

Statistical Decisions under Ambiguity

Jörg Stoye*

New York University

July 29, 2009

Abstract

This paper provides unified axiomatic foundations for the most common optimality criteria in statistical decision theory. It considers a decision maker who faces a number of possible models of the world (possibly corresponding to true parameter values). Every model generates objective probabilities, and von Neuman-Morgenstern expected utility applies where these obtain, but no probabilities of models are given. This is the classic problem captured by Wald's (1950) device of risk functions.

In an Anscombe-Aumann environment, I characterize Bayesianism (as a backdrop), the statistical minimax principle, the Hurwicz criterion, minimax regret, and the "Pareto" preference ordering that rationalizes admissibility. Two interesting findings are that c-independence is not crucial in characterizing minimax principle and that the axiom which picks minimax regret over maximin utility is von Neumann-Morgenstern independence.

Keywords: statistical decisions, risk functions, ambiguity, maximin utility, minimax regret.

JEL classification codes: C44, D81.

*This paper incorporates material from chapter 4 of my doctoral dissertation submitted at Northwestern University. I am indebted to Peter Klibanoff and Chuck Manski for their advice and to Karl Schlag for especially helpful comments. I also thank Itai Sher, an associate editor and two referees, and audiences in Bonn (Verein für Socialpolitik annual meeting, Max Planck Institute for Research on Collective Goods), Florence (European University Institute), London (Econometric Society 9th World Congress), and Montreal (Canadian Economics Association annual meeting). Of course, any and all errors are mine. Financial support through the Robert Eisner Memorial Fellowship, Department of Economics, Northwestern University, as well as a Dissertation Year Fellowship, The Graduate School, Northwestern University, is gratefully acknowledged. Address: Jörg Stoye, Department of Economics, New York University, 19 W. 4th St., New York, NY 10012, j.stoye@nyu.edu.

1 Introduction

“[T]here are two types of uncertainty: one as to the hypothesis, which is expressed by saying that the hypothesis is known to belong to a certain class or model, and one as to the future events or observations given the hypothesis, which is expressed by a probability distribution.” (Arrow 1951, p. 418)

This paper provides axiomatic foundations for optimality criteria in statistical decision theory. It therefore extends a literature that began and peaked in the 1950’s. The meantime saw dramatic progress in decision theory, but much of it was in a form geared toward economic theorists rather than users of statistical decision theory. In order to address the latter audience, this paper deviates from much other recent work both in the framework used and in the type of results generated.

As to the framework, I assume that a decision maker simultaneously faces nonprobabilistic ambiguity or “Knightian uncertainty” (about the true model) and probabilistic uncertainty or risk (given the true model). This idea is expressed in the preceding quotation. It is formalized by using an Anscombe-Aumann (1963) setting not just for simplicity, but as the appropriate model. The decision maker resolves risk by expected utility, but may react differently to ambiguity. This captures what statisticians do when they write down *risk functions*. Different criteria in statistical decision theory differ in the way these risk functions are evaluated. I axiomatize Bayesian and maximin utility (i.e. statistical minimax), α -maximin (a.k.a. the Hurwicz criterion), and minimax regret preferences as well as a “Pareto” preference that rationalizes the concept of admissibility. All of these are characterized within an axiomatic system that consists entirely of well-known axioms and weakenings thereof. Joint imposition of all axioms generates a contradiction, and the different optimality criteria are characterized by careful relaxation of the axioms in different directions.

The remainder of this paper is structured as follows. Section 2 is devoted to explaining the setup and notation and to further elaborating its relation to statistical decisions as well as differences to the existing literature. Section 3 contains the axiomatic treatment; section 4 concludes. Appendices collect proofs and demonstrate tightness of the main theorem.

2 The Decision Theoretic Framework

2.1 Setup and Notation

Consider a set \mathcal{S} of states of the world s , endowed with an algebra Σ of events E, F , etc.; a set \mathcal{X} of outcomes x ; and a set \mathcal{F} of possible acts (e.g. decision rules or estimators) f, g , etc. The objects \mathcal{S} , \mathcal{X} , and Σ are only restricted as follows: For all results, there must be at least three events and two

outcomes, and for one specific result (theorem 1(v)), Σ must be infinite. An act f is a Σ -measurable, finite step function from states s onto probability measures $f(s) \in \Delta\mathcal{X}$; $\Delta(\cdot)$ denotes the set of finite probability mixtures over the argument. I embed $\Delta\mathcal{X}$ in \mathcal{F} by writing p^* for the constant act defined by $p^*(s) = p \in \Delta\mathcal{X}$ for every s . Mixtures between acts are written in the usual way and are identified with statewise mixtures, i.e. $h = \lambda f + (1 - \lambda)g$ is the act generated by performing f with probability λ and g otherwise and is characterized by $h(s) = \lambda f(s) + (1 - \lambda)g(s)$ for every s . The decision maker can choose from a nonempty, finite menu $M \subseteq \mathcal{F}$. She can randomize, thus her set of feasible acts is really the convex hull ΔM of M . The notation $\lambda M + (1 - \lambda)g$ denotes the menu generated by replacing every element f of M with the analog mixture. The object to be axiomatized is a family of preference relations \succsim_M , where \succsim_M is a binary relation on $\Delta M \times \Delta M$ from which \succ_M and \sim_M are derived in the usual manner. All in all, the setting is just as in Gilboa and Schmeidler (1989) as well as many other references, with the existence of menus being made explicit. I also define preferences that condition on a state of the world being realized, thus $f|s \succsim_M g|s$ means that the decision maker would prefer f over g in menu M if she knew that s has occurred. As is standard in the literature, a subtle notion of dynamic consistency (or state independent utility) is imposed by equating choice conditional on states with choice from the corresponding constant acts.

Maintained axioms will ensure that von Neumann-Morgenstern expected utility applies to constant acts p^* , thus there exists $U : \mathcal{X} \rightarrow \mathbb{R}$ s.t. those acts are ranked according to $\int U(x)dp^*$. The range of U can be bounded or unbounded; in particular, best and worst possible outcomes may or may not exist.¹ One can then think (without loss of generality, as will be shown) in terms of “utility acts” $u \circ f$ defined by $u \circ f(s) = \int U(x)df(s)$. An act will be called *admissible* within a menu M if no element of M generates strictly higher utility $u \circ f$ in every state. This definition is slightly more restrictive than is common in statistics, more on which below.

As this paper is motivated by statistical decision problems, it is of interest to clarify the relation to statistical terminology (see also Hirano (2008) and Stoye (2009a)). The primitive objects in statistical decision theory are a parameter $\theta \in \mathbb{R}^j$, a sample space \mathcal{Z} with typical element z , a decision rule $\delta : \mathcal{Z} \rightarrow \mathcal{A}$, where \mathcal{A} collects primitive actions available to the decision maker, and a loss function $L : \mathbb{R}^j \times \mathcal{A} \rightarrow \mathbb{R}^+$. To translate to the present setup, identify decision rules δ with acts f , loss L with (the negative of) utility U , and states s with couplets $(\theta, P(z)) \in \mathbb{R}^j \times \Delta\mathcal{Z}$, then the *risk function* of a decision rule is a mapping $r : \mathcal{S} \rightarrow \mathbb{R}$ that maps states s onto risks $r(\delta, s) \equiv \int L(\theta, \delta)dP(z)$ and that corresponds to utility acts. In particular, these utility acts are of intrinsic interest here, whereas in standard decision theory, they are generally considered auxiliary and are hidden in the proofs.

