Theoretical Economics 17 (2022), 507–519 1555-7561/20220507

What were you thinking? Decision theory as coherence test

Itzhak Gilboa
School of Economics, Tel Aviv University and Department of Economics and Decision Sciences, HEC, Paris

Larry Samuelson
Department of Economics, Yale University

Decision theory can be used to test the logic of decision making—one may ask whether a given set of decisions can be justified by a decision-theoretic model. Indeed, in principal–agent settings, such justifications may be required—a manager of an investment fund may be asked what beliefs she used when valuing assets and a government may be asked whether a portfolio of rules and regulations is coherent. In this paper we ask which collections of uncertain-act evaluations can be simultaneously justified under the maxmin expected utility criterion by a single set of probabilities. We draw connections to the fundamental theorem of finance (for the special case of a Bayesian agent) and revealed-preference results.

Keywords. Decision theory, revealed preference, coherence, maxmin expected utility.

JEL classification. D8.

1. Introduction

People often make decisions on behalf of others. This is the case with democratically elected leaders, hedge fund managers, and professionals hired by clients. How can one tell if these decision makers are successful? Past experience is a natural answer, but assessing a decision by its outcome is not a trivial task. The appropriate counterfactuals for comparison may not be obvious and the realized outcome may not be indicative of the ex ante quality of the decision.

When considering several decisions together, one can find situations in which decision makers would obviously be deemed incompetent or corrupt. Cyclical choices, inconsistent with the existence of a preference order, can signal incompetence. Decisions that can only be justified by personal aggrandizement can indicate corruption. Decisions that cannot be justified by a common model of the underlying uncertainty

Itzhak Gilboa: tzachigilboa@gmail.com
Larry Samuelson: Larry.Samuelson@yale.edu
Gilboa gratefully acknowledges ISF Grants 1077/17 and 1443/20, the Investissements d'Avenir ANR-11-IDEX-0003/Labex ECODEC ANR-11-LABX-0047, the AXA Chair for Decision Sciences at HEC, and the Foerder Institute at Tel-Aviv University. We are grateful to Rossella Argenziano for comments, and to the two referees.

 https://doi.org/10.3982/TE4707

may indicate confusion or inconsistency. People may accordingly question the decision maker, asking "What were you thinking when you made these decisions? Can you provide a coherent account of what you were doing?"[1]

In this paper we study the justification of decisions. We consider a model in which states of the world and utilities are given, and for each of a set of uncertain acts there is a certainty equivalent. We ask when there are beliefs that can justify these certainty equivalents, simultaneously for all acts.[2] In the Bayesian case, in which "beliefs" are modeled by a single probability measure and preferences are defined by expected utility, a version of the fundamental theorem of finance can be reinterpreted as providing the answer. We therefore focus here on a non-Bayesian model in which "beliefs" are sets of probabilities and preferences follow the maxmin expected utility rule. Our results can be interpreted in a positive light—we identify circumstances under which one can justify a collection of decisions as coherent. Our analysis also has a normative flavor—one can view our results as supporting calls for transparency, demanding that officials who make decisions for others provide protocols formulated in the language of decision theory to account for their decisions.

Section 2 presents our main result, establishing conditions for the existence of a set of probability vectors, such that the collection of act/certainty-equivalent pairs is consistent with maxmin expected utility maximization. Section 3 explains how our work is related to the literature. Section 4 provides discussion.

2. The coherence criterion

2.1 *The model*

Let there be a set of states $\mathcal{S} = \{1, \dots, S\}$, indexed by s, and a set of acts $\mathcal{A} = \{1, \dots, A\}$, indexed by a. Act a yields x_s^a in state s, and thus corresponds to the vector $x^a = (x_s^a)_{s=1}^S \in \mathbb{R}^S$. For each act x^a, there is a certainty equivalent $c^a \in \mathbb{R}$. Whenever we can do so without confusion, we will also use c^a to denote a vector in $\mathbb{R}^S$, each of whose elements equals c^a, trusting the context to clarify the usage. As the name suggests, a certainty equivalent c^a is "certain" in the sense that it is independent of s. We interpret the value $x_s^a \in \mathbb{R}$ of an uncertain act as the utility generated by an outcome associated with that state, or the expected utility of a lottery, or a sum of money accruing to a risk-neutral person, or some other measure of utility. Similarly, the value c^a, taken in each state by a certainty equivalent, is measured in utility terms.

According to our interpretation, each act x^a is chosen from some set of feasible acts. Decision makers are required to explicitly specify their subjective certainty equivalents of all acts available to them in each decision problem, in such a way that all act/certainty-equivalent pairs can be simultaneously justified by a single model of subjective beliefs.

[1]See Berger et al. (2021) for such an analysis of COVID-19 decision making.

[2]We thus give the decision maker the benefit of the doubt: as long as there are beliefs that can justify decisions with the accepted utility function, she is assumed to be neither incompetent nor corrupt. In the formal analysis we do not question the plausibility of such beliefs. We return to this point in the Section 4.

