THE DECEPTIVE BEAUTY OF MONOTONICITY, AND THE MILLION-DOLLAR QUESTION: ROW-FIRST OR COLUMN-FIRST AGGREGATION?

Chen Li, C.li@ese.eur.nl
Kirsten I.M. Rohde, Rohde@ese.eur.nl
Peter P. Wakker, Wakker@ese.eur.nl
Erasmus School of Economics, Erasmus University Rotterdam

January, 2023

This paper proposes a unified framework for optimization over two or more components (risk/time; risk/welfare; etc). Using a century-old theorem on macro-micro aggregation, we show that many existing debates, on incentive compatibility of random incentives, hedging confoundings in ambiguity measurements, equity in Harsanyi’s veil of ignorance, multiattribute risk aversion, and many others, all concern the same bifurcation question “row-first or column-first aggregation?” For a single component, behavioral models typically relax separability while maintaining monotonicity. For two or more components, this is, surprisingly, no longer possible. Then at least one monotonicity must be violated. The question of which one is equivalent to the above bifurcation question. Our analysis clarifies many ongoing debates in many fields, including the aforementioned ones. We provide diagnoses and techniques for overcoming undesirable violations of monotonicity. A mathematical online appendix shows how our framework can be used theoretically to generalize many well-known preference axiomatizations.

JEL-CODES: D81, C91
KEYWORDS: risk and time, hedging under ambiguity, ex ante fairness, aggregation, separability, discounted expected utility, random incentive system

1. INTRODUCTION

It may come as a surprise, but many existing debates in the literature, from Harsanyi's veil of ignorance to incentive compatibility of the random incentive system or multiattribute risk aversion, share a common hidden cause. That is, when the decision problem involves multiple components, in which order should we aggregate the components? As a simple example, consider choices from probability distributions over intertemporal outcome streams. Here risk is one component and time the other. Should we first aggregate over time, taking the present value of each possible outcome stream, and next aggregate over the probabilities, taking a certainty equivalent? Or should we first aggregate over risk, taking the certainty equivalent of the probability distribution at each timepoint, and only then aggregate over time, taking the present value? Whereas the order of aggregation does not matter in classical (rational) models, it becomes essential in behavioral generalizations, not only for time and risk but for every situation where two or more components are involved (persons, commodities, production inputs, health attributes, and so on).

For decisions with a single component, behavioral decision models typically relax strong separability assumptions while preserving monotonicity, e.g., in nonexpected utility for risk or equity models for welfare. We show that when two or more components are involved, relaxing (strong) separability while maintaining monotonicity is no longer possible. Relaxing separability then comes at the cost of giving up monotonicity regarding at least one of the components. The question of which monotonocity to give up, is equivalent to the question of which component to aggregate over first.

There have by now been 100s, or even 1000s, of discussions of the order of aggregation in many fields, showing its importance and ubiquity. Many references will follow later, but the literature is too broad to cite or survey completely. We provide a unifying framework to study the aforementioned problem concerning the proper order of aggregation. Most prior treatments focused on single domains and implicitly chose the order of aggregation without discussing it. Several authors, cited later, did critically discuss the order of aggregation and did compare different domains, but the fundamental nature of the problem and its universality have not been observed before.

Our paper builds on Nataf's (1948) century-old theorem on macro-micro aggregation. It leads to the paradoxical result that when two or more components are involved, relaxing separability has to go hand in hand with relaxing monotonicity. This has vast implications for all modern behavioral approaches involving aggregation over multiple components. We show that many behavioral paradoxes or puzzles in different domains, even though exhibiting different symptoms on the surface, do share the same underlying cause: the middle ground of partial rationality in the form of monotonicity with relaxed separability, which exists for uni-component problems, is not available for multiple components. This fundamental problem comes as an implication of Nataf's theorem. As a result, behavioral approaches face a bifurcation question (give up column monotonicity or row monotonicity; see Figure 3 and Theorem 5), that is a more fundamental and serious problem than was thought before. Even though our results derive from known mathematical theorems, the vast impact for behavioral approaches, ranging from hedging in ambiguity meassurements to ex ante/ex post equality in Harsanyi's veil of ignorance, has not been understood before. We provide general suggestions for how to determine and generate the proper order of aggregation, providing a unified road map for many fields.

This paper is organized as follows. The first part (Sections 2-3) starts with a new and thought-provoking preference axiomatization of discounted expected utility (DEU). We show that unobjectionable technical and standard axioms, such as monotonicity (presented in Section 2.2), when assumed for both time and risk (more than one component) are strong enough to give rise to DEU, the workhorse of classical economics, satisfying complete separability. This result is surprising: how can unobjectionable and widely accepted axioms characterize a widely falsified model? This sets our quest to uncover the underlying cause shared by many seemingly unrelated but essentially similar paradoxes and debates. Section 3 presents the formal framework of this paper and our basic theorem, a modern version of Nataf's theorem, and explains what underlies the paradox of Theorem 1.

The novelty of our theorems is in their simplicity and appeal, and not in mathematical generality. Our preference conditions can be stated verbally and are accessible to nonspecialists more than any preceding axiomatization of DEU. Although many authors, cited later, used advanced implications of the preceding results in various decision theories, their basic impact for empirical and theoretical work, specified in Section 4 and applied in the rest of this paper, has not been observed before. Mongin & Pivato (2015 Proposition 1) and other papers presented more general theorems. Details are in Online Appendix A. We do not seek for mathematical generality but for applied relevance and conceptual implications.

A moral of the story follows in Section 4, where we consider behavioral generalizations of classical models. We identify the culprit that gives rise to the paradoxical axiomatization of DEU and we further explain the problem of bifurcation. With the culprit identified, Section 5 gives general suggestions of how to deal with the absence of the middle ground.

Section 6 elaborates on two widely debated issues upon which our results shed new lights (monotonicity in the Anscombe-Aumann framework and validity of the random incentive system), and briefly mentions several others. Then a conclusion and proofs follow. To limit the size of this paper, the mathematical power of our aggregation results, generalizing several well-known preference axiomatizations with simplified proofs, is presented in Online Appendix C. For instance, Gul's (1992) axiomatization of subjective expected utility readily follows as a corollary of our Theorem 1 and thus, essentially, of Nataf (1948).

2. DISCOUNTED EXPECTED UTILITY: A PARADOX TO REVEAL THE UNDERLYING PROBLEM

This section considers the aggregation problem by providing an appealing, but paradoxical, axiomatization of discounted expected utility.

2.1. Definitions for Uncertainty and Time

We consider choices between "actstreams," i.e., matrices as in Figure 1. Here, if state of nature s^i obtains then, at timepoint t_j one receives money x_j^i. Columns designate acts, i.e., maps from states to outcomes, and rows similarly are outcome streams. An actstream gives a stream of acts or, equivalently an act yielding streams.

FIGURE 1. An actstream

$$\begin{array}{c} \\ s^1 \\ \vdots \\ s^i \\ \vdots \\ s^m \end{array} \begin{array}{c} \begin{array}{ccccc} t_1 & \cdots & t_j & \cdots & t_n \end{array} \\ \begin{bmatrix} x_1^1 & \cdots & & \cdots & x_n^1 \\ \vdots & & & & \vdots \\ \vdots & & x_j^i & & \vdots \\ \vdots & & & & \vdots \\ x_1^m & \cdots & & \cdots & x_n^m \end{bmatrix} \end{array}$$

Preferences $\succcurlyeq$ are over actstreams. Any outcome stream can be identified with the matrix having that outcome stream in each row, i.e., the degenerate lottery giving that outcome stream with certainty. Any act can be identified with the matrix having that act in the first column, and outcome 0 elsewhere. This way, preferences are generated over acts and over streams.

Expected utility (*EU*) holds if there exist positive *probabilities* $p^1, \ldots, p^m$ and a *utility function* U ($U: \mathbb{R} \to \mathbb{R}$ continuous and strictly increasing) such that preferences over acts (elements of $\mathbb{R}^m$) are represented by *expected utility*

$$(\alpha^1, \ldots, \alpha^m) \mapsto \sum_{i=1}^m p^i \left(U(\alpha^i)\right) \tag{1}$$

Discounted utility (*DU*) holds if there exist *discount factors* $0 < d_j$ ($j = 1, \ldots, n$) and a utility function U such that preferences over streams are represented by *discounted utility* (*DU*)

$$(\beta_1, \ldots, \beta_n) \mapsto \sum_{j=1}^n d_j \left(U(\beta_j)\right) \tag{2}$$

Constant discounting can readily be obtained by adding an appropriate preference condition that guarantees the same discount rate over time.