¹Puppe and Schlag (2009) relax an assumption made here, namely that the outcome space is state independent. The weakest structure actually needed to generate results is identified in their paper. A simple sufficient condition is that there exists constant acts p^* and q^* with $p^* \succ_{\{p^*, q^*\}} q^*$.

The following preference orderings will be axiomatized.

Definition 1 Bayesianism (“Subjective Expected Utility”)

There exists a probability measure π over \mathcal{S} (the “prior”) s.t.

$$f \succsim_M g \iff \int u \circ f(s) d\pi \geq \int u \circ g(s) d\pi.$$

Definition 2 Maximin Utility

$$f \succsim_M g \iff \min_{s \in \mathcal{S}} u \circ f(s) \geq \min_{s \in \mathcal{S}} u \circ g(s).$$

Definition 3 α -Maximin Utility (“Hurwicz Criterion”)

There exists $\alpha \in [0, 1]$ s.t.

$$f \succsim_M g \iff \alpha \max_{s \in \mathcal{S}} u \circ f(s) + (1 - \alpha) \min_{s \in \mathcal{S}} u \circ f(s) \geq \alpha \max_{s \in \mathcal{S}} u \circ g(s) + (1 - \alpha) \min_{s \in \mathcal{S}} u \circ g(s).$$

Definition 4 Minimax Regret

$$f \succsim_M g \iff \max_{s \in \mathcal{S}} \{ \max_{f^* \in M} u \circ f^*(s) - u \circ f(s) \} \leq \max_{s \in \mathcal{S}} \{ \max_{f^* \in M} u \circ f^*(s) - u \circ g(s) \}.$$

Definition 5 Strict Pareto

$$f \succ_M g \iff u \circ f(s) > u \circ g(s), \forall s \in \mathcal{S}$$

$$f \sim_M g \iff f \not\succ g \wedge g \not\succ f.$$

The Bayesian model is well known not only to theorists, but also in statistical applications, where it corresponds to the ranking of decision rules by Bayes risk. The focus of this paper is on criteria that avoid probabilistic treatment of ambiguity however. The best-known among these is probably maximin utility, i.e. the statistical minimax principle, which was introduced to statistical decision theory by Wald’s work culminating in Wald (1950). Minimax regret was first suggested in Savage’s (1951) reading of Wald (1950), is favorably discussed (with interesting examples) by Savage (1954) and Berger (1985), and recently saw some revival in econometrics due to Manski (2004); see Stoye (2009b) for references. Of course, it is menu-dependent: preference between f and g may depend on the menu from which the two can be chosen, and it can happen that f is chosen from the menu $\{f, g\}$, yet g is chosen from $\{f, g, h\}$.² Finally, strict Pareto is interesting because it selects the admissible acts (in this paper’s strict sense of the term) as choice correspondence.

²For a striking example, assume there are three states $\{s_1, s_2, s_3\}$, identify acts f with vectors $(u \circ f(s_1), u \circ f(s_2), u \circ f(s_3))$, and consider the acts $f \equiv (1, 2, 3)$ and $g \equiv (3, 4, 2)$. It can be verified that $g \succ_{\{f, g\}} f$, that the MR-act is to choose g with probability $2/3$ and that the pure act f is the least preferred choice in $\Delta(\{f, g\})$. If one adds $h \equiv (-10, -10, 5)$ to the picture, then the ranking is $f \succ_{\{f, g, h\}} g \succ_{\{f, g, h\}} h$, and the MR-act is to choose f with certainty, i.e. an act that was feasible yet *worst* absent h . Arrow (1951), Chernoff (1954), and Milnor (1954) provide further examples.

2.2 Comparison to Related Literature

The criteria proposed here minimize utility, maximize regret, or compare the performance of two acts uniformly over the entire state space. This contrasts with other work in decision theory, which uses sets of priors. In particular, maxmin expected utility (Gilboa and Schmeidler (1989)) is represented by $\min_{\pi \in C} \int u \circ f(s) d\pi$, where C is a revealed set of priors. No claims are made regarding the psychological accuracy of the model, and C should not be interpreted as describing the agent’s actual beliefs.

This is all as it should be from a revealed preference point of view: In most descriptive applications, agent’s beliefs are unobservable except insofar as they are revealed through choice behavior. But this concern does not apply to statistical decision theory. Statistician’s beliefs are certainly observable to themselves and usually specified explicitly before actions are contemplated. A clean example of this is the progression from Manski’s (2000) description of the ambiguity inherent in a decision problem to his use of minimax regret (Manski (2004, 2005)) to solve that problem. What’s more, if maximin-type decision rules are advocated, they tend to compute some worst-case contingency over the entire state space rather than over a set of priors (e.g., Wald (1950), Manski (2004)). This is just what is modelled here; indeed, it is pretty much unavoidable for a frequentist.

Accommodating these concerns comes at a cost because it requires more axioms and renders the specification of \mathcal{S} a more sensitive matter: Σ is really the algebra of plausible, and not merely conceivable, events, where a characterization of “plausible” exceeds the scope of this paper. The benefit is that subject to this limitation, the present axioms characterize what statistical decision makers actually do.³

In the case of minimax regret, framing the discussion in terms of preferences sacrifices an “as if”-perspective in yet another sense: The minimax regret preference ordering, and hence any axiomatization of it, entails statements that are vacuous in terms of observable choices, like the second part of $f \succ_{\{f,g,h\}} g \succ_{\{f,g,h\}} h$. A revealed preference characterization should, therefore, focus on choice correspondences as in Hayashi (2008). This difference does not much affect substantive results: Stoye (2007, theorem 3) shows that the present characterization of minimax regret has a close analog in the language of choice correspondences. I here stick with preferences for several reasons: First, in modelling statistical decision makers, it would appear that preferences and not choices are primary; statisticians

³The idea of taking sets of priors to be non-behavioral has recently surfaced in other decision theoretic work. Gajdos et al. (2004) take as given a set of priors that contains a reference prior and then axiomatize maximin utility with respect to a behavioral C that is constrained by the exogenous objects. In Gajdos et al. (2008), C is furthermore centered on an endogenously identified reference prior. Ahn (2008) and Olszewski (2007) take the sets of lotteries induced by acts as primitives, thus avoiding state spaces altogether. Neither C nor \mathcal{S} explicitly appear in their approach, but a non-behavioral take on beliefs is implied; indeed, Olszewski (2007) uses the phrase “objective ambiguity.” Their approach does not permit a discussion of minimax regret or statewise dominance, however, because these use state-space information. Finally, in research subsequent to earlier versions of this paper, Puppe and Schlag (2009) follow the present approach.

want to choose what they prefer and not vice versa.⁴ Second, a focus on preferences generates tight links to the existing literature. Indeed, one selling point of this paper is that the trade-off between the above optimality criteria will be cast as trade-off between very familiar axioms. Finally, statistical decision rules are potentially randomized, and this possibility is not just theoretic – with finite sample spaces, most non-Bayesian criteria yield optimal statistical decision rules that randomize conditional on nonzero events in sample space. Randomization is trivially accommodated in the present framework but poses significant difficulties for representation results about choice correspondences: It implies that these correspondences are defined only on convex sets, drastically reducing the domain on which axioms can be used. Accordingly adapted axiomatizations are the subject of ongoing research.

3 Axiomatic Treatment

3.1 Axioms and a Representation Result

This section is devoted to introducing axioms and stating the representation theorem. The following axioms will be maintained throughout.

Axiom 1 *Order*

\succsim_M is complete and reflexive.

Axiom 2 *Transitive Extension of Monotonicity (TM)*

For any f, g , and menu $M \supset \{f, g\}$ s.t. $f|s \succsim_M g|s$ for all states $s \in \mathcal{S}$, one has $h \succsim_M f \Rightarrow h \succsim_M g$ and $g \succsim_M h \Rightarrow f \succsim_M h$.