If beliefs are restricted to be Bayesian, the question boils down to the following: given a (finite) collection of act/certainty-equivalent pairs, $(x^a, c^a)_{a=1}^A$, when is there a probability $p \in \Delta(S)$ (the set of probability distributions over the states $\mathcal{S}$) such that the expected payoff of x^a is c^a, for each act a? The answer is that this is possible, i.e., there exists p such that $p \cdot x^a = c^a$ for all a, if and only if there exists no collection of real numbers $(\lambda^a)_{a=1}^A$ satisfying[3]

$$\sum_{a=1}^{A} \lambda^a c^a \gg \sum_{a=1}^{A} \lambda^a x^a. \tag{1}$$

This is a no-arbitrage condition: if the certainty equivalents are interpreted as prices of uncertain assets in a financial market, this condition states that no linear combination of the assets can be strictly dominated by the corresponding linear combination of their prices.[4] The statement that (1) is necessary and sufficient for the existence of Bayesian beliefs consistent with the collection of act/certainty-equivalent pairs, $(x^a, c^a)_{a=1}^A$ is a version of the fundamental theorem of finance (Ross (2005, Chapter 1)). The standard interpretation of the fundamental theorem of finance takes linear pricing as a fact about the operations of markets and views the no-arbitrage condition as an equilibrium constraint, to derive the existence of prices satisfying $p \cdot x^a = c^a$ as a result. By contrast, we reverse the roles: we consider the condition $p \cdot x^a = c^a$ as defining the rules of the game (of justifying decisions), and read the no-arbitrage condition as a result that characterizes the evaluations for which such a probability p exists.

Bayesian justifications might be too restrictive. Suppose there are two states of the world, corresponding to the long run effects of global warming. An investment fund manager needs to make a choice among

$$\{x^b = (1, 0), x^d = (0.4, 0.4), x^f = (0, 1)\}.$$

Clearly, no Bayesian beliefs could justify the selection of d out of $\{b, d, f\}$. Indeed, if the manager would value these acts by

$$c^b, c^f \leq 0.4 \qquad c^d = 0.4,$$

we would find that the coefficients $\lambda^b = \lambda^f = -0.5$, $\lambda^d = 1$ yield

$$\sum_{a=1}^{A} \lambda^a c^a \geq 0 \gg \sum_{a=1}^{A} \lambda^a x^a = (-0.1, -0.1).$$

However, if the manager is unsure about the probability of the two states, she might find act d safer than both b and f. If called upon to defend her decision, she might respond as follows.

[3]We take $\gg$ to mean that every element of the vector on the left is strictly larger than its counterpart on the right.

[4]As the coefficients λ^a are unconstrained by sign, the converse statement is equivalent.

> I know that the collection of evaluations I listed cannot be justified by a single probability, but I did not have a single probability that I could trust. Indeed, were I to rely on one such probability, you would have asked me where I got it from—no one can quantify precisely the effects of global warming. Instead, I used a model that allows for ambiguity, and, with some ambiguity aversion, I can justify the choice of d.

Similarly, an individual may be more comfortable assessing a sequence of decisions by appealing to bounds on probabilities than by making point estimates of probabilities. We therefore modify the basic question, allowing the notion of "justification" to recognize the fact that probabilities may not be known. In this paper, we deal with a specific model of non-Bayesian decision making, namely maxmin expected utility. We therefore ask, "When is there a *set* of probability vectors such that the putative certainty equivalent for each act is the minimum (over the set of probabilities) expected utility of that act?"

2.2 *The coherence result*

The Appendix proves the following theorem.

Theorem 1. *Let there be given acts* $(x^a)_{a=1}^A$ *and a vector of certainty equivalents* $(c^a)_{a=1}^A$. *The following statements are equivalent.*

(A) There exists a convex, closed set of probabilities on S, $P \subset \Delta(\mathcal{S})$ *such that, for each act* $a \in \mathcal{A}$,

$$\min_{p\in P} p \cdot x^a = c^a.$$

(B) There is no act $\tilde{a}$ *and collection of real numbers* $(\lambda^a)_{a=1}^A$ *such that* $\lambda^a \geq 0$ *for* $a \neq \tilde{a}$ *such that*

$$\sum_{a=1}^{A} \lambda^a c^a \gg \sum_{a=1}^{A} \lambda^a x^a.$$

Condition (A) deals with a set of probabilities, and hence the dimensionality of the problem is in general infinite. It is easy to see—as noted in the first steps of the proof—that the condition can be reduced to finitely many probability vectors (one for each act).

It is clear that condition (B) is a relaxation of the corresponding condition for the Bayesian case, since the latter prohibits the inequality (1) for any collection of real numbers $(\lambda^a)_{a=1}^A$, while (B) prohibits the inequality only for such collections that have at most one strictly negative element.