Discounted expected utility (*DEU*) holds if there exist probabilities, discount factors, and a utility function U such that preferences over actstreams are represented by their *discounted expected utility* (*DEU*)

$$\sum_{i=1}^m p^i \sum_{j=1}^n d_j \left(U(x_j^i)\right) \tag{3}$$

DEU has the following implications:

(1) EU holds for uncertainty preferences.

(2) DU holds for intertemporal preferences.

(3) EU and DU use the same utility function.

Each of these implications has often been criticized on normative grounds. For instance, numerous debates on cardinal utility (Moscati 2018) and on the difference between risky and riskless utility (Fleurbaey 2010 p. 675; Keeney & Raiffa 1976) have been advanced, challenging implication (3). The three implications have also been extensively criticized on empirical grounds, so much that DEU may qualify as the most falsified decision model (Attema 2012; Starmer 2000). In DEU, the order of aggregation, first over time or first over risk, is immaterial: first computing discounted utilities and then computing their expectation gives the same result as first computing expected utilities and then their discounted value. In Sections 2.2 and 2.3, we present the axioms needed to axiomatize DEU.

2.2 "Unobjectionable" Axioms

AXIOM 1. Weak ordering: transitivity and completeness (including reflexivity).

AXIOM 2. Continuity: the usual (Euclidean) continuity on $\mathbb{R}^{m\times n}$.

AXIOM 3. *Outcome monotonicity*: strictly increasing any x_j^i strictly improves the actstream.

AXIOM 4. *Act monotonicity*: at any timepoint, replacing the act there by a weakly [strictly] preferred act leads to a weakly [strictly] preferred actstream.

AXIOM 5. *Stream monotonicity*: at any state, replacing the stream there by a weakly [strictly] preferred stream leads to a weakly [strictly] preferred actstream.

2.3. Objectionable Axioms

This section is supposed to list the critical axioms, to be added to the preceding ones, needed to axiomatize DEU. Given the strong separabilities over states and timepoints involved in DEU, so widely falsified empirically, strong axioms may be expected to come. However, there will be none. This section does not provide any further axiom. That is, the axioms in Section 2.2 suffice to give DEU! This may come as a surprise. How can the least objectionable axioms be equivalent to the most objectionable model? How can such seemingly weak preference conditions have such strong implications? The paradox becomes most salient from the theorem in the following section.

2.4. Axiomatization of Discounted Expected Utility

THEOREM 1. The following two statements are equivalent.

(i) Discounted expected utility holds.

(ii) Weak ordering, continuity, and monotonicity with respect to outcomes, acts, and streams hold. □

Because of its simplicity, we claim that Theorem 1 provides the most appealing axiomatization of DEU presently available.[1] However, a question yet to be resolved is how such seemingly weak preference conditions can have such strong implications. Resolving the paradox will lead us to the surprising fact that, for decision problems involving multiple components, the combination of seemingly weak rationality requirements for each component easily gives rise to strong rationality requirements in full force when the components are combined. Before touching on the crux of the problem, Section 3 introduces a unified framework that we need to showcase the problem underlying this paradox, and many other debates or paradoxes of the same nature.

In many contexts, extensions to infinite components are desirable. Online Appendix B shows that this can readily be achieved using standard tools from mathematical measure theory (e.g., Theorem 9). The important point to note is that our intuitive axioms, mainly the monotonicities, remain unaffected in this process. Only the technical continuity is modified. Thus, these modifications do not affect the practical implications discussed in the main part of this paper.

3. GENERAL DEFINITIONS AND THEOREM

Our general framework considers preferences $\succcurlyeq$ over *matrices*

[1] The theorem can readily be extended to risk. For example, if all s^i have known probabilities $1/m$ implying symmetry (and subjective probabilities p^i=1/m) we obtain all equal-probability distributions. Online Appendix B shows that extensions to all probability distributions readily follow.

FIGURE 2. A matrix

$$\begin{array}{c} \\ r^1 \\ \vdots \\ r^i \\ \vdots \\ r^m \end{array} \begin{array}{c} \begin{array}{ccccc} c_1 & \cdots & c_j & \cdots & c_n \end{array} \\ \begin{bmatrix} x_1^1 & \cdots & & \cdots & x_n^1 \\ \vdots & & & & \vdots \\ \vdots & & x_j^i & & \vdots \\ \vdots & & & & \vdots \\ x_1^m & \cdots & & \cdots & x_n^m \end{bmatrix} \end{array}$$

There are two *components*: (1) a finite *row set* $R = \{r^1, \ldots, r^m\}$ with its *attributes* being m *rows* and (2) a finite *column set* $C = \{c_1, \ldots, c_n\}$ with its *attributes* being n *columns*. Before, the components concerned uncertainty and time, with m states and n timepoints as their attributes, respectively. In general, components can also designate persons, commodities, production inputs, health attributes, and so on. Wakker (2010 Appendix D) gives many other examples. For simplicity, we continue to assume that the outcome set is $\mathbb{R}$. In some examples outcomes may concern nonmonetary commodities.[2] We assume $m, n > 1$ fixed. Throughout, superscripts do not designate powers. $\mathbb{R}$ is the *outcome* space, say monetary. *Rows* $\in \mathbb{R}^n$ map C to $\mathbb{R}$ and *columns* $\in \mathbb{R}^m$ map R to $\mathbb{R}$. A *matrix* $\in \mathbb{R}^{m \times n}$ (actstream before) maps $R \times C$ to $\mathbb{R}$. It specifies a row $(x_1^i, \ldots, x_n^i)$ for each r^i and a column $(x_j^1, \ldots, x_j^n)$ for each c_j. Continuity of preference is as usual, referring to the Euclidean topology.

We throughout assume that all decisions are made at one fixed timepoint, preceding all timepoints of a time component if present. The decision timepoint also precedes any information about the resolution of risk or uncertainty if an uncertainty component is present. Thus, if the true state was determined prior to the decision, the decision maker does not know which it is. Prior resolution of uncertainty is only a matter of perception and never of strategic relevance. Dynamic decision principles and updating play no role in this paper.

---

[2] Mathematical extensions of our theorems to connected topological outcome spaces (e.g., convex sets of commodity bundles) are straightforward.

Outcome monotonicity is defined as before (strictly increasing any cell strictly improves the matrix). We generalize the other monotonicity conditions. We now consider underlying preferences $\succcurlyeq^i$ over rows $(x_1^i, \ldots, x_n^i)$ and $\succcurlyeq_j$ over columns $(x_j^1, \ldots, x_j^m)$ that may depend on i and j, respectively, and we will consider monotonicity for preferences $\succcurlyeq$ over matrices with respect to such underlying preferences. In particular, the preferences $\succcurlyeq^i$ and $\succcurlyeq_j$ can be derived from $\succcurlyeq$ over matrices by keeping "outside cells" fixed. This procedure works smoothly if proper separability/monotonicity conditions hold. For consistency, we maintain the monotonicity terminology.

DEFINITION 2. A subset of cells is *separable* if preferences over those cells, while keeping the outcomes at all other cells fixed, are independent of the levels where the other cells are kept fixed. *Weak separability of rows*, or *row monotonicity*, holds if each row is separable; that is, for each i, preferences over rows $(x_1^i, \ldots, x_n^i)$ by keeping all other rows fixed are independent of the levels where the other rows are kept fixed. *Weak separability of columns*, or *column monotonicity*, holds if each column is separable. *Complete separability* holds if each subset of cells is separable. □

Row monotonicity means that improving a row improves the matrix, and column monotonicity is similar. Complete separability can concern any subset of cells, also if this subset is not a union of rows and/or columns. *Additive utility* (*AU*) holds if preferences over matrices are represented by

$$\sum_{i=1}^{m} \sum_{j=1}^{n} V_j^i(x_j^i) \qquad (4)$$

for strictly increasing continuous functions $V_j^i(x_j^i)$. It readily implies complete separability. The following well-known result is basic to this paper. It has been known as the "theorem of aggregation." Its history is discussed later.

THEOREM 3 [of aggregation] The following two statements are equivalent.

(i) Additive utility holds.