Axiom 3 *Nontriviality*

$$\exists f, g, M : f \succ_M g.$$

Axiom 4 *Mixture Continuity (Archimedean Property)*

$$f \succ_M g \succ_M h \implies \exists \lambda, \gamma \in (0, 1) : \lambda f + (1 - \lambda)h \succ_M g \succ_M \gamma f + (1 - \gamma)h.$$

Axiom 5 *Independence for Constant Acts (ICA)*

Let the menu M consist of constant acts only, then

$$p^* \succsim_M q^* \iff \lambda p^* + (1 - \lambda)r^* \succsim_{\lambda M + (1 - \lambda)r^*} \lambda q^* + (1 - \lambda)r^*$$

⁴One might take issue here if one's interest in revealed preference derives from a conviction that choices are a fundamentally superior foundation of economic theory, rather than just what's typically observable. This view is controversial to say the least, though. This author agrees with Sen (1973) that the "rationale of the revealed-preference approach lies in the assumption of revelation and not in doing away with the notion of underlying preferences, despite occasional noises to the contrary." See also Hausman (2000).

for any $\lambda \in (0, 1]$ and constant act r^* .

Axiom 6 Independence of Irrelevant Alternatives for Constant Acts

Let the menus M and N consist of constant acts only, then

$$p^* \succsim_M q^* \iff p^* \succsim_N q^*.$$

Axiom 7 Independence of Never Strictly Optimal Alternatives (INA)

Call an act $h \in M$ strictly potentially optimal in M if there exists a state $s \in \mathcal{S}$ s.t. $h|s \succ_M f|s$ for any $f \in M$. For any h that is not strictly potentially optimal in M ,

$$f \succsim_M g \iff f \succsim_{M \setminus \{h\}} g.$$

The reader will find some of these axioms familiar but weaker than usual, in particular when they are applied to non-constant acts. For example, transitive extension of monotonicity implies the usual definition of monotonicity (by letting $f = g$ and also using reflexivity) but implies transitivity only on the sub-domain of constant acts. Other than the axioms' general weakness, one aspect of interest is the formulation of independence, where the third act was mixed into the entire menu. To see why, recall that independence is frequently motivated by the following thought experiment: An agent prefers p^* over q^* , but then she is told that her choice will be actualized only if heads occur in a previous coin toss; otherwise, r^* will occur whatever her intentions. Then it can be argued that this information should not reverse her preferences, hence these should obey independence. But in the thought experiment, r^* would be mixed into all options.

It is easy to see that the axioms imply a von Neumann-Morgenstern representation of preferences on menus that consist of constant acts. What's more, the following statement holds.

Lemma 1 *Let axioms 1-7 hold. Then*

(i) *There exists a nonconstant (and unique up to positive affine transformation) function $U : X \rightarrow \mathbb{R}$ s.t.*

$$p^* \succsim_M q^* \iff \int U(x) dp^* \geq \int U(x) dq^*$$

for any menu M and constant acts $p^*, q^* \in M$.

(ii) *Define*

$$u \circ f(s) = \int U(x) df(s).$$

Let f, \bar{f}, g , and \bar{g} be s.t. $u \circ f = u \circ \bar{f}$ and $u \circ g = u \circ \bar{g}$. Then

$$f \succsim_{M \cup \{f, g\}} g \iff \bar{f} \succsim_{M \cup \{\bar{f}, \bar{g}\}} \bar{g}, \forall M.$$

In words, not only does expected utility apply to constant acts, but attention can be restricted to utility acts. Results of this sort stand at the beginning of many representation theorems in Anscombe-Aumann settings. This one, while mostly preliminary to what follows, is of some independent interest for two reasons. First, it relies on substantially weaker axioms than is typical. Second, recall that utility acts correspond to risk functions, a basic building block of statistical decision theory. One now sees that a statistician's focus on them as descriptions of a decision rule's performance can be justified by the above axioms. Most of these are very weak, and while the implication in lemma 1 is generally one-sided, use of a risk function even implies the perhaps least compelling among them, namely independence for constant acts. I will, therefore, take axioms 1-7 for granted and cast the choice between different decision criteria as choice between different strengthenings of them.

Controversial discussions of Bayesianism show that some decision makers will want to avoid prior probabilistic judgment about states. Axioms that exclude such judgment were first proposed by Arrow and Hurwicz (1972, written 20 years prior) and were formalized for a context similar to the present one, but with a finite state space, by Milnor (1954). The below formulation reflects their further adaptation to the state space considered here and compares to Cohen and Jaffray (1980).

Axiom 8 Symmetry

For any menu M , let $E, F \in \Sigma \setminus \{\emptyset\}$ be disjoint events s.t. for any $f \in M$, $u \circ f$ is constant on E as well as F . Define f' by

$$u \circ f'(s) = \begin{cases} u \circ f(s)|_{s \in E}, & s \in F \\ u \circ f(s)|_{s \in F}, & s \in E \\ u \circ f(s) & \text{otherwise} \end{cases} .$$

Let M' be the menu generated by replacing every act $f \in M$ with f' . Then

$$f \succsim_M g \iff f' \succsim_{M'} g'.$$

The idea behind symmetry is that a preference ordering should not impose prior beliefs by implicitly assigning different likelihoods to different events. This is implausible if one has available, and wishes to consider, sharp prior information about states. Indeed, a Bayesian analysis would then seem appropriate. If no prior information about states exists, the restriction makes sense since in its absence, a decision criterion would be sensitive to arbitrary manipulations of the state space, either by relabeling states or by duplicating some via conditioning on trivial events. These considerations should be especially clear to frequentists since they are closely related to standard objections against “noninformative” priors.

Other axioms to be imposed are strengthenings of axioms 1-7 which recover a number of well-known axioms.

Axiom 9 Independence of Irrelevant Alternatives (IIA)

$$f \succsim_M g \iff f \succsim_N g$$

for all menus M, N .

Axiom 10 Independence

$$f \succsim_M g \iff \lambda f + (1 - \lambda)h \succsim_{\lambda M + (1 - \lambda)h} \lambda g + (1 - \lambda)h, \forall \lambda \in (0, 1).$$

Axiom 11 C-Independence

Let p^* be constant, then

$$f \succsim_M g \iff \lambda f + (1 - \lambda)p^* \succsim_{\lambda M + (1 - \lambda)p^*} \lambda g + (1 - \lambda)p^*, \forall \lambda \in (0, 1].$$

Axiom 12 Ambiguity Aversion

$$f \sim_M g \implies \lambda f + (1 - \lambda)g \succsim_M g, \forall \lambda \in (0, 1].$$

Axiom 13 Acyclicity

There exists no strict preference cycle, i.e. no set of acts f_1, f_2, \dots, f_k and menu $M \supseteq \{f_1, f_2, \dots, f_k\}$ s.t.

$$f_1 \succ_M f_2 \succ_M \dots \succ_M f_k \succ_M f_1.$$

Axiom 14 Transitivity

For any acts f, g, h and menu $M \supseteq \{f, g, h\}$,

$$f \succsim_M g \succsim_M h \implies f \succsim_M h.$$

As all of these axioms are familiar, a lengthy discussion would be redundant.⁵ Note that ambiguity aversion is frequently seen as weakening independence, but this is true here only in conjunction with IIA. Also, two considerations may make acyclicity interesting relative to transitivity. First, well-known discussions highlight that the strong intuitive appeal of transitivity is really one of acyclicity, or at most of transitive strict preference (which is implied but not imposed here); numerous plausible stories lead to chains like $f \sim g \sim h \succ f$. Second, if a concern for transitivity is mainly motivated by a desire to ensure nonempty choice correspondences, then acyclicity is the appropriate axiom because in the present setting, it is essentially equivalent to nonemptiness of choice correspondences (Bergstrom 1975).