It is straightforward to see that condition (A) implies condition (B): assume (A), and let there be given $\tilde{a}$ and $(\lambda^a)_{a=1}^A$ as in (B). Let $\tilde{p} \in P$ be such that $c^{\tilde{a}} = \tilde{p} \cdot x^{\tilde{a}}$. For any other $a \neq \tilde{a}$ we have $c^a \leq \tilde{p} \cdot x^a$ and $\lambda^a \geq 0$. Hence $\lambda^a c^a \leq \tilde{p} \cdot \lambda^a x^a$ for all a (with equality for $\tilde{a}$), which is incompatible with the vector domination $\sum_{a=1}^A \lambda^a c^a \gg \sum_{a=1}^A \lambda^a x^a$. The main point of Theorem 1 is, therefore, that the converse direction also holds.

2.3 *Interpretation*

2.3.1 *The debate game* Consider the following "debate game." There is a decision maker, say, an investment fund manager, and a skeptic. They play a two-person zero-sum game in which the manager tries to justify a collection of decisions, and the skeptic tries to debunk them. It is helpful to think of the acts as assets, backed by evaluations of the type "asset x^a is worth c^a." One way to define the game's outcome is to declare the manager victorious if she can exhibit a set of probabilities, according to which the minimal expected value of each act equals its certainty equivalent, thereby verifying (A). Another possible definition would make the skeptic the winner if he can produce a violation of (B). The result states that these two specifications of the debate game are equivalent.

The debate game interpretation can also apply to the version of the fundamental theorem of finance mentioned above. In that case, the manager is more limited in her choices, as she can only point to a single probability that should simultaneously justify the certainty equivalents of all acts. Correspondingly, the skeptic in the Bayesian game can declare victory as soon as he finds any set of coefficients $(\lambda^a)_{a=1}^A$ for which $\sum_{a=1}^A \lambda^a c^a \gg \sum_{a=1}^A \lambda^a x^a$—whatever their signs. By contrast, the current game is clearly biased in favor of the manager. Viewed from her perspective, she has more freedom in defining "beliefs,"allowing for sets of probabilities that need not be singletons. In the financial-markets interpretation, this leads to a set P of possible asset prices instead of single price vector (as in the fundamental theorem); similar ideas appear in the formulation of coherent risk measures (e.g., Artzner, Delbaen, Eber, and Heath (1999)). Correspondingly, the skeptic has less freedom in looking for presumed inconsistencies, as condition (B) does not allow for all linear combinations of assets (and certainty equivalents). Indeed, the acts $\{x^b = (1, 0), x^d = (0.4, 0.4), x^f = (0, 1)\}$ can all be evaluated by $c^a = 0.4$, justified by the set of probabilities $\{(p, 1 - p)|0.4 \leq p \leq 0.6\}$. Correspondingly, the weights $\lambda^b = \lambda^f = -0.5$, $\lambda^d = 1$ cannot be used by the skeptic to prove his point, as only one negative value is allowed by condition (B).

2.3.2 *A geometric interpretation* We illustrate the portfolio tests of Theorem 1, focusing on the case of two states in order to allow convenient representations. The case $S = 2$ is rather special: with two states, one probability vector (maximizing the probability of state 1) is used to evaluate all assets whose maximal value is obtained in state 2, and a second probability vector is used to evaluate all assets whose maximal value is obtained in state 1. In other words, the indifference curves of a maxmin expected value manager are linear and parallel above as well as below the diagonal, with the former steeper than the latter. Moreover, the two-state case is special also in a Bayesian analysis, because when $S = 2$, a single act evaluation suffices to pinpoint the probability used by the decision maker (and, in our case, the probability she uses below/above the diagonal). As a result, if any two of the decision maker's evaluations are consistent with each other, all of them are also consistent as a set. By contrast, with more than two states, both in the Bayesian and the maxmin analysis one needs to examine sets of evaluations that are larger than pairs. Nonetheless, some insight is gained by considering this case.