(ii) Weak ordering, continuity, and monotonicity with respect to outcomes, rows, and columns hold. □

It is obvious that Statement (i) implies Statement (ii), and even complete separability. For the reversed implication, it is clear that row and column monotonicity preclude particular interactions between cells. However, the tradeoffs directly precluded this way are only few. The surprising point of Theorem 3 is that, in this setting with multiple components, all interactions are precluded “indirectly” after all, also for the many subsets of cells besides rows and columns. This was, essentially, Nataf’s (1948) finding, although his proof has sometimes been criticized for being inaccessible. Nowadays, the result can readily be obtained as one of the many surprising implications of Gorman’s (1968) strong result. Hence, we will not give a separate proof. This result also explains the paradox of Theorem 1, and underlies the bifurcation result and absence of middle grounds derived later.

The analysis in Theorem 1 and preceding sections concerned the special case where $\succcurlyeq^i$ and $\succcurlyeq_j$ were independent of i and j, respectively, implying that the V_j^i can be taken proportional. In general contexts we use the terms *uniform row monotonicity* instead of stream monotonicity and *uniform column monotonicity* instead of act monotonicity for these special cases. The difference between DEU and AU, or between uniform and general monotonicity/separability, or between Theorems 1 and 3, never plays a role in any of the conceptual debates later in this paper. The generalizations (“state- and time-dependence”) increase the applicability of our results.

Whenever risk, the most-studied component in the literature and also in this paper, is involved we let it correspond with rows r^i, which then are states with known probabilities. We then refer to row monotonicity as *risk monotonicity*.

Our results can easily be extended to more than two components. In particular, each component may itself combine several components. Section 6.2 will illustrate this point.

4. FORMALIZING THE RESTRICTIVENESS OF MONOTONICITY: A BIFURCATION

As with DEU, the complete separability of many classical economic theories has been challenged on normative and descriptive grounds. It is therefore natural for researchers to think of behavioral relaxations of complete separability to increase the models’ validity. Common strategies of behavioral relaxations were originally developed for single components, first risk, and later uncertainty, time, welfare, and so on. Even though different relaxations focus on different psychological insights or behavioral patterns, they typically share the same structure. That is, they operate in a *middle ground*, where complete separability is relaxed but weak separability is still maintained. For instance, nonexpected utility models aim to maintain mononicity (thus weak separability) or, equivalently, stochastic dominance, but search for proper relaxation of complete separability (the "sure-thing principle" or mixture-independence). Kahneman & Tverky's (1979) original prospect theory is an example that, by violating stochastic dominance, fell out of the middle ground. It was, therefore, replaced by a rank-dependent version that satisfies monotonicity and hence is in the middle ground (Tversky & Kahneman 1992). In this section, we will show how Theorem 3 implies that such middle grounds do not exist for multi-component decision problems.

For our analyses, weak ordering and continuity will be taken as unobjectionable and thus will not be discussed further. We focus on the monotonicity conditions. We follow the terminological convention that for any product structure, monotonicity with respect to that structure refers to changes in single attributes. Thus, for outcome monotonicity, we treat the space of matrices as if one component, i.e., one product space with $m \times n$ attributes, being cells. We then consider changes in single cells. The underlying relation, $\geq$ on money, is "physical" and objective, and outcome monotonicity is usually unobjectionable. It is a version of weak separability, and is not very restrictive. Then changes within single attributes are not ordinally affected by other attributes, but changes within nonsingle subsets of attributes, involving (cardinal) tradeoffs between attributes, can still be.

We next consider row and column monotonicity. We now take the space of matrices as a combination of two product structures. That is, we take it as comprising two components—the topic of this paper. Row monotonicity refers to an m fold product space, with rows, element of $\mathbb{R}^n$, as single attributes. This monotonicity again concerns a change in a single attribute, now a row. Similarly, column monotonicity refers to an n fold product space with columns, element of $\mathbb{R}^m$, as single attributes. It again concerns a change in a single attribute, now a column.

Row monotonicity refers to underlying preference relations $\succcurlyeq^i$ over rows that are subjective. They are not beyond doubt, and can readily be impacted by other variables, as many examples below will show. Therefore, monotonicity now is debatable.[3] This holds similarly for column monotonicity.

In our quest for behavioral relaxations of complete separability, we will first reformulate the monotonicity conditions in terms of functional forms that can represent preferences. For quantitative optimizations with two (or more) components, recursive procedures are commonly used because they are tractable. They can occur in two ways, i.e., two orders of aggregation. In *row-monotonic aggregation* one first, for every row r^i, aggregates over the columns c_j there. There then exist *row-functions* R^i and a *column-function* $\overline{C}$, all continuous and strictly increasing in each coordinate, such that preferences are represented by

$$\overline{C}\left(R^1(x_1^1,\dots,x_n^1),\ \dots\dots, R^i\left(x_1^i,\dots,x_n^i\right),\ \dots\dots, R^m(x_1^m,\dots,x_n^m)\right) \tag{5}$$

The second procedure uses *column-monotoni*c aggregation. Now one first, for every column c_j aggregates over the rows r^i there. There then exist *column-functions* C_j and a *row-function* $\overline{R}$, all continuous and strictly increasing in each coordinate, such that preferences are represented by

$$\overline{R}\left(C_1(x_1^1,\dots,x_1^m),\ \dots\dots, C_j\left(x_j^1,\dots,x_j^m\right),\ \dots\dots, C_n(x_n^1,\dots,x_n^m)\right) \tag{6}$$

Under uniform row monotonicity, we can take all R^i in Eq. 5 the same, i.e., independent of i, and under uniform column monotonicity, we can take all C_j in Eq. 6 the same, independent of j.

In the literature, terminologies of row-first and column-first aggregation are popular, as used in our title. Unfortunately, those terms have been used interchangeably, with several linguistic requirements and conventions going in opposite directions. Ambiguities cannot be avoided then. In our formal analysis we, therefore, chose the terms row-monotonic and column-monotonic aggregation instead.

The two procedures do not seem to be very restrictive because they involve many functions that can be chosen independently and with almost no restrictions imposed on those functions. The following observation shows that orders of aggregation, i.e.,

---

[3] Bommier (2017) weakened row monotonicity to hold only with stochastic dominance as underlying preference, so as to maintain objectivity.

aggregation monotonicities, are, indeed, quantitative versions of preference monotonicities. The observation is standard in consumer theory.

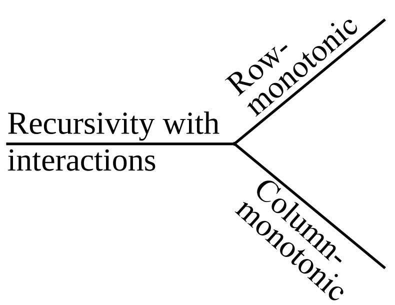
OBSERVATION 4. Given weak ordering, continuity, and outcome monotonicity, column-monotonic aggregation can be used if and only if column monotonicity holds. Row-monotonic aggregation can be used if and only if row monotonicity holds. □

As for the proof, row- (column-)monotonic aggregation can be derived from the corresponding preference condition by taking constant-equivalent functions for the functions $R^i, C_j, R,$ and C. The rest is straightforward.

The following result, mathematically a corollary of Nataf's (1948) Theorem 3, has vast and paradoxical implications for behavioral theories. Figure 3 illustrates it.

OBSERVATION 5 [Bifurcation]. If one wants to adopt a behavioral model with interactions (i.e. relaxing complete separability), and for tractability reasons use a recursive model, then the two routes available, row-monotonic and column-monotonic aggregation, are exclusive, and one faces a bifurcation. □

FIGURE 3. Bifurcation with no middle ground



The paradoxical point is that there is no middle ground. The moment one commits to the, ordinal, weak separability of both components, one is committed to their cardinal, complete, separability. And the moment one commits to one, all the interactions allowed by the other are precluded. The following example illustrates how this paradox can lead researchers astray.

EXAMPLE 6 [Paradoxical absence of the middle ground]. A researcher, facing actstreams as in Figure 1, wants to relax complete separability and allow for interactions between risky states s^i but not between timepoints t_j. She aims at a middle ground with risk monotonicity, i.e., stochastic dominance, kept. Thus, in Eq. 3 (DEU), the right summation, DU over columns is kept, but the left summation, EU, is replaced by a nonexpected utility formula that satisfies stochastic dominance. This replacement adapts Dejarnette et al.'s (2020) generalized local bilinear utility to our context.