This paper's main result is a characterization of all decision rules introduced above. It will be stated now, then related to existing findings, and then discussed in some more depth.

⁵Gilboa (2009) provides an excellent discussion of most axioms.

Theorem 1 *Characterization of Preference Orderings*

- (i) *A preference ordering fulfils axioms 1-7, IIA, independence, and transitivity iff it is Bayesian.*
- (ii) *A preference ordering fulfils axioms 1-7, symmetry, IIA, c-independence, and transitivity iff it is α -maximin utility.*
- (iii) *A preference ordering fulfils axioms 1-7, symmetry, IIA, ambiguity aversion, and transitivity iff it is maximin utility.*
- (iv) *A preference ordering fulfils axioms 1-7, symmetry, independence, ambiguity aversion, and transitivity iff it is minimax regret.*
- (v) *A preference ordering fulfils axioms 1-7, symmetry, IIA, independence, and acyclicity iff it is strict Pareto.*
- (vi) *There exists no preference ordering that fulfils axioms 1-7, symmetry, IIA, independence, and transitivity.*

Part (i) of this result is due to Anscombe and Aumann (1963) and is provided as a backdrop. Parts (ii)-(iv) owe some debt to Milnor (1954). Innovations include the embedding in an Anscombe-Aumann setup via lemma 1 and the substantial weakening of several axioms as well as restrictions on the environment. An important benefit of this is alignment: The environment as well as some axioms mirror Gilboa and Schmeidler (1989), a fact that will be useful in the discussion. An additional adjustment concerns the fact that minimax regret takes differences between expected utilities, thus seemingly presuming cardinally measurable utility. In Milnor (1954), this feature is due to an axiom that uses utility differences and therefore transparently introduces this presumption. Here, it is generated by independence, a trick that was anticipated by Chernoff (1954) but is missing in the subsequent literature (e.g., Borodin and El-Yaniv 1998). I am not aware of a precursor to (v).⁶

Some intuitions for important parts of the proof go as follows. For parts (ii) and (iii), the first major step is that by symmetry (together with monotonicity), the evaluation of an act f can only depend on its extremal utility values $\{\min_{s \in \mathcal{S}} u \circ f(s), \max_{s \in \mathcal{S}} u \circ f(s)\}$. C-independence then implies that preferences must be represented by a criterion function that is linear in these, leading directly to α -minimax. While ambiguity aversion would now select $\alpha = 0$ and hence the minimax principle, this reasoning does not inform a tight proof. A more instructive intuition is that ambiguity aversion precludes any role for $\max_{s \in \mathcal{S}} u \circ f(s)$ because the mixing of acts may differentially affect their best-case and worst-case performances. Ambiguity aversion therefore leads to minimax without the use of c-independence.

⁶In terms of preference orderings characterized, (ii) also relates to Ghirardato et al. (2004), (iii) to Gilboa and Schmeidler (1989), and (v) to Bewley (2002). All of these are technically quite different, use behavioral sets of priors, and will be discussed later.

An intuition for the characterization of minimax regret is that independence can be used to reduce all choice problems to choice from menus whose ex post utility frontier is constant with a value of zero. Thus, consider any menu M and let the “oracle” act \bar{f}_M be s.t. $u \circ \bar{f}_M(s) = \max_{f \in M} u \circ f(s)$. Then for any acts f, g in M , we have $f \succsim_M g$ iff $f/2 - \bar{f}_M/2 \succsim_{M/2 - \bar{f}_M/2} g/2 - \bar{f}_M/2$ – but the menu $(M/2 - \bar{f}_M/2)$ has said utility frontier of zero. On the accordingly reduced domain, INA effectively implies IIA. The preference ordering therefore must be maximin on the reduced domain and, undoing the transformation, minimax regret in general. A similar trick does not work for α -MU because the c -independence axiom is not invariant under the transformation. Finally, if IIA is imposed, then independence can be similarly used to argue that preference between f and g can only depend on utility differences $(u \circ f(s) - u \circ g(s))_{s \in \mathcal{S}}$. Symmetry and transitive monotonicity furthermore imply that only the maximal and minimal such difference can matter. But if some strictly negative value of $\min_{s \in \mathcal{S}} \{u \circ f(s) - u \circ g(s)\}$ can be compensated by a sufficiently high value of the corresponding maximum, then a strict preference cycle can be constructed.

3.2 Discussion

I conclude this section with a substantive discussion of some aspects of theorem 1. A brief verbal summary of the theorem goes as follows: While all of the above axioms may have some merit, imposing them simultaneously leads to a contradiction, thus something has to give. Dropping symmetry leads to Bayesianism; weakening independence leads to maximin or to the Hurwicz criterion; weakening IIA leads to minimax regret; and weakening transitivity leads to strict Pareto.

The characterization of minimax can be usefully compared to Gilboa and Schmeidler (1989) because the axioms used are precisely theirs, plus symmetry (and less c -independence). The result, on the other hand, differs from theirs by identifying the set of priors with \mathcal{S} . This feature can, therefore, be attributed to symmetry. In addition, and perhaps contrary to common wisdom, c -independence is *not* crucial when thinking about the foundations of statistical minimax. Of course, this observation is testament to the power of symmetry. The characterization of α -maximin utility can be compared to one in Ghirardato et al. (2004), but the effect of symmetry cannot be isolated quite as neatly. Using only c -independence, Ghirardato et al. (2004, theorem 4) characterize α -maximin utility with a behavioral set of priors and the feature that α can vary between acts in complex, although not unconstrained, ways. This property is also encountered in an axiomatization by Arrow and Hurwicz (1972), but is alien to the criterion’s original (Hurwicz 1951) and more common definition. To achieve constant α , Ghirardato et al. (2004) propose an additional axiom that seems to introduce a vestige of symmetry: It requires that an act’s evaluation only depend on the range of its image in utility space,

and it follows from the present axioms only after symmetry has been added.⁷ Symmetry again turns out to be very powerful in that it resolves these issues.

Parts (iv) and (v) establish another finding of methodological interest: Independence does not enforce Bayesianism and can be reconciled with symmetry, a salient example being minimax regret. Even the observation of mere consistency between minimax regret and independence is missing from most verbal discussions of minimax regret versus maximin utility, which tend to revolve around examples like those in Rawls (1999) and Berger (1985). Using independence to characterize minimax regret is one of this paper’s main contributions. Finally, it is of interest to note that the optimality criterion underlying admissibility can be characterized within the present axiomatic framework, and that this can be done by introducing a well-known and easily comprehended axiom, namely acyclicity.

Strict Pareto is not a very decisive criterion, and it might seem intuitively desirable to sharpen it to what could be called “weak Pareto” as follows: If $u \circ f(s) \geq u \circ g(s)$ for all states s , with strict inequality for at least one state s^* , then $f \succ g$. Indeed, the resulting choice correspondence reflects how statisticians typically understand admissibility. This preference violates both symmetry and continuity, however, and would therefore require a very different axiomatization. What’s more, the violation of these axioms is instructive: The above sharpening appears plausible at first glance, but together with continuity, implies positive willingness to pay for a price that is received only in state s^* , thus s^* cannot be null (nor can any other state, by the same logic). Thus, the sharpening is substantively at tension with the goal of removing any notion of prior probabilities.⁸

4 Summary and Outlook

This paper investigated the foundations of statistical decision theory for situations of simultaneous (“model”) ambiguity and (“estimation”) uncertainty. The purpose was to explore the theoretical foundations of decision criteria that treat uncertainty but not ambiguity in a probabilistic fashion. The axiomatic discussion differs from previous contributions by being more applied, using the full structure of a real-world problem to give results that are tightly specified for this problem. An Anscombe-Aumann setup is used because its distinction between probabilistic and nonprobabilistic uncertainty

⁷See Eichberger et al. (2008), however, for some caveats that limit the comparability between Ghirardato et al. (2004) and the present result. In particular, the set of priors in Ghirardato et al. (2004) is not identified separately from the (non-constant) α . Olszewski (2007) provides a very different treatment that achieves constant α .