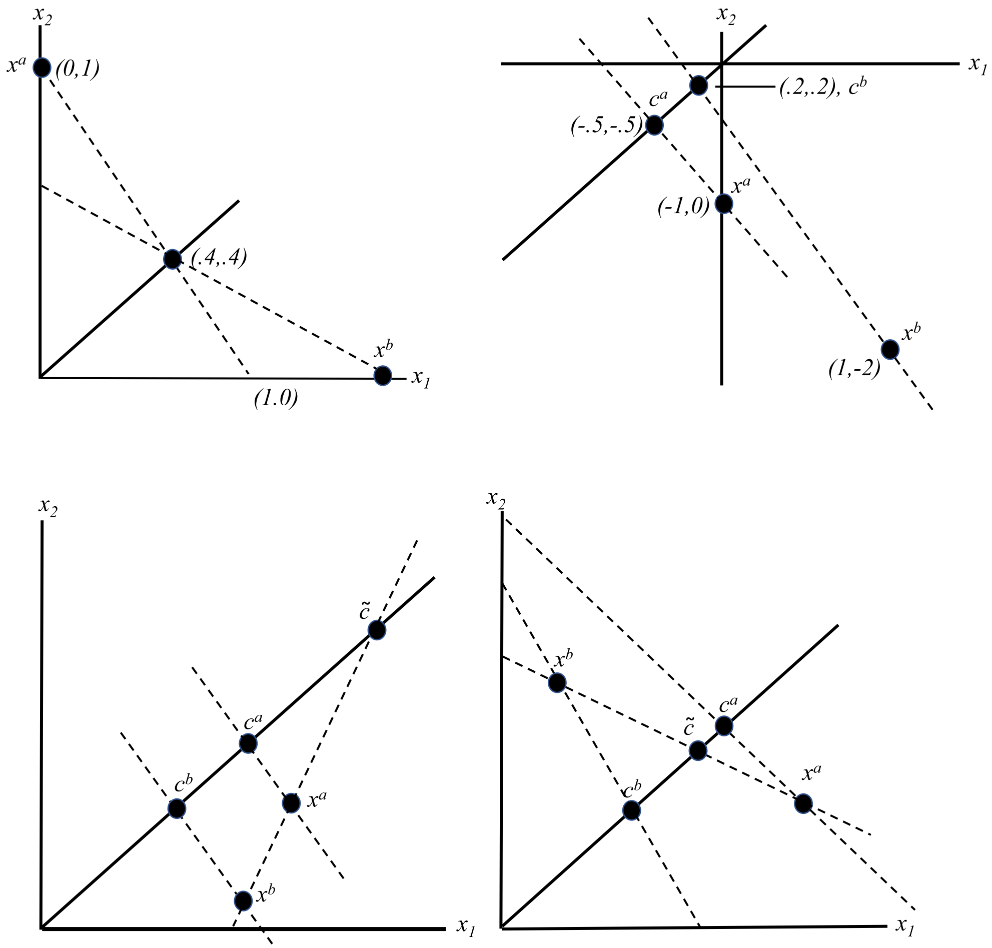

FIGURE 1. Illustration of Theorem 1 with two states.

The top left panel of Figure 1 illustrates how the manager can motivate the choice of x^d from the set $\{x^b = (1, 0), x^d = (0.4, 0.4), x^f = (0, 1)\}$ introduced in Section 2.1. Appealing to the set of probabilities $\{(p, 1 - p)|0.4 \leq p \leq 0.6\}$, asset x^b is evaluated according to $(0.4, 0.6)$, giving (via the indifference curve shown) a certainty equivalent 0.4, while probability $(0.6, 0.4)$ gives the same certainty equivalent for x^f. Noting that the certainty equivalent of x^d is obviously also 0.4, this ensures (A), which we have noted immediately implies (B).

For a first look at how condition (B) implies condition (A), consider the top right panel of Figure 1. These certainty equivalents are inconsistent with maxmin expected utility and hence with condition (A), as seen in the fact that the corresponding indifference curves are not parallel ($x^a = (0, -1)$ is evaluated with a higher probability on state 2 than is $x^b = (1, -2)$). Letting a take the role of $\tilde{a}$, and setting $\lambda^a = \lambda^{\tilde{a}} = -3$ and $\lambda^b = 1$, we get

$$\sum_{a=1}^{A} \lambda^a c^a = (1.3, 1.3) \gg (1, 1) = \sum_{a=1}^{A} \lambda^a x^a,$$

which is a violation of (B). Intuitively, the portfolio obtained by buying one unit of asset b and selling short three units of asset a gives the (perfectly hedged) payoff $(1, 1)$, which is strictly dominated by the corresponding portfolio of the certainty equivalents (given by 1.3).

For a more general exposition of how condition (B) implies condition (A), suppose we have two acts, x^a and x^b, that are on the same side of the diagonal, and assume that they are considered equivalent to c^a and c^b, as in the bottom left panel of Figure 1. Assume for simplicity that $(x^a - x^b)$ is not parallel to the diagonal, and let $\tilde{c}$ be the act defined by the intersection with the diagonal of the line defined by x^a and x^b. Hence, $\tilde{c} = \lambda x^a + (1 - \lambda)x^b$ for some $\lambda \notin [0, 1]$.[5] Assume without loss of generality that $\lambda > 1$ (again, this corresponds to the bottom left panel of Figure 1).

Given these acts and certainty equivalents, apply condition (B) with $\tilde{a} = a$ and $\lambda^a = -\lambda$ (with $\lambda > 1$) and $\lambda^b = \lambda - 1 > 0$. The condition says that the vector $(\lambda - 1)(x^b - c^b)$ cannot be dominated by $\lambda(x^a - c^a)$, which means that for some state $s = 1, 2$ we have $(\lambda - 1)(x^b_s - c^b) \geq \lambda(x^a_s - c^a)$, which we can rearrange to give (substituting the definition of $\tilde{c}$ to obtain the second inequality)

$$\begin{aligned} \lambda x^a_s + (1 - \lambda)x^b_s &\leq \lambda c^a + (1 - \lambda)c^b \\ \tilde{c}_s &\leq \lambda c^a + (1 - \lambda)c^b. \end{aligned}$$

Similarly, we can set $\tilde{a} = b$, and for $\lambda^a = \lambda > 1$ and $\lambda^b = 1 - \lambda < 0$ obtain the conclusion that for some state s', the converse inequality holds. Because $\tilde{c}$ is on the diagonal, $\tilde{c}_1 = \tilde{c}_2 = \lambda c^a + (1 - \lambda)c^b$ follows. In other words, the point $\tilde{c}$ lies on the continuation of the segment $[x^a, x^b]$ as well as on the continuation of the segment $[c^a, c^b]$, and it is at the same relative distance on both. Explicitly, we have the similarity of triangles

$$\Delta_{\tilde{c}, x^a, c^a} \sim \Delta_{\tilde{c}, x^b, c^b},$$

which, by Thales's theorem, implies that the segments $[x^a, c^a]$ and $[x^b, c^b]$ are parallel. Thus, there exists a vector p that satisfies $p \cdot (x^a - c^a) = p \cdot (x^b - c^b) = 0$, giving (A).