At first sight, we seem to have achieved the desired middle ground, with cardinal but no ordinal interactions between states s^i. However, the bifurcation of Observation 5 shows that it cannot be. Act monotonicity must be violated. Figure 4 gives an example, for $d_1 = d_2 = 1$, $P(s^1) = P(s^2) = 0.5$, U linear ($U(\alpha) = \alpha$), and any nonexpected utility model with overweighting of the worst outcome. In the example, preferences over columns at timepoint t_1 depend on the levels at which the outcomes at timepoint t_2 are fixed—timepoint t_1 is affected by t_2. Even though the adapted formula seems to have preserved separable discounting, in reality, interactions between timepoints are still happening under the cover. Nonseparable discounting is not only possible here (Dejarnette et al. 2020 p. 630) but it even is unavoidable. Further, interactions even occur at the elementary level of single timepoints. Possibly unbeknownst to the researcher, she has introduced interactions between timepoints after all.

FIGURE 4. Violation of act monotonicity

$$\begin{array}{c} \\ s^1 \\ s^2 \end{array}\begin{array}{cc} t_1 & t_2 \\ \hline 1 & 0 \\ 0 & 2 \end{array} \succ \begin{array}{c} \\ s^1 \\ s^2 \end{array}\begin{array}{cc} t_1 & t_2 \\ \hline 0 & 0 \\ 1 & 2 \end{array}$$

$$\begin{array}{c} \\ s^1 \\ s^2 \end{array}\begin{array}{cc} t_1 & t_2 \\ \hline 1 & 2 \\ 0 & 0 \end{array} \prec \begin{array}{c} \\ s^1 \\ s^2 \end{array}\begin{array}{cc} t_1 & t_2 \\ \hline 0 & 2 \\ 1 & 0 \end{array}$$

In behavioral approaches, the choice in the bifurcation presented in Figure 3 is mostly made implicitly (Andreoni & Sprenger 2012; Machina 2014 Eq. 6 & footnote

11 & p. 3821 l. -3). As we have shown, the choice is critical though and explicit arguments for the monotonicity assumed are desirable. Several papers did discuss this point explicitly, including Bommier, Kochov, & le Grand (2017 Sections 3 and 6), Dejarnette et al. (2020 Section 4), Epper & Fehr-Duda (2021), Marinacci (2015 p. 1026), and Onay & Öncüler (2009). But the critical nature of the issue (Figure 3) has not been noted before. Now it is clear why in the numerous discussions in the literature no-one ever proposed a middle ground: because it doesn't exist.

The following claim, mathematically a corollary of Observation 5 and, thus, of Nataf (1948), highlights another practical implication of essentially Nataf's theorem. While not fully formalized, dividing logical implications over assumptions, it shows the true face of monotonicity, signalling the alarming restrictiveness of resursive procedures. They may inadvertently preclude many relevant interactions. The "at least" clause below is because of the interactions precluded by both monotonicities.

INTERPRETATION 7 [Precluding many interactions]. Given weak ordering, continuity, outcome monotonicity, and $m = n$, row monotonicity precludes at least half of the possible interactions (violations of complete separability), and so does column monotonicity. Each condition precludes all interactions allowed by the other. □

For $m \neq n$ the implications should be divided modulo m, n, but the situation is similarly alarming,

The following example will serve as lead example in Section 5. It applies our framework to Harsanyi's (1955) utilitarianism, a version of DEU.

EXAMPLE 8 [Welfare and risk]. Columns c_j refer to persons and rows r^i to risk, i.e., states with known probabilities p^i. For simplicity, we assume $p^i = 1/m$ for all i. Harsanyi's (1955) utilitarian model is AU with $V_k^i = U_k/m$ for all i, k, where U_k is the utility function of person k. It is a column-dependent generalization of DEU. Preferences over a matrix are of a benevolent social planner with no own stakes. Harsanyi's Pareto optimality is column monotonicity, and his expected utility for the social planner implies uniform risk (= row) monotonicity.

An axiomatization of Harsanyi's model readily follows from Theorem 3 by adding symmetry of rows (implying uniformity of row preference and EU with equal probabilities). The extension to general probabilities follows from Theorem 9 in Online Appendix B. This result is more general than Harsanyi's (1955) axiomatization in weakening his assumption of expected utility. In return, Harsanyi did not need continuity in outcomes, and could handle subdomains of the matrix space.[4] Harsanyi's axiomatization was received as a paradox because many were misled by the "hidden" restrictiveness of column monotonicity. But with our preceding theorems and Interpretation 7, his result comes as no surprise.

Harsanyi's model has often been criticized for ignoring inequality aversion, which involves interactions between different persons (= columns). In the following Figures 5 and 6 we assume symmetry of c_1 and c_2 (i.e., anonymity) and also of r^1 and r^2 (which have probability 0.5).

FIGURE 5

$$\begin{array}{c} \\ r^1 \\ r^2 \end{array}\begin{array}{cc} c_1 & c_2 \\ \hline 0 & 1 \\ 1 & 0 \end{array} \sim \begin{array}{c} \\ r^1 \\ r^2 \end{array}\begin{array}{cc} c_1 & c_2 \\ \hline 1 & 1 \\ 0 & 0 \end{array}$$

$$\begin{array}{c} \\ r^1 \\ r^2 \end{array}\begin{array}{cc} c_1 & c_2 \\ \hline 0 & 0 \\ 1 & 1 \end{array} \sim \begin{array}{c} \\ r^1 \\ r^2 \end{array}\begin{array}{cc} c_1 & c_2 \\ \hline 1 & 0 \\ 0 & 1 \end{array}$$

FIG. 5a. Column monotonicity

$$\begin{array}{c} \\ r^1 \\ r^2 \end{array}\begin{array}{cc} c_1 & c_2 \\ \hline 0 & 1 \\ 1 & 0 \end{array} \prec \begin{array}{c} \\ r^1 \\ r^2 \end{array}\begin{array}{cc} c_1 & c_2 \\ \hline 1 & 1 \\ 0 & 0 \end{array}$$

$$\begin{array}{c} \\ r^1 \\ r^2 \end{array}\begin{array}{cc} c_1 & c_2 \\ \hline 0 & 0 \\ 1 & 1 \end{array} \succ \begin{array}{c} \\ r^1 \\ r^2 \end{array}\begin{array}{cc} c_1 & c_2 \\ \hline 1 & 0 \\ 0 & 1 \end{array}$$

FIG. 5b. Column monotonicity violated

Broome (1991 p. 185) proposed Figure 5 as a criticism of Harsanyi's utilitarianism. In all matrices, both persons always receive $1_{0.5}0$. Hence, under column monotonicity ("Pareto optimality"), all matrices are indifferent, and so are they under Harsanyi's utilitarianism (Fig. 5a). Broome pointed out that, to the contrary, the strict preferences in Fig. 5b are plausible under inequality aversion. The dispreferred matrices certainly, under both r^1 and r^2, give inequality, and the preferred matrices certainly (for every row) give equality. The preference over the

[4] Harsanyi did not explicity introduce persons as different attributes, but his domain can be remodeled accordingly, turning it into a subdomain of AA. Thus, AA's theorem is a corollary of Harsanyi's

first column is affected by the second here, and column monotonicity and Harsanyi's utilitarianism are violated. Row monotonicity may still hold.

Diamond (1967) proposed Figure 6 as criticism of Harsanyi's utilitarianism. In all matrices, both rows (states) give the good outcome to one of the two persons which, by symmetry, is equivalent. Hence, by row symmetry, all matrices are indifferent, and so are they under Harsanyi's utilitarianism (Fig. 6a). Diamond pointed out that, to the contrary, the strict preferences in Fig. 6b are plausible under inequality aversion. In the dispreferred matrices, one person certainly receives the good outcome and the other person certainly not, giving inequality. In the preferred matrices there is equality in the sense that both persons receive the same lottery, $1_{0.5}0$. Diamond emphasized that the sure-thing principle is violated, which in this simple case is equivalent to our row monotonicity. The preference over the first row is affected by the second here, and expected utility and Harsanyi's utilitarianism are violated. Column monotonicity may still hold.

Our bifurcation result suggests that the above two examples are the only two tractable recursive ways to deviate from Harsanyi (1955), and that the interactions are detectable at the basic level of single persons or risks, as occurring in both figures. Broome's example fits in the upper route in Figure 3 and Diamond's in the lower one. The examples are each other's dual by interchanging rows and columns. Hence, recursive models that allow for ex-ante as well as ex-post inequality aversion, will have to give up on both row- and column monotonicity.