⁸Bewley (2002), when axiomatizes a multiple priors analog of weak Pareto by directly imposing just this refinement. A side effect is that the behavioral set of priors must be interior to $\Delta\mathcal{S}$. This is inconsistent with symmetry, thus the resolution is not available here. It also means that the resulting preference ordering fails to rationalize admissibility: To pick the admissible acts as choice correspondence, the set of priors would have to be maximal, the exact set that is excluded.

captures an important feature of the intended applications. The characterizations link all objects in the decision rule to the decision maker's environment and therefore apply to maximin-criteria as they are actually used in many applications. They also work in frameworks that relax many of standard axioms, notably transitivity and IIA. The use of risk functions as summaries of a decision rule's performance was justified from a baseline set of weak axioms. Numerous known decision criteria were characterized by strengthening these axioms in different directions, with joint imposition of all axioms encountering a contradiction.

The examination of decision rules of this type offers rich opportunities for further research, some of which were recently exploited. Revealed preference axiomatizations of the minimax regret choice correspondence were provided by Hayashi (2008) and Stoye (2007), and their extensions to convex sets (i.e., randomization) is the object of current research by this author. Puppe and Schlag (2009) investigate possible relaxations of state independence of the outcome space. Finally, it would be of obvious interest – if challenging – to connect research on non-Bayesian statistical decision rules to analyses of ambiguity aversion in dynamic settings.

A Proofs

Lemma 1

(i) Follows straightforwardly by restricting attention to menus consisting of constant acts.

(ii) Define $M' \equiv M \cup \{f, g, \bar{f}, \bar{g}\}$. By the hypothesis and part (i), $f|s \sim_M \bar{f}|s$ and $g|s \sim_M \bar{g}|s$ for any s and M , hence no element of $\{f, g, \bar{f}, \bar{g}\}$ is strictly potentially optimal in M' . By axiom INA, it follows that $f \succsim_{M \cup \{f, g\}} g \Leftrightarrow f \succsim_{M'} g$ and $\bar{f} \succsim_{M \cup \{\bar{f}, \bar{g}\}} \bar{g} \Leftrightarrow \bar{f} \succsim_{M'} \bar{g}$, thus it suffices to show $f \succsim_{M'} g \Leftrightarrow \bar{f} \succsim_{M'} \bar{g}$. But both $f \succsim_{M'} g \Rightarrow \bar{f} \succsim_{M'} \bar{g}$ and $\bar{f} \succsim_{M'} \bar{g} \Rightarrow f \succsim_{M'} g$ follow from repeated application of transitive monotonicity.

Theorem 1

Preliminaries. Denote by $\mathcal{U} \subseteq \mathbb{R}$ the convex hull of the range of U . Fix any M . By finiteness of lotteries P , $u \circ f$ is finite. By finiteness of acts, $\max/\min_{s \in \mathcal{S}} \{u \circ f(s)\}$ exist and are finite as well. Recalling that U is nonconstant, assume w.l.o.g. that $[-1, 1] \subseteq \mathcal{U}$. By defining appropriate mappings from s to distributions over \mathcal{X} , one can generate acts that correspond to any pre-assigned, finite, Σ -measurable mapping from \mathcal{S} to \mathcal{U} . In particular, there is a constant act p_0^* with utility value 0. Finally, for any act f and scalar γ , γf denotes the act with $u \circ \gamma f(s) = \gamma u \circ f(s)$, which exists because it can be generated by mixture with p_0^* .

(i) A proof within (essentially) the present setting is given by Kreps (1988, theorem 7.17).

Preliminaries to (ii) and (iii). Since IIA is imposed, the menu subscript on the preference ordering can be dropped. Fix a partition of \mathcal{S} into three nonempty events $E^*, F^*, G^* \in \Sigma$. Fix any act f and define $(\underline{f}, \overline{f}) \equiv (\min_{s \in \mathcal{S}} u \circ f(s), \max_{s \in \mathcal{S}} u \circ f(s))$. For scalars u, v and events $E \in \Sigma$, let $u_E v$ denote the act that achieves utility u on event E and utility v otherwise. Define $f^* \equiv \underline{f}_{E^*} \overline{f}$. I will now show that $f^* \sim f$.

Consider any two events $E, F \in \Sigma \setminus \{\emptyset, \mathcal{S}\}$ and scalars $u, v \in \mathcal{U}$, then $u_E v \sim u_F v$. To see this, assume first that E and F are non-nested, then the claim follows from symmetry, used by identifying the events E and F in the axiom with $E \cap F^c$ respectively $F \cap E^c$ here. Now assume E and F are nested, then the claim can be similarly established by first exchanging the consequences of E and F^c (thus $u_E v \sim u_{F^c} v$), then the consequences of F and F^c (thus $u_{F^c} v \sim u_F v$), then using transitivity.

Consider any act f s.t. $u \circ f$ is not constant. Let $\underline{E}[\overline{E}] \in \Sigma$ be the event on which $u \circ f(s) = \underline{f}[\overline{f}]$. Then monotonicity implies that $\underline{f}_{\underline{E}} \overline{f} \succsim f \succsim \overline{f}_{\overline{E}} \underline{f}$. But $\underline{f}_{\underline{E}} \overline{f} \sim \overline{f}_{\overline{E}} \underline{f}$ by the previous paragraph's conclusion, hence $f \sim \underline{f}_{\underline{E}} \overline{f}$ by transitivity. Using the previous paragraph's conclusion and transitivity again, one finds $\underline{f}_{\underline{E}} \overline{f} \sim f^*$ and finally $f \sim f^*$. It therefore suffices to characterize preferences over acts of the form f^* ; these acts will be called *standardized*. Standardized acts can be described by vectors $(\underline{f}, \overline{f})$ that summarize their outcomes over E^* respectively $\{F^*, G^*\}$. This two-dimensional notation will be used whenever sufficient.

From here, the proofs take different directions.

(ii) C-independence implies that preferences are homothetic: $f \succsim_M g \Leftrightarrow \lambda f + (1-\lambda)p_0^* \succsim_{\lambda M + (1-\lambda)p_0^*} \lambda g + (1-\lambda)p_0^*$, but $\lambda f + (1-\lambda)p_0^* = \lambda f$. Preferences can, therefore, be generated by extending preferences over acts s.t. $-1 < \underline{f} \leq \overline{f} < 1$.