We cannot simply apply this argument when we have two acts x^a and x^b on different sides of the diagonal, as these acts may then be evaluated by different probabilities. Assume, without loss of generality, that $x^a_1 > x^a_2$, $x^b_1 < x^b_2$ with $x^a_1 \geq c^a \geq x^a_2$ and $x^b_1 \leq c^b \leq x^b_2$, as in the bottom right panel of Figure 1. Let $\lambda \in (0, 1)$ be such that $\tilde{c} = \lambda x^a + (1 - \lambda)x^b$ is on the diagonal. Then, setting $\lambda^a = \lambda$, $\lambda^b = (1 - \lambda)$, condition (B) implies that $\tilde{c}$ cannot be dominated by $\lambda c^a + (1 - \lambda)c^b$. In other words

$$\lambda x^a_1 + (1 - \lambda)x^b_1 = \lambda x^a_2 + (1 - \lambda)x^b_2 \geq \lambda c^a + (1 - \lambda)c^b,$$

[5] We know that $\lambda \notin [0, 1]$, as both x^a and x^b are below the diagonal.

which we can rearrange to give

$$\frac{x_1^a - c^a}{c^a - x_2^a} \geq \frac{c^b - x_1^b}{x_2^b - c^b},$$

which means that the segment $[x^b, c^b]$ is steeper than the segment $[x^a, c^a]$. Equivalently, for the unique probability vectors p^a, p^b such that $p^a \cdot (x^a - c^a) = p^b \cdot (x^b - c^b) = 0$ we have $p_1^a \leq p_1^b$ (and $p_2^a \geq p_2^b$). This implies that we can think of c^a as the minimum of $(p^a \cdot x^a, p^b \cdot x^a)$ and of c^b as the minimum of $(p^a \cdot x^b, p^b \cdot x^b)$, giving the result.

3. Relationship to the literature

The type of question we ask is formally equivalent to revealed preference analysis: given a set of decisions, can they be justified by a particular decision model?

The initial forays into revealed preference theory (e.g., Samuelson (1938) and Houthakker (1950)) were designed to clarify the conceptual foundations of utility maximization. These papers argued that an analyst could dispense with the interpretation of utility maximization as a process by which decisions are made, along with Edgeworth's (1881, Appendix III) hedonimeter, and still reasonably model a person as maximizing utility. The goal is to characterize "what economic models say about the observable world" (Chambers and Echenique (2016, p. xiii)).

By contrast, our motivation is to discuss the coherence of decisions rather than the content of economic models. Our question is not "under which conditions would an economist have to discard a certain class of models," but rather "under which conditions will a decision maker be considered incompetent (or even corrupt)?"

The first generation of revealed-preference papers, examined in the first six chapters of Chambers and Echenique (2016), addresses cases in which there is no uncertainty. A next generation of papers examines settings in which an agent is choosing under uncertainty, but objective probabilities are given (Border (1992), Chambers and Echenique (2016, Chapter 8), Chambers, Liu, and Martinez (2014), Fishburn (1975), Kubler, Selden, and Wei (2014), Lin (2019), Polisson, Quah, and Renou (2020, Section IB)). Fishburn (1975), focusing on expected utility maximization, highlights the usefulness of separation theorems, which implicitly lie behind our arguments, while the criterion derived by Border (1992) is closest to the spirit of our results.

The natural next step is to assume that neither probabilities nor utilities are given. Richter and Shapiro (1978), in a setting of two states, ask what restrictions on the set of possible distributions are imposed by a set of binary rankings of lotteries, given that no structure is imposed on the utility function.[6] The fundamental result here is by Echenique and Saito (2015) (see Chambers and Echenique (2016, Section 8.2) for an exposition). Echenique and Saito assume that x^a takes on monetary values, and offer a characterization of expected utility maximization that is similar to the result in

[6]Their finding is that what might be viewed as a folk theorem, stating (very roughly) that for any set P of probabilities defined by a finite set of integer polynomial inequalities, one can find a finite set of prizes and a finite set of binary rankings of lotteries over those prizes such that the rankings are consistent with expected utility maximization if and only the attendant probability is drawn from the set P.