FIGURE 6

$$\begin{array}{c|cc} & c_1 & c_2 \\ \hline r^1 & 0 & 1 \\ r^2 & 1 & 0 \end{array} \sim \begin{array}{c|cc} & c_1 & c_2 \\ \hline r^1 & 1 & 0 \\ r^2 & 1 & 0 \end{array}$$

$$\begin{array}{c|cc} & c_1 & c_2 \\ \hline r^1 & 0 & 1 \\ r^2 & 0 & 1 \end{array} \sim \begin{array}{c|cc} & c_1 & c_2 \\ \hline r^1 & 1 & 0 \\ r^2 & 0 & 1 \end{array}$$

FIG. 6a. Row monotonicity

$$\begin{array}{c|cc} & c_1 & c_2 \\ \hline r^1 & 0 & 1 \\ r^2 & 1 & 0 \end{array} \succ \begin{array}{c|cc} & c_1 & c_2 \\ \hline r^1 & 1 & 0 \\ r^2 & 1 & 0 \end{array}$$

$$\begin{array}{c|cc} & c_1 & c_2 \\ \hline r^1 & 0 & 1 \\ r^2 & 0 & 1 \end{array} \prec \begin{array}{c|cc} & c_1 & c_2 \\ \hline r^1 & 1 & 0 \\ r^2 & 0 & 1 \end{array}$$

FIG. 6b. Row monotonicity violated

□

We will use our framework to analyze several existing debates and/or paradoxes. Ultimately, they all evolve around what is the proper choice in the bifurcation of Figure 3, and many valuable interactions may have been lost inadvertently.

Before discussing applications, we first tackle the natural question of which route in the bifurcation to choose. General guidelines, considerations, and ways to avoid undesirable violations of monotonicity are presented in the next section. They help shed new light on many existing problems.

5. GUIDELINES, DIAGNOSES, AND WAYS TO AVOID VIOLATIONS OF MONOTONICITY

This section provides guidelines for researchers when facing the choice in the bifurcation question of Figure 3. To decide on which separability to keep and which to give up, one needs to rank the plausibility of each separability condition. In general, separability is most plausible for uncertainty and risk because there can be no physical interactions between mutually exclusive events (Broome 1991 Section 5.3; Dejarnette et al. 2020 p. 632). Within uncertainty, it is more convincing for risk than for ambiguous events (Wakker 2010 Section 10.4). Next, interactions are less likely to occur between different persons at different locations than within one person at different timepoints. Thus, the [risk > ambiguity > welfare > time] plausibility ordering regarding separability/monotonicity results, with risk monotonicity usually being the most convincing. For time, payment in consumption is more separable than payment in money. For commodities or attributes, separability is less plausible than for uncertainty, but can take any remaining place in the ordering depending on the nature of the attributes. With time and risk involved, risk-monotonic aggregation is most plausible (Abdellaoui et al. 2019). Yet, time-(column-)monotonic aggregation has sometimes been adopted.

Researchers often add a component not for its own interest, but as an auxiliary tool to facilitate the analysis of other components. According to our conceptual analysis above, risk is most suited to serve as such a tool because separability there is very plausible. Risk has indeed mostly been used for this purpose. This is what Harsanyi (1955) did for welfare (our Example 8), essentially exploiting the paradox of our Theorem 1.[5] Other examples include Anscombe & Aumann (1963) and Keeney & Raiffa (1976). The restrictive results of Theorems 1 and 3 were often convenient in these applications.

Empirically, it is also plausible that decision makers mostly adopt one of the two recursive procedures. Again, it is for tractability reasons, but now from the psychological perspective of the decision maker instead of conventional theoretical modeling for the researcher. Nevertheless, some interactions and spillover effects due to the presence of other attributes and stimuli can still be expected. Hence, empirically, people will be close to one of the two routes in Figure 3, but with small deviations.

Whether a version of monotonicity is satisfied, and which route is chosen in the bifurcation in Figure 3, can be manipulated by stimuli and their framings. We explain three manipulation techniques below. They can be used to avoid undesirable violations. For example, spillover effects in preference measurements, hedging effects in ambiguity measurements, and perceptions of unfairness, are violations of monotonicity that may be undesirable. The first manipulation technique is the *framing technique*. In general, a two-stage display of matrices will enhance one kind of monotonicity. Thus, Fig. 7a enhances row monotonicity and Fig. 7b enhances column monotonicity. Similarly, in Figure 1, the framing "For each i, at state s^i you receive stream $(x_1^i, \ldots, x_n^i)$" enhances row-monotonicity, similar to Fig. 7a.

---

[5] Undoubtedly, Harsanyi (1955) devised his result independently without relating it to the preceding Nataf (1948).

FIGURE 7. Two-stage displays of matrices

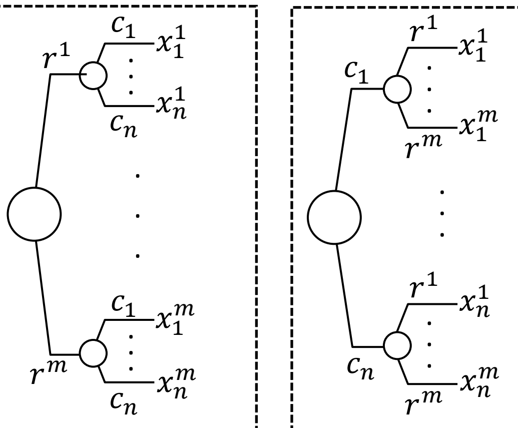


FIG. 7a. Visual display enhancing row-monotonic aggregation

FIG. 7b. Visual display enhancing column-monotonic aggregation

In Figure 5, if a social planner wants the dispreferred matrices in Fig. 5b to be accepted for some good extraneous reason, then the framing of Fig. 7b (with c_js designating persons) is best suited to enhance the column monotonicity of Fig. 5a. In Figure 6, if a social planner wants the dispreferred matrices in Fig. 6b to be accepted for some good extraneous reason, the framing of Fig. 7a (with c_js designating persons) is best suited to enhance the row monotonicity of Fig. 6a.

The second manipulation technique, the *timing technique*, can be used if risk or uncertainty is involved, and concerns the perceived timing of the resolution of uncertainty—early, before decision time, or late, after decision time. Early resolution of uncertainty enhances a perception as in Fig. 7a (with the r^js uncertain events) and row-monotonic aggregation. In Figure 6, it leads to Fig. 6a. Late resolution of uncertainty enhances a perception as in Fig. 7b and column-monotonic aggregation. In Figure 5, it leads to Fig. 5a. Thus, the perception of fairness can be manipulated by manipulating prior or late resolution. We stress that this paper only considers situations where, if resolution takes place before the decision time, then the decision maker knows this but does not know the result of the resolution. It is, therefore, of no strategic relevance here and only concerns perception. This timing technique has been widely discussed and tested in the welfare literature and other fields (Section 6.3). Onay & Öncüler (2009) tested the two different framings in Figure 7 for actstreams.

The third technique, the *partial-info technique*, provides only partial information. For example, in Figures 5 and 6, the two persons c_1, c_2 may not be informed about the outcomes that the other person receives. This enhances separability of the columns and, hence, column-monotonic aggregation. There then is less room for inequality aversion because the persons themselves cannot perceive it.

One can avoid our bifurcation by considering subdomains. Our analysis as yet made the idealized assumption, common in decision theory and preference axiomatizations, that we deal with a full domain. That is, all matrices are conceivable. This is essential for our theorems. Several studies on risk and time only considered actstreams with one nonzero outcome, in which case the order of aggregation is immaterial under many behavioral models (Baucells & Heukamp 2012; Ida & Goto 2009). Similarly, McCarthy, Mikkola, & Thomas (2020) and Pivato (2013) considered incomplete preferences, Alon & Gayer (2016) imposed Pareto optimality only if agreement on probabilities and utilities, and Halevy (2008) considered a restricted (comonotonic) domain where both orders of aggregation can hold for behavioral theories. For principled discussions of decision principles this escape route is not very convincing. If conditions deemed appropriate cannot survive extension to all possibilities, then this remains a point of concern. The issue arises and will be discussed in several applications in the next section.