Let $\alpha \equiv \inf\{v : (v, v) \succsim (0, 1)\}$, then monotonicity implies that $\alpha \leq 1$. Suppose by contradiction that $\alpha < 0$, then repeated uses of monotonicity and transitivity yield $\frac{1}{2}(\alpha, \alpha) \succsim (0, 1) \succsim (0, 0) \Rightarrow \frac{1}{2}(\alpha, \alpha) \succsim (0, 0)$, but this contradicts the expected utility representation for constant acts. Hence, $\alpha \geq 0$. Now suppose by contradiction that $(\alpha, \alpha) \succ (0, 1)$. Monotonicity, transitivity, and the expected utility representation for constant acts yield $(0, 1) \succ (-1, -1)$. By continuity, there then exists $\gamma < 1$ s.t. $(\gamma\alpha - (1-\gamma), \gamma\alpha - (1-\gamma)) \succ (0, 1)$, contradicting the definition of α . A slight variation on this argument excludes $(\alpha, \alpha) \prec (0, 1)$, hence $(\alpha, \alpha) \sim (0, 1)$. By using c-independence, where p^* is identified with (α, α) , this conclusion can be extended to every point on the ray $R_\alpha \equiv \{\lambda(0, 1) + (1-\lambda)(\alpha, \alpha) : \lambda \geq 0\}$, hence this ray is part of an indifference set.

Consider now any $u \in (-1, 1)$ and define $R_u \equiv R_\alpha + (u, u) - (\alpha, \alpha)$, the ray through (u, u) that is parallel to R_α . R_u is contained in an indifference set. To see this, suppose $u < \alpha$, then the claim

follows from c-independence, where p^* is identified with $(1, 1)$ and λ is identified with $\frac{1-\alpha}{1-u}$. A similar argument applies if $u > \alpha$.

Since the collection $\{R_u\}_{u \in (-1, 1)}$ partitions the preference domain, every vector in that domain has been mapped onto exactly one ray. By the expected utility representation for constant acts, any two different rays constitute strictly ordered indifference sets. It is now easily verified that the ordering is α -MU with α as defined in this proof.

(iii) For this paragraph only, consider acts that are measurable on $\{E^*, F^*, G^*\}$ but not necessarily constant on $\{F^*, G^*\}$; denote these acts by triples (u, v, w) with the obvious interpretation. Previous arguments imply that $(\underline{f}, \bar{f}, \underline{f}) \sim (\underline{f}, \underline{f}, \bar{f}) \sim (\underline{f}, \bar{f}, \bar{f})$, thus ambiguity aversion yields $(\underline{f}, \frac{\underline{f} + \bar{f}}{2}, \frac{\underline{f} + \bar{f}}{2}) \succsim (\underline{f}, \underline{f}, \bar{f}) \sim (\underline{f}, \bar{f}, \bar{f})$ and monotonicity then $(\underline{f}, \frac{\underline{f} + \bar{f}}{2}, \frac{\underline{f} + \bar{f}}{2}) \sim (\underline{f}, \bar{f}, \bar{f})$. By induction over n (and using transitivity), the argument can be extended to show $(\underline{f}, \underline{f} + \frac{\bar{f} - \underline{f}}{2^n}, \underline{f} + \frac{\bar{f} - \underline{f}}{2^n}) \sim (\underline{f}, \bar{f}, \bar{f})$ for any natural number n . As the sequence $\{2^{-n}\}$ is dense at 0, monotonicity and transitivity then jointly imply that $(\underline{f}, \underline{f} + \lambda(\bar{f} - \underline{f}), \underline{f} + \lambda(\bar{f} - \underline{f})) \sim (\underline{f}, \bar{f}, \bar{f})$ for any $\lambda \in (0, 1]$.

Now return attention to standardized acts. The acts in the previous paragraph's conclusion are standardized, hence $(\underline{f}, \underline{f} + \lambda(\bar{f} - \underline{f})) \sim (\underline{f}, \bar{f})$ for any $\lambda \in (0, 1]$. It remains to extend this conclusion to $\lambda = 0$. Suppose by contradiction that $(\underline{f}, \bar{f}) \succ (\underline{f}, \underline{f})$. Assume first that \underline{f} is not the minimal element of \mathcal{U} . Then there exists an act (u, u) with $u < \underline{f}$, and the expected utility representation for constant acts implies that $(\underline{f}, \underline{f}) \succ (u, u)$. By continuity, there must then exist $\delta \in (0, 1)$ s.t. $(\delta \underline{f} + (1 - \delta)u, \delta \bar{f} + (1 - \delta)u) \succ (\underline{f}, \underline{f})$. Recall that

$$(\delta \underline{f} + (1 - \delta)u, \delta \underline{f} + (1 - \delta)u + \gamma) \sim (\delta \underline{f} + (1 - \delta)u, \delta \bar{f} + (1 - \delta)u), \forall \gamma > 0,$$

thus the left side of the above indifference is strictly preferred to $(\underline{f}, \underline{f})$ for any $\gamma > 0$; but this contradicts monotonicity for γ small enough. It follows that $(\underline{f}, \bar{f}) \precsim (\underline{f}, \underline{f})$ and, by monotonicity, $(\underline{f}, \bar{f}) \sim (\underline{f}, \underline{f})$.

Assume now that \underline{f} is the minimal element of \mathcal{U} , then $\underline{f} \leq -1$ by the range normalization of U . Suppose by contradiction that $(\underline{f}, 0) \succ (\underline{f}, \underline{f})$. Monotonicity, transitivity, and the expected utility representation for constant acts yield $(1, 1) \succ (\underline{f}, 0)$. By continuity, there then exists $\delta \in (0, 1)$ s.t. $(\underline{f}, 0) \succ (\delta + (1 - \delta)\underline{f}, \delta + (1 - \delta)\underline{f})$. This is consistent with monotonicity only if $\delta + (1 - \delta)\underline{f} < 0$. Now, $\delta + (1 - \delta)\underline{f}$ is not a minimal element of \mathcal{U} , hence the previous paragraph's finding and transitivity jointly imply $(\underline{f}, 0) \succ (\delta + (1 - \delta)\underline{f}, 0)$, contradicting monotonicity. Hence $f \succsim g \Leftrightarrow \underline{f} \geq \underline{g}$ as required.

Existence of three events, which has been exploited in this section, is necessary: If Σ has only two atoms, α -maximin utility with $\alpha = 0.5$ is ambiguity averse.

(iv) Fix any acts f and g in menu M . Define f' by

$$u \circ f'(s) = \frac{1}{2} \left(u \circ f(s) - \max_{f^* \in M} \{u \circ f^*(s)\} \right)$$

and g' similarly. I claim that $f \succsim_M g \Leftrightarrow f' \succsim_{\{f', g', p_0^*\}} g'$. To see this, define act \bar{f}_M by $u \circ \bar{f}_M(s) = -\max_{f \in M} \{u \circ f(s)\}$. One can then write

$$f \succsim_M g \Leftrightarrow \frac{1}{2}f + \frac{1}{2}\bar{f}_M \succsim_{\frac{1}{2}M + \frac{1}{2}\bar{f}_M} \frac{1}{2}g + \frac{1}{2}\bar{f}_M \Leftrightarrow f' \succsim_{\{f', g', p_0^*\}} g'.$$

Here, the first equivalence uses independence, and the second equivalence uses INA, noting that $\max_{f \in \frac{1}{2}M + \frac{1}{2}\bar{f}_M} u \circ f(s) = 0$ for every s .