Chambers and Echenique (2016, Section 8.1.3) for objective probabilities. In particular, the necessary and sufficient condition is a restriction on products of ratios of risk-neutral prices. The risk-neutral prices under uncertainty play the role otherwise played by objective probabilities. Polisson, Quah, and Renou (2020) note that each observation from a budget set can give rise to infinitely many revealed-preference implications, and develop conditions under which a finite number of conditions suffice to determine whether the data are consistent with a model of choice. In contrast to the latter two papers, our data set consists of acts with known utility values and their certainty equivalents.

Chambers and Echenique (2016, Section 8.4) examine the rationalization of choice between acts by formulations in which (as with maxmin expected utility) the probability with which the expected utility of an act is evaluated can depend on the act. The key result is their Proposition 8.9, stating that a collection of revealed preferences $\succsim_R$ has an expected utility representation in which probabilities can depend on acts if and only if $\succsim_R$ satisfies a quite weak condition known as uniform monotonicity,[7] which is characterized in Bossert and Suzumura (2012).[8] The proof relies on the ability to assign to any act x an arbitrary probability distribution by which it is to be evaluated. In contrast, maxmin expected utility puts restrictions on which probabilities are used to evaluate which acts.

Finally, Chambers, Echenique, and Saito (2016) also assume that utilities are given and focus on the existence of probabilities. They suppose that the data consist of a collection $\{x^a, p^a\}_{a=1}^A$, where p^a is a price vector and the presumption is that the state-contingent plan x^a was chosen from the set $\{x | p^a x \leq p^a x^a\}$. Chambers, Echenique, and Saito (2016) show (Theorem 2) that the data can be characterized by maxmin expected utility if and only if the Arrow–Debreu prices attached to the assets, when normalized, can be interpreted as probabilities satisfying the maxmin expected utility optimality conditions.

One might view this paper and Chambers, Echenique, and Saito (2016) as providing revealed-preference tools to ascertain when a collection of decisions inconsistent with Bayesian expected utility maximization can nonetheless be rationalized by some "reasonable" (in this case, maxmin expected utility) model. Complementary analyses include (among others) Caplin and Dean (2015) (expected utility with information costs), Masatlioglu, Nakajima, and Ozbay (2012) (consideration sets), and Ok, Ortoleva, and Riella (2015) (reference dependence).

---

[7]Choices satisfy uniform monotonicity if, for two acts x and y, if $x(\omega) \succsim_R y(\omega')$ for all states ω and ω' (think of this as the indication that the worst outcome under x, taken as a constant act, is better than the best outcome under y, also taken as a constant act), then $x \succsim_R y$.

[8]Bossert and Suzumura (2012) assume there is uncertainty, no probabilities are given, and choice sets are arbitrary sets of state-contingent consumption plans, taking either monetary values or identifying commodity allocations. They are concerned with when there exists a preference relation exhibiting certain general properties that rationalize the data, but do not investigate whether these preferences might correspond to something like expected utility maximization or maxmin expected utility maximization.

4. Discussion

A textbook application of normative decision theory suggests that the decision maker specify her problem—including decision variables, objective functions, constraints, and beliefs—and invoke an algorithm to solve it. This scenario fits many decisions, characterized by "obvious" models gleaned from copious experience and applied to sharply defined problems. In many other scenarios, there is no obvious model of the problem. A venture capitalist wondering how much to invest in a new technology may have no idea how to assign probabilities to the various outcomes.

This paper argues that decision theory is useful even in such vague problems by acting as a *coherence test* (as do Gilboa, Postlewaite, Samuelson, and Schmeidler (2018) and Gilboa, Rouziou, and Sibony (2018)). We can imagine the venture capitalist making up her mind by whatever mix of intuition, advice, imitation, calculation, and guesswork she has available, and then checking her judgment by asking whether she can justify the decision using a model. We assume that the set of acts, the utility function, the constraints, and the states of the world are relatively straightforward, and ask whether there exist beliefs that justify the tentative choice.

The quest to justify a decision by a formal model may help individuals reach decisions that they end up liking better than those suggested by their intuition. More importantly, when we consider agents who are part of institutions, the need for justification may be an essential part of responsible, transparent decision making. As suggested in the Section 1, people may trust the intuition and the ability to "read" the markets of a manager who invests money on their behalf, but nonetheless may not be happy with just any decision. They may well ask, "But why did you do that? *What were you thinking* exactly?" The present paper provides the tools to determine which decisions can be justified if called upon to do so.

Our analysis assumes that the utility function is fully specified, whereas there are no constraints on beliefs. Both assumptions are somewhat extreme. One may wish to allow for a range of utility functions, for example, allowing for different weights on an economy's gross domestic product (GDP) and inequality, while ruling out other variables, such as a president's chances of reelection. Similarly, some ranges of beliefs are reasonable while others would be too outlandish to endorse. For example, it is difficult to justify an investment in a preemptive strike on Mars, even though *some* beliefs make it an optimal choice. While our formal analysis is silent on the content of beliefs, such constraints can be embedded in the model in the form of act–evaluation pairs that the decision maker is constrained to accept.