6. APPLICATIONS

This section illustrates several applications of our results. We elaborate on two applications in our area of expertise, ambiguity, in Sections 6.1 and 6.2, and indicate others briefly in Section 6.3. In all examples in this section, rows r^i model risky events. The common theme of the examples, and of this paper, is that the seemingly innocuous conditions of weak separability for each component are more restrictive than had been understood before.

6.1. Monotonicity in the Anscombe-Aumann Framework for Ambiguity

We show that the well-known Anscombe-Aumann (AA) framework also made a bifurcation choice and apply our results to it. In Figure 2, *roulette events* (rows) $r^1, \ldots, r^m$ partition the universal event and have known probabilities. *Horse events* (columns) $c_1, \ldots, c_n$ also partition the universal event but are ambiguous. The AA framework adopts column-monotonic aggregation, using the same expected utility functional (C_j in Eq. 6) for each column. Thus, uniform column monotonicity, called *horse monotonicity* here, is assumed (Observation 4). This implies that only the marginal distributions given every horse c_j matter. This is characteristic of the modern AA framework. By $\alpha_p\beta$ we denote a lottery, i.e., probability distribution, yielding α with probability p and β with probability $1 - p$.

We first assume a full domain where all matrices are available, as for instance in Machina (2014) who assumed simultaneity of the horse and roulette events. Figure 8 displays ambiguity aversion as commonly assumed in the literature. The rows have 0.5 probability each. The indifference follows from the AA assumptions: each horse yields lottery $(1_{0.5}0)$, with no ambiguity. The strict preferences reflect ambiguity aversion. They reveal a violation of risk (row) monotonicity: preferences over the first row are affected by the second row, and rows interact. Having committed to horse monotonicity, the common AA framework has to give up risk monotonicity (and conditioning on risky events), as shown by Observation 5. However, as pointed out in Section 5, in general, risk (row) monotonicity is more plausible than column monotonicity. It suggests that the common AA framework chose the less plausible bifurcation route. Jaffray (1992, personal communication) emphasized this view and recommended risk monotonicity for ambiguity, adopting it in all his works (e.g., Jaffray 1989). Eichberger & Pasichnichenko (2021) and Martins-da-Rocha & Rosa (2021) followed Jaffray's approach. The early Keeney & Raiffa (1976) provided a rich toolbox for interactions in this approach. In agreement with the timing technique, it then works best to let the resolution of the roulette events precede those of the horse events. Remarkably, Bommier (2017) kept the original AA framework but nevertheless developed a reversed order of aggregation. Saito (2015 Figure 1) used the framing of Fig. 7b to enhance horse monotonicity. Subjects are usually told that the risk uncertainty is resolved after the horse monotonicity, again, to enhance horse monotonicity (timing technique). Oechssler & Roomets (2021) used Fig. 7b but still found much risk monotonicity, providing strong evidence *against* horse monotonicity.

FIGURE 8. Violation of risk monotonicity in the Anscombe-Aumann framework

$$\begin{array}{c} \\ r_1 \\ r_2 \end{array}\!\begin{array}{c} c^1 \quad c^2 \\ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \end{array} \sim \begin{array}{c} \\ r_1 \\ r_2 \end{array}\!\begin{array}{c} c^1 \quad c^2 \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array} \succ \begin{array}{c} \\ r_1 \\ r_2 \end{array}\!\begin{array}{c} c^1 \quad c^2 \\ \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ r_1 \\ r_2 \end{array}\!\begin{array}{c} c^1 \quad c^2 \\ \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \end{array} \prec \begin{array}{c} \\ r_1 \\ r_2 \end{array}\!\begin{array}{c} c^1 \quad c^2 \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{array}$$

In the version of the AA framework most popular today, the domain of matrices considered is restricted. The lotteries for different horses are assumed to be stochastically independent.[6] Then the matrices in Figure 8 can no more be used and we escape from the violation of risk monotonicity there. Theorem 3 still shows that correlations between horses cannot be added without violating risk monotonicity (or sacrificing one of the other conditions), which remains a worrisome issue. In particular, we cannot add risk prior to, or simultaneously with the horse race and have EU there, because this would automatically assume away ambiguity attitudes. Several authors observed this impossibility and investigated ways to relax other assumptions in particular multistage setups, including Ke & Zhang (2020) and Saito (2015). Our analysis with the elementary risk monotonicity instead of EU is more basic.

Although the modern version of the AA framework escapes from the "counterexample" of Figure 8, the underlying problem, weak separability of horse events which does not fit well with their ambiguity, remains. We, therefore, illustrate this problem through another implication, a variation of Figure 8 that uses only stimuli within the restricted domain assumed by the modern version of AA. In Figure

[6] Equivalently, they can be taken as unspecified, e.g., by taking them as conditional on a horse ("statewise randomization"; Ke & Zhang 2020; Saito 2015 p 1248 "second problem"). Compare Figures 9 and 10 below. The essence is that they are considered to be uninformative. Thus, subjects may only be informed about the outcome realized for the winning horse and the roulette resolution there (partial-info technique). Further, infinite risk models are mostly assumed, for one reason to allow for mixing. These points do not impact the conceptual issues discussed here.

10 below, for each matrix the two columns are stochastically independent. We take outcome α such that $\alpha \sim 8_{0.5}0$, i.e., it gives the indifference in Fig. 9a. Under expected value maximization, $\alpha = 4$, but in general it depends on the risk attitude.

All columns in Figure 9 are indifferent. By AA's horse monotonicity, all matrices should be indifferent. However, under ambiguity aversion the strict preferences are plausible. For the dispreferred matrices all outcomes are ambiguous whereas for the other matrices none is. Figure 10 displays the same choices as in Figure 9 but now using the matrix notation of this paper, with stochastic independence of the two columns for each matrix.

FIGURE 9. Violation of horse monotonicity due to ambiguity aversion

$$\begin{array}{cc} c_1 & c_2 \end{array} \qquad\qquad \begin{array}{cc} c_1 & c_2 \end{array}$$
$$\begin{bmatrix} 8_{0.5}0 & 8_{0.5}0 \end{bmatrix} \sim \begin{bmatrix} \alpha & \alpha \end{bmatrix}$$

FIG. 9a

$$\begin{array}{cc} c_1 & c_2 \end{array} \qquad\qquad \begin{array}{cc} c_1 & c_2 \end{array}$$
$$\begin{bmatrix} 8_{0.5}0 & \alpha \end{bmatrix} \prec \begin{bmatrix} \alpha & \alpha \end{bmatrix}$$

FIG. 9b

$$\begin{array}{cc} c_1 & c_2 \end{array} \qquad\qquad \begin{array}{cc} c_1 & c_2 \end{array}$$
$$\begin{bmatrix} 8_{0.5}0 & 8_{0.5}0 \end{bmatrix} \succ \begin{bmatrix} \alpha & 8_{0.5}0 \end{bmatrix}$$

FIG. 9c

FIGURE 10. Violation of horse monotonicity due to ambiguity aversion using matrices. $P(r^j) = 0.25$ for all j.

$$\begin{array}{c} \\ r^1 \\ r^2 \\ r^3 \\ r^4 \end{array} \begin{array}{c} \begin{array}{cc} c_1 & c_2 \end{array} \\ \begin{bmatrix} 8 & 8 \\ 8 & 0 \\ 0 & 8 \\ 0 & 0 \end{bmatrix} \end{array} \sim \begin{array}{c} \\ r^1 \\ r^2 \\ r^3 \\ r^4 \end{array} \begin{array}{c} \begin{array}{cc} c_1 & c_2 \end{array} \\ \begin{bmatrix} \alpha & \alpha \\ \alpha & \alpha \\ \alpha & \alpha \\ \alpha & \alpha \end{bmatrix} \end{array}$$

FIG. 10a

$$\begin{array}{c} \\ r^1 \\ r^2 \\ r^3 \\ r^4 \end{array} \begin{array}{c} \begin{array}{cc} c_1 & c_2 \end{array} \\ \begin{bmatrix} 8 & \alpha \\ 8 & \alpha \\ 0 & \alpha \\ 0 & \alpha \end{bmatrix} \end{array} \prec \begin{array}{c} \\ r^1 \\ r^2 \\ r^3 \\ r^4 \end{array} \begin{array}{c} \begin{array}{cc} c_1 & c_2 \end{array} \\ \begin{bmatrix} \alpha & \alpha \\ \alpha & \alpha \\ \alpha & \alpha \\ \alpha & \alpha \end{bmatrix} \end{array}$$