It therefore suffices to characterize the menu-independent preference ordering \succeq , defined by $f \succeq g \Leftrightarrow f' \succsim_{\{f, g, p_0^*\}} g'$, over Σ -measurable, bounded acts with nonpositive utility range. Noting that by axiom INA, one can merge menus of the form $\{f, g, p_0^*\}$ (for different f, g) without affecting preferences, one can straightforwardly establish that axioms on \succsim_M restrict \succeq to be ambiguity averse, monotone, complete, transitive, nontrivial, and symmetric. Hence, part (iii) of this proof implies that \succeq is the maximin utility ordering. It follows that

$$\begin{aligned} f \succsim_M g &\Leftrightarrow f' \succeq g' \\ &\Leftrightarrow \min_{s \in \mathcal{S}} u \circ f'(s) \geq \min_{s \in \mathcal{S}} u \circ g'(s) \\ &\Leftrightarrow \min_{s \in \mathcal{S}} \left\{ \frac{1}{2}(u \circ f(s) - \max_{f^* \in M} \{u \circ f^*(s)\}) \right\} \geq \min_{s \in \mathcal{S}} \left\{ \frac{1}{2}(u \circ g(s) - \max_{f^* \in M} \{u \circ f^*(s)\}) \right\} \\ &\Leftrightarrow \min_{s \in \mathcal{S}} \left\{ \max_{f^* \in M} \{u \circ f^*(s)\} - u \circ f(s) \right\} \leq \min_{s \in \mathcal{S}} \left\{ \max_{f^* \in M} \{u \circ f^*(s)\} - u \circ g(s) \right\}, \end{aligned}$$

i.e. the preference ordering is minimax regret.

(v) IIA is imposed, so the reference to a menu can be dropped from notation. Fix any menu M and pair of acts f and g . Let the act $g \ominus f$ be characterized by $u \circ (g \ominus f)(s) = \frac{1}{2}(u \circ g(s) - u \circ f(s))$; this act is finite if f and g are. Let the act h be characterized by $u \circ h(s) = -u \circ f(s)$. Then repeated uses of independence and the observation that $\frac{1}{2}f + \frac{1}{2}h = p_0^*$ yield

$$f \succsim g \Leftrightarrow \frac{1}{2}f + \frac{1}{2}h \succsim \frac{1}{2}g + \frac{1}{2}h \Leftrightarrow p_0^* \succsim \frac{1}{2}g + \frac{1}{2}h \Leftrightarrow p_0^* \succsim g \ominus f.$$

It suffices, therefore, to characterize preferences relative to p_0^* . Similar to previous definitions, let act $u_E v$ achieve utility u on event E and v otherwise, define $(\underline{f}, \bar{f}) = (\min_{s \in \mathcal{S}} \{u \circ f(s)\}, \max_{s \in \mathcal{S}} \{u \circ f(s)\})$, and let $\underline{E}[\bar{E}] \in \Sigma$ be the event on which $\underline{f}[\bar{f}]$ is achieved. Then $\underline{f}_{\underline{E}} \bar{f} \geq f \geq \bar{f}_{\bar{E}} \underline{f}$, where \geq stands for strict Pareto in terms of u , and repeated uses of transitive monotonicity yield

$$\begin{aligned} \bar{f}_{\bar{E}} \underline{f} \succ p_0^* &\implies f \succ p_0^* \\ f \succ p_0^* &\implies \underline{f}_{\underline{E}} \bar{f} \succ p_0^*. \end{aligned}$$

Symmetry implies that preferences between $\underline{f}_E \bar{f}$ and p_0^* cannot depend on the identity of event E (the precise argument much resembles the one from the preliminaries to (ii) and (iii)). In particular,

$$\bar{f}_E \underline{f} \succsim p_0^* \iff \underline{f}_E \bar{f} \succsim p_0^* \iff \underline{f}_E \bar{f} \succsim p_0^*,$$

where $E \in \Sigma \setminus \{\emptyset, \mathcal{S}\}$ is some pre-assigned event. Taken together, these findings imply that

$$f \succsim p_0^* \iff \underline{f}_E \bar{f} \succsim p_0^*.$$

Independence, used with p_0^* in the place of h , furthermore implies that $f \succsim p_0^* \iff \gamma f \succsim p_0^*$ for any scalar $\gamma > 0$. Hence, it is w.l.o.g. to postulate the existence of utility acts with arbitrary range.

Suppose $(-k)_{E1} \succ p_0^*$ for some $k > 0$. Let $n = \min\{m \in \mathbb{N} : mk \geq 1\} + 1$. Partition \mathcal{S} into n events $E_1, \dots, E_n \in \Sigma \setminus \{\emptyset\}$ and consider the following utility table for n acts $\{f_1, \dots, f_n\}$:

	E_1	E_2	\dots	E_{n-1}	E_n
$\mathbf{u} \circ \mathbf{f}_1$	0	k	\dots	$(m-1)k$	mk
$\mathbf{u} \circ \mathbf{f}_2$	k	$2k$	\dots	mk	0
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
$\mathbf{u} \circ \mathbf{f}_n$	mk	0	\dots	$(m-2)k$	$(m-1)k$

Then $f_1 \succ f_2 \succ \dots \succ f_n$. Note the construction requires existence of n events, which is given because Σ is infinite. Infiniteness of Σ is needed, else the example for necessity of acyclicity given in appendix B will fail for α small enough.

(vi) From part (v), the ordering would have to be strict Pareto, which is intransitive.

B Tightness of Theorem 1

This section establishes that axioms are individually necessary. The examples given are of preference orderings that fulfil all axioms of a certain axiomatization except the one whose necessity is to be shown.

For All Orderings Necessity of nontriviality: Consider $f \sim g, \forall f, g$.

(i) **Bayesianism** Necessity of completeness: Let $f \succ^* g$ iff $\int u \circ f(s) d\pi > \int u \circ g(s) d\pi$ for all $\pi \in C$, where $C \subseteq \Delta \mathcal{S}$ is closed and convex.

Necessity of continuity: Let $\pi \in \Delta \mathcal{S}$ be a prior that assigns zero probability to event $E \in \Sigma$. Consider the Bayesian criterion with prior π , except that indifferences are broken lexicographically according to $u \circ f(s^*)$, some $s^* \in \Sigma$.

Necessity of transitivity: Consider strict Pareto.

Necessity of IIA: Consider minimax regret.

Necessity of independence: Consider maximin utility.

For All Maximin Orderings Necessity of completeness: For any ordering \succsim , let \succsim^* be the incomplete ordering that agrees with \succsim except that if $f \sim g$ and (f, g) is not ordered by weak strict Pareto, then f and g are not comparable under \succsim^* . (Notice the example violates sequential continuity but not the Archimedean axiom.)

(ii) α -Maximin Utility Necessity of monotonicity: Consider

$$f \succsim g \iff \min_{s \in \mathcal{S}} u \circ f(s) - \frac{1}{2} \max_{s \in \mathcal{S}} u \circ f(s) \geq \min_{s \in \mathcal{S}} u \circ g(s) - \frac{1}{2} \max_{s \in \mathcal{S}} u \circ g(s).$$

Necessity of continuity: For any acts f and g , let $\{E_i\}_{i=1}^I \subset \Sigma$ be a partition of \mathcal{S} that renders both f and g measurable. Let $u_1 \leq u_2 \leq \dots \leq u_I$ be a nondecreasing ordering of $(u \circ f(s) : s \in E_i)_{i=1}^I$ and let $v_1 \leq v_2 \leq \dots \leq v_N$ be defined analogously but with respect to g . Define $f \succ g \iff \exists i^* \leq I : u_i = v_i, i < i^*, u_{i^*} > v_{i^*}$, the “leximin” criterion. (Clearly the criterion does not depend on choice of $\{E_i\}_{i=1}^I$.)

Necessity of transitivity: Consider strict Pareto.

Necessity of IIA: Consider minimax regret.

Necessity of symmetry: Consider Bayesianism.

Necessity of c-independence: Consider α -maximin utility with act-dependent α : $\alpha = \phi(\bar{f})$, and $\phi : \mathbb{R} \mapsto [0, 1]$ is continuous and nondecreasing.

(iii) Maximin Utility Necessity of ambiguity aversion: Consider α -maximin utility.

For all other axioms, see (i).