We believe that in many cases there are obvious benefits to protocols that require justification of decisions by decision-theoretic models. Before a pension fund invests in exotic assets, before a president decides to bomb or invade another country, before countries embrace or abandon climate change policies, we would ask that they put the tools provided in this paper to work and identify at least one set of coherent probabilities rationalizing their decision. Clearly, one can also point to drawbacks of such protocols, ranging from cumbersome bureaucracy to corruption. We view our contribution as raising the question and laying some theoretical foundations for decision protocols that might be needed should one choose to implement them.

Appendix: Proof

Let there be given acts $(x^a)_{a=1}^A$ and certainty equivalents $(c^a)_{a=1}^A$.

It suffices to show that, for each $\tilde{a} \in A$, the following statements are equivalent.[9]

(A.$\tilde{a}$) There exists $p^{\tilde{a}} \in \Delta(\mathcal{S})$ such that

$$p^{\tilde{a}} \cdot x^a \geq c^a \quad \forall a \in A$$
$$p^{\tilde{a}} \cdot x^{\tilde{a}} = c^{\tilde{a}}.$$

(B.$\tilde{a}$) There is no collection of real numbers $(\lambda^a)_{a=1}^A$ such that $\lambda^a \geq 0$ for $a \neq \hat{a}$ and

$$\sum_{a=1}^{A} \lambda^a \big(x^a - c^a\big) \ll 0. \tag{2}$$

Indeed, if (A) holds, then we can choose, for each $\tilde{a} \in A$, $p^{\tilde{a}} \in P \subseteq \Delta(\mathcal{S})$ such that $p^{\tilde{a}} \cdot x^a \geq c^a$ for all a with equality for $\tilde{a}$, so that (A.$\tilde{a}$) would hold. Conversely, if (A.$\tilde{a}$) holds, then taking P to be the convex hull of $(p^{\tilde{a}})_{\tilde{a} \in A}$ would yield (A). Condition (B) is clearly equivalent to (B.$\tilde{a}$) holding for all $\tilde{a} \in A$.

To see that (A.$\tilde{a}$) and (B.$\tilde{a}$) are equivalent for each $\tilde{a} \in A$, fix such an $\tilde{a}$. Construct a two-person zero-sum "debate" game. Player I's (the decision maker) set of strategies is $\mathcal{S}$. Player II (the skeptic) has $A + 1$ strategies: A are denoted by a^+ for each $a \in \mathcal{A}$ and another one is $\tilde{a}^-$. The payoff to player I, should she pick $s \in \mathcal{S}$, is $x_s^a - c^a$ if player II plays a^+ and it is $c^{\tilde{a}} - x_s^{\tilde{a}}$ if player II plays $\tilde{a}^-$. Intuitively, the skeptic can claim that any asset a is overvalued (by choosing the strategy a^+). Player I earns the positive payoff $x_s^a - c^a$ if she points to a state s where the payoff to act a exceeds its certainty equivalent and earns the negative payoff $x_s^a - c^a$ if she points to a state in which the payoff to act a falls short of its certainty equivalent. The skeptic can also claim the asset $\tilde{a}$ (but only that asset) is undervalued (by choosing $\tilde{a}^-$).

We argue that each of (A.$\tilde{a}$) and (B.$\tilde{a}$) is equivalent to the claim that the value of the game is zero. This ensures that (A.$\tilde{a}$) and (B.$\tilde{a}$) are equivalent, which in turn ensures that conditions (A) and (B) are equivalent.

First, note that player II can place probability of 0.5 on each of $\tilde{a}^-$ and $\tilde{a}^+$, in which case the payoff to player I is zero. Hence the value of the game is bounded above by zero. We thus only need to show that each of (A.$\tilde{a}$) and (B.$\tilde{a}$) is equivalent to the claim that player I can guarantee herself a zero payoff.

Starting with (A.$\tilde{a}$), if (as in (A.$\tilde{a}$)) there exists $p^{\tilde{a}} \in \Delta(\mathcal{S})$ such that

$$p^{\tilde{a}} \cdot x^a \geq c^a \quad \forall a \in A$$
$$p^{\tilde{a}} \cdot x^{\tilde{a}} = c^{\tilde{a}},$$

[9] We thank an anonymous referee who suggested to us a simplification of the original proof. The proof that follows is not precisely the one suggested by the referee, but it is in the same spirit: apply a duality argument to each act separately. Earlier version of the paper used a duality argument for all acts simultaneously, resulting in a significantly longer proof.

then $p^{\tilde{a}}$ guarantees an expected payoff of at least zero to player I. Conversely, if there exists a mixed strategy of player I, $p \in \Delta(\mathcal{S})$, that guarantees her zero, then we have $p(x^a - c^a) \geq 0$ for all a, whereas equality has to hold for $\tilde{a}$ (because the expected payoff of p given that player II plays $\tilde{a}^-$ is the negative of the expected payoff given $\tilde{a}^+$), ensuring condition (A.$\tilde{a}$).