FIG. 10b

$$\begin{array}{c} \\ r^1 \\ r^2 \\ r^3 \\ r^4 \end{array} \begin{array}{c} \begin{array}{cc} c_1 & c_2 \end{array} \\ \begin{bmatrix} 8 & 8 \\ 8 & 0 \\ 0 & 8 \\ 0 & 0 \end{bmatrix} \end{array} \succ \begin{array}{c} \\ r^1 \\ r^2 \\ r^3 \\ r^4 \end{array} \begin{array}{c} \begin{array}{cc} c_1 & c_2 \end{array} \\ \begin{bmatrix} \alpha & 8 \\ \alpha & 0 \\ \alpha & 8 \\ \alpha & 0 \end{bmatrix} \end{array}$$

FIG. 10c

Many authors discussed horse monotonicity in the AA framework, both theoretically and empirically. Besides those cited before, these include Hill (2019, state-consistency), Machina (2014 p. 3835 3rd bulleted point), Oechssler & Roomets (2021), and Schneider & Schonger (2018). More general discussions of the role of the timing of uncertainty include Battigalli et al. (2017), Berger & Eeckhoudt (2021), Calford (2021), Chandrasekher et al. (2022), Eichberger, Grant, & Kelsey (2016), Grant et al. (2010), Kochov (2015), Oechssler, Rau, & Roomets (2019), Siniscalchi (2022), and Strzalecki (2013). Our results present the issue in its most basic and general form, showing that the issue is more acute than what has been observed before.

6.2. *Validity of the Random Incentive System and Hedging for Ambiguity*

Debates about validity of the random incentive system (*RIS*) and, in particular, a hedging confound for ambiguity measurements, are again special cases of the bifurcation question. To see this, let each row of our matrices specify one of m choice situations in an experiment. We assume that in each a subject chooses an option that is an n-dimensional object. It may be a commodity bundle, an outcome stream, a welfare allocation over n persons, an act assigning outcomes to n states of nature, and so on. One choice situation (row) will be randomly selected for real implementation. Matrices are strategies, specifying a choice for each choice situation in the experiment.

It trivially follows that incentive compatibility of the RIS is equivalent to risk (row) monotonicity ("isolation") where the underlying preferences are the true preferences in the choice situations (Cohen, Jaffray, & Said 1987, appendix; Cox, Sadiraj, & Schmidt 2014). Row-monotonic aggregation together with EU for risk is sufficient for this. Obviously, EU is not necessary here, as explained by Bardsley et al. (2010 p. 269) and many others. Nevertheless, there have been widespread misunderstandings about this point in the literature. Adding column monotonicity, uniformity of monotonicities, and a full domain would, by Theorem 1, indeed give EU for risk. Misunderstandings about the misleading restrictiveness of these extra assumptions (Section 4) may underlie the widespread misunderstandings.

As explained before, our theorems assumed full domains and complete preferences, whereas in many applications only subdomains are relevant or available. Nevertheless, the domains are often rich enough for our results to provide new insights. This point is further illustrated in this application.

In the RIS, the more (row-)risk interacts with the components of interest, the worse validity is. This is especially problematic for the measurement of ambiguity, where risk generates direct contrast effects. Nevertheless, the RIS is commonly used there too because no better alternative has been established. We discuss a particular problem in detail: hedging in RIS.

We assume two Ellsberg urns: a known urn K contains 50 White and 50 Black balls, and an unknown ambiguous urn A contains 100 balls, each White or Black, but in unknown proportions. From both urns a ball will be drawn at random. W_K denotes the event that the ball drawn from urn K is white, and B_K, W_A, and B_A are similar.

$(B_A\colon 101)$ denotes a gamble yielding €101 if the ball drawn from urn A is black and nothing otherwise. Other gambles are denoted similarly. An experiment concerns $m = 2$ choice situations for a given subject. The first, r^1, reveals the preference $(B_A\colon 101) \succcurlyeq (B_K\colon 100)$; the second, r^2, reveals the preference $(W_A\colon 101) \succcurlyeq (W_K\colon 100)$. The RIS randomly selects r^1 or r^2 for real implementation, each with probability 0.5. Can we conclude that there is virtually no ambiguity aversion? Hedging, explained below, has often been advanced as a counfound invalidating this measurement.

Three components can be distinguished concerning: the color from K, the color from A, and the selection from $\{r^1, r^2\}$. We can nevertheless use our techniques for two components, by combining the first two into one. We thus define four c_j as in Figure 11. The figure illustrates the two observed preferences. In the usual RIS, we do not consider binary choices between any two matrices, but choices from matrices (strategies) that combine all possible experimental choices. However, by standard revealed preference techniques, they still imply preferences between matrices as in Figure 11. Our subject can also choose the two matrices that result from combining the upper row of one matrix in Figure 11 with the lower row of the other matrix. But these extra options do not affect the following reasoning. Further, they are rarely chosen.

If we assume column monotonicity, then the preferences in Figure 11 simply follow from stochastic dominance: all columns of the preferred matrix stochastically dominate those of the dispreferred matrix ($101_{0.5}0 > 100_{0.5}0$). In the left matrix, the outcomes under r_2 provide a kind of hedge against those under r_1, same in the right matrix, which explains the term hedging. The preferences then do not speak to ambiguity attitudes in any sense. It has often been observed that, under column monotonicity and ambiguity nonneutrality, validity of the RIS may be violated. Our Observation 5 shows, more strongly, that it must *necessarily* be violated. Under column monotonicity the RIS cannot be used to measure ambiguity attitudes. Further, violations already exist at the basic level of a single measurement in the experiment.

The bifurcation in Figure 3 amounts to validity versus invalidity (including hedging) of the RIS. The hedging in the above example involved event complementarity, an extreme case of hedging (Hartmann 2021). Of course, besides hedging, any other reason to violate risk monotonicity discussed before can invalidate RIS.

FIGURE 11. Hedging in ambiguity

$$\begin{array}{c} \\ r_1 \\ r_2 \end{array}\begin{array}{cccc} B_K\cap B_A = c_1 & B_K\cap W_A = c_2 & W_K\cap B_A = c_3 & W_K\cap W_A = c_4 \\ \end{array}$$

$$\begin{matrix} & c_1 & c_2 & c_3 & c_4 \\ r_1 & 101 & 0 & 101 & 0 \\ r_2 & 0 & 101 & 0 & 101 \end{matrix} \succcurlyeq \begin{matrix} & c_1 & c_2 & c_3 & c_4 \\ r_1 & 100 & 100 & 0 & 0 \\ r_2 & 0 & 0 & 100 & 100 \end{matrix}$$

Discussions of hedging under ambiguity include Agranov & Ortoleva (2022), Bade (2015), Cerreia-Vioglio et al. (2019), and Oechssler, Rau, & Roomets (2019). Baillon, Halevy, & Li (2022) provided the first empirical demonstration and reviewed further literature. Their presentation of stimuli enhanced violations of risk (row) monotonicity, demonstrating the potential severity of the basic problem. In empirical studies using RIS, stimuli should therefore be framed so as to minimize such violations. The techniques of Section 5 apply here. Thus, Johnson et al.'s (2021) Prince, an implementation of RIS to maximize validity, selected the real choice situation prior to the experiment rather than after as usually done. As explained before (the timing technique), their prior selection works better to enhance the desired row monotonicity. They further used framing (e.g., as in Fig. 7b) and partial-info techniques as best as possible. Baillon, Halevy, & Li (2022), Cox, Sadiraj, & Schmidt (2014), and Oechssler & Roomets (2021) also investigated the timing technique.

Regarding the partial-info technique for RIS, when facing a choice situation, subjects are usually not yet informed about the choice situations that come after, precluding all "backward" interactions. To reduce "forward" interactions, each choice situation may be presented on a different page or screen, so that subjects can only know about preceding choice situations from memory. In general, full understanding of strategies in an experiment is humanly impossible. Validity of the RIS can therefore be expected to be good (Bardsley et al. 2010 Section 6.5, "behavioral incentive compatibility"). Some limited interactions between different choice situations can nevertheless occur (reviewed by Johnson et al. 2021). Unfortunately, alternatives to the RIS are not easy to devise.

6.3. Implications for Other Domains

There are numerous cases where aggregation over two or more components is central besides those considered before. Our analysis pertains to all those cases, underscoring the unity in many debates. Broome (1991) provided deep discussions. We briefly mention some further cases.