(iv) Minimax Regret Necessity of monotonicity: Consider

$$\begin{aligned} f \succsim g &\iff \max_{s \in \mathcal{S}} \left\{ \max_{f^* \in M} u \circ f^*(s) - u \circ f(s) \right\} - \frac{1}{2} \min_{s \in \mathcal{S}} \left\{ \max_{f^* \in M} u \circ f^*(s) - u \circ f(s) \right\} \\ &\leq \max_{s \in \mathcal{S}} \left\{ \max_{g^* \in M} u \circ g^*(s) - u \circ g(s) \right\} - \frac{1}{2} \min_{s \in \mathcal{S}} \left\{ \max_{g^* \in M} u \circ g^*(s) - u \circ g(s) \right\}. \end{aligned}$$

Necessity of transitivity: Consider strict Pareto.

Necessity of symmetry: Consider Bayesianism.

Necessity of independence: Consider maximin utility.

Necessity of ambiguity aversion: Consider “minimin regret,” i.e.

$$f \succsim g \iff \min_{s \in \mathcal{S}} \left\{ \max_{f^* \in M} u \circ f^*(s) - u \circ f(s) \right\} \leq \min_{s \in \mathcal{S}} \left\{ \max_{f^* \in M} u \circ f^*(s) - u \circ g(s) \right\}.$$

Necessity of INA: Consider “maximin joy” (found in earlier versions of Hayashi (2008)), i.e.

$$f \succsim g \iff \min_{s \in \mathcal{S}} \left\{ u \circ f(s) - \min_{f^* \in M} u \circ f^*(s) \right\} \geq \min_{s \in \mathcal{S}} \left\{ u \circ g(s) - \min_{f^* \in M} u \circ f^*(s) \right\}.$$

(v) **Strict Pareto** Necessity of IIA: Consider minimax regret.

Necessity of independence: Consider maximin utility.

Necessity of symmetry: Consider Bayesianism.

Necessity of transitive extension of monotonicity: Let $f \succ g$ iff $f \ominus g$ (as defined in the proof) is constant with $u(f \ominus g)(s) > 0$.

Necessity of acyclicity: Consider

$$f \succ g \iff \max_{s \in \mathcal{S}} \{u \circ g(s) - u \circ f(s)\} < \alpha \max_{s \in \mathcal{S}} \{u \circ f(s) - u \circ g(s)\}$$

for some $\alpha \in (0, 1)$.

Necessity of continuity: In this case, continuity was only needed for lemma 1(i).

References

- [1] Ahn, D.S. (2008): “Ambiguity Without a State Space,” *Review of Economic Studies* 75: 3-28.
- [2] Anscombe, F.J. and R.J. Aumann (1963): “A Definition of Subjective Probability,” *Annals of Mathematical Statistics* 34: 199-205.
- [3] Arrow, K.J. (1951): “Alternative Approaches to the Theory of Choice in Risk-Taking Situations,” *Econometrica* 19: 404-437.
- [4] Arrow, K.J. and L. Hurwicz (1972): “An Optimality Criterion for Decision-Making under Ignorance,” in C.F. Carter and J.L. Ford (Eds.), *Uncertainty and Expectations in Economics: Essays in Honour of G.L.S. Shackle*. Basil Blackwell.
- [5] Berger, J.O. (1985): *Statistical Decision Theory and Bayesian Analysis (2nd Edition)*. Springer Verlag.
- [6] Bergstrom, T.C. (1975): “Maximal Elements of Acyclic Relations on Compact Sets,” *Journal of Economic Theory* 10: 403-404.
- [7] Bewley, T. (2002): “Knightian Decision Theory: Part I,” *Decisions in Economics and Finance* 25: 79–110.
- [8] Borodin, A. and R. El-Yaniv (1998): *Online Computation and Competitive Analysis*. Cambridge University Press.
- [9] Chernoff, H. (1954): “Rational Selection of Decision Functions,” *Econometrica* 22: 422-443.

- [10] Cohen, M. and J.-Y. Jaffray (1980): “Rational Behavior under Complete Ignorance,” *Econometrica* 48: 1281-1299.
- [11] Eichberger, J., S. Grant, D. Kelsey, and G.A. Koshevoy (2008): “Differentiating Ambiguity: An Expository Note,” *Economic Theory* 36: 327-336.
- [12] Gajdos, T., J.-M. Tallon, and J.-C. Vergnaud (2004): “Decision Making with Imprecise Probabilistic Information,” *Journal of Mathematical Economics* 40: 647-681.
- [13] Gajdos, T., T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud (2008): “Attitude Toward Imprecise Information,” *Journal of Economic Theory* 140: 27-65.
- [14] Ghirardato, P., F. Maccheroni, and M. Marinacci (2004): “Differentiating Ambiguity and Ambiguity Attitude,” *Journal of Economic Theory* 118: 133-173.
- [15] Gilboa, I. (2009): *Theory of Decision under Uncertainty*. Cambridge University Press.
- [16] Gilboa, I. and D. Schmeidler (1989): “Maxmin Expected Utility with Non-unique Prior,” *Journal of Mathematical Economics* 18: 141-153.
- [17] Hausman, D. (2000): “Revealed Preference, Belief, and Game Theory,” *Economics and Philosophy* 16: 99-115.
- [18] Hayashi, T. (2008): “Regret Aversion and Opportunity-dependence,” *Journal of Economic Theory* 139: 242-268.
- [19] Hirano, K. (2008): “Decision Theory in Econometrics,” in S.N. Durlauf and L.E. Blume (eds.) *The New Palgrave Dictionary of Economics*.
- [20] Hurwicz, L. (1951): “Some Specification Problems and Applications to Econometric Models,” *Econometrica* 19: 343-344.
- [21] Kreps, D.M. (1988): *Notes on the Theory of Choice*. Westview Press.
- [22] Manski, C.F. (2000): “Identification Problems and Decisions Under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice,” *Journal of Econometrics* 95: 415-442.
- [23] — (2004): “Statistical Treatment Rules for Heterogeneous Populations,” *Econometrica* 72: 1221-1246.
- [24] — (2005): *Social Choice with Partial Knowledge of Treatment Response*. Princeton University Press.
- [25] Milnor, J. (1954): “Games Against Nature,” in R.M. Thrall, C.H. Coombs, R.L. Davis (eds.), *Decision Processes*. Wiley.

- [26] Olszewski, W. (2007): “Preferences over Sets of Lotteries,” *Review of Economic Studies* 74: 567-598.
- [27] Puppe, C. and K.H. Schlag (2009): “Choice under Complete Uncertainty when Outcome Spaces are State-Dependent,” *Theory and Decision* 66: 1-16.
- [28] Rawls, J. (1999): *A Theory of Justice (Revised Edition)*. Harvard University Press.
- [29] Savage, L.J. (1951): “The Theory of Statistical Decision,” *Journal of the American Statistical Association* 46: 55-67.
- [30] — (1954): *The Foundations of Statistics*. New York: Wiley.
- [31] Schmeidler, D. (1989): “Subjective Probability and Expected Utility without Additivity,” *Econometrica* 57: 571-587.
- [32] Sen, A.K. (1973): “Behaviour and the Concept of Preference,” *Economica* 40: 241-259.
- [33] Stoye, J. (2007): “Axioms for Minimax Regret Choice Correspondences,” mimeo, New York University.
- [34] — (2009a): “Minimax,” in S.N. Durlauf and L.E. Blume (eds.) *The New Palgrave Dictionary of Economics*.
- [35] — (2009b): “Minimax Regret,” in S.N. Durlauf and L.E. Blume (eds.) *The New Palgrave Dictionary of Economics*.
- [36] Wald, A. (1950): *Statistical Decision Functions*. Wiley.