is, for any measurable set A, $q(A) = 0$ implies $p(A) = 0$. The term $\frac{dp}{dq}$ denotes the Radon–Nikodym derivative (density) of p with respect to q, which exists if and only if p is absolutely continuous with respect to q.[20] It is immediate from the definition of the ϕ–divergence that $D_\phi(p \,\|\, q) \geq 0$, with equality if $p = q$. Moreover, if ϕ is strictly convex, then $D_\phi(p \,\|\, q) = 0$ if and only if $p = q$. *Relative entropy* (or *Kullback–Leibler divergence*) is the special case of a ϕ–divergence where $\phi(t) = t\ln(t) - t + 1$. In this case, Equation (5) simplifies to

$$R(p \,\|\, q) = \begin{cases} \displaystyle\int \ln\left(\frac{dp}{dq}\right) dp & \text{if } p \ll q, \\ \infty & \text{otherwise.} \end{cases} \tag{6}$$

Ben-Tal and Teboulle (1987, 2007) provided an explicit dual characterization of variational preferences (Maccheroni, Marinacci, and Rustichini (2006)) with a divergence cost function as the supremum of a set of expected utilities under the reference measure, where the supremum is taken over a set of possible Bernoulli utility indices. The following proposition extends their result to permit a nondecreasing convex transformation k of the divergence term.

Proposition 2 (Divergence Duality)**.** *Fix any $\mu \in \triangle(S)$, any ϕ–divergence $D_\phi(\cdot \,\|\, \cdot)$, and any function $u : Z \to \mathbb{R}$. Also, fix any nondecreasing, convex, and lower semicontinuous function $k : \mathbb{R}_+ \to [0, \infty]$ such that $k(0) = 0$ and k is finite on some interval $[0, \varepsilon)$.*[21] *Then, there exists a set Ψ satisfying Assumption 1 such that, for any simple act $f : S \to Z$,*

$$\sup_{\psi \in \Psi} \int_S \psi(f(s))\, d\mu(s) = \inf_{\eta \in \triangle(S)} \left[\int_S u(f(s))\, d\eta(s) + k(D_\phi(\eta \,\|\, \mu)) \right].$$

The next corollary illustrates the flexibility of the transformation k in Proposition 2 by highlighting two special cases. The first is where $k(x) = \theta x$ for some scalar $\theta > 0$. The second is where we fix a scalar $\kappa > 0$ and take $k(x) = 0$ if $x \leq \kappa$, and $k(x) = +\infty$ if $x > \kappa$.

[20] Formally, $\frac{dp}{dq}$ is the integrable function that satisfies $p(A) = \int_A \frac{dp}{dq}\, dq$ for any measurable set A.

[21] We adopt the convention that $k(\infty) = \infty$. Thus, for any function k as in the statement of the proposition, if $D_\phi(r \,\|\, p) = \infty$ then $k(D_\phi(r \,\|\, p)) = \infty$.

Corollary 2. *Fix any $\mu \in \triangle(S)$, any ϕ–divergence $D_\phi(\cdot \,\|\, \cdot)$, and any function $u : Z \to \mathbb{R}$. Given a set Ψ, define V by Equation (4).*

1. *Fix any scalar $\theta > 0$. There exists a set Ψ such that*

$$V(f) = \ln\left(\inf_{\eta \in \triangle(S)} \int_S u(f(s))\, d\eta(s) + \theta D_\phi(\eta \,\|\, \mu) \right).$$

2. *Fix a scalar $\kappa > 0$, and define $\mathcal{D}(\mu, \kappa) = \{\eta \in \triangle(S) : D_\phi(\eta \,\|\, \mu) \leq \kappa\}$. There exists a set Ψ such that*

$$V(f) = \ln\left(\inf_{\eta \in \mathcal{D}(\mu,\kappa)} \int_S u(f(s))\, d\eta(s) \right).$$

Divergence preferences have typically been considered in the context of ambiguity rather than risk. For example, the formula inside the logarithm in part 1 of Corollary 2 was analyzed by Maccheroni, Marinacci, and Rustichini (2006) as a special case of their model of variational preferences. The formula inside the logarithm in part 2 of the corollary is a multiple prior representation (Gilboa and Schmeidler (1989)). However, while these types of functional forms have received the most attention in the ambiguity literature, they also have a natural interpretation in the context of objective risk.[22] The pessimistic distortion of the (objective and idiosyncratic) probability measure μ to some other measure η in these formulas allows over-weighting of bad outcomes, similar to the interpretation of rank-dependent utility in the previous section. In fact, this similarity between the two models is not merely conceptual: Although the classes of divergence and RDU preferences are distinct, we provide an example in Appendix B showing there is actually some overlap between the two classes of preferences.

In the proof of Proposition 2, we provide an explicit formula for the set of fitness functions Ψ using the Fenchel conjugate from convex analysis (see Proposition 3 in Appendix A.3). The following example is based on these explicit formulas.

Example 3 (Divergence Fitness Functions)**.** Suppose $Z \subset \mathbb{R}$, and consider the following parametric class of fitness functions involving two parameters, where the first ($\gamma \in \mathbb{R}$) is a target level of consumption and the second ($\alpha \geq 0$) determines the sensitivity to gains and losses: Define $\psi_{\gamma,\alpha} : \mathbb{R} \to [-\infty, \infty)$ by

$$\psi_{\gamma,\alpha}(z) = \gamma + \alpha - \alpha \exp\left(\frac{\gamma - z}{\alpha} \right) - c(\alpha)$$

if $\alpha > 0$. For $\alpha = 0$, let $\psi_{\gamma,\alpha}(z) = \gamma$ for $z \geq \gamma$ and $\psi_{\gamma,\alpha}(z) = -\infty$ for $z < \gamma$. The components of $\psi_{\gamma,\alpha}(z)$ have a simple interpretation: $\gamma + \alpha - \alpha \exp\left(\frac{\gamma - z}{\alpha}\right)$ is a gain-loss function that is strictly concave, takes the value γ at $z = \gamma$, and takes values strictly below z for $z \neq \gamma$.

[22] Moreover, this class of divergence preferences is propabilistically sophisticated by Proposition 2. Maccheroni, Marinacci, and Rustichini (2006) previously observed this for the case where k is linear.

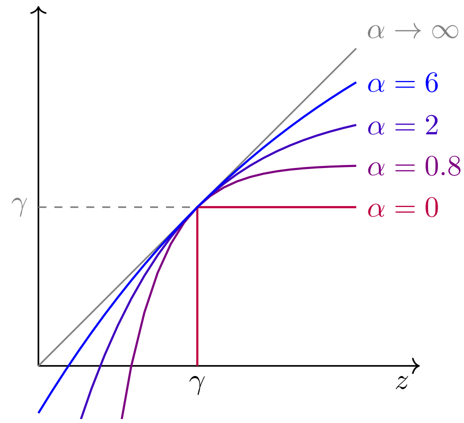


Figure 2: Illustration of gain-loss function $\gamma + \alpha - \alpha \exp\left(\frac{\gamma - z}{\alpha}\right)$ in Example 3.

The parameter α determines the sensitivity to gains and losses, with larger α leading to decreased sensitivity. See Figure 2 for an illustration. Finally, $c : \mathbb{R}_+ \to [0, \infty]$ is a function that determines the "cost" of increasing α (decreasing sensitivity to gains and losses). Assume that c is nondecreasing and convex, with $c(0) = 0$. ▲

Claim 2. *Define $\psi_{\gamma,\alpha}$ as in Example 3. Then, for any simple act $f : S \to \mathbb{R}$,*

$$\max_{\gamma \in \mathbb{R}} \max_{\alpha \geq 0} \int_S \psi_{\gamma,\alpha}(f(s))\, d\mu(s) = \inf_{\eta \in \triangle(S)} \left[\int_S f(s)\, d\eta(s) + k(R(\eta \,\|\, \mu)) \right],$$

where $R(\eta \,\|\, \mu)$ is the relative entropy defined in Equation (6) and special cases of the functions c and k are related as follows, where $\kappa, \theta > 0$:

1. *If $c(\alpha) = \kappa\alpha$, then $k(x) = 0$ for $x \leq \kappa$ and $k(x) = \infty$ otherwise.*
2. *If $c(\alpha) = 0$ for $\alpha \leq \theta$ and $c(\alpha) = \infty$ otherwise, then $k(x) = \theta x$.*

Claim 2 relates the class of fitness functions in Example 3 to divergence preferences and, moreover, illustrates two instances of the duality between the cost function c in the fitness functions and the transformation k that is applied to the divergence.[23]

6 Ambiguity Aversion and Non-Expected Utility

We observed in Sections 4 and 5 that our model of adaptive preferences nests as special cases rank-dependent utility and divergence preferences in the context of risk and a version of the smooth model in the context of ambiguity. In this section, we expand our analysis to

[23]More generally, the function c is the Fenchel conjugate of k (by Proposition 3 in the appendix).

special cases of our representation that simultaneously incorporate both ambiguity aversion and non-expected utility for risk.[24] One motivation for simultaneously considering both is the empirically observed correlation between ambiguity aversion and violations of the independence axiom (Dean and Ortoleva (2019)). Another motivation for this generality is that within our model of adaptive preferences, there are some attitudes toward ambiguity that are incompatible with a single fitness function ψ, but can be accommodated with non-trivial adaptation, in which case violations of expected utility are implied. Section 6.4 provides an example of such a pattern based on the experimental findings of Abdellaoui et al. (2011).

An impediment to the analysis of special cases of our general representation in Equation (2) is that it has a logarithm between the two layers of integration. For example, our results for rank-dependent utility and divergence preferences in the previous section assumed that there was no common uncertainty, and it is not immediately obvious how those results might be extended to the general case of both common and idiosyncratic uncertainty. The main result of this section, Theorem 2, is a duality result that recasts our representation in a form that facilitates the analysis of these and other special cases. We then study several special cases in detail in Sections 6.1 and 6.2, and we briefly discuss comparative statics that link risk and uncertainty aversion across some of those special cases in Section 6.3.

Our results will involve the relative entropy of one probability measure with respect to another, as defined in Equation (6). In what follows, for any probability measure $p \in \triangle(\Omega)$, let

$$M(p) = \{q \in \triangle(\Omega) : q \ll p \text{ and } R(p \,\|\, q) < \infty\}.$$

In particular, since $R(p \,\|\, q) < \infty$ requires that $p \ll q$, if $q \in M(p)$ then the measures p and q are mutually absolutely continuous, that is, both $p \ll q$ and $q \ll p$.[25] When necessary to avoid confusion, we will denote the marginal distribution of μ on Ω by μ_Ω.

Theorem 2. *Suppose Ψ satisfies Assumption 1, and fix $\mu \in \triangle(\Omega \times S)$. For any random action plan $\rho \in \mathcal{R}(\mathcal{F})$, the function V defined by Equation (2) can be equivalently expressed as*

$$V(\rho) = \inf_{q \in M(\mu_\Omega)} \left[\ln\left(\mathbb{E}_\rho\left[\sup_{\psi \in \Psi} \int_\Omega \int_S \psi(f(\omega, s))\, d\mu(s|\omega)\, dq(\omega) \right] \right) + R(\mu_\Omega \,\|\, q) \right]. \tag{7}$$

For intuition, we highlight the key steps in the proof: First, using duality techniques related to those employed in the literature on large deviations in statistics (cf. Dupuis and

[24] There are relatively few models in the axiomatic decision theory literature that combine ambiguity aversion and non-expected utility for risk; see, for example, Segal (1987), Dean and Ortoleva (2017) and Izhakian (2017).

[25] Note that it is possible to have $R(p \,\|\, q) = \infty$ even if $p \ll q$, so $M(p)$ may be a strict subset of the set of all measures that are mutually absolutely continuous with respect to p.

Ellis (1997)), we show that Equation (2) can be equivalently expressed as

$$V(\rho) = \sup_{\tau \in \mathcal{R}(\Psi|\mathcal{F})} \inf_{q \in M(\mu_\Omega)} \left[\ln\left(\int_\Omega \int_S \mathbb{E}_{\tau \otimes \rho}\big[\psi(f(\omega, s))\big]\, d\mu(s|\omega)\, dq(\omega) \right) + R(\mu_\Omega \,\|\, q) \right].$$

This expression is not yet amenable to analysis, as we would like to reverse the order of the supremum and infimum in order to further simplify it and connect with existing functional forms. The next step in the proof is to do just that by leveraging a particular version of the von Neumann–Sion minimax theorem (von Neumann (1928), Sion (1958)) that is due to Tuy (2004). Then, after we switch the order of the supremum and infimum, the supremum over τ applies to the expression inside the logarithm, which is linear in τ. Therefore, optimization over adaptation plans τ can be reduced to the deterministic selection of a fitness function ψ following every act f that realizes under ρ, giving Equation (7). This final observation will greatly simplify the analysis of the model since it eliminates randomization over ψ from the formula for long-run growth rates.

Despite the resemblance, the functional in Equation (7) with a single fitness function $\Psi = \{\psi\}$ is not a variational representation (Maccheroni, Marinacci, and Rustichini (2006)). The distinction is the logarithm around the integral in the first term. In fact, in the case of a single fitness function, taking the exponential transformation of the representation in Equation (7) establishes it as a special case of the confidence preferences studied by Chateauneuf and Faro (2009), where confidence in a prior q is measured by $\exp(R(\mu_\Omega \,\|\, q))$. More generally, this no-adaptation case is also nested by the general representation for uncertainty-averse preferences proposed by Cerreia-Vioglio et al. (2011).

Turning to the specifics of our functional form, relative entropy has appeared in a number of representations for ambiguity-averse preferences, perhaps most notably in the multiplier preferences introduced by Hansen and Sargent (2001) and studied axiomatically by Strzalecki (2011),[26] and also within a version of confidence preferences in Chateauneuf and Faro (2012). However, in these models, the entropy term used is $R(q \,\|\, \mu_\Omega)$ rather than $R(\mu_\Omega \,\|\, q)$. While relative entropy is often interpreted as a "distance" between the two distributions involved, it is not a distance function in the metric sense, because it is not symmetric. To interpret the subtle difference in the context of the representation in Equation (7), suppose the decision-maker takes as the reference measure μ_Ω the empirical frequencies in a large sample of independently realized states $\omega \in \Omega$, but worries that the data is actually generated by the measure q on Ω. Of course, the larger the sample, the closer to zero the probability that it would be generated by $q \neq \mu_\Omega$. The theory of large deviations establishes that the rate at which this probability vanishes increases in $R(\mu_\Omega \,\|\, q)$ (see, e.g., Cover and Thomas (2006, Section 11.4)). The representation suggests, therefore, that the decision-maker is less confident in a measure q the faster it becomes implausible with growing sample size.

[26]Hansen and Sargent (2001) interpret their representation in terms of a concern about robustness to model misspecification. Our approach provides a related perspective on concern for robustness in contexts where uncertainty about ω can be interpreted as model uncertainty.

In order to describe the special cases of the next two subsections, it will be convenient to define a measure $\mu \otimes q$ on $\Omega \times S$ with marginal q on Ω and conditional distribution $\mu(\cdot|\omega)$ on S. That is, for any event E in the product σ-algebra $\mathcal{B}_\Omega \otimes \mathcal{B}_S$, let

$$\mu \otimes q(E) = \int_\Omega \int_S \mathbf{1}[(\omega, s) \in E] \, d\mu(s|\omega) \, dq(\omega).$$

With this definition in hand, Equation (7) can be written as

$$V(\rho) = \inf_{q \in M(\mu_\Omega)} \left[\ln\left(\mathbb{E}_\rho \left[\sup_{\psi \in \Psi} \int_{\Omega \times S} \psi(f(\omega, s)) \, d(\mu \otimes q)(\omega, s) \right] \right) + R(\mu_\Omega \,\|\, q) \right]. \tag{8}$$

6.1 Nesting Rank-Dependent Utility

Proposition 1 linked our adaptive model to RDU preferences in the special case of no common uncertainty ($\Omega = \{\omega\}$), in which case the state space was effectively S. The next corollary follows directly from that result by taking the state space to be $S' = \Omega \times S$ and the measure to be $\mu' = \mu \otimes q \in \triangle(\Omega \times S)$. Note that this application is only possible because we first apply Theorem 2 to remove the logarithm from between the two layers of integration.

Corollary 3. *Suppose $Z \subset \mathbb{R}$. Fix $\mu \in \triangle(\Omega \times S)$, and fix any bounded nondecreasing function $u : Z \to \mathbb{R}$ and any function $\varphi : [0, 1] \to [0, 1]$ that is nondecreasing, concave, and onto. Then, there exists a set Ψ of functions $\psi : Z \to \mathbb{R}$ satisfying Assumption 1 such that the function V defined by Equation (2) can be equivalently expressed as*

$$V(\rho) = \inf_{q \in M(\mu_\Omega)} \left[\ln\left(\mathbb{E}_\rho \left[\int_Z u(z) \, d(\varphi \circ F_{f,\mu\otimes q})(z) \right] \right) + R(\mu_\Omega \,\|\, q) \right],$$

where

$$F_{f,\mu\otimes q}(z) = \int_{\Omega \times S} \mathbf{1}[f(\omega, s) \leq z] \, d(\mu \otimes q)(\omega, s)$$

is the cumulative distribution function of f given $\mu \otimes q$.

This representation illustrates the simplicity of analyzing the combination of ambiguity aversion, non-expected-utility risk preferences, and random choice when working with the dual formula in Equation (8) and its special cases. In this application, the RDU representation inside the logarithm generates aversion to any kind of uncertainty, while ambiguity aversion (roughly speaking, the *additional* aversion to uncertainty from Ω) is captured by the outer part of the representation—the confidence preferences within which the RDU representation is embedded. The outer part is fixed across genotypes, even if those differ in terms of Ψ and hence in terms of their attitudes towards risk.[27] Random choice of acts is also easy

[27] There is some empirical evidence that risk aversion and additional aversion to ambiguity indeed have little correlation in the population (Chapman et al. (2019)).

to analyze in this representation, since the expectation with respect to ρ appears inside the confidence preferences (reflecting the hedging benefits of randomization) but outside of the RDU formula.

6.2 Nesting Divergence Preferences

Proposition 2 linked our adaptive model to a general class of divergence preferences in the special case of no common uncertainty ($\Omega = \{\omega\}$), in which case the state space was effectively S. In parallel our analysis of rank-dependent utility in the previous section, the next corollary follows directly from Proposition 2 by taking the state space to be $S' = \Omega \times S$ and the measure to be $\mu' = \mu \otimes q \in \triangle(\Omega \times S)$. Again, this application is only possible because of Theorem 2.

Corollary 4. *Fix any $\mu \in \triangle(\Omega \times S)$, any ϕ–divergence $D_\phi(\cdot \,\|\, \cdot)$, and any function $u : Z \to \mathbb{R}$. Also, fix any nondecreasing, convex, and lower semicontinuous function $k : \mathbb{R}_+ \to [0, \infty]$ such that $k(0) = 0$ and k is finite on some interval $[0, \varepsilon)$. Then, there exists a set Ψ satisfying Assumption 1 such that the function V in Equation (2) can be equivalently expressed as*

$$V(\rho) = \inf_{q \in M(\mu_\Omega)} \left[\ln\left(\mathbb{E}_\rho\left[\inf_{\eta \in \triangle(\Omega \times S)} \int_{\Omega \times S} u(f(\omega, s))\, d\eta(\omega, s) + k(D_\phi(\eta \,\|\, \mu \otimes q)) \right] \right) + R(\mu_\Omega \,\|\, q) \right].$$

The representation in Corollary 4 can be further specialized by considering specific functional forms for k, just as in Corollary 2 from Section 5.2. This value function embeds a general divergence representation inside confidence preferences. To see how it captures ambiguity aversion, note that the measure η ultimately used to evaluate an act may be more pessimistic than $\mu \otimes q$ on $\Omega \times S$, which in turn may be more pessimistic than μ only on Ω. Hence, compared to μ, there is more "opportunity" for η to be pessimistic about Ω than about S.