Violations of risk monotonicity as in Fig. 6b can be due to correlation aversion. This has been extensively studied in many domains, including intertemporal choice (Rohde & Yu 2022), multiattribute utility theory (multivariate risk aversion; Tsetlin & Winkler 2009) and consumer theory with t_1 and t_2 competing or completing commodities. For temporal ambiguity (unknown probabilities for r^1 and r^2), Kochov (2015) emphasized the plausibility of stream monotonicity and correlation preference. Epstein & Halevy (2019) considered an interesting case: both rows and columns refer to events with known probabilities, but their correlation is ambiguous. Then ambiguity aversion gives Figure 6b with reversed preferences.

Andreoni & Sprenger (2012) considered actstreams for risk, with probabilities of the states given. Their "direct preference for certainty" is exactly our uniform risk (stream/row) monotonicity. Contrary to their suggestions, the violations found were violations of any existing risk theory[7] rather than only of prospect theory. In their quantitative evaluations, Andreoni and Sprenger implicitly assumed risk monotonicity by taking the upper branch in the bifurcation of Figure 3, and were criticized (e.g., for absence of required correlations) by Cheung (2015), Epper & Fehr-Duda (2015), and Miao & Zhong (2015). Similar violations of uniform risk monotonicity had been found before, by Abdellaoui, Diecidue, & Öncüler (2011), Bleichrodt & Pinto (2009), and others.

Many papers studied infinite-dimensional recursive temporal lotteries that do not readily fit into the finitistic framework of this paper. See, for instance, Bommier, Kochov, & le Grand (2017) and their references.

For uncertainty, McCarthy, Mikkola, & Thomas (2020) and Zimper (2008) observed an equivalence of weak and complete separability, somewhat in the spirit of

[7] The utility of gambling theory (Diecidue, Schmidt, & Wakker 2004) accommodated these violations. However, this theory is not very tractable or suited for applications.

our Theorem 3. They considered only one component and then imposed weak separability on all possible decompositions of the component.

Besides the references mentioned before, numerous papers examined the timing technique, theoretically and empirically. Again, we bring a unification of these analyses. We mention some. For time and risk, see Bommier, Kochov, & le Grand (2017), Dejarnette et al. (2020 Section 4), Epper & Fehr-Duda (2021), and Onay & Öncüler (2009). For welfare and risk (where timing of resolution of uncertainty is only one way to manipulate an ex post or ex ante viewpoint), see Fleurbaey (2010), Miao & Zhong (2018), Rohde & Rohde (2015), and Saito (2013). Berger & Emmerling (2020) examined the overall effect of inequality aversion in separate components under different orders of aggregation. They provided a unifying framework of their results for several kinds of components. Applications for deliberate randomization in choice include Agranov & Ortoleva (2022), Cerreia-Vioglio et al. (2019), and Miao & Zhong (2018).

7. CONCLUSION

We considered optimization over two (or more) components. We proposed a unified framework to reexamine many puzzling issues under this light, including monotonicity and hedging in the Anscombe-Aumann framework of ambiguity, equity in Harsanyi's utilitarianism, and incentive compatibility of the random incentive system. We showed that Nataf's century-old theorem underlies the aforementioned debates and paradoxes. They all concern a common cause: the seemingly innocuous conditions of weak separability for each component become surprisingly restrictive when combined. We also explicitly state the main implication of our result, the bifurcation problem, which has often been implicitly dealt with or assumed away in the literature. Our analysis shows that the problem is more fundamental and acute than what was thought before. We provided guidelines for making a conscious choice of optimization when dealing with more than one component.

APPENDIX. PROOFS

As explained in the introduction and in Online Appendix A, Theorem 3 follows from Mongin & Pivato (2015 Proposition 1). We next prove Theorem 1. Statement (i) readily implies Statement (ii). We assume Statement (ii), and derive Statement (i). By Theorem 3, we obtain an AU representation. We derive proportionality of the V_j^i in the AU representation. We can let all V_j^i take value 0 at 0. The AU representation is a state- and time-dependent version of DEU. Gorman's uniqueness result is at this state- and time-dependent stage: the functions V_j^i, all "grounded" at 0, can jointly be replaced by $\lambda \times V_j^i$ for any $\lambda > 0$, independent of i and j, and by no other functions.

By act monotonicity, the n arrays $(V_j^1, \dots, V_j^m)$ through their sum all represent the same preference relation over acts ("column"). Hence, by standard uniquess, these n arrays of functions, grounded at 0, are proportional to each other. That is, each $(V_j^1, \dots, V_j^m)$ is d_j times $(V_1^1, \dots, V_1^m)$ for positive $d_2, \dots, d_n$, where we set $d_1 = 1$. Similarly, by stream monotonicity, the m arrays $(V_1^i, \dots, V_n^i)$ though their sum all represent the same preference relation over streams ("rows"), and each is q^i times $(V_1^1, \dots, V_n^1)$ for positive $q^2, \dots, q^m$ with $q^1 = 1$. We can normalize the q^is to sum to 1, and denote them p^i. All V_j^is are proportional to each other and to one function that can be denoted U. For U we can take V_1^1 or any other V_j^i.

For completeness, we give the uniqueness results of Theorem 1. By Gorman's aforementioned uniqueness result at the state- and time-dependent stage, now at this state- and time-independent stage we have: U is unique up to a positive factor (scale), and the d_js are unique up to one other common positive factor. Because of normalization, the p^js are unique. We can relax the requirement $U(0) = 0$ and add any constant, after which U is also unique up to location.

REFERENCES

Abdellaoui, Mohammed, Enrico Diecidue, & Ayse Öncüler (2011) "Risk Preferences at Different Time Periods: An Experimental Investigation," *Management Science* 57, 975–987.

Abdellaoui, Mohammed, Emmanuel Kemel, Amma Panin, & Ferdinand M. Vieider (2019) "Measuring Time and Risk Preferences in an Integrated Framework," *Games and Economic Behavior* 115, 459–469.

Agranov, Marina & Pietro Ortoleva (2022) "Revealed Preferences for Randomization: An Overview," *American Economic Review, Papers and Proceedings* 112, 426–430.

Alon, Shiri & Gabrielle Gayer (2016) "Utilitarian Preferences with Multiple Priors," *Econometrica* 84, 1181–1201.

Andreoni, James & Charles Sprenger (2012) "Risk Preferences Are not Time Preferences," *American Economic Review* 102, 3357–3376.

Anscombe, Frank J. & Robert J. Aumann (1963) "A Definition of Subjective Probability," *Annals of Mathematical Statistics* 34, 199–205.

Attema, Arthur E. (2012) "Developments in Time Preference and Their Implications for Medical Decision Making," *Journal of the Operational Research Society* 63, 1388–1399.

Bade, Sophie (2015) "Randomization Devices and the Elicitation of Ambiguity-Averse Preferences," *Journal of Economic Theory* 159, 221–235.

Baillon, Aurélien, Yoram Halevy, & Chen Li (2022) "Randomize at Your Own Risk: On the Observability of Ambiguity Aversion," *Econometrica* 90, 1085–1107.

Bardsley, Nicholas, Robin P. Cubitt, Graham Loomes, Peter Moffat, Chris Starmer, & Robert Sugden (2010) "*Experimental Economics; Rethinking the Rules*." Princeton University Press, Princeton, NJ.

Battigalli, Pierpaolo, Simone Cerreia-Vioglio, Fabio Maccheroni, & Massimo Marinacci (2017) "Mixed Extensions of Decision Problems under Uncertainty," *Economic Theory* 63, 827–866.

Baucells, Manel & Franz H. Heukamp (2012) "Probability and Time Tradeoff," *Management Science* 58, 831–842.

Berger, Loïc & Louis Eeckhoudt (2021) "Risk, Ambiguity, and the Value of Diversification," *Management Science* 67(3):1639–1647.

Berger, Loïc & Johannes Emmerling (2020) "Welfare as Equity Equivalents," *Journal of Economic Surveys* 34, 727–752.

Bleichrodt, Han & José Luis Pinto (2009) "New Evidence of Preference Reversals in Health Utility Measurement," *Health Economics* 18, 713–726.

Bommier, Antoine (2017) "A Dual Approach to Ambiguity Aversion," *Journal of Mathematical Economics* 71, 104–118.

Bommier, Antoine, Asen Kochov, & François le Grand (2017) "On Monotone Recursive Preferences," *Econometrica* 85, 1433–1466.

Broome, John R. (1991) "Desire, Belief and