Next consider (B.$\tilde{a}$). As we know that the value of the game is nonpositive, it is zero if and only if it is not negative. Moreover, the value is negative if and only if player II has a mixed strategy that drives player I's payoff strictly below zero for every state s. Hence, the value of the game is zero if and only if there does not exist a collection of probabilities $(\pi^{a^+})_{a^+\in A}$ and $\pi^{\tilde{a}^-}$ (i.e., nonnegative numbers with $[\sum_{a^+\in A}\pi^{a^+}] + \pi^{\tilde{a}^-} = 1$) such that for every s,

$$\left[\sum_{a^+\in A}\pi^{a^+}\left(x_s^{a^+} - c_s^{a^+}\right)\right] + \pi^{\tilde{a}^-}\left(c_s^{\tilde{a}^-} - x_s^{\tilde{a}^-}\right) < 0,$$

which (noting that multiplying $(\pi^{a^+})_{a^+\in A}$ and $\pi^{\tilde{a}^-}$ by a positive constant preserves the inequality) is equivalent to the claim that there exists no collection of nonnegative numbers $(\pi^{a^+})_{a^+\in A}$ and $\pi^{\tilde{a}^-}$ such that

$$\left[\sum_{a^+\in A}\pi^{a^+}\left(x^{a^+} - c^{a^+}\right)\right] + \pi^{\tilde{a}^-}\left(c^{\tilde{a}^-} - x^{\tilde{a}^-}\right) \ll 0.$$

Letting $\lambda^a = \pi^{a^+}$ for all $a \neq \tilde{a}$ and letting $\lambda^{\tilde{a}} = \pi^{\tilde{a}^+} - \pi^{\tilde{a}^-}$, this is equivalent to (2).

References

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath (1999), "Coherent measures of risk." *Mathematical Finance*, 9, 203–228. [511]

Berger, Loïc, Nicolas Berger, Valentina Bosetti, Itzhak Gilboa, Lars Peter Hansen, Christopher Jarvis, Massimo Marinacci, and Richard D. Smith (2021), "Rational Policy during a Pandemic." *Proceedings of the National Academy of Sciences*, 118(4). [508]

Border, Kim C. (1992), "Revealed preference, stochastic dominance, and the expected utility hypothesis." *Econometrica*, 56, 20–42. [514]

Bossert, Walter and Kotaro Suzumura (2012), "Revealed preference and choice under uncertainty." *SERIEs*, 3, 247–258. [515]

Caplin, Andrew and Mark Dean (2015), "Revealed preference, rational inattention, and costly information acquisition." *American Economic Review*, 105, 2183–2203. [515]

Chambers, Christopher and Federico Echenique (2016), *Revealed Preference Theory*. Cambridge University Press, Cambridge. [514, 515]

Chambers, Christopher, Federico Echenique, and Kota Saito (2016), "Testing theories of financial decision making." *Proceedings of the National Academy of Science*, 113, 4003–4008. [515]

Chambers, Christopher, Ce Liu, and Seung-Keun Martinez (2014), "A test for risk-averse expected utility." *Journal of Economic Theory*, 163, 775–785. [514]

Echenique, Federico and Kota Saito (2015), "Savage in the market." *Econometrica*, 83, 1467–1495. [514]

Edgeworth, Frances Y. (1881), *Mathematical Psychics*. Kegan Paul, London. [514]

Fishburn, Peter C. (1975), "Separation theorems and expected utility." *Journal of Economic Theory*, 11, 16–34. [514]

Gilboa, Itzhak, Andrew Postlewaite, Larry Samuelson, and David Schmeidler (2018), "Economics: Between prediction and criticism." *International Economic Review*, 59, 367–390. [516]

Gilboa, Itzhak, M. Rouziou, and O. Sibony (2018), "Decision theory made relevant: Between the software and the shrink." *Research in Economics*, 72, 240–250. [516]

Houthakker, Hendrik S. (1950), "Revealed preference and the utility function." *Economica*, 17, 159–174. [514]

Kubler, Felix, Larry Selden, and Xiao Wei (2014), "Asset demand based tests of expected utility maximization." *Journal of Economic Theory*, 104, 3459–3480. [514]

Lin, Yi-Hsuan (2019), *Revealed 'Betweenness' Preferences Over Lotteries*. Technical report, Intitute of Economics, Academia Sinica. [514]

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay (2012), "Revealed attention." *American Economic Review*, 102, 2183–2205. [515]

Ok, Efe, Pietro Ortoleva, and Gil Riella (2015), "Revealed (p)reference theory." *American Economic Review*, 105, 299–321. [515]

Polisson, Matthew, John K.-H. Quah, and Ludovic Renou (2020), "Revealed preferences over risk and uncertainty." *American Econnomic Review*, 110, 1782–1820. [514, 515]

Richter, Marcel K. and Leonard Shapiro (1978), "Revelations of a gambler." *Journal of Mathematical Economiccs*, 5, 229–244. [514]

Ross, Stephen A. (2005), *Neoclassical Finance*. Princeton University Press, Princeton. [509]

Samuelson, Paul A. (1938), "A note on the pure theory of consumer's behavior." *Economica*, 5, 61–71. [514]

Co-editor Federico Echenique handled this manuscript.

Manuscript received 21 January, 2021; final version accepted 14 July, 2021; available online 26 July, 2021.