6.3 Comparative Statics

We briefly mention comparative statics that compare the behavior of individuals with different sets of fitness functions Ψ:[28] Suppose that all conceivable genotypes perform equally well when facing deterministic outcomes (no uncertainty). In terms of the model of adaptive preferences, this means that the upper envelope of Ψ is the same for all those genotypes. In this case, one can show that individual A with adaptive preferences for Ψ_A is more risk averse than an individual B with Ψ_B if and only if individual A is also more uncertainty averse than B. For example, in the representations of Corollaries 3 and 4 the upper envelope

[28] We have taken the set Ψ as given throughout. To compare individuals with different Ψ, it is important to understand how Ψ is determined. One possibility is that different choice situations involve different sets of hidden actions. Another possibility is that Ψ itself is subject to constrained evolutionary optimization.

of Ψ is u, and hence holding fixed u, individuals with either of these two types of preferences who can be ranked in terms of risk aversion will be ranked the same way in terms of overall uncertainty aversion.

6.4 Source Preferences

Adaptive preferences imply that any ambiguity attitude that cannot be captured by the smooth model of ambiguity aversion in Equation (3) must go hand in hand with violations of expected utility. We illustrate this connection via the example of source preferences.

Abdellaoui et al. (2011) compare behavior under known (risky) and unknown (ambiguous) sources of uncertainty. In their experiment, the known uncertainty comes from betting on the color of a ball drawn from an urn with eight balls of eight different colors. The unknown uncertainty comes from betting on the color of a ball drawn from an urn with eight balls with the same colors, but unknown composition in the sense that some colors might appear several times and others might be absent. Based on symmetry it is clear that each color should be considered equally likely even for the unknown urn (just as the two colors are usually revealed to be considered equally likely in the ambiguous two color Ellsberg urn), but nonetheless an ambiguity adverse decision maker would prefer to bet on the known urn.

Abdellaoui et al. (2011) interpret their data through the lens of *source functions* w_K and w_U, which are probability weighting functions for the known and unknown sources, respectively. That is, fixing outcomes $x > y$, the gamble that yields x with probability p and y otherwise is evaluated as if the probability assigned to x is $w(p)$ instead of p. For a representative subject,[29] they find that for $p > 0.5$ the source function w_U is systematically lower than w_K, while for small p there is no significant difference. To be more specific, consider the function $\alpha(p) := w_K(w_U^{-1}(p))$. While the experiment determines preferences via certainty equivalents, transitivity of preferences implies that the representative subject is indifferent between the gamble with unknown probability p and the one with known probability $\alpha(p)$, which we therefore call the *risk equivalent* of p, given x and y. Their findings then imply that $\alpha(0) = 0$, $\alpha(1) = 1$ and $\alpha(p)$ is continuous and strictly increasing, but is not convex everywhere, is close to the identity function for small p, and lies below for large p.

Here, we will not attempt to calibrate our model of adaptive preferences to match their data.[30] Our modest aim in this section is merely to demonstrate via the following two claims that (i) it cannot match the qualitative pattern of $\alpha(p)$ just described if Ψ is a singleton (no adaptation), but that (ii) it can match this pattern if Ψ has two or more elements. To that end, let $\alpha_\Psi(p)$ be the risk equivalent implied by our model for a particular Ψ.

[29] They use a parametric family of functions to approximate individual source functions, and report the median parameter values.

[30] In particular, some individuals in their data are ambiguity loving for small p, while our model always implies ambiguity aversion.

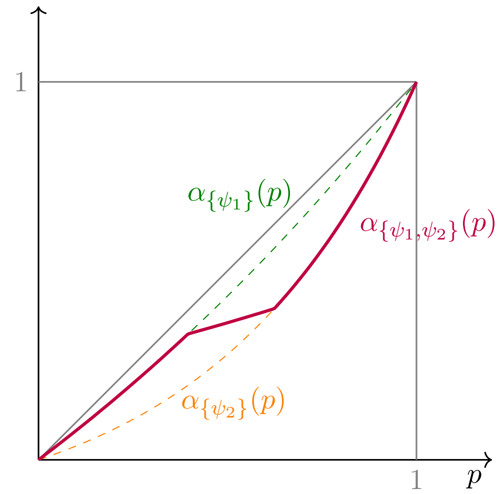


Figure 3: $\alpha_{\{\psi_1,\psi_2\}}$ for $\psi_1(x) = 1.2$, $\psi_1(y) = 0.7$, $\psi_2(x) = 2$, $\psi_2(y) = 0.3$

Claim 3. *For any strictly increasing fitness function ψ, the risk equivalent $\alpha_{\{\psi\}}(p)$ is continuous, strictly increasing and strictly convex in p, and satisfies $\alpha_{\{\psi\}}(0) = 0$ and $\alpha_{\{\psi\}}(1) = 1$. Furthermore, if $\psi_2(x) > \psi_1(x) > \psi_1(y) > \psi_2(y)$, then $\alpha_{\{\psi_1\}}(p) > \alpha_{\{\psi_2\}}(p)$ for all $p \in (0, 1)$.*

Claim 4. *If $\psi_2(x) > \psi_1(x) > \psi_1(y) > \psi_2(y)$, then there are $p_* < p^* \in (0, 1)$ such that $\alpha_{\{\psi_1,\psi_2\}}(p) = \alpha_{\{\psi_1\}}(p)$ for all $p \leq p_*$, $\alpha_{\{\psi_1,\psi_2\}}(p) = \alpha_{\{\psi_2\}}(p)$ for all $p \geq p^*$, and $\alpha_{\{\psi_1,\psi_2\}}(p)$ is convex for all $p \geq p_*$.*

Figure 3 illustrates via an example how our model generates the pattern in Claim 4. Obviously, the fact that Ψ contains two elements that do not dominate each other immediately implies that preferences over gambles with known probabilities (risk) must violate expected utility. To continue the example in Figure 3, suppose $\Psi = \{\psi_1, \psi_2\}$ where $\psi_1(z) = 0.5\sqrt{z+2}$ and $\psi_2(z) = \sqrt{z + 0.1}$. Then it is easy to verify that for $x = 4$ and $y = 0$ the two fitness functions generate the values in the figure (rounded to the first decimal) and that for risky gambles the common ratio violation of expected utility proposed in Allais (1953) ensues: Getting $w = 3$ for sure is better than getting $x = 4$ with probability $p = 0.8$ and $y = 0$ otherwise, but getting $x = 4$ with probability $p = 0.2$ and $y = 0$ otherwise is better than getting $w = 3$ with probability $p = 0.25$ and $y = 0$ otherwise.

7 Realism of the Evolutionary Model

Section 7.1 discusses two assumptions that are implicit in our formulation of the evolutionary model and that are commonly made in economic contexts. Section 7.2 concludes by discussing the interpretation of adaptive preferences in the context of phenotypic flexibility.

7.1 Simplifying Assumptions

Corollary 1 shows that the long-run growth rate is optimized by choosing the action plan $\rho \in \mathcal{R}(A)$ that maximizes V, assuming the decision problem A is faced by the genotype repeatedly *in every period*. In fact, this assumption is unnecessarily strong and is made solely for ease of exposition. As can be seen in the proof of Theorem 1, aggregate fitness in each period affects the population size multiplicatively, which provides a degree of separability for choice problems that appear at different times. For example, if the genotype faces an infinite sequence of decision problems $(A_t)_{t\in\mathbb{N}}$, then attaining the highest possible long-run growth rate requires that individuals maximize adaptive preferences from any decision problem A that repeats with fixed frequency within this sequence.[31]

Another assumption in our model is that time is divided into discrete time periods. Robatto and Szentes (2017) made the surprising observation that correlation aversion disappears in the continuous-time limit of this basic model. Further extending this line of research, Robson and Samuelson (2019) allowed fertility and mortality rates to vary with age in order to separate the assumption of continuous time from the assumption that new organisms can reproduce immediately after birth, and they found that correlation aversion can be recovered even in continuous time. Investigating the implications of different timing and age structures in our context of hidden actions and updating could be an interesting avenue for future research. In this paper, we stick to discrete time with age-independent fertility and mortality rates as is common in evolutionary models in economics.

7.2 Phenotypic Flexibility in Evolutionary Biology

While our approach is inspired by evolutionary biology, we hope that our insights might in turn also be useful in biological contexts where phenotypic flexibility plays a role, as we now explain in more detail. Evolutionary success appears to be greatly enhanced by the ability of organisms of a particular genotype to adapt their phenotype to the environment. Adopting the terminology proposed by Piersma and Drent (2003), we use *phenotypic flexibility* to refer to the rapid and apparently purposeful variation in phenotype expressed by individual reproductively mature organisms throughout their life. This is in contrast to *developmental plasticity*, environmentally induced variations that occur only during development.[32]

While developmental plasticity has long been a focus of evolutionary biologists, the role of phenotypic flexibility in the evolutionary process has only recently attracted significant attention. According to Piersma and Drent (2003):

[31] The assumption that all individuals of the genotype face the same decision problem at the same time is also implicit in our model, and this assumption can be relaxed as well. If, instead, there is a distribution of decision problems within the population, then this uncertainty can be encoded into the state spaces in our model.

[32] Piersma and Drent (2003) use *phenotypic plasticity* as an umbrella term that includes both phenotypic flexibility and developmental plasticity.

> When environmental conditions change rapidly [...] individuals that can show continuous but reversible transformations in behaviour, physiology and morphology might incur a selective advantage. There are now several studies documenting substantial but reversible phenotypic changes within adult organisms.

Striking examples among vertebrates include various species of amphibious fish that adjust to life on land with reversible and rapid (sometimes within minutes) changes to their muscle tissue, breathing organs, and skin properties (Wright and Turko (2016) provide a survey), or marine iguanas on the Galapagos islands that can shrink their overall body length by up to 20% (6.8 cm) in what appears to be a reversible, rapid, and strategic response to food scarcity during an El Niño weather pattern (Wikelski and Thom (2000)). A familiar example that can be viewed as phenotypic flexibility in humans and other mammals is the adjustment of the makeup of muscle tissue in response to changes in functional demands (Flück (2006)), for instance, from a more or less active lifestyle.

Of course, the evolutionary benefit of phenotypic flexibility is that different phenotypes may perform better in different situations, and hence have different fitness functions ψ. For instance, each possible phenotype might be tailored to a specific range of outcomes, such as the amount of available food for the iguanas in the example above (see Figure 1a for a collection of fitness functions with this feature). Or one phenotype might be a specialist with high fitness for a small range of outcomes, while the other is a generalist, with lower peak fitness that is more robust to the outcome (see Figure 2 of Example 3 for a collection of fitness functions that capture this trade-off once the cost c from the example is incorporated).

Biologists in the studies above directly observe variations in individual phenotypes over time. In economic applications, in contrast, phenotypes, such as the determinants of risk and ambiguity preferences in our model, are notoriously hard to observe. Economists instead rely on preferences that are revealed from observable choice data. Respecting this limitation, our model predictions concern only observable choices between outcome-relevant actions (f), treating the phenotype and resulting fitness function (ψ) as unobservable. As a consequence, our model does not distinguish between the case where adaptation is due to a biological change (phenotypic flexibility) or a strategic but hidden choice of action, and it is equally relevant and applicable under either interpretation of the set of fitness functions Ψ.

A Proofs

A.1 Proof of Lemma 1

Note that

$$\ln(N^i(T)) = \ln(N^i(0)) + \sum_{t=1}^{T} \ln(\lambda_t^i),$$

and therefore

$$\ln\left(\frac{N^A(T)}{N^B(T)}\right) = \ln\left(\frac{N^A(0)}{N^B(0)}\right) + \sum_{t=1}^{T} \ln(\lambda_t^A) - \sum_{t=1}^{T} \ln(\lambda_t^B).$$

Since α^A and α^B are the long-run growth rates of these two genotypes, we have

$$\frac{1}{T}\Big[\sum_{t=1}^{T} \ln(\lambda_t^A) - \sum_{t=1}^{T} \ln(\lambda_t^B)\Big] \to \alpha^A - \alpha^B \quad \text{a.s.}$$

Since $\alpha^A - \alpha^B > 0$, this implies

$$\ln\left(\frac{N^A(T)}{N^B(T)}\right) \to \infty \quad \text{a.s.}$$

Therefore, $N^A(T)/N^B(T) \to \infty$ almost surely as $T \to \infty$. This completes the proof.

A.2 Proof of Proposition 1

Since u is bounded, there exist $a, b \in \mathbb{R}$ such that $u(Z) \subset [a, b]$. The following two lemmas provide key steps in our construction.

Lemma 2. *Suppose $\varphi : [0,1] \to [0,1]$ is nondecreasing, concave, and onto. Define a function $W : \triangle([a,b]) \to \mathbb{R}$ by*

$$W(\eta) = \int_a^b x \, d(\varphi \circ F_\eta)(x),$$

where $F_\eta(x) = \eta([a,x])$ is the cumulative distribution function for the measure η. Then, there exists a set Φ of nondecreasing and concave continuous functions $\phi : [a,b] \to \mathbb{R}$ such that

$$W(\eta) = \sup_{\phi \in \Phi} \int_Z \phi(z) \, d\eta(z).$$

Proof. It can be shown that W is convex using similar arguments to those in Section S.2.1 of the Supplementary Material of Sarver (2018) (alternatively, see Wakker (1994) or Chatterjee and Krishna (2011)). It is also not difficult to show that W is continuous in the topology of weak convergence. Finally, since φ is concave, the function W respects second-order stochastic dominance by Theorem 2 in Yaari (1987).[33] In light of these conditions, we can apply Proposition 1 from Sarver (2018) to obtain a set Φ with the claimed properties. ■

[33] This was also proved by Chew, Karni, and Safra (1987) in the special case where φ is Lipschitz continuous.

Lemma 3. *Fix a set Ψ of functions $\psi : Z \to [-\infty, \infty)$ that is pointwise bounded above. Then, for any $\mu \in \triangle(S)$ and any simple act $f : S \to Z$,*

$$\sup_{\psi \in \Psi} \int_S \psi(f(s))\, d\mu(s) = \sup_{\psi \in \mathrm{cl}(\Psi)} \int_S \psi(f(s))\, d\mu(s),$$

where the closure is taken with respect to the product topology (i.e., the topology of pointwise convergence) on $[-\infty, \infty]^Z$.

Proof. Fix any $\mu \in \triangle(S)$ and any simple act $f : S \to Z$. Since f is a simple act, there exists a finite partition $\mathcal{E} \subset \mathcal{B}_S$ such that f is measurable with respect to $\mathcal{E}$. For each $E \in \mathcal{E}$, let $z_E = f(s)$ for some $s \in E$. Since f is $\mathcal{E}$-measurable, the value z_E does not depend on the exact choice of $s \in E$. Define a function $G : [-\infty, \infty)^Z \to \mathbb{R}$ by

$$G(\psi) = \int_S \psi(f(s))\, d\mu(s) = \sum_{E \in \mathcal{E}} \psi(z_E)\, \mu(E),$$

and let $\gamma = \sup_{\psi \in \Psi} G(\psi)$. Note that γ is finite since the functions in Ψ are pointwise bounded above. Now, fix any $\psi \in \mathrm{cl}(\Psi)$. By the definition of the closure, there exists a net $(\psi_\alpha)_{\alpha \in D}$ in Ψ that converges to ψ.[34] Note that since $\psi_\alpha \in \Psi$ for each α, we must have $G(\psi_\alpha) \leq \gamma$. Since convergence is preserved under scalar multiples and finite sums, $\psi_\alpha \to \psi$ implies that $G(\psi_\alpha) \to G(\psi)$ and hence $G(\psi) \leq \gamma$. Since this is true for all $\psi \in \mathrm{cl}(\Psi)$, we have

$$\sup_{\psi \in \mathrm{cl}(\Psi)} \int_S \psi(f(s))\, d\mu(s) = \sup_{\psi \in \mathrm{cl}(\Psi)} G(\psi) = \gamma,$$

as desired. ■

Proof of Proposition 1. Take Φ as in Lemma 2 for the function φ, and let $\Psi = \{\phi \circ u : \phi \in \Phi\}$. Fix any $\mu \in \triangle(S)$ and any simple act $f : S \to Z$, and let η be the distribution of utility values induced by μ, f, and u. Formally,

$$\eta = \mu \circ f^{-1} \circ u^{-1} \in \triangle([a, b]).$$

Then, we have

$$\begin{aligned}
\sup_{\psi \in \Psi} \int_S \psi(f(s))\, d\mu(s) &= \sup_{\phi \in \Phi} \int_S \phi(u(f(s)))\, d\mu(s) \\
&= \sup_{\phi \in \Phi} \int_a^b \phi(x)\, d\eta(x) && \text{(change of variables)} \\
&= \int_a^b x\, d(\varphi \circ F_\eta)(x) && \text{(Lemma 2)} \\
&= \int_Z u(z)\, d(\varphi \circ F_{f,\mu})(z).
\end{aligned}$$

[34]It is well known that the product topology on an uncountable product space cannot be completely described by sequential convergence, as such spaces are not metrizable. Hence, we must use nets.

The last equality is essentially another application of the change of variables formula, but there are a few subtleties. One needs to show that if ν^u is the probability measure over utility values with cumulative distribution function $\varphi \circ F_\eta$ and if ν^z is the probability measure over outcomes in Z with cumulative distribution function $\varphi \circ F_{f,\mu}$, then $\nu^u = \nu^z \circ u^{-1}$. This is not true for arbitrary u, but it can be shown to hold whenever u is nondecreasing.

Note that since $W(\eta) = x$ when $\eta(\{x\}) = 1$, we must have $\phi(x) \leq x$ for all $x \in [a, b]$ and $\phi \in \Phi$. Now, for any $\psi \in \Psi$ there exists $\phi \in \Phi$ such that $\psi = \phi \circ u$. Thus, $\psi(z) = \phi(u(z)) \leq b$ for all $z \in Z$, so the set Ψ is bounded above. Moreover, taking the closure of Ψ does not alter the values in the equality above by Lemma 3, so we can assume that Ψ is closed without loss of generality. ■

A.3 Proof of Proposition 2

Some basic definitions and results from functional analysis will be used frequently in this proof. If X is a Banach space, we use X^* to denote the space of all continuous linear functionals on X (the norm dual of X). For $x \in X$ and $x^* \in X^*$, we use $\langle x^*, x\rangle$ to denote the duality pairing $x^*(x)$.

Given a function $F : X \to (-\infty, \infty]$, the *effective domain* of F is the set

$$\operatorname{dom}(F) = \{x \in X : F(x) < \infty\}.$$

The function F is *proper* if $\operatorname{dom}(F) \neq \emptyset$, that is, if it is not identically equal to ∞. The *(Fenchel) conjugate* of F is the function $F^* : X^* \to [-\infty, \infty]$ defined by

$$F^*(x^*) = \sup_{x \in X} \big[\langle x^*, x\rangle - F(x)\big]. \tag{9}$$

Note that if F is proper, then $F^*(x^*) > -\infty$ for all $x^* \in X^*$. Finally, given a set $C \subset X$, we define δ_C by $\delta_C(x) = 0$ if $x \in C$ and $\delta_C(x) = \infty$ if $x \notin C$. This is the indicator function commonly used in functional analysis. Note that

$$(\delta_C)^*(x^*) = \sup_{x \in C} \langle x^*, x\rangle.$$

In this proof, we will work with the L^1 and L^∞ spaces of functions. That is, given a probability space $(\Omega, \mathcal{B}_\Omega, p)$, the space $L^1(\Omega, \mathcal{B}_\Omega, p)$ is the set of all (equivalence classes of) integrable functions, and the space $L^\infty(\Omega, \mathcal{B}_\Omega, p)$ is the set of all (equivalence classes of) essentially bounded functions.[35] When the reference probability space is understood, we will sometimes denote these spaces simply as L^1 and L^∞, respectively. It is a standard result that these are Banach spaces (when endowed with the L^1 and L^∞ norms, respectively) and that $(L^1)^* = L^\infty$, with the duality pairing

$$\langle X, Y\rangle = \int_\Omega X(\omega)Y(\omega)\, dp(\omega)$$

for $Y \in L^1$, $X \in L^\infty$.

[35] Note that in this proof we use Ω to denote an arbitrary probability space, not necessarily the space of common states as in the main text.

Proposition 3. *Fix any probability space $(\Omega, \mathcal{B}_\Omega, p)$. Let $D_\phi(\cdot \,\|\, \cdot)$ be a ϕ–divergence, and fix any nondecreasing, convex, and lower semicontinuous function $k : \mathbb{R} \to (-\infty, \infty]$ such that $k(0) = 0$ and k is finite on some interval $(-\varepsilon, \varepsilon)$. Then, for any random variable $X \in L^\infty(\Omega, \mathcal{B}_\Omega, p)$,*

$$\inf_{q \in \triangle(\Omega)} \left[\int_\Omega X(\omega)\, dq(\omega) + k(D_\phi(q \,\|\, p)) \right] = \max_{\gamma \in \mathbb{R}} \max_{\alpha \geq 0} \int_\Omega \psi_{\gamma,\alpha}(X(\omega))\, dp(\omega),$$

where $\psi_{\gamma,\alpha} : \mathbb{R} \to [-\infty, \infty)$ is defined for $\gamma \in \mathbb{R}$ and $\alpha \geq 0$ by[36]

$$\begin{aligned}
\psi_{\gamma,\alpha}(x) &= \gamma - (\alpha\phi)^*(\gamma - x) - k^*(\alpha) \\
&= \begin{cases} \gamma - \alpha\phi^*\left(\frac{\gamma - x}{\alpha}\right) - k^*(\alpha) & \text{if } \alpha > 0 \\ \gamma + \sup(\mathrm{dom}(\phi)) \cdot (x - \gamma) - k^*(0) & \text{if } \alpha = 0 \text{ and } x < \gamma \\ \gamma + \inf(\mathrm{dom}(\phi)) \cdot (x - \gamma) - k^*(0) & \text{if } \alpha = 0 \text{ and } x \geq \gamma. \end{cases}
\end{aligned}$$

Note that in Proposition 2, we took k to be a function from $\mathbb{R}_+$ to $[0, \infty]$; however, we can treat k as a nondecreasing, convex, and lower semicontinuous function from $\mathbb{R}$ into $[0, \infty]$ by taking $k(x) = 0$ for $x < 0$. Similarly, our definition of a divergence requires ϕ to be a continuous convex function mapping from $\mathbb{R}_+$ to $\mathbb{R}_+$, but we can treat ϕ as lower semicontinuous convex function defined on all of $\mathbb{R}$ by taking $\phi(y) = \infty$ for $y < 0$, and hence

$$\phi^*(x) = \sup_{y \in \mathbb{R}_+} \big[xy - \phi(y) \big].$$

Proposition 2 then follows as a special case of Proposition 3 where the state space is $\widehat{\Omega} = S$, the probability measure is $p = \mu \in \triangle(S)$, and $X : S \to \mathbb{R}$ is defined by

$$X(s) = u(f(s)).$$

Note that since f is a simple act and u is real-valued, X is bounded. Thus, by Proposition 3,

$$\inf_{\eta \in \triangle(S)} \left[\int_S u(f(s))\, d\eta(s) + k(D_\phi(\eta \,\|\, \mu)) \right] = \max_{\gamma \in \mathbb{R}} \max_{\alpha \geq 0} \int_S \psi_{\gamma,\alpha}(u(f(s)))\, d\mu(s).$$

Take Ψ to be the closure of the set

$$\{\psi_{\gamma,\alpha} \circ u : \gamma \in \mathbb{R}, \alpha \geq 0\},$$

where the closure is taken with respect to the topology of pointwise convergence on the extended reals. Then, Ψ satisfies Assumption 1 and the arguments above together with Lemma 3 (which allows us to take the closure) establish that the equality in the statement of the proposition holds.

Therefore, all that remains is to prove Proposition 3. Our proof will be based on the following

[36] Recall that we require the function ϕ in the definition of a divergence to be finite on some interval $[\alpha, \beta]$ where $\alpha < 1 < \beta$, and hence $\sup(\mathrm{dom}(\phi)) > 1$ and $\inf(\mathrm{dom}(\phi)) < 1$, so $\psi_{\gamma,\alpha}$ for $\alpha = 0$ is a piecewise linear and concave gain-loss function. In particular, if ϕ is finite-valued on all of $\mathbb{R}_+$, then $\sup(\mathrm{dom}(\phi)) = \infty$ and $\inf(\mathrm{dom}(\phi)) = 0$.

three lemmas. The first two lemmas closely parallel the proof strategy used by Ben-Tal and Teboulle (2007, Theorem 4.2) who provide a similar result for the case when $k(x) = x$, that is, when there is no transformation of the divergence term.

Lemma 4. *Fix any probability space $(\Omega, \mathcal{B}_\Omega, p)$. Let $H : L^1 \to (-\infty, \infty]$ be a convex and lower semicontinuous function, and suppose there exist $\alpha < 1 < \beta$ such that $Y \in L^1$ and $\alpha \leq Y(\omega) \leq \beta$ for all $\omega \in \Omega$ implies $H(Y) < \infty$. Then, for any $X \in L^\infty$,*

$$\inf_{\substack{Y \in L^1: \\ \int Y(\omega)\, dp(\omega) = 1}} \left[\int_\Omega X(\omega) Y(\omega)\, dp(\omega) + H(Y) \right] = \max_{\gamma \in \mathbb{R}} \left[\gamma - H^*(\gamma - X) \right]$$

Proof. The proof of this result replicates the first steps in the proof of Theorem 4.2 in Ben-Tal and Teboulle (2007), but we include it for completeness. Denote by v the value of the left side of the equation in the statement of the lemma:

$$v \equiv \inf_{\substack{Y \in L^1: \\ \int Y(\omega)\, dp(\omega) = 1}} \left[\int_\Omega X(\omega) Y(\omega)\, dp(\omega) + H(Y) \right].$$

The Lagrangian dual of this convex minimization problem is given by

$$\begin{aligned} w &\equiv \sup_{\gamma \in \mathbb{R}} \inf_{Y \in L^1} \left[\int_\Omega X(\omega) Y(\omega)\, dp(\omega) + H(Y) + \gamma \left(1 - \int_\Omega Y(\omega)\, dp(\omega) \right) \right] \\ &= \sup_{\gamma \in \mathbb{R}} \left[\gamma + \inf_{Y \in L^1} \left(H(Y) + \int_\Omega (X(\omega) - \gamma) Y(\omega)\, dp(\omega) \right) \right] \\ &= \sup_{\gamma \in \mathbb{R}} \left[\gamma - \sup_{Y \in L^1} \left(\int_\Omega (\gamma - X(\omega)) Y(\omega)\, dp(\omega) - H(Y) \right) \right] \\ &= \sup_{\gamma \in \mathbb{R}} \left[\gamma - H^*(\gamma - X) \right]. \end{aligned}$$

It remains only to show that $v = w$, that is, there is no duality gap. The convex duality result in Corollary 4.8 of Borwein and Lewis (1992) shows that there is no duality gap and there is attainment of a solution in the dual problem if the following constraint qualification condition is satisfied:[37]

(CQ) There exist $\alpha < \beta$ such that $\alpha \leq Y(\omega) \leq \beta$ implies $H(Y) < \infty$, and there exists some $Y \in L^1$ with $\alpha < Y(\omega) < \beta$ that satisfies the constraint $\int_\Omega Y(\omega)\, dp(\omega) = 1$.

Given the assumptions in the statement of the lemma, this condition is satisfied by taking Y identically equal to 1. This completes the proof. ∎

[37]Borwein and Lewis (1992) define the *quasi relative interior* of a set C to be the set of all points $x \in C$ such that the closure of the cone generated by $C - x$ is a subspace. In the context of our minimization problem, their constraint qualification condition requires that there is a function Y in the quasi relative interior of the set $\operatorname{dom}(H) \equiv \{Y \in L^1 : H(Y) < \infty\}$ that satisfies the constraint $\int_\Omega Y(\omega)\, dp(\omega) = 1$. It can be shown that if $\{Y \in L^1 : \alpha \leq Y \leq \beta\} \subset \operatorname{dom}(H)$ then any $Y \in L^1$ with $\alpha < Y(\omega) < \beta$ is in the quasi relative interior of $\operatorname{dom}(H)$ (see Example 3.11(i) in Borwein and Lewis (1992)).

Lemma 5. *Fix any probability space $(\Omega, \mathcal{B}_\Omega, p)$, and fix any proper convex and lower semicontinuous function $\phi : \mathbb{R} \to (-\infty, \infty]$. Define a functional $J : L^1 \to (-\infty, \infty]$ by*

$$J(Y) = \int_\Omega \phi(Y(\omega))\, dp(\omega).$$

Then, J is a proper convex and lower semicontinuous functional, and the Fenchel conjugate $J^* : L^\infty \to (-\infty, \infty]$ of J is given by

$$J^*(X) = \int_\Omega \phi^*(X(\omega))\, dp(\omega).$$

Proof. See the corollary to Theorem 2 in Rockafellar (1968). ■

Fix any proper convex and lower semicontinuous function $\phi : \mathbb{R} \to (-\infty, \infty]$ that is finite on an open interval containing 1. Then, defining J as in Lemma 5 and setting $H = J$ in Lemma 4, we obtain the following dual formula:

$$\inf_{\substack{Y \in L^1: \\ \int Y(\omega)\, dp(\omega) = 1}} \left[\int_\Omega X(\omega) Y(\omega)\, dp(\omega) + J(Y) \right] = \max_{\gamma \in \mathbb{R}} \int_\Omega \big[\gamma - \phi^*(\gamma - X(\omega)) \big]\, dp(\omega).$$

This is precisely Theorem 4.2 in Ben-Tal and Teboulle (2007). To extend their result to $H = k \circ J$, we need the following lemma.

Lemma 6. *Fix any probability space $(\Omega, \mathcal{B}_\Omega, p)$, and fix any convex and lower semicontinuous function $\phi : \mathbb{R}_+ \to [0, \infty]$ such that $\phi(1) = 0$ and there exists some $\alpha < 1 < \beta$ such that ϕ is finite on the interval $[\alpha, \beta]$. Also, fix any nondecreasing, convex, and lower semicontinuous function $k : \mathbb{R} \to (-\infty, \infty]$ such that k is finite on some interval $(-\varepsilon, \varepsilon)$. Define $J : L^1 \to (-\infty, \infty]$ by*

$$J(Y) = \int_\Omega \phi(Y(\omega))\, dp(\omega),$$

and define $H : L^1 \to (\infty, \infty]$ by $H = k \circ J$. Then, for any $X \in L^\infty$,

$$H^*(X) = \min_{\alpha \geq 0} \big[(\alpha J)^*(X) + k^*(\alpha) \big], \tag{10}$$

where

$$(\alpha J)^*(X) = \int_\Omega (\alpha \phi)^*(X(\omega))\, dp(\omega)$$

and where $(\alpha\phi)^*(x) = \alpha \phi^*(\frac{x}{\alpha})$ for $\alpha > 0$ and

$$(0\phi)^*(x) = \begin{cases} \inf(\operatorname{dom}(\phi)) \cdot x & \text{if } x \leq 0 \\ \sup(\operatorname{dom}(\phi)) \cdot x & \text{if } x > 0. \end{cases}$$

Proof. To obtain the formula for the conjugate of the composition of two functions, we appeal to

Theorem 2 of Hiriart-Urruty (2006):[38] Since k and J are both lower semicontinuous and convex, k is nondecreasing, and there exists a function $Y \in L^1$ such that $J(Y) \in \text{int}(\text{dom}(k))$ (namely, Y identically equal to 1), his theorem implies that the Fenchel conjugate of $k \circ J$ is given by Equation (10), when one sets $(0J) = \delta_{\text{dom}(J)}$. For $\alpha > 0$, we therefore have

$$(\alpha J)^*(X) = \int_\Omega (\alpha\phi)^*(X(\omega))\, dp(\omega) = \int_\Omega \alpha\phi^*\left(\frac{X(\omega)}{\alpha}\right) dp(\omega),$$

where the first equality follows from Lemma 5 and the second equality follows directly from the definition of the conjugate.

It remains only to establish the formula for $(0J)^*$. By the definition of the conjugate,

$$(0J)^*(X) = \sup_{Y \in L^1} \left[\langle X, Y\rangle - \delta_{\text{dom}(J)}(Y)\right] = \sup_{Y \in \text{dom}(J)} \int_\Omega X(\omega)Y(\omega)\, dp(\omega).$$

Now, fix any $X \in L^\infty$ and let $E = \{\omega \in \Omega : X(\omega) > 0\}$. Note that a necessary (but not sufficient) condition for $Y \in \text{dom}(J)$ is that $0 \le \inf(\text{dom}(\phi)) \le Y \le \sup(\text{dom}(\phi)) \le \infty$ almost surely. Therefore,

$$\sup_{Y \in \text{dom}(J)} \int_\Omega X(\omega)Y(\omega)\, dp(\omega) \le \int_\Omega X(\omega)\widehat{Y}(\omega)\, dp(\omega), \tag{11}$$

where $\widehat{Y} : \Omega \to [0, \infty]$ is defined by

$$\widehat{Y}(\omega) = \begin{cases} \inf(\text{dom}(\phi)) & \text{if } \omega \notin E \\ \sup(\text{dom}(\phi)) & \text{if } \omega \in E. \end{cases}$$

Note that the integral on the right is well-defined since X is bounded and $0 \le \inf(\text{dom}(\phi)) \le 1$, but possibly infinite since we could have $\sup(\text{dom}(\phi)) = \infty$. The proof is completed by showing that the supremum attains this bound, so that Equation (11) holds equality (with both sides possibly being $+\infty$). Since it may be that $\widehat{Y} \notin \text{dom}(J)$ (e.g., if $\phi(y) = \infty$ for $y = \inf(\text{dom}(\phi))$ or $y = \sup(\text{dom}(\phi))$), we will approximate $\widehat{Y}$ using a sequence: Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence in $\text{dom}(J)$ defined by

$$Y_n(\omega) = \begin{cases} \underline{y}_n & \text{if } \omega \notin E \\ \overline{y}_n & \text{if } \omega \in E, \end{cases}$$

where $(\underline{y}_n)_{n\in\mathbb{N}}$ is a monotonically decreasing sequence in $\text{dom}(\phi)$ with $\underline{y}_n \to \inf(\text{dom}(\phi))$, and $(\overline{y}_n)_{n\in\mathbb{N}}$ is a monotonically increasing sequence in $\text{dom}(\phi)$ with $\overline{y}_n \to \sup(\text{dom}(\phi))$ (with the standard convention that $\overline{y}_n$ diverges to $+\infty$ in the case of $\sup(\text{dom}(\phi)) = \infty$). Notice that by construction, the function Z_n defined by $Z_n(\omega) = X(\omega)Y_n(\omega)$ is bounded, and $Z_n \uparrow Z$ where $Z(\omega) = X(\omega)\widehat{Y}(\omega)$. Therefore, by the monotone convergence theorem (e.g., Theorem 4.3.2 in

[38] Hiriart-Urruty (2006) provides a concise treatment of this problem, but earlier, more general results about conjugates of compositions of convex functions exist, e.g., Kutateladze (1979, Theorem 3.7.1) or Combari, Laghdir, and Thibault (1996, Theorem 3.4(ii)).

Dudley (2002)), we have

$$\int_\Omega X(\omega)Y_n(\omega)\,dp(\omega) \to \int_\Omega X(\omega)\widehat{Y}(\omega)\,dp(\omega),$$

as desired. ■

Proof of Proposition 3. Note that $D_\phi(q\,\|\,p) = \infty$ whenever q is not absolutely continuous with respect to p. Thus, we can restrict attention to $q \ll p$, and we can therefore express the divergence using Radon–Nikodym derivatives $Y = \frac{dq}{dp} \in L^1(\Omega, \mathcal{B}_\Omega, p)$:

$$\begin{aligned}
&\inf_{q\in\triangle(\Omega)} \left[\int_\Omega X(\omega)\,dq(\omega) + k(D_\phi(q\,\|\,p))\right] \\
&\quad = \inf_{q\ll p} \left[\int_\Omega X(\omega)\frac{dq}{dp}(\omega)\,dp(\omega) + k\bigg(\int_\Omega \phi\Big(\frac{dq}{dp}(\omega)\Big)dp(\omega)\bigg)\right] \\
&\quad = \inf_{\substack{Y\in L^1:\\ \int Y(\omega)\,dp(\omega)=1}} \left[\int_\Omega X(\omega)Y(\omega)\,dp(\omega) + k\bigg(\int_\Omega \phi(Y(\omega))\,dp(\omega)\bigg)\right].
\end{aligned}$$

Note that for $Y \in L^1$ to be a Radon-Nikodym derivative, we must have $\int_\Omega Y(\omega)\,dp(\omega) = 1$ and $Y \geq 0$ a.s. The first constraint is stated explicitly in the equation above, and since $\phi(y) = \infty$ for $y < 0$, the second constraint becomes superfluous.

As before, define $J : L^1 \to (-\infty, \infty]$ by

$$J(Y) = \int_\Omega \phi(Y(\omega))\,dp(\omega),$$

and define $H : L^1 \to (\infty, \infty]$ by $H = k \circ J$. Note that J is convex and lower semicontinuous by Lemma 5, and therefore H is convex and lower semicontinuous given our assumptions on k. We also assumed that there is an interval $(-\varepsilon, \varepsilon)$ on which k is finite. Since $\phi : \mathbb{R}_+ \to [0, \infty]$ is convex and finite on some interval $[\alpha, \beta]$ for $\alpha < 1 < \beta$, it is necessarily continuous on (α, β). Therefore, since $\phi(1) = 0$, there exists some $\alpha' < 1 < \beta'$ such that $\alpha' \leq y \leq \beta'$ implies $0 \leq \phi(y) < \varepsilon$. Thus, $\alpha \leq Y(\omega) \leq \beta$ for all $\omega \in \Omega$ implies $0 \leq J(Y) < \varepsilon$ and hence $H(Y) < \infty$. Therefore,

$$\begin{aligned}
&\inf_{\substack{Y\in L^1:\\ \int Y(\omega)\,dp(\omega)=1}} \left[\int_\Omega X(\omega)Y(\omega)\,dp(\omega) + k\bigg(\int_\Omega \phi(Y(\omega))\,dp(\omega)\bigg)\right] \\
&\qquad = \max_{\gamma\in\mathbb{R}} \left[\gamma - H^*(\gamma - X)\right] \\
&\qquad = \max_{\gamma\in\mathbb{R}} \max_{\alpha\geq 0} \left[\gamma - (\alpha J)^*(\gamma - X) - k^*(\alpha)\right],
\end{aligned}$$

where the first equality follows from Lemma 4 and the second equality follows from Lemma 6. Then, using the formula for $(\alpha J)^*$ from Lemma 6, we have that for any $X \in L^\infty$, $\gamma \in \mathbb{R}$, and $\alpha \geq 0$,

$$\begin{aligned}
\gamma - (\alpha J)^*(\gamma - X) - k^*(\alpha) &= \gamma - \int_\Omega (\alpha\phi)^*(\gamma - X(\omega))\, dp(\omega) - k^*(\alpha) \\
&= \int_\Omega \big[\gamma - (\alpha\phi)^*(\gamma - X(\omega)) - k^*(\alpha)\big]\, dp(\omega),
\end{aligned}$$

where

$$\gamma - (\alpha\phi)^*(\gamma - x) - k^*(\alpha) = \begin{cases} \gamma - \alpha\phi^*\big(\frac{\gamma - x}{\alpha}\big) - k^*(\alpha) & \text{if } \alpha > 0 \\ \gamma - \sup(\mathrm{dom}(\phi)) \cdot (\gamma - x) - k^*(0) & \text{if } \alpha = 0 \text{ and } \gamma - x > 0 \\ \gamma - \inf(\mathrm{dom}(\phi)) \cdot (\gamma - x) - k^*(0) & \text{if } \alpha = 0 \text{ and } \gamma - x \leq 0, \end{cases}$$

where is precisely the formula for $\psi_{\gamma,\alpha}(x)$ from the statement of the proposition. This completes the proof. ■

A.4 Proof of Claim 2

Define ϕ by $\phi(y) = y\ln(y) - y + 1$ for $y > 0$ and $\phi(0) = 1$, so that the ϕ–divergence is precisely relative entropy: $D_\phi(\eta \,\|\, \mu) = R(\eta \,\|\, \mu)$. It is standard that the Fenchel conjugate of ϕ (see Equation (9)) is $\phi^* : \mathbb{R} \to \mathbb{R}$ defined by

$$\phi^*(x) = e^x - 1.$$

Therefore, Proposition 3 gives the equality stated in Claim 2 for precisely the class of parameterized fitness functions defined in Example 3 when $c(\alpha) = k^*(\alpha)$. Note also that:

1. If $k(x) = 0$ for $x \leq \kappa$ and $k(x) = \infty$ otherwise, then $k^*(\alpha) = \kappa\alpha$ for all $\alpha \geq 0$.
2. If $k(x) = \theta x$ for all $x \geq 0$ (and $k(x) = 0$ otherwise), then $k^*(\alpha) = 0$ if $0 \leq \alpha \leq \theta$ and $k^*(\alpha) = \infty$ if $\alpha > \theta$.

A.5 Proof of Theorem 2

The following proposition will be central to our first step in the proof of Theorem 2. Given some $p \in \triangle(\Omega)$, recall that $M(p) = \{q \in \triangle(\Omega) : q \ll p \text{ and } R(p \,\|\, q) < \infty\}$. In particular, since $R(p \,\|\, q) < \infty$ requires that $p \ll q$, the measures q and p are mutually absolutely continuous whenever $q \in M(p)$.

Proposition 4. *Fix any measurable space $(\Omega, \mathcal{B}_\Omega)$. Suppose $X : \Omega \to [-\infty, \infty)$ is measurable and bounded above, and let $p \in \triangle(\Omega)$. Then,*

$$\int_\Omega \ln(X(\omega))\, dp(\omega) = \inf_{q \in M(p)} \left[\ln\left(\int_\Omega X(\omega)\, dq(\omega)\right) + R(p \,\|\, q)\right]. \tag{12}$$

In addition, if X is bounded away from zero, that is, if $X(\omega) \geq \varepsilon > 0$ for all $\omega \in \Omega$, then the infimum in Equation (12) is uniquely attained by the measure q_0 with Radon–Nikodym derivative

$$\frac{dq_0}{dp}(\omega) = \frac{1}{X(\omega) \displaystyle\int_\Omega \frac{1}{X(\hat{\omega})}\, dp(\hat{\omega})}.$$

Proposition 4 restricts to $q \in M(p)$, thereby ensuring that we do not encounter terms of the form $-\infty + \infty$. That is, while the first term inside the infimum in Equation (12) could take the value $-\infty$, the second term $R(p \,\|\, q)$ will necessarily be finite. Proposition 4 is based on dual formulas for relative entropy that are related to those commonly invoked in the theory of large deviations (e.g., Dupuis and Ellis (1997)), although the result itself is distinct from any known results of which we are aware. The complete proof of the proposition is relegated to Section S2 of the Supplementary Appendix.

Using Proposition 4, the following lemma provides the first step in our proof of Theorem 2.

Lemma 7. *Suppose Ψ is a nonempty set of functions $\psi : Z \to [-\infty, \infty)$ that is pointwise bounded above, and fix $\mu \in \triangle(\Omega \times S)$. For any random action $\rho \in \triangle_s(\mathcal{F})$, the function V defined by Equation (2) can be equivalently expressed as*

$$V(\rho) = \sup_{\tau \in \mathcal{R}(\Psi|\mathcal{F})} \inf_{q \in M(\mu_\Omega)} \left[\ln\left(\int_\Omega \int_S \mathbb{E}_{\tau \otimes \rho}\big[\psi(f(\omega, s))\big]\, d\mu(s|\omega)\, dq(\omega) \right) + R(\mu_\Omega \,\|\, q) \right].$$

Proof. For a given $\rho \in (\triangle_s(\mathcal{F}))$ and $\tau \in \mathcal{R}(\Psi|\mathcal{F})$, define $X : \Omega \to [-\infty, \infty)$ by

$$X(\omega) = \int_S \mathbb{E}_{\tau \otimes \rho}\big[\psi(f(\omega, s))\big]\, d\mu(s|\omega).$$

To verify that X is bounded above, recall that $\rho \in \triangle_s(\mathcal{F})$ has finite support and each $f \in \text{supp}(\rho)$ is a simple act. This implies that only finitely many realizations of z occur with positive probability. Since the set Ψ is pointwise bounded above, this implies that there exists $\kappa \in \mathbb{R}$ such that $\psi(f(\omega, s)) \leq \kappa$ for all $\omega \in \Omega$, $s \in S$, $\psi \in \Psi$ and $f \in \text{supp}(\rho)$. Therefore, $X(\omega) \leq \kappa$ for all ω. Applying Proposition 4 to this function, we obtain

$$\begin{aligned}
&\int_\Omega \ln\left(\int_S \mathbb{E}_{\tau \otimes \rho}\big[\psi(f(\omega, s))\big]\, d\mu(s|\omega) \right) d\mu(\omega) \\
&\quad = \int_\Omega \ln(X(\omega))\, d\mu_\Omega(\omega) \\
&\quad = \inf_{q \in M(\mu_\Omega)} \left[\ln\left(\int_\Omega X(\omega)\, dq(\omega) \right) + R(\mu_\Omega \,\|\, q) \right] \\
&\quad = \inf_{q \in M(\mu_\Omega)} \left[\ln\left(\int_\Omega \int_S \mathbb{E}_{\tau \otimes \rho}\big[\psi(f(\omega, s))\big]\, d\mu(s|\omega)\, dq(\omega) \right) + R(\mu_\Omega \,\|\, q) \right].
\end{aligned}$$

Thus, when V is defined by Equation (2), we have

$$\begin{aligned} V(\rho) &= \sup_{\tau\in\mathcal{R}(\Psi|\mathcal{F})} \int_\Omega \ln\biggl(\int_S \mathbb{E}_{\tau\otimes\rho}\bigl[\psi(f(\omega,s))\bigr]\,d\mu(s|\omega)\biggr)d\mu(\omega) \\ &= \sup_{\tau\in\mathcal{R}(\Psi|\mathcal{F})}\ \inf_{q\in M(\mu_\Omega)} \biggl[\ln\biggl(\int_\Omega\int_S \mathbb{E}_{\tau\otimes\rho}\bigl[\psi(f(\omega,s))\bigr]\,d\mu(s|\omega)\,dq(\omega)\biggr) + R(\mu_\Omega \,\|\, q)\biggr]. \end{aligned}$$

This completes the proof. ■

The next proposition will be central to the second step in the proof of Theorem 2.

Proposition 5. *Fix a measure $\mu \in \triangle(\Omega\times S)$, and suppose Ξ is a nonempty set of functions $\xi : \Omega\times S \to [-\infty,\infty)$ with the following properties:*

1. Closedness: *When the set of extend reals $[-\infty,\infty]$ is endowed with its usual topology and $[-\infty,\infty]^{\Omega\times S}$ is endowed with the product topology (i.e., the topology of pointwise convergence), Ξ is a closed subset of this space.*
2. Finite measurability: *There exists a finite partition $\mathcal{E}\subset\mathcal{B}_\Omega\otimes\mathcal{B}_S$ of $\Omega\times S$ such that every $\xi\in\Xi$ is measurable with respect to $\mathcal{E}$.*
3. Pointwise boundedness: $\sup_{\xi\in\Xi}\xi(\omega,s)<\infty$ *for every $(\omega,s)\in\Omega\times S$.*

Then,

$$\begin{aligned} &\sup_{\xi\in\mathrm{co}(\Xi)}\ \inf_{q\in M(\mu_\Omega)} \biggl[\ln\biggl(\int_\Omega\int_S \xi(\omega,s)\,d\mu(s|\omega)\,dq(\omega)\biggr) + R(\mu_\Omega\,\|\,q)\biggr] \\ &= \inf_{q\in M(\mu_\Omega)} \biggl[\ln\biggl(\sup_{\xi\in\Xi}\int_\Omega\int_S \xi(\omega,s)\,d\mu(s|\omega)\,dq(\omega)\biggr) + R(\mu_\Omega\,\|\,q)\biggr]. \end{aligned}$$

Proposition 5 is based on an application of an extension of the von Neumann–Sion Minimax Theorem due to Tuy (2004). Despite the reliance on these established tools and techniques, the complete proof of this proposition is quite involved and is therefore relegated to Section S3 of the Supplementary Appendix.

Proceeding with the proof of Theorem 2, fix any $\rho\in\triangle_s(\mathcal{F})$, and let $B=\mathrm{supp}(\rho)$. Since ρ is a simple lottery over acts, B is a finite set of acts. We will define Ξ to be the set of individual expected fitness functions that are attainable given the fixed random choice of act under the random action ρ together with some deterministic adaptation plan. That is, we are focusing for now on adaptations plans τ that place probability one on some fitness function $\psi_f\in\Psi$ following each $f\in B$.

Formally, deterministic adaptation plans are denoted by $(\psi_f)_{f\in B}\in\Psi^B$, or (ψ_f) for short.[39] Define a mapping $J:\Psi^B\to[-\infty,\infty]^{\Omega\times S}$ by

$$J\Bigl[(\psi_{\hat f})_{\hat f\in B}\Bigr](\omega,s) = \int_B \psi_f(f(\omega,s))\,d\rho(f) \tag{13}$$

[39]Note that since (ψ_f) is an element of Ψ^B rather than $\Psi^{\mathcal{F}}$, the value of ψ_f is unspecified for $f\in\mathcal{F}\setminus B$. However, since acts $f\notin B$ are chosen with probability zero, expected individual fitness is fully determined by the values of ψ_f for $f\in B$.

for $(\omega, s) \in \Omega \times S$. Define Ξ to be the range of J, that is,

$$\Xi = \Big\{ J[(\psi_f)] \in [-\infty, \infty]^{\Omega \times S} : (\psi_f) \in \Psi^B \Big\}. \tag{14}$$

In other words, Ξ is the set of all functions ξ that take the form

$$\xi(\omega, s) = \int_B \psi_f(f(\omega, s))\, d\rho(f)$$

for some deterministic adaptation plan $(\psi_f)_{f \in B}$. The next two lemmas show that taking the convex hull of Ξ generates precisely the set of individual expected fitness functions that can be attained through random adaptation plans and that the set Ξ is closed. Indeed, the use of deterministic action plans above was precisely in order to ensure that Ξ is closed. The proofs of these two lemmas are based on standard arguments and are relegated to Sections S4 and S5 of the Supplementary Appendix.

Lemma 8. *Define Ξ as in Equation (14). For any random adaptation plan $\tau \in \mathcal{R}(\Psi|\mathcal{F})$, define $\xi^\tau : \Omega \times S \to [-\infty, \infty)$ by*

$$\xi^\tau(\omega, s) = \mathbb{E}_{\tau \otimes \rho}\big[\psi(f(\omega, s))\big] = \int_{\mathcal{F}} \int_{\Psi} \psi(f(\omega, s))\, d\tau(\psi|f)\, d\rho(f).$$

Then,

$$\mathrm{co}(\Xi) = \big\{ \xi^\tau : \tau \in \mathcal{R}(\Psi|\mathcal{F}) \big\}.$$

Lemma 9. *The set Ξ defined in Equation (14) is a closed subset of $[-\infty, \infty]^{\Omega \times S}$.*

We now verify that the set Ξ defined in Equation (14) satisfies the three conditions from Proposition 5:

- Lemma 9 already showed that this set is closed, which establishes first condition.
- We now show that Ξ satisfies the second condition (finite measurability) from Proposition 5. Since each $f \in \mathcal{F}$ is a simple act, and since the set of acts B in the support of ρ is finite, there exists a finite partition $\mathcal{E} \subset \mathcal{B}_\Omega \otimes \mathcal{B}_S$ of $\Omega \times S$ such that every act $f \in B$ is measurable with respect to $\mathcal{E}$. We claim that every function in Ξ is measurable with respect to $\mathcal{E}$. To see this, fix any $\xi \in \Xi$. Then, there exists $(\psi_f) \in \Psi^B$ such that

  $$\xi(\omega, s) = \int_B \psi_f(f(\omega, s))\, d\rho(f).$$

  Fix any $E \in \mathcal{E}$ and $(\omega, s), (\omega', s') \in E$. By construction of the partition $\mathcal{E}$, we must have $f(\omega, s) = f(\omega', s')$ for any $f \in B = \mathrm{supp}(\rho)$. Therefore,

  $$\xi(\omega, s) = \int_B \psi_f(f(\omega, s))\, d\rho(f) = \int_B \psi_f(f(\omega', s'))\, d\rho(f) = \xi(\omega', s'),$$

  as claimed. Thus, the second condition of Proposition 5 is satisfied.

- To verify the third condition (pointwise boundedness) in Proposition 5, note that since B is a finite set of simple acts, there is a finite set $\widehat{Z} \subset Z$ such that $f(\omega, s) \in \widehat{Z}$ for all $f \in B$ and $(\omega, s) \in \Omega \times S$. Recall that the set Ψ is pointwise bounded above, so $\sup_{\psi\in\Psi} \psi(z) < \infty$ for all $z \in Z$. Therefore, and any $(\omega, s) \in \Omega \times S$,

$$\begin{aligned}\sup_{\xi\in\Xi} \xi(\omega, s) &= \sup_{(\psi_f)\in\Psi^B} \int_B \psi_f(f(\omega, s))\, d\rho(f) \\ &\leq \int_B \sup_{\psi\in\Psi} \psi(f(\omega, s))\, d\rho(f) \leq \max_{z\in\widehat{Z}} \sup_{\psi\in\Psi} \psi(z) < \infty,\end{aligned}$$

where the last inequality follows from the finiteness of $\widehat{Z}$. Thus, Ξ satisfies condition 3.

We are now ready to apply Lemma 7 and Proposition 5. Define V as in Equation (2). Then, we have

$$\begin{aligned}V(\rho) &= \sup_{\tau\in\mathcal{R}(\Psi|\mathcal{F})} \inf_{q\in M(\mu_\Omega)} \left[\ln\left(\int_\Omega \int_S \mathbb{E}_{\tau\otimes\rho}\big[\psi(f(\omega, s))\big]\, d\mu(s|\omega)\, dq(\omega)\right) + R(\mu_\Omega \,\|\, q)\right] \\ &= \sup_{\xi\in\mathrm{co}(\Xi)} \inf_{q\in M(\mu_\Omega)} \left[\ln\left(\int_\Omega \int_S \xi(\omega, s)\, d\mu(s|\omega)\, dq(\omega)\right) + R(\mu_\Omega \,\|\, q)\right] \\ &= \inf_{q\in M(\mu_\Omega)} \left[\ln\left(\sup_{\xi\in\Xi} \int_\Omega \int_S \xi(\omega, s)\, d\mu(s|\omega)\, dq(\omega)\right) + R(\mu_\Omega \,\|\, q)\right] \\ &= \inf_{q\in M(\mu_\Omega)} \left[\ln\left(\sup_{\xi\in\Xi} \int_{\Omega\times S} \xi(\omega, s)\, d(\mu\otimes q)(\omega, s)\right) + R(\mu_\Omega \,\|\, q)\right],\end{aligned}$$

where the first equality follows from Lemma 7, the second from Lemma 8, the third from Proposition 5, and the fourth from the definition of the measure $\mu \otimes q$. Simple manipulations of the term inside the logarithm yield

$$\begin{aligned}&\sup_{\xi\in\Xi} \int_{\Omega\times S} \xi(\omega, s)\, d(\mu\otimes q)(\omega, s) \\ &\quad= \sup_{(\psi_f)\in\Psi^B} \int_{\Omega\times S} \int_B \psi_f(f(\omega, s))\, d\rho(f)\, d(\mu\otimes q)(\omega, s) \\ &\quad= \sup_{(\psi_f)\in\Psi^B} \int_B \int_{\Omega\times S} \psi_f(f(\omega, s))\, d(\mu\otimes q)(\omega, s)\, d\rho(f) \\ &\quad= \int_B \sup_{\psi\in\Psi} \int_{\Omega\times S} \psi(f(\omega, s))\, d(\mu\otimes q)(\omega, s)\, d\rho(f) \\ &\quad= \mathbb{E}_\rho\left[\sup_{\psi\in\Psi} \int_{\Omega\times S} \psi(f(\omega, s))\, d(\mu\otimes q)(\omega, s)\right],\end{aligned}$$

and hence

$$V(\rho) = \inf_{q\in M(\mu_\Omega)} \left[\ln\left(\mathbb{E}_\rho\left[\sup_{\psi\in\Psi} \int_{\Omega\times S} \psi(f(\omega, s))\, d(\mu\otimes q)(\omega, s)\right]\right) + R(\mu_\Omega \,\|\, q)\right].$$

Since this is true for any $\rho \in (\triangle_s(\mathcal{F}))$, the proof is complete.

A.6 Proof of Claims 3 and 4

Proof of Claim 3. Given ψ, let $a := \psi(x)$ and $b = \psi(y)$. Then $\alpha_{\{\psi\}}(p)$ satisfies

$$\ln(\alpha_{\{\psi\}}(p)a + (1-\alpha_{\{\psi\}}(p))b) = p\ln a + (1-p)ln(b)$$
$$\iff$$
$$\alpha_{\{\psi\}}(p) = \frac{a^p b^{1-p} - b}{a-b}$$

Continuous differentiability of $\alpha_{\{\psi\}}(p)$ is then immediate, and with $a > b$ it is straight forward to check that the first and second derivatives of $\alpha_{\{\psi\}}(p)$ are positive, establishing increasingness and convexity. Direct evaluation yields $\alpha_{\{\psi\}}(0) = 0$ and $\alpha_{\{\psi\}}(1) = 1$.

Consider now ψ_1 and ψ_2 with $\psi_2(x) > \psi_1(x) > \psi_1(y) > \psi_2(y)$ as in the claim. To establish that $\alpha_{\{\psi_1\}}(p) > \alpha_{\{\psi_2\}}(p)$ for all $p \in (0,1)$, it suffices to verify that for any strictly increasing ψ with a and b as defined above and for all $p \in (0,1)$

$$\frac{\partial\alpha_{\{\psi\}}(p)}{\partial a} < 0 \text{ and } \frac{\partial\alpha_{\{\psi\}}(p)}{\partial b} > 0.$$

Indeed, with some simple algebra,

$$\frac{\partial\alpha_{\{\psi\}}(p)}{\partial a} < 0$$
$$\iff$$
$$a^{1-p}b^p < (1-p)a + pb$$
$$\iff$$
$$(1-p)\ln a + p\ln b < \ln((1-p)a + pb)$$

which is true by the convexity of the logarithm. That $\frac{\partial\alpha_{\{\psi\}}(p)}{\partial b} > 0$ follows analogously.

■

Proof of Claim 4. For $\Psi = \{\psi_1, \psi_2\}$ with $\psi_2(x) > \psi_1(x) > \psi_1(y) > \psi_2(y)$ as in the claim, let $\alpha_\Psi(p)$ be the risk equivalent of p from optimally choosing $\psi \in \Psi$. It is clear that $\alpha_\Psi(p)$ must be continuous, strictly increasing and that $\alpha_\Psi(0) = 0$ and $\alpha_\Psi(1) = 1$.

To further analyze $\alpha_\Psi(p)$, consider the hypothetical where ψ_i with $i \in \{1,2\}$ is the fitness function used to evaluate the gamble xpy when p is generated from the unknown urn, and ψ_j with $j \in \{1,2\}$ is the one used to evaluate the gamble $x\alpha y$ when α is generated from the known urn. Replicating the derivation in the proof of Claim 3 yields that the risk equivalent of p under those fitness functions is

$$\alpha_{i,j}(p) = \frac{\psi_i(x)^p\psi_i(y)^{1-p} - \psi_j(y)}{\psi_j(x) - \psi_j(y)}$$

which is increasing, convex and continuous.

Since $\psi_1(y) > \psi_2(y)$, there is $p_U \in (0,1)$ such that $p\ln\psi_i(x) + (1-p)\ln\psi_i(y)$ is maximized for $i = 1$ if and only if $p \leq p_U$. Further, since $\alpha_\Psi(p)$ is continuous, strictly increasing and satisfies $\alpha_\Psi(0) = 0$, there is $p_K \in (0,1)$ such that $\ln(\alpha_{\psi_j}(p)\psi_j(x) + (1-\alpha_{\psi_j}(p))\psi_j(y))$ is maximized for $j = 1$ if and only if $p \leq p_K$. Thus $p_* := \min\{p_U, p_K\}$ is the largest $p \in (0,1)$ such that $\alpha_\Psi(p) = \alpha_{1,1}(p) = \alpha_{\{\psi_1\}}(p)$. Analogously, since $\psi_2(x) > \psi_1(x)$, $p^* := \max\{p_U, p_K\}$ is the smallest $p \in (0,1)$ such that $\alpha_\Psi(p) = \alpha_{2,2}(p) = \alpha_{\{\psi_2\}}(p)$.

To find the value of $\alpha_\Psi(p)$ for $p \in (p_*, p^*)$, it remains to establish the order of p_K and p_U. By Claim 3, $\alpha_{\{\psi_1\}}(p) > \alpha_{\{\psi_2\}}(p)$ for all $p \in (0,1)$. First, since $\alpha_\Psi(p)$ is continuous, it must bet that $p_K \neq p_U$. Second, note that

$$\begin{aligned}
\alpha_{1,2}(0) &= \frac{\psi_1(y) - \psi_2(y)}{\psi_2(x) - \psi_2(y)} > 0 \\
\alpha_{1,2}(1) &= \frac{\psi_1(x) - \psi_2(y)}{\psi_2(x) - \psi_2(y)} < 1 \\
\alpha_{2,1}(0) &= \frac{\psi_2(y) - \psi_1(y)}{\psi_1(x) - \psi_1(y)} < 0 \\
\alpha_{2,1}(1) &= \frac{\psi_2(x) - \psi_1(y)}{\psi_1(x) - \psi_1(y)} > 1
\end{aligned}$$

To show that $p_U < p_K$, suppose to the contrary that $p_K < p_U$. Then

$$\alpha_\Psi(p) = \begin{cases} \alpha_{\{\psi_1\}}(p) & \text{if } p \leq p_K \\ \alpha_{2,1}(p) & \text{if } p \in (p_K, p_U) \\ \alpha_{\{\psi_2\}}(p) & \text{if } p \geq p_U \end{cases}$$

and p_K is the intersection of $\alpha_{\{\psi_1\}}(p)$ with $\alpha_{2,1}(p)$, while p_U is the intersection of $\alpha_{\{\psi_2\}}(p)$ with $\alpha_{2,1}(p)$. But since $\alpha_{2,1}(0) < \alpha_{\{\psi\}}(0)$ and $\alpha_{2,1}(1) > \alpha_{\{\psi\}}(1)$ for $\psi \in \{\psi_1, \psi_2\}$, and since $\alpha_{2,1}(p)$ is continuous and increasing, it must intersect the smaller function $\alpha_{\{\psi_2\}}(p)$ before the larger function $\alpha_{\{\psi_1\}}(p)$, and hence $p_U < p_K$, a contradiction to the assumption that $p_K < p_U$.

Thus indeed $p_U < p_K$ and

$$\alpha_\Psi(p) = \begin{cases} \alpha_{\{\psi_1\}}(p) & \text{if } p \leq p_U \\ \alpha_{1,2}(p) & \text{if } p \in (p_U, p_K) \\ \alpha_{\{\psi_2\}}(p) & \text{if } p \geq p_K \end{cases}$$

Finally, since $\alpha_{1,2}(p)$ intersects $\alpha_{\{\psi_1\}}(p)$ and $\alpha_{\{\psi_2\}}(p)$ from above, the only non-convexity arises in p_U. This established the claim for $p_* = p_U$ and $p^* = p_K$. ■

B Overlap Between Divergence and RDU Preferences

As we noted in Section 5.2, while the class of divergence preferences is generally distinct from the class of rank-dependent utility preferences, there is some overlap, as the following claim demonstrates.

Claim 5. *Define ψ_γ as in Example 2 for some $0 \le \alpha < 1 < \beta$, and define φ as in Claim 1. Fix any k satisfying the assumptions of Proposition 2. Then, for any simple act $f : S \to \mathbb{R}$,*

$$\inf_{\eta \in \triangle(S)} \left[\int_S f(s)\, d\eta(s) + k(D_\phi(\eta \,\|\, \mu)) \right] = \max_{\gamma \in \mathbb{R}} \int_S \psi_\gamma(f(s))\, d\mu(s) = \int_Z z\, d(\varphi \circ F_{f,\mu})(z),$$

where $\phi : \mathbb{R}_+ \to [0, \infty]$ is defined by $\phi(t) = 0$ for $t \in [\alpha, \beta]$ and $\phi(t) = \infty$ otherwise.

Note that the divergence $D_\phi(\eta \,\|\, \mu)$ defined in Claim 5 takes only the values 0 and $+\infty$, so the equality in the claim holds for any admissible function k (since we must have $k(0) = 0$ and $k(\infty) = \infty$).

Proof. Define ϕ by $\phi(y) = 0$ if $y \in [\alpha, \beta]$ and $\phi(y) = \infty$ otherwise. If is standard that the Fenchel conjugate of ϕ (see Equation (9)) is $\phi^* : \mathbb{R} \to \mathbb{R}$ defined by

$$\phi^*(x) = \begin{cases} \alpha x & \text{if } x \le 0 \\ \beta x & \text{if } x > 0 \end{cases}$$

Notice also that for this function, we have $(\hat{\alpha}\phi)^*(x) = \phi^*(x)$ for all $\hat{\alpha} \ge 0$, so the formula for $\psi_{\gamma,\hat{\alpha}}$ in Proposition 3 reduces to

$$\begin{aligned} \psi_{\gamma,\hat{\alpha}}(x) &= \gamma - \phi^*(\gamma - x) - k^*(\hat{\alpha}) \\ &= \begin{cases} \gamma + \beta(x - \gamma) - k^*(\hat{\alpha}) & \text{if } x < \gamma \\ \gamma + \alpha(x - \gamma) - k^*(\hat{\alpha}) & \text{if } x \ge \gamma \end{cases} \end{aligned}$$

for all $\hat{\alpha} \ge 0$. Since $\min_{\hat{\alpha} \ge 0} k^*(\hat{\alpha}) = 0$ (e.g., $k^*(0) = 0$ if we take $k(x) = 0$ for $x \le 0$), maximization over $\hat{\alpha}$ eliminates this term, and this family of fitness functions reduces to the parametric class ψ_γ from Example 2. ■

References

Abdellaoui, M., A. Baillon, L. Placido, and P. P. Wakker (2011): "The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation," *American Economic Review*, 101, 695–723 (pages 19, 23).

Allais, M. (1953): "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axioms de l'Ecole Americaine," *Econometrica*, 21, 503–546 (page 24).

Athreya, K. B. and P. E. Ney (1972): *Branching Processes*, New York: Springer-Verlag (page 8).

Ben-Tal, A. and M. Teboulle (1987): "Penalty Functions and Duality in Stochastic Programming via φ-Divergence Functionals," *Mathematics of Operations Research*, 12, 224–240 (page 16).

Ben-Tal, A. and M. Teboulle (2007): "An Old-New Concept of Convex Risk Measures: The Optimized Certainty Equivalent," *Mathematical Finance*, 17, 449–476 (pages 14, 16, 31, 32).

Bianchi, F., C. L. Ilut, and M. Schneider (2018): "Uncertainty Shocks, Asset Supply and Pricing over the Business Cycle," *Review of Economic Studies*, 85, 810–854 (page 2).

Börgers, T. and R. Sarin (1997): "Learning Through Reinforcement and Replicator Dynamics," *Journal of Economic Theory*, 77, 1–14 (page 4).

Borwein, J. M. and A. S. Lewis (1992): "Partially Finite Convex Programming, Part I: Quasi Relative Interiors and Duality Theory," *Mathematical Programming*, 57, 15–48 (page 31).

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011): "Uncertainty Averse Preferences," *Journal of Economic Theory*, 146, 1275–1330 (page 20).

Chapman, J., M. Dean, P. Ortoleva, E. Snowberg, and C. Camerer (2019): "Econographics," working paper (page 21).

Chateauneuf, A. and J. H. Faro (2009): "Ambiguity Through Confidence Functions," *Journal of Mathematical Economics*, 45, 535–558 (page 20).

Chateauneuf, A. and J. H. Faro (2012): "On the Confidence Preferences Model," *Fuzzy Sets and Systems*, 188, 1–15 (page 20).

Chatterjee, K. and R. V. Krishna (2011): "A Nonsmooth Approach to Nonexpected Utility Theory under Risk," *Mathematical Social Sciences*, 62, 166–175 (pages 14, 27).

Chetty, R. and A. Szeidl (2007): "Consumption Commitments and Risk Preferences," *Quarterly Journal of Economics*, 122, 831–877 (page 2).

Chetty, R. and A. Szeidl (2016): "Consumption Commitments and Habit Formation," *Econometrica*, 84, 855–890 (page 2).

Chew, S. H., E. Karni, and Z. Safra (1987): "Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities," *Journal of Economic Theory*, 42, 370–381 (page 27).

Combari, C., M. Laghdir, and L. Thibault (1996): "A Note on Subdifferentials of Convex Composite Functionals," *Archiv der Mathematik*, 67, 239–252 (page 33).

Cover, T. M. and J. A. Thomas (2006): *Elements of Information Theory*, 2nd ed., Hoboken, New Jersey: John Wiley & Sons (page 20).

Dean, M. and P. Ortoleva (2017): "Allais, Ellsberg, and Preferences for Hedging," *Theoretical Economics*, 12, 377–424 (page 19).

Dean, M. and P. Ortoleva (2019): "The Empirical Relationship Between Nonstandard Economic Behaviors," *Proceedings of the National Academy of Sciences*, 116, 16262–16267 (page 19).

Dow, J. and S. R. Werlang (1992): "Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio," *Econometrica*, 60, 197–204 (page 2).

Dudley, R. M. (2002): *Real Analysis and Probability*, 2nd ed., Cambridge, United Kingdom: Cambridge University Press (pages 5, 34).

Dupuis, P. and R. S. Ellis (1997): *A Weak Convergence Approach to the Theory of Large Deviations*, New York: John Wiley & Sons (pages 19, 36).

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643–669 (page 11).

Epstein, L. G. and T. Wang (1994): "Intertemporal Asset Pricing under Knightian Uncertainty," *Econometrica*, 62, 283–322 (page 2).

Ergin, H. and F. Gul (2009): "A Theory of Subjective Compound Lotteries," *Journal of Economic Theory*, 144, 899–929 (page 10).

Ergin, H. and T. Sarver (2015): "Hidden Actions and Preferences for Timing of Resolution of Uncertainty," *Theoretical Economics*, 10, 489–541 (page 2).

Flück, M. (2006): "Functional, Structural and Molecular Plasticity of Mammalian Skeletal Muscle in Response to Exercise Stimuli," *Journal of Experimental Biology*, 209, 2239–2248 (page 26).

Friedman, M. (1953): "The Methodology of Positive Economics," in *Essays in Positive Economics*, ed. by M. Friedman, University of Chicago Press (page 4).

Gabaix, X. and D. Laibson (2001): "The 6D Bias and the Equity-Premium Puzzle," *NBER Macroeconomics Annual*, 16, 257–312 (page 2).

Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mmathematical Economics*, 18, 141–153 (page 17).

Grossman, S. J. and G. Laroque (1990): "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods," *Econometrica*, 58, 25–51 (page 2).

Halevy, Y. and V. Feltkamp (2005): "A Bayesian Approach to Uncertainty Aversion," *Review of Economic Studies*, 72, 449–466 (page 12).

Hansen, L. P. and T. J. Sargent (2001): "Robust Control and Model Uncertainty," *American Economic Review*, 91, 60–66 (pages 3, 20).

Hansen, L. P. and T. J. Sargent (2008): *Robustness*, Princeton University Press (page 3).

Hiriart-Urruty, J.-B. (2006): "A Note on the Legendre-Fenchel Transform of Convex Composite Functions," in *Nonsmooth Mechanics and Analysis*, ed. by P. Alart, O. Maisonneuve, and R. T. Rockafellar, Springer, 35–46 (page 33).

Ilut, C. L. and M. Schneider (2014): "Ambiguous Business Cycles," *American Economic Review*, 104, 2368–99 (page 2).

Izhakian, Y. (2017): "Expected Utility with Uncertain Probabilities Theory," *Journal of Mathematical Economics*, 69, 91–103 (page 19).

Klibanoff, P., M. Marinacci, and S. Mukerji (2005): "A Smooth Model of Decision Making Under Ambiguity," *Econometrica*, 73, 1849–1892 (pages 10–12).

Kreps, D. M. and E. L. Porteus (1979): "Temporal von Neumann-Morgenstern and Induced Preferences," *Journal of Economic Theory*, 20, 81–109 (page 2).

Kutateladze, S. S. (1979): "Convex Operators," *Russian Mathematical Surveys*, 34, 181–214 (page 33).

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74, 1447–1498 (pages 16, 17, 20).

Machina, M. J. (1984): "Temporal Risk and the Nature of Induced Preferences," *Journal of Economic Theory*, 33, 199–231 (page 2).

Marinacci, M. (2015): "Model Uncertainty," *Journal of the European Economic Association*, 13, 1022–1100 (page 12).

Nau, R. F. (2006): "Uncertainty Aversion with Second-Order Utilities and Probabilities," *Management Science*, 52, 136–145 (page 10).

Nelson, R. R. and S. G. Winter (1982): *An Evolutionary Theory of Economic Change*, Harvard University Press (page 4).

von Neumann, J. (1928): "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, 100, 295–320 (page 20).

Piersma, T. and J. Drent (2003): "Phenotypic Flexibility and the Evolution of Organismal Design," *Trends in Ecology & Evolution*, 18, 228–233 (page 25).

Robatto, R. and B. Szentes (2017): "On the Biological Foundation of Risk Preferences," *Journal of Economic Theory*, 172, 410–422 (pages 1, 25).

Robson, A. J. (1996): "A Biological Basis for Expected and Non-Expected Utility," *Journal of Economic Theory*, 68, 397–424 (pages 1, 3, 8–10).

Robson, A. J. and L. Samuelson (2011): "The Evolutionary Foundations of Preferences," in *Handbook of Social Economics*, vol. 1, Elsevier, 221–310 (page 10).

Robson, A. J. and L. Samuelson (2019): "Evolved Attitudes to Idiosyncratic and Aggregate Risk in Age-Structured Populations," *Journal of Economic Theory*, 181, 44–81 (pages 2, 25).

Rockafellar, R. T. (1968): "Integrals Which Are Convex Functionals," *Pacific Journal of Mathematics*, 24, 525–539 (page 32).

Sarver, T. (2018): "Dynamic Mixture-Averse Preferences," *Econometrica*, 86, 1347–1382 (pages 2, 3, 14, 15, 27).

Schlag, K. H. (1998): "Why Imitate, and If So, How?: A Boundedly Rational Approach to Multi-armed Bandits," *Journal of Economic Theory*, 78, 130–156 (page 4).

Segal, U. (1987): "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach," *International Economic Review*, 28, 175–202 (page 19).

Sion, M. (1958): "On General Minimax Theorems," *Pacific Journal of Mathematics*, 8, 171–176 (page 20).

Strzalecki, T. (2011): "Axiomatic Foundations of Multiplier Preferences," *Econometrica*, 79, 47–73 (page 20).

Tuy, H. (2004): "Minimax Theorems Revisited," *Acta Mathematica Vietnamica*, 29, 217–229 (pages 20, 37).

Wakker, P. (1994): "Separating Marginal Utility and Probabilistic Risk Aversion," *Theory and Decision*, 36, 1–44 (pages 14, 27).

Wikelski, M. and C. Thom (2000): "Marine Iguanas Shrink to Survive El Niño," *Nature*, 403, 37 (page 26).

Wright, P. A. and A. J. Turko (2016): "Amphibious Fishes: Evolution and Phenotypic Plasticity," *Journal of Experimental Biology*, 219, 2245–2259 (page 26).

Yaari, M. E. (1987): "The Dual Theory of Choice under Risk," *Econometrica*, 55, 95–115 (page 27).

For Online Publication

Supplementary Appendix

In this supplement, we provide proofs of Propositions 4 and 5 and Lemmas 8 and 9 from Appendix A.5 of the main paper. We restate the results below for ease of reference.

S1 Restatement of Results

As in the main text, let $\triangle(\Omega)$ denote the set of all countably additive probability measures on the space Ω. Given some $p \in \triangle(\Omega)$, recall that $M(p) = \{q \in \triangle(\Omega) : q \ll p \text{ and } R(p \,\|\, q) < \infty\}$. In particular, since $R(p \,\|\, q) < \infty$ requires that $p \ll q$, the measures q and p are mutually absolutely continuous whenever $q \in M(p)$.

Proposition 4. *Fix any measurable space $(\Omega, \mathcal{B}_\Omega)$. Suppose $X : \Omega \to [-\infty, \infty)$ is measurable and bounded above, and let $p \in \triangle(\Omega)$. Then,*

$$\int_\Omega \ln(X(\omega))\, dp(\omega) = \inf_{q \in M(p)} \left[\ln\left(\int_\Omega X(\omega)\, dq(\omega) \right) + R(p \,\|\, q) \right]. \tag{S1}$$

In addition, if X is bounded away from zero, that is, if $X(\omega) \geq \varepsilon > 0$ for all $\omega \in \Omega$, then the infimum in Equation (S1) is uniquely attained by the measure q_0 with Radon–Nikodym derivative

$$\frac{dq_0}{dp}(\omega) = \frac{1}{X(\omega) \displaystyle\int_\Omega \frac{1}{X(\hat\omega)}\, dp(\hat\omega)}. \tag{S2}$$

Proposition 4 restricts to $q \in M(p)$, thereby ensuring that we do not encounter terms of the form $-\infty + \infty$. That is, while the first term inside the infimum in Equation (S1) could take the value $-\infty$, the second term $R(p \,\|\, q)$ will necessarily be finite.

Proposition 5. *Fix a measure $\mu \in \triangle(\Omega \times S)$, and suppose Ξ is a nonempty set of functions $\xi : \Omega \times S \to [-\infty, \infty)$ with the following properties:*

1. Closedness: *When the set of extend reals $[-\infty, \infty]$ is endowed with its usual topology and $[-\infty, \infty]^{\Omega \times S}$ is endowed with the product topology (i.e., the topology of pointwise convergence), Ξ is a closed subset of this space.*
2. Finite measurability: *There exists a finite partition $\mathcal{E} \subset \mathcal{B}_\Omega \otimes \mathcal{B}_S$ of $\Omega \times S$ such that every $\xi \in \Xi$ is measurable with respect to $\mathcal{E}$.*
3. Pointwise boundedness: $\sup_{\xi \in \Xi} \xi(\omega, s) < \infty$ *for every $(\omega, s) \in \Omega \times S$.*

Then,

$$\begin{aligned}&\sup_{\xi\in\mathrm{co}(\Xi)}\inf_{q\in M(\mu_\Omega)}\left[\ln\left(\int_\Omega\int_S \xi(\omega,s)\,d\mu(s|\omega)\,dq(\omega)\right)+R(\mu_\Omega\,\|\,q)\right]\\&=\inf_{q\in M(\mu_\Omega)}\left[\ln\left(\sup_{\xi\in\Xi}\int_\Omega\int_S \xi(\omega,s)\,d\mu(s|\omega)\,dq(\omega)\right)+R(\mu_\Omega\,\|\,q)\right].\end{aligned}\tag{S3}$$

Lemma 8. *Define Ξ as in Equation (14). For any random adaptation plan $\tau\in\mathcal{R}(\Psi|\mathcal{F})$, define $\xi^\tau:\Omega\times S\to[-\infty,\infty)$ by*

$$\xi^\tau(\omega,s)=\mathbb{E}_{\tau\otimes\rho}\big[\psi(f(\omega,s))\big]=\int_{\mathcal{F}}\int_\Psi \psi(f(\omega,s))\,d\tau(\psi|f)\,d\rho(f).$$

Then,

$$\mathrm{co}(\Xi)=\big\{\xi^\tau:\tau\in\mathcal{R}(\Psi|\mathcal{F})\big\}.$$

Lemma 9. *The set Ξ defined in Equation (14) is a closed subset of $[-\infty,\infty]^{\Omega\times S}$.*

S2 Proof of Proposition 4

The proof proceeds in three steps. We first prove Equation (S1) for random variables X that are bounded above and satisfy $X(\omega)\geq\varepsilon>0$ for all $\omega\in\Omega$. We then extend the result to all bounded $X\geq 0$. Finally, we extend to any X that is bounded above.[40]

Step 1: Suppose that X that is bounded above and satisfies $X(\omega)\geq\varepsilon>0$ for all $\omega\in\Omega$. Then, $\ln(X)$ is a bounded function, and it is therefore integrable. Fix any measures $p,q\in\triangle(\Omega)$ with $p\ll q$ and define a measure p_0 by its Radon–Nikodym derivative

$$\frac{dp_0}{dq}(\omega)=\frac{X(\omega)}{\displaystyle\int_\Omega X(\hat\omega)\,dq(\hat\omega)}.\tag{S4}$$

Since X is strictly positive, p_0 and q are mutually absolutely continuous. In particular, since $p\ll q$,

[40]The first two steps in our proof employ similar techniques to the proofs of Propositions 1.4.2 and 4.5.1 in Dupuis and Ellis (1997), although the details are quite different.

this implies $p \ll p_0$. Thus, $\frac{dp}{dp_0}$ exists and $\frac{dp}{dq} = \frac{dp}{dp_0} \cdot \frac{dp_0}{dq}$. Note that

$$\begin{aligned}
&\int_\Omega \ln(X)\, dp - R(p \,\|\, q) \\
&\quad = \int_\Omega \ln(X)\, dp - \int_\Omega \ln\left(\frac{dp}{dq}\right) dp \\
&\quad = \int_\Omega \ln(X)\, dp - \int_\Omega \ln\left(\frac{dp}{dp_0}\right) dp - \int_\Omega \ln\left(\frac{dp_0}{dq}\right) dp \\
&\quad = \int_\Omega \ln(X)\, dp - \int_\Omega \ln\left(\frac{dp}{dp_0}\right) dp - \int_\Omega \ln(X)\, dp + \ln\left(\int_\Omega X\, dq\right) \\
&\quad = -R(p \,\|\, p_0) + \ln\left(\int_\Omega X\, dq\right).
\end{aligned}$$

By Lemma 1.4.1 in Dupuis and Ellis (1997), $R(p \,\|\, p_0) \geq 0$, with equality if and only if $p = p_0$. Therefore,

$$\int_\Omega \ln(X)\, dp \leq \ln\left(\int_\Omega X\, dq\right) + R(p \,\|\, q),$$

with equality if and only if $p = p_0$. It is not difficult to show that Equations (S2) and (S4) are dual in the sense that $p = p_0$ if and only if $q = q_0$. Therefore, given p, if we set $q = q_0$ then the above holds with equality. Moreover, since X is bounded and $1/X \leq 1/\varepsilon$,

$$R(p \,\|\, q_0) = \int_\Omega \ln\left(\frac{dp}{dq_0}\right) dp = \int_\Omega \ln(X)\, dp + \ln\left(\int_\Omega \frac{1}{X}\, dp\right) < \infty,$$

which implies $q_0 \in M(p)$. Hence the infimum in Equation (S1) is attained at q_0.

Step 2: Consider now any bounded $X \geq 0$. Define a sequence of random variables $(X_n)_{n \in \mathbb{N}}$ by $X_n(\omega) = \max\{X(\omega), 1/n\}$. By step 1, we know that Equation (S1) holds for each X_n and for any p. Using this, together with the fact that $X_n \geq X$ for all n, we have

$$\begin{aligned}
\int_\Omega \ln(X_n)\, dp &= \inf_{q \in M(p)} \left[\ln\left(\int_\Omega X_n\, dq\right) + R(p \,\|\, q)\right] \\
&\geq \inf_{q \in M(p)} \left[\ln\left(\int_\Omega X\, dq\right) + R(p \,\|\, q)\right].
\end{aligned}$$

Since $\int \ln(X_1) dp < \infty$ and $\ln(X_n) \downarrow \ln(X)$, the monotone convergence theorem for extended real-valued functions (e.g., Theorem 4.3.2 of Dudley (2002)) implies

$$\begin{aligned}
\int_\Omega \ln(X)\, dp &= \lim_{n \to \infty} \int_\Omega \ln(X_n)\, dp \\
&\geq \inf_{q \in M(p)} \left[\ln\left(\int_\Omega X\, dq\right) + R(p \,\|\, q)\right].
\end{aligned}$$

Note that these terms could take the value $-\infty$.

To prove the opposite inequality, note that for any n and any $q \in M(p)$, Equation (S1) applied to the function X_n implies

$$\int_\Omega \ln(X_n)\, dp \le \ln\biggl(\int_\Omega X_n\, dq\biggr) + R(p \,\|\, q).$$

Since both sides of this inequality are finite for all n, we can again take the limit as $n \to \infty$ and apply the monotone convergence theorem to obtain

$$\int_\Omega \ln(X)\, dp \le \ln\biggl(\int_\Omega X\, dq\biggr) + R(p \,\|\, q).$$

Since this is true for all $q \in M(p)$, we have

$$\int_\Omega \ln(X)\, dp \le \inf_{q \in M(p)} \biggl[\ln\biggl(\int_\Omega X\, dq\biggr) + R(p \,\|\, q)\biggr].$$

Thus, Equation (S1) holds for any bounded $X \ge 0$.

Step 3: Finally, consider any X that is bounded above. Let $X^+(\omega) = \max\{X(\omega), 0\}$. Since we have adopted the standard convention that $\ln(x) = -\infty$ for any $x \le 0$, we have $\ln(X^+(\omega)) = \ln(X(\omega))$ for all ω. Therefore, since Equation (S1) holds for X^+ by step 2,

$$\begin{aligned}
\int_\Omega \ln(X)\, dp = \int_\Omega \ln(X^+)\, dp \\
&= \inf_{q \in M(p)} \biggl[\ln\biggl(\int_\Omega X^+\, dq\biggr) + R(p \,\|\, q)\biggr] \\
&\ge \inf_{q \in M(p)} \biggl[\ln\biggl(\int_\Omega X\, dq\biggr) + R(p \,\|\, q)\biggr].
\end{aligned}$$

To establish the opposite inequality, we consider two cases. Let $A = \{\omega \in \Omega : X(\omega) \le 0\}$. The first case is when $p(A) > 0$. Then, $\int_\Omega \ln(X)dp = -\infty$, so the above must hold with equality. The second case is when $p(A) = 0$. Then, $q(A) = 0$ for all $q \in M(p)$, since any $q \in M(p)$ must be absolutely continuous with respect to p. Therefore, $\int_\Omega X\, dq = \int_\Omega X^+ dq$ for all $q \in M(p)$ and hence

$$\inf_{q \in M(p)} \biggl[\ln\biggl(\int_\Omega X\, dq\biggr) + R(p \,\|\, q)\biggr] = \inf_{q \in M(p)} \biggl[\ln\biggl(\int_\Omega X^+\, dq\biggr) + R(p \,\|\, q)\biggr].$$

Thus, the equality is established for both cases, which completes the proof.

S3 Proof of Proposition 5

Our proof will rely on a version of the von Neumann–Sion Minimax Theorem. von Neumann (1928) proved that when $F : C \times D \to \mathbb{R}$ is a bilinear function and C and D are finite-dimensional simplexes,

$$\sup_{x \in C} \inf_{y \in D} F(x, y) = \inf_{y \in D} \sup_{x \in C} F(x, y).$$

Perhaps the most important and well-known extension of von Neumann's result is due to Sion (1958), who showed that the same conclusion can be derived under the weaker assumptions that C and D are convex subsets of topological vector spaces, one of these sets is compact, F is quasiconcave and upper semicontinuous in x, and F is quasiconvex and lower semicontinuous in y. Sion's result is not quite strong enough for our purposes, since in our application it may be that neither C nor D is compact and since F may not be lower semicontinuous in y. We will therefore rely on the following generalization of the von Neumann–Sion Theorem, which is due to Tuy (2004).

Theorem S1 (von Neumann–Sion–Tuy Minimax Theorem)**.** *Let C be a closed and convex subset of a topological vector space, and let D be a convex subset of a topological vector space. Suppose $F : C \times D \to \mathbb{R}$ satisfies the following conditions:*

1. *For every $y \in D$, the function $x \mapsto F(x, y)$ is quasiconcave and upper semicontinuous on C.*
2. *For every $x \in C$ and $y, y' \in D$, the function $\lambda \mapsto F(x, \lambda y + (1 - \lambda)y')$ is quasiconvex and lower semicontinuous on $[0, 1]$.*
3. *There exists some $\eta < \inf_{y \in D} \sup_{x \in C} F(x, y)$ and a nonempty finite set $L \subset D$ such that the set $C^L_\eta = \{x \in C : \min_{y \in L} F(x, y) \geq \eta\}$ is compact.*

Then,

$$\sup_{x \in C} \inf_{y \in D} F(x, y) = \inf_{y \in D} \sup_{x \in C} F(x, y).$$

This result is a special case of Theorem 2 in Tuy (2004). His result requires that F be what he calls *α-connected*. This condition is implied by our assumptions that C is closed and convex, D is convex, F is quasiconcave and upper semicontinuous in x, and $\lambda \mapsto F(x, \lambda y + (1 - \lambda)y')$ is quasiconvex in λ for all x, y, y'. His result also requires the lower semicontinuity property that we assumed in condition 2.[41] The final assumption needed for his result is condition 3.[42] For completeness and ease of reference, we include a complete proof of Theorem S1 in Section S6.

Note that the theorem of Sion (1958) follows as a corollary to this result: If F is quasiconvex and lower semicontinuous in y then condition 2 is implied, and if D is compact then condition 3 is implied (given that F is upper semicontinuous in x).

We now proceed with the proof of Proposition 5. Fix any measure $\mu \in \triangle(\Omega \times S)$, and fix any convex set Ξ satisfying the properties described in the statement of the proposition. We proceed in several steps. Using the second property of Ξ from the statement of the proposition, we know that

[41] Note that the assumption of lower semicontinuity in y in every line segment (that is, lower semicontinuity of the mapping $\lambda \mapsto F(x, \lambda y + (1 - \lambda)y')$ for all x, y, y') in condition 2 is in general weaker than assuming lower semicontinuity in y. However, the assumption of quasiconvexity in y in every line segment (that is, quasiconvexity of the mapping $\lambda \mapsto F(x, \lambda y + (1 - \lambda)y')$ for all x, y, y') in condition 2 is equivalent to quasiconvexity in y. Also, note that we have switched the roles of C and D compared to Tuy (2004).

[42] Strictly speaking, Theorem 2 in Tuy (2004) assumes that C^L_η is compact for $\eta = \sup_{x \in C} \inf_{y \in D} F(x, y)$ and shows that $\eta < \inf_{y \in D} \sup_{x \in C} F(x, y)$ leads to a contradiction. As is evident from his proof, our condition 3 is sufficient to obtain the same result.

there exists a finite partition $\mathcal{E}$ of $\Omega \times S$ such that every $\xi \in \Xi$ is measurable with respect to $\mathcal{E}$. We can enumerate the elements of this partition as

$$\mathcal{E} = \{E_i : i \in N\},$$

where N is a finite index set. For each $i \in N$, fix an arbitrary element $(\omega_i, s_i) \in E_i$. Since each $\xi \in \Xi$ is measurable with respect $\mathcal{E}$, we know that $\xi(\omega, s) = \xi(\omega_i, s_i)$ for all $i \in N$ and $(\omega, s) \in E_i$. Consider the mapping

$$\xi \mapsto \theta^{\xi} = (\xi(\omega_i, s_i))_{i \in N}$$

from Ξ into $[-\infty, \infty]^N$. It is easy to see that this mapping is a homeomorphism from Ξ to the set

$$\Theta = \{\theta^{\xi} : \xi \in \Xi\} \subset [-\infty, \infty]^N.$$

In other words, the set of functions Ξ is topologically equivalent to the set of vectors Θ.

As in the main paper, for any $q \in M(\mu_\Omega)$, define the measure $\mu \otimes q$ on $\Omega \times S$ to have marginal q on Ω and conditional distribution $\mu(\cdot|\omega)$ on S. That is, for any event E in the product σ-algebra $\mathcal{B}_\Omega \otimes \mathcal{B}_S$, let

$$\mu \otimes q(E) = \int_\Omega \int_S \mathbf{1}[(\omega, s) \in E] \, d\mu(s|\omega) \, dq(\omega).$$

Define a function $H : [-\infty, \infty)^N \times M(\mu_\Omega) \to \mathbb{R}_+$ by

$$H(\theta, q) = \max\left\{0, \sum_{i \in N} \theta_i \cdot \mu \otimes q(E_i)\right\} \exp(R(\mu_\Omega \,\|\, q)).$$

Lemma S1. *The set Θ and function H satisfy the following conditions:*

1. *When $[-\infty, \infty]^N$ is endowed with the product topology (i.e., the topology of pointwise convergence), Θ is compact.*
2. *There exists $\kappa \in \mathbb{R}$ such that $\theta_i \leq \kappa$ for all $\theta \in \Theta$ and $i \in N$.*
3. *Equation (S3) from the statement of the proposition is equivalent to the following:*

$$\sup_{\theta \in \mathrm{co}(\Theta)} \inf_{q \in M(\mu_\Omega)} H(\theta, q) = \inf_{q \in M(\mu_\Omega)} \sup_{\theta \in \Theta} H(\theta, q). \tag{S5}$$

Proof. Since Ξ is a closed subset of $[-\infty, \infty]^{\Omega \times S}$ by the first property in the statement of the proposition and since Ξ and Θ are homeomorphic, Θ is closed. In addition, since $[-\infty, \infty]^N$ is a compact space when endowed with the product topology,[43] this implies that Θ is compact. Since Ξ is pointwise bounded above by the third property in the statement of the proposition, we have

$$\sup_{\theta \in \Theta} \theta_i = \sup_{\xi \in \Xi} \xi(\omega_i, s_i) < \infty$$

[43] It is easy to see that the set of extended reals $[-\infty, \infty]$ is compact in its usual topology (see Example 2.75 in Aliprantis and Border (2006)), and hence $[-\infty, \infty]^N$ endowed with the product topology is compact by the Tychonoff Product Theorem (Theorem 2.61 in Aliprantis and Border (2006)).

for all $i \in N$. In particular, since N is finite, there exists $\kappa \in \mathbb{R}$ such that $\theta_i \leq \kappa$ for all $\theta \in \Theta$ and $i \in N$. To establish the third condition, note that[44]

$$\begin{aligned}
\ln\left[\sup_{\theta\in\mathrm{co}(\Theta)} \inf_{q\in M(\mu_\Omega)} H(\theta,q)\right] \\
&= \sup_{\theta\in\mathrm{co}(\Theta)} \inf_{q\in M(\mu_\Omega)} \left[\ln\left(\sum_{i\in N} \theta_i \cdot \mu\otimes q(E_i)\right) + R(\mu_\Omega \,\|\, q)\right] \\
&= \sup_{\xi\in\mathrm{co}(\Xi)} \inf_{q\in M(\mu_\Omega)} \left[\ln\left(\sum_{i\in N} \xi(\omega_i,s_i) \cdot \mu\otimes q(E_i)\right) + R(\mu_\Omega \,\|\, q)\right] \\
&= \sup_{\xi\in\mathrm{co}(\Xi)} \inf_{q\in M(\mu_\Omega)} \left[\ln\left(\int_{\Omega\times S} \xi(\omega,s)\, d(\mu\otimes q)(\omega,s)\right) + R(\mu_\Omega \,\|\, q)\right].
\end{aligned}$$

Similarly,

$$\begin{aligned}
\ln\left[\inf_{q\in M(\mu_\Omega)} \sup_{\theta\in\Theta} H(\theta,q)\right] \\
&= \inf_{q\in M(\mu_\Omega)} \left[\ln\left(\sup_{\theta\in\Theta}\sum_{i\in N} \theta_i \cdot \mu\otimes q(E_i)\right) + R(\mu_\Omega \,\|\, q)\right] \\
&= \inf_{q\in M(\mu_\Omega)} \left[\ln\left(\sup_{\xi\in\Xi}\int_{\Omega\times S} \xi(\omega,s)\, d(\mu\otimes q)(\omega,s)\right) + R(\mu_\Omega \,\|\, q)\right].
\end{aligned}$$

Thus, Equation (S3) is equivalent to Equation (S5). ■

Next, we show that we can remove any indices $i \in N$ that correspond to probability zero events. By definition, q and μ_Ω must be mutually absolutely continuous for any $q \in M(\mu_\Omega)$, and hence $\mu \otimes q$ and μ are also mutually absolutely continuous. Thus, for any $i \in N$ and $q \in M(\mu_\Omega)$,

$$\mu\otimes q(E_i) = 0 \iff \mu(E_i) = 0.$$

We can therefore remove any events $E_i \in \mathcal{E}$ that occur with zero probability under μ, since such events must also occur with zero probability under $\mu \otimes q$ for any $q \in M(\mu_\Omega)$. That is, consider the index set $M \subset N$ given by

$$M = \{i \in N : \mu(E_i) > 0\}.$$

Define the projection function $P_M : [-\infty,\infty]^N \to [\infty,\infty]^M$ by $P_M(\theta) = (\theta_i)_{i\in M}$, and set

$$\Theta' = P_M(\Theta) = \{\theta' = P_M(\theta) : \theta \in \Theta\}.$$

Define a function $F : [-\infty,\infty)^M \times M(\mu_\Omega) \to \mathbb{R}_+$ by

$$F(\theta,q) = \max\left\{0, \sum_{i\in M} \theta_i \cdot \mu\otimes q(E_i)\right\} \exp(R(\mu_\Omega \,\|\, q)).$$

[44] To deal with vectors θ and functions ξ that can take the value $-\infty$, we adopt the notational convention throughout that $\ln(x) = -\infty$ for any $x \in [-\infty, 0]$. Hence $\ln(\max\{0,x\}) = \ln(x)$ for all $x \in [-\infty,\infty)$.

Lemma S2. *The set Θ' and function F satisfy the following conditions:*

1. *When $[-\infty, \infty]^M$ is endowed with the product topology, Θ' is compact (hence closed).*
2. *There exists $\kappa \in \mathbb{R}$ such that $\theta_i \leq \kappa$ for all $\theta \in \Theta'$ and $i \in M$.*
3. *Equation (S5) is equivalent to the following:*

$$\sup_{\theta \in \mathrm{co}(\Theta')} \inf_{q \in M(\mu_\Omega)} F(\theta, q) = \inf_{q \in M(\mu_\Omega)} \sup_{\theta \in \Theta'} F(\theta, q). \tag{S6}$$

Proof. The projection function P_M is continuous when $[-\infty, \infty]^N$ and $[-\infty, \infty]^M$ are endowed with their product topologies. Therefore, the set Θ' is compact, as it is the image of the compact set Θ under the continuous function P_M. Since $[-\infty, \infty]^M$ is a Hausdorff space, compact subsets of of this space are closed (Lemma 2.32 in Aliprantis and Border (2006)). Hence, Θ' is closed. The second condition follows directly from the second condition in Lemma S1. To establish the third condition, recall from above that μ and $\mu \otimes q$ are mutually absolutely continuous for any $q \in M(\mu_\Omega)$. This implies that for any $\theta \in [-\infty, \infty)^N$, if we take $\theta' = P_M(\theta) \in [-\infty, \infty)^M$, then $H(\theta, q) = F(\theta', q)$ for all $q \in M(\mu_\Omega)$. Therefore, Equations (S5) and (S6) are equivalent. ■

We now show that we can remove any $\theta \in \Theta'$ such that $\theta_i = -\infty$ for some $i \in M$, thereby reducing this set to a subset of the Euclidean space $\mathbb{R}^M$. Formally, let

$$\Theta'' = \{\theta \in \Theta' : \theta_i > -\infty, \ \forall i \in M\}.$$

Note that it is possible to have $\Theta'' = \emptyset$.

Lemma S3. *The set Θ'' and function F satisfy the following conditions:*

1. *When $\mathbb{R}^M$ is endowed with the Euclidean topology, Θ'' is closed.*
2. *There exists $\kappa \in \mathbb{R}$ such that $\theta_i \leq \kappa$ for all $\theta \in \Theta''$ and $i \in M$.*
3. *Equation (S6) holds either if $\Theta'' = \emptyset$, or if $\Theta'' \neq \emptyset$ and*

$$\sup_{\theta \in \mathrm{co}(\Theta'')} \inf_{q \in M(\mu_\Omega)} F(\theta, q) = \inf_{q \in M(\mu_\Omega)} \sup_{\theta \in \Theta''} F(\theta, q). \tag{S7}$$

4. *Fix any $q \in M(\mu_\Omega)$. When restricted to $\mathbb{R}^M$ (endowed with the Euclidean topology), the mapping $\theta \mapsto F(\theta, q)$ is continuous, nondecreasing, quasiconcave, and quasiconvex.*

Proof. Since Θ' is a closed subset of $[-\infty, \infty]^M$ (endowed with the product topology of the extended reals), it is easy to verify that Θ'' is a closed subset of $\mathbb{R}^M$ (endowed with the Euclidean topology). Note, however, that Θ'' need not be compact. Next, the second condition follows directly from the second condition in Lemma S2. To establish the third condition, note that if $\theta \in \mathrm{co}(\Theta')$ has $\theta_i = -\infty$ for some $i \in M$, then for any $q \in M(\mu_\Omega)$,

$$\sum_{i \in M} \theta_i \cdot \mu \otimes q(E_i) = -\infty,$$

and hence $F(\theta, q) = 0$. Thus, if $\Theta'' = \emptyset$, then $F(\theta, q) = 0$ for all $\theta \in \mathrm{co}(\Theta')$ and $q \in M(\mu_\Omega)$, so Equation (S6) holds trivially. In the alternative case of $\Theta'' \neq \emptyset$, it is immediate that Equations (S6) and (S7) are equivalent.

To verify the fourth condition, fix any $q \in M(\mu_\Omega)$. Note that the mapping

$$\theta \mapsto \sum_{i \in M} \theta_i \cdot \mu \otimes q(E_i)$$

is continuous, nondecreasing, and linear. Therefore, the mapping $\theta \mapsto F(\theta, q)$ is continuous, nondecreasing, quasiconcave, and quasiconvex (though it is obviously no longer linear). ■

To apply the minimax theorem, we need the set over which the supremum is being taking to be closed an convex. That is, we will want to show that we can replace $\mathrm{co}(\Theta'')$ with $\mathrm{cl}(\mathrm{co}(\Theta''))$ on the left side of Equation (S7) and replace Θ'' with $\mathrm{cl}(\mathrm{co}(\Theta''))$ on the right side without affecting either of these values. The next two lemmas show that this is possible for the set Θ'' and function F in question.

Lemma S4. *Suppose $Y \subset \mathbb{R}^M$ is closed, and suppose there exists $\kappa \in \mathbb{R}$ such that $y_i \leq \kappa$ for all $y \in Y$ and $i \in M$. Then, for any $y \in \mathrm{cl}(\mathrm{co}(Y))$ there exists $y' \in \mathrm{co}(Y)$ such that $y' \geq y$ (that is, $y'_i \geq y_i$ for all $i \in M$).*

Proof. Suppose $y \in \mathrm{cl}(\mathrm{co}(Y))$. There there exists a sequence (y_n) in $\mathrm{co}(Y)$ such that $y_n \to y$. Let m be the cardinality of the set M. By Caratheodory's Convexity Theorem (Theorem 5.32 in Aliprantis and Border (2006)), every element of $\mathrm{co}(Y)$ can be written as a convex combination of at most $m+1$ vectors from Y. Therefore, each y_n can be written as

$$y_n = \sum_{j=1}^{m+1} \alpha_n^j y_n^j,$$

where $y_n^j \in Y$ for all $n \in \mathbb{N}$ and $j \in \{1, \ldots, m+1\}$, and $\alpha_n = (\alpha_n^1, \ldots, \alpha_n^{m+1}) \in [0,1]^{m+1}$ satisfies $\alpha_n^1 + \cdots + \alpha_n^{m+1} = 1$ for all $n \in \mathbb{N}$. Since $[0,1]^{m+1}$ is compact, (α_n) has a convergent subsequence. With slight abuse of notation, denote this subsequence again by (α_n). That is, we can assume without loss of generality that $\alpha_n \to \alpha$ for some $\alpha = (\alpha^1, \ldots, \alpha^{m+1}) \in [0,1]^{m+1}$.

We claim that the sequence (y_n^j) in Y is bounded for all j such that $\alpha^j > 0$. For suppose to the contrary that (y_n^j) is unbounded. Then, since Y is bounded above by κ, this would imply there there exists some subsequence $(y_{n_k}^j)$ and some dimension $i \in M$ such that $y_{i,n_k}^j \to -\infty$. However, since $\alpha^j > 0$ and $y_{i,n_k}^{j'} \leq \kappa$ for all j', this implies $y_{i,n_k} \to -\infty$, contradicting the fact that this subsequence converges to $y_i \in \mathbb{R}$. Thus, (y_n^j) must be bounded.

Therefore, by passing to subsequences if necessary, it is without loss of generality to assume that (y_n^j) converges for all j for which $\alpha^j > 0$. Denote the limits of these sequences by y^j, respectively, and let

$$y' = \sum_{\substack{j \in \{1,\ldots,m+1\}: \\ \alpha^j > 0}} \alpha^j y^j.$$

Since Y is closed, each of these y^j is in Y, and hence $y' \in \mathrm{co}(Y)$. Now, for every $n \in \mathbb{N}$ and $i \in M$,

$$y_{i,n} = \sum_{\substack{j\in\{1,\ldots,m+1\}:\\ \alpha^j>0}} \alpha_n^j y_{i,n}^j + \sum_{\substack{j\in\{1,\ldots,m+1\}:\\ \alpha^j=0}} \alpha_n^j y_{i,n}^j \leq \sum_{\substack{j\in\{1,\ldots,m+1\}:\\ \alpha^j>0}} \alpha_n^j y_{i,n}^j + \sum_{\substack{j\in\{1,\ldots,m+1\}:\\ \alpha^j=0}} \alpha_n^j \kappa,$$

since $y_{i,n}^j \leq \kappa$. Taking limits, the left side of this inequality converges to y_i and the right side converges to y_i'. Thus, $y \leq y'$, as claimed. ■

Lemma S5. *If $\Theta'' \neq \emptyset$, Equation (S7) is equivalent to the following:*

$$\sup_{\theta\in\mathrm{cl}(\mathrm{co}(\Theta''))} \inf_{q\in M(\mu_\Omega)} F(\theta, q) = \inf_{q\in M(\mu_\Omega)} \sup_{\theta\in\mathrm{cl}(\mathrm{co}(\Theta''))} F(\theta, q). \tag{S8}$$

Proof. The function F in nondecreasing in θ by Lemma S3. Therefore, for any $\theta, \theta' \in \mathbb{R}^M$ and $q \in M(\mu_\Omega)$, $\theta' \geq \theta$ implies $F(\theta', q) \geq F(\theta, q)$. Therefore,

$$\theta' \geq \theta \implies \inf_{q\in M(\mu_\Omega)} F(\theta', q) \geq \inf_{q\in M(\mu_\Omega)} F(\theta, q).$$

Also, since Θ'' is closed and bounded above by Lemma S3, Lemma S4 implies for any $\theta \in \mathrm{cl}(\mathrm{co}(\Theta''))$ there exists $\theta' \in \mathrm{co}(\Theta'')$ such that $\theta' \geq \theta$. Therefore,

$$\sup_{\theta\in\mathrm{co}(\Theta'')} \inf_{q\in M(\mu_\Omega)} F(\theta, q) = \sup_{\theta\in\mathrm{cl}(\mathrm{co}(\Theta''))} \inf_{q\in M(\mu_\Omega)} F(\theta, q).$$

This establishes that the left sides of Equations (S7) and (S8) are the same.

To see that the right sides of these equations are also the same, first fix any $\theta \in \mathrm{co}(\Theta'')$. Thus, $\theta = \sum_{j=1}^{m} \alpha^j \theta^j$ for some $m \in \mathbb{N}$ and $\theta^j \in \Theta''$, $j \in \{1, \ldots, m\}$. Since for any $q \in M(\mu_\Omega)$, the mapping $\theta \mapsto F(\theta, q)$ is quasiconvex by Lemma S3, this implies that $F(\theta) \leq F(\theta^j)$ for some j. Therefore,

$$\sup_{\theta\in\Theta''} F(\theta, q) = \sup_{\theta\in\mathrm{co}(\Theta'')} F(\theta, q)$$

for every $q \in M(\mu_\Omega)$. By the same arguments used above, it is also true that

$$\sup_{\theta\in\mathrm{co}(\Theta'')} F(\theta, q) = \sup_{\theta\in\mathrm{cl}(\mathrm{co}(\Theta''))} F(\theta, q).$$

Combining these observations, we see that the right sides of Equations (S7) and (S8) are the same. ■

We are almost ready to apply the minimax theorem to prove that Equation (S8) holds whenever $\Theta'' \neq \emptyset$. First, the following lemma will be used to establish some of the necessary properties of the mapping $q \mapsto F(\theta, q)$.

Lemma S6. *Suppose $X : \Omega \to \mathbb{R}$ is measurable and bounded, and fix any $p \in \triangle(\Omega)$. Then, for any $q, q' \in M(p)$, the mapping*

$$\lambda \mapsto \max\left\{0, \int_\Omega X \, d(\lambda q + (1-\lambda)q')\right\} \exp\big(R\big(p \,\big\|\, \lambda q + (1-\lambda)q'\big)\big)$$

is quasiconvex and lower semicontinuous on the interval $[0, 1]$.

Proof. Our proof will make use of the Donsker–Varadhan variational formula (see, for example, Lemma 1.4.3 in Dupuis and Ellis (1997)), which states that for any $p, r \in \triangle(\Omega)$,

$$R(p \,\|\, r) = \sup_{Y \in B_b(\Omega)} \left[\int_\Omega Y \, dp - \ln\left(\int_\Omega \exp(Y) \, dr \right) \right],$$

where $B_b(\Omega)$ denotes the space of all bounded Borel measurable real functions on Ω. Therefore,

$$\exp(R(p \,\|\, r)) = \sup_{Y \in B_b(\Omega)} \frac{\exp\left(\int_\Omega Y \, dp\right)}{\int_\Omega \exp(Y) \, dr},$$

and hence

$$\max\left\{0, \int_\Omega X \, dr\right\} \exp(R(p \,\|\, r)) = \max\left\{0, \sup_{Y \in B_b(\Omega)} \frac{\exp\left(\int_\Omega Y \, dp\right) \int_\Omega X \, dr}{\int_\Omega \exp(Y) \, dr}\right\}.$$

We will show for any $X, Y \in B_b(\Omega)$, $p \in \triangle(\Omega)$, and $q, q' \in M(p)$, the function $h : [0, 1] \to \mathbb{R}$ defined by

$$h(\lambda) = \frac{\exp\left(\displaystyle\int_\Omega Y \, dp\right) \displaystyle\int_\Omega X \, d(\lambda q + (1 - \lambda) q')}{\displaystyle\int_\Omega \exp(Y) \, d(\lambda q + (1 - \lambda) q')}$$

is quasiconvex and lower semicontinuous. This will establish the claim in the statement of the lemma, since the supremum of a set of quasiconvex and lower semicontinuous functions retains these properties.

Continuity of the function h in λ is immediate. To see that h is quasiconvex, fix any $\gamma \in \mathbb{R}$ and fix any $\lambda_1, \lambda_2 \in [0, 1]$ such that $h(\lambda_1) \leq \gamma$ and $h(\lambda_2) \leq \gamma$. Suppose without loss of generality that $\lambda_1 \leq \lambda_2$. We need to show that $h(\lambda) \leq \gamma$ for any $\lambda \in (\lambda_1, \lambda_2)$. Note that $h(\lambda_i) \leq \gamma$ is equivalent to

$$\exp\left(\int_\Omega Y \, dp \right) \int_\Omega X \, d(\lambda_i q + (1 - \lambda_i) q') \leq \gamma \int_\Omega \exp(Y) \, d(\lambda_i q + (1 - \lambda_i) q').$$

Any $\lambda \in (\lambda_1, \lambda_2)$ can be written as $\alpha \lambda_1 + (1 - \alpha) \lambda_2$ for $\alpha = (\lambda_2 - \lambda)/(\lambda_2 - \lambda_1)$. Therefore, we have

$$\begin{aligned}
&\exp\left(\int_\Omega Y \, dp \right) \int_\Omega X \, d(\lambda q + (1 - \lambda) q') \\
&= \alpha \exp\left(\int_\Omega Y \, dp \right) \int_\Omega X \, d(\lambda_1 q + (1 - \lambda_1) q') + (1 - \alpha) \exp\left(\int_\Omega Y \, dp \right) \int_\Omega X \, d(\lambda_2 q + (1 - \lambda_2) q') \\
&\leq \alpha \gamma \int_\Omega \exp(Y) \, d(\lambda_1 q + (1 - \lambda_1) q') + (1 - \alpha) \gamma \int_\Omega \exp(Y) \, d(\lambda_2 q + (1 - \lambda_2) q') \\
&= \gamma \int_\Omega \exp(Y) \, d(\lambda q + (1 - \lambda) q'),
\end{aligned}$$

which implies $h(\lambda) \leq \gamma$. This establishes that h is quasiconvex, which completes the proof. ■

The following lemma applies Theorem S1 to prove that Equation (S8) holds whenever $\Theta'' \neq \emptyset$. In light of Lemmas S1, S2, S3, and S5, this will establish Equation (S3) and complete the proof of Proposition 5.

Lemma S7. *If $\Theta'' \neq \emptyset$, then Equation (S8) is satisfied.*

Proof. We only need to establish that the assumptions of Theorem S1 are satisfied for the sets $C = \mathrm{cl}(\mathrm{co}(\Theta''))$, $D = M(\mu_\Omega)$, and for the function F defined above.

Note that C is a closed and convex subset of $\mathbb{R}^M$ by definition. It is also straightforward to show that the set D is convex. To see that condition 1 is satisfied, recall that for any $q \in D$, the mapping $\theta \mapsto F(\theta, q)$ is continuous and quasiconcave on C by Lemma S3.

Next, fix any $\theta \in C$ and define $X : \Omega \to \mathbb{R}$ by[45]

$$X(\omega) = \int_S \sum_{i \in M} \theta_i \cdot \mathbf{1}[(\omega, s) \in E_i] \, d\mu(s|\omega).$$

Then, for any $q \in M(\mu_\Omega)$,

$$\begin{aligned}
&\max\left\{0, \int_\Omega X \, dq\right\} \exp(R(\mu_\Omega \,\|\, q)) \\
&\quad = \max\left\{0, \int_{\Omega \times S} \sum_{i \in M} \theta_i \cdot \mathbf{1}[(\omega, s) \in E_i] \, d(\mu \otimes q)(\omega, s)\right\} \exp(R(\mu_\Omega \,\|\, q)) \\
&\quad = \max\left\{0, \sum_{i \in M} \theta_i \cdot \mu \otimes q(E_i)\right\} \exp(R(\mu_\Omega \,\|\, q)) \\
&\quad = F(\theta, q).
\end{aligned}$$

Therefore, Lemma S6 applied to this random variable X and to $p = \mu_\Omega$ implies that for any $q, q' \in D$, the mapping $\lambda \mapsto F(\theta, \lambda q + (1-\lambda) q')$ is quasiconvex and lower semicontinuous on $[0, 1]$. Thus, condition 2 in Theorem S1 are satisfied.

Finally, we show that either condition 3 holds for $L = \{\mu_\Omega\}$ and some $\eta > 0$, or Equation (S8) holds trivially with both sides of the equality equal to zero. Thus, there are two cases to consider. The first case is when

$$\inf_{q \in D} \sup_{\theta \in C} F(\theta, q) > 0.$$

In this case, fix any $\eta > 0$ that is strictly less than this value and take $L = \{\mu_\Omega\}$. The set

$$C_\eta^{\mu_\Omega} \equiv \{\theta \in C : F(\theta, \mu_\Omega) \geq \eta\}$$

is closed since C is closed and F is continuous in θ. Given this, and since C is a subset of the finite-dimensional Euclidean space $\mathbb{R}^M$, the set $C_\eta^{\mu_\Omega}$ is compact if and only if it is bounded. By

[45] Note that the sets $\{s \in S : (\omega, s) \in E_i\}$ are measurable for each $\omega \in \Omega$ and $i \in M$ by Lemma 4.46 in Aliprantis and Border (2006), and hence the function being integrated is indeed measurable.

Lemma S3, there exists $\kappa \in \mathbb{R}$ such that $\theta_i \leq \kappa$ for all $\theta \in C$ and $i \in M$. Let

$$\beta \equiv \min_{i \in M} \mu(E_i) > 0.$$

Then, for any $\theta \in C$ and $i \in M$,

$$\sum_{i' \in M} \mu(E_{i'})\theta_{i'} \leq \mu(E_i)\theta_i + (1 - \mu(E_i))\kappa \leq \beta\theta_i + (1 - \beta)\kappa.$$

Thus, since $R(\mu_\Omega \,\|\, \mu_\Omega) = 0$ and since $\eta > 0$, for any $\theta \in C_\eta^{\mu_\Omega}$ and $i \in M$, we have

$$\begin{aligned} 0 < \eta \leq F(\theta, \mu_\Omega) = \sum_{i' \in M} \mu(E_{i'})\theta_{i'} \leq \beta\theta_i + (1 - \beta)\kappa \\ \implies \theta_i > -\frac{(1 - \beta)\kappa}{\beta}. \end{aligned}$$

Therefore, the set $C_\eta^{\mu_\Omega}$ is bounded above by κ and bounded below by $-(1 - \beta)\kappa/\beta$. This implies that $C_\eta^{\mu_\Omega}$ is bounded, hence compact. Thus, all of the assumptions of Theorem S1 are satisfied, so we can conclude that Equation (S8) holds.

The second case is when

$$\inf_{q \in D} \sup_{\theta \in C} F(\theta, q) = 0.$$

In this case, since $F \geq 0$ and since

$$\sup_{\theta \in C} \inf_{q \in D} F(\theta, q) \leq \inf_{q \in D} \sup_{\theta \in C} F(\theta, q),$$

Equation (S8) must hold with both sides equal to zero. Thus, in either case, the equation is satisfied. This completes the proof. ∎

S4 Proof of Lemma 8

Fix any $\xi \in \mathrm{co}(\Xi)$. By the definition of Ξ and the definition of the convex hull, there exists $n \in \mathbb{N}$ and $(\psi_f^1), \ldots, (\psi_f^n) \in \Psi^B$ and $\alpha_1, \ldots, \alpha_n \geq 0$ with $\alpha_1 + \cdots + \alpha_n = 1$ such that

$$\begin{aligned} \xi(\omega, s) &= \sum_{i=1}^n \alpha_i \int_B \psi_f^i(f(\omega, s))\, d\rho(f) \\ &= \int_B \sum_{i=1}^n \alpha_i \psi_f^i(f(\omega, s))\, d\rho(f) \\ &= \int_{\mathcal{F}} \int_\Psi \psi(f(\omega, s))\, d\tau(\psi|f)\, d\rho(f), \end{aligned}$$

where we define $\tau \in \mathcal{R}(\Psi|\mathcal{F})$ for each $f \in B$ by[46]

$$\tau(\psi|f) = \sum_{i=1}^{n} \alpha_i \mathbf{1}[\psi = \psi_f^i].$$

Thus, $\xi = \xi^\tau$.

Conversely, suppose $\xi = \xi^\tau$ for some $\tau \in \mathcal{R}(\Psi|\mathcal{F})$. Since $\tau(\cdot|f)$ has finite support for all f, and since B is finite, the product measure on Ψ^B generated by these measures also has finite support. That is, there exists a product measure ν on Ψ^B with finite support, defined by

$$\nu\Big((\psi_f)_{f\in B}\Big) = \prod_{f\in B} \tau(\psi_f|f).$$

We can enumerate the elements of the support of this measure as

$$\operatorname{supp}(\nu) = \big\{(\psi_f^1), \ldots, (\psi_f^n)\big\}.$$

Thus,

$$\begin{aligned}
\xi^\tau(\omega, s) &= \int_{\mathcal{F}} \int_{\Psi} \psi(f(\omega, s))\, d\tau(\psi|f)\, d\rho(f) \\
&= \int_B \int_{\Psi^B} \psi_f(f(\omega, s))\, d\nu\Big((\psi_{\hat{f}})_{\hat{f}\in B}\Big)\, d\rho(f) \\
&= \sum_{i=1}^{n} \nu\Big((\psi_{\hat{f}}^i)_{\hat{f}\in B}\Big) \int_B \psi_f^i(f(\omega, s))\, d\rho(f),
\end{aligned}$$

and hence $\xi^\tau \in \operatorname{co}(\Xi)$.

S5 Proof of Lemma 9

The set $[-\infty, \infty]$ is a compact Hausdorff space when endowed with its usual topology.[47] By the Tychonoff Product Theorem (Theorem 2.61 in Aliprantis and Border (2006)), the set $[-\infty, \infty]^Z$ endowed with the product topology (also know as the topology of pointwise convergence) is compact. Since $\Psi \subset [-\infty, \infty]^Z$ is closed, it is also compact. Applying the Tychonoff Product Theorem again, the set Ψ^B is compact in the product topology.

We next show that the mapping $J : \Psi^B \to [-\infty, \infty]^{\Omega\times S}$ defined in Equation (13) is continuous when $[-\infty, \infty]^{\Omega\times S}$ is endowed with the product topology. To see this, fix any net $(\psi_f^\alpha)_{\alpha\in D}$ in Ψ^B that converges to some $(\psi_f) \in \Psi^B$. We will show that $J[(\psi_f^\alpha)]$ converges to $J[(\psi_f)]$.[48] First, by the

[46]We can define $\tau(\cdot|f)$ arbitrarily for $f \in \mathcal{F} \setminus B$.

[47]The topology on $[-\infty, \infty]$ is generated by sets of the form (a, b), $[-\infty, c)$ and $(c, \infty]$ for $a, b, c \in \mathbb{R}$. It is easy to see that under this topology, $[-\infty, \infty]$ is Hausdorff (meaning that for any two distinct points x, y there exist neighborhoods U of x and V of y such that $U \cap V = \emptyset$) and compact. Indeed, $[-\infty, \infty]$ is often referred to as the two-point compactification of $\mathbb{R}$ (see Example 2.75 in Aliprantis and Border (2006)).

[48]It is well known that the product topology on an uncountable product space cannot be completely

definition of the product topology, convergence of the net (ψ_f^α) implies that $\psi_f^\alpha(z) \to \psi_f(z)$ for all f and z. In particular, $\psi_f^\alpha(f(\omega,s)) \to \psi_f(f(\omega,s))$ for all $f \in B$ and $(\omega,s) \in \Omega \times S$. Therefore, since convergence is preserved under scalar multiples and finite sums,

$$\sum_{f\in B} \psi_f^\alpha(f(\omega,s))\rho(f) \to \sum_{f\in B} \psi_f(f(\omega,s))\rho(f)$$

for all ω and s. Thus, $J[(\psi_f^\alpha)] \to J[(\psi_f)]$ in the topology of pointwise convergence on $[-\infty,\infty]^{\Omega\times S}$.

Therefore, the set $\Xi = J[\Psi^B]$ is compact, since it is the image of the compact set Ψ^B under the continuous function J. Moreover, since $[-\infty,\infty]^{\Omega\times S}$ is a Hausdorff space, compact subsets of this space are closed (Lemma 2.32 in Aliprantis and Border (2006)). Thus, Ξ is closed.

S6 Proof of Theorem S1

This proof directly replicates the arguments in Tuy (2004) and is included only for ease of reference. Throughout the proof, define δ and γ as follows:

$$\delta \equiv \sup_{x\in C} \inf_{y\in D} F(x,y) \quad \text{and} \quad \gamma \equiv \inf_{y\in D} \sup_{x\in C} F(x,y).$$

Also, for any $\alpha \in \mathbb{R}$ and $y \in D$ define

$$C_\alpha(y) = \{x \in C : F(x,y) \geq \alpha\}.$$

Lemma S8. *Suppose C is a closed and convex subset of a topological vector space, and suppose D is a convex subset of a topological vector space. If $F : C \times D \to \mathbb{R}$ satisfies conditions 1 and 2 in Theorem S1 and if $\gamma > -\infty$, then for any $\alpha < \gamma$ and any $y, y' \in D$,*

$$C_\alpha(y) \cap C_\alpha(y') \neq \emptyset.$$

Proof. Fix any $\alpha < \gamma$, and denote $C_\alpha(y)$ simply by $C(y)$ for ease of notation. Note that $\alpha < \gamma$ implies that $\sup_{x\in C} F(x,y) > \alpha$ for all $y \in D$. Thus, $C(y)$ is nonempty for all $y \in D$. In addition, since C is closed and convex and since $x \mapsto F(x,y)$ is quasiconcave and upper semicontinuous for every y by condition 1, $C(y)$ is closed and convex for every y. Finally, for any $x \in C$ and $y, y' \in D$, the quasiconvexity of the mapping $\lambda \mapsto F(x, \lambda y + (1-\lambda)y')$ assumed in condition 2 implies that, for every $\lambda \in [0,1]$,

$$F(x, \lambda y + (1-\lambda)y') \leq \max\{F(x,y), F(x,y')\}$$

and therefore

$$C(\lambda y + (1-\lambda)y') \subset C(y) \cup C(y'). \tag{S9}$$

described by sequential convergence, as such spaces are not metrizable. Although B is finite, Z could be uncountable. Hence we must use nets to establish the continuity of J.

Suppose, contrary to the claim in the lemma, that there exists $y, y' \in D$ such that

$$C(y) \cap C(y') = \emptyset.$$

We will show that this leads to a contradiction. For any $\lambda \in [0,1]$, let $y_\lambda = \lambda y + (1-\lambda)y'$. For any $\lambda \in [0,1]$, note that we cannot have both $C(y_\lambda) \cap C(y) \neq \emptyset$ and $C(y_\lambda) \cap C(y') \neq \emptyset$. For if this were true, then we would have $C(y_\lambda) = E_y \cup E_{y'}$ where $E_y = C(y_\lambda) \cap C(y)$ and $E_{y'} = C(y_\lambda) \cap C(y')$ are two nonempty, closed, and disjoint sets, which is impossible since $C(y_\lambda)$ is convex and hence connected. Therefore, for every $\lambda \in [0,1]$, one and only one of the following alternatives holds:

$$C(y_\lambda) \subset C(y) \quad \text{or} \quad C(y_\lambda) \subset C(y').$$

Denote by M_y and $M_{y'}$ the set of all $\lambda \in [0,1]$ such that $C(y_\lambda) \subset C(y)$ and $C(y_\lambda) \subset C(y')$, respectively. Then, $0 \in M_y$, $1 \in M_{y'}$, and $M_y \cup M_{y'} = [0,1]$ by the preceding arguments. In additional, by Equation (S9),

$$C(y_\lambda) \subset C(y_{\lambda_1}) \cup C(y_{\lambda_2}) \quad \forall \lambda \in [\lambda_1, \lambda_2],$$

and hence $\lambda \in M_y$ implies $[0,\lambda] \subset M_y$ and $\lambda \in M_{y'}$ implies $[\lambda, 1] \subset M_{y'}$. Let $\lambda^* = \sup M_y = \inf M_{y'}$, where the second equality holds because $M_y \cap M_{y'} = \emptyset$ and $M_y \cup M_{y'} = [0,1]$.

Suppose without loss of generality that $\lambda^* \in M_y$ (the argument is analogous for $\lambda^* \in M_{y'}$). We cannot have $\lambda^* = 1$ since this would imply $C(y) \subset C(y')$. Therefore, $0 \leq \lambda^* < 1$. Since $\alpha < \gamma \leq \sup_{x \in C} F(x, y_{\lambda^*})$, there is some $\bar{x} \in C$ such that $F(\bar{x}, y_{\lambda^*}) > \alpha$. Since the mapping $\lambda \mapsto F(\bar{x}, y_\lambda)$ is lower semicontinuous by condition 2, there exists $\varepsilon > 0$ such that $F(\bar{x}, y_{\lambda^*+\varepsilon}) > \alpha$, and hence $\bar{x} \in C(y_{\lambda^*+\varepsilon})$. But since $\bar{x} \in C(y_{\lambda^*}) \subset C(y)$, this implies $C(y_{\lambda^*+\varepsilon}) \subset C(y)$, that is, $\lambda^* + \varepsilon \in M_y$, contradicting the definition of λ^*. Thus, $C(y) \cap C(y') = \emptyset$ leads to a contradiction. ■

Lemma S9. *Suppose C is a closed and convex subset of a topological vector space, and suppose D is a convex subset of a topological vector space. If $F : C \times D \to \mathbb{R}$ satisfies conditions 1 and 2 in Theorem S1 and if $\gamma > -\infty$, then for any $\alpha < \gamma$ and any finite set $L \subset D$,*

$$\bigcap_{y \in L} C_\alpha(y) \neq \emptyset. \tag{S10}$$

Proof. We prove by induction. We know from Lemma S8 that Equation (S10) holds if $|L| = 2$. We now show that if this equation holds for all C, D, L as in the statement of the lemma when $|L| = k$ then it also holds for all such C, D, L when $|L| = k+1$.

Let $L = \{y^1, \ldots, y^k, y^{k+1}\} \subset D$. Let $C' = C_\alpha(y^{k+1})$, and let

$$\gamma' \equiv \inf_{y \in D} \sup_{x \in C'} F(x, y).$$

Note that C' is nonempty since $\alpha < \gamma$, and it is closed and convex since the mapping $x \mapsto F(x, y)$ is quasiconcave and upper semicontinuous for every y by condition 1. Also, since

$$C_{\alpha'}(y) \cap C' \supset C_{\alpha'}(y) \cap C_{\alpha'}(y^{k+1}) \neq \emptyset$$

for every $y \in D$ and $\alpha' \in (\alpha, \gamma)$ by Lemma S8, we have

$$\sup_{x \in C'} F(x, y) \geq \alpha'$$

for all $y \in D$. Hence $\gamma' \geq \alpha'$. Since this is true for all $\alpha' \in (\alpha, \gamma)$, it must be that $\gamma' = \gamma$. Now, applying the induction hypothesis to the sets C' and D and to $L' = \{y^1, \ldots, y^k\}$, we have

$$\bigcap_{y \in L'} C'_\alpha(y) \neq \emptyset,$$

where $C'_\alpha(y) = \{x \in C' : F(x, y) \geq \alpha\}$. But this implies

$$\bigcap_{y \in L} C_\alpha(y) \neq \emptyset.$$

This completes the proof. ■

Using these lemmas, we now complete the proof of Theorem S1. First, it is immediate that $\delta \leq \gamma$. Also, this implies that $\delta = \gamma$ if either $\delta = \infty$ or $\gamma = -\infty$, so it remains only to consider the case of $\delta < \infty$ and $\gamma > -\infty$. For any $\alpha \in \mathbb{R}$ and $L \subset D$, let

$$C_\alpha^L \equiv \{x \in C : F(x, y) \geq \alpha \ \forall y \in L\} = \bigcap_{y \in L} C_\alpha(y).$$

By condition 3, there exists some $\eta < \gamma$ and some finite set $L \subset D$ such that the set C_η^L is compact. Fix this set L and fix any $\alpha \in (\eta, \gamma)$. For any $y \in D$, define

$$C_\alpha^L(y) = \{x \in C_\alpha^L : F(x, y) \geq \alpha\} = \bigcap_{y' \in L \cup \{y\}} C_\alpha(y').$$

Note that $C_\alpha^L(y)$ is closed for all $y \in D$ since $x \mapsto F(x, y)$ is quasiconcave by condition 1. By Lemma S9, C_α^L is nonempty and, moreover, the sets $C_\alpha^L(y)$ for $y \in D$ have the finite intersection property. Since the sets $C_\alpha^L(y)$ for $y \in D$ are all contained in the compact set C_η^L, this implies that

$$\bigcap_{y \in D} C_\alpha^L(y) = \bigcap_{y \in D} C_\alpha(y) \neq \emptyset.$$

Therefore, taking any element $\bar{x}$ from this set, we have $F(\bar{x}, y) \geq \alpha$ for all $y \in D$ and hence $\inf_{y \in D} F(\bar{x}, y) \geq \alpha$. Therefore $\delta \geq \alpha$. Since this is true for $\alpha \in (\eta, \gamma)$, we must have $\delta \geq \gamma$, which completes the proof.

References for Supplementary Appendix

Aliprantis, C. and K. Border (2006): *Infinite Dimensional Analysis*, 3rd ed., Berlin, Germany: Springer-Verlag (pages S6, S8, S9, S12, S14, S15).

Dudley, R. M. (2002): *Real Analysis and Probability*, 2nd ed., Cambridge, United Kingdom: Cambridge University Press (page S3).

Dupuis, P. and R. S. Ellis (1997): *A Weak Convergence Approach to the Theory of Large Deviations*, New York: John Wiley & Sons (pages S2, S3, S11).

von Neumann, J. (1928): "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, 100, 295–320 (page S4).

Sion, M. (1958): "On General Minimax Theorems," *Pacific Journal of Mathematics*, 8, 171–176 (page S5).

Tuy, H. (2004): "Minimax Theorems Revisited," *Acta Mathematica Vietnamica*, 29, 217–229 (pages S5, S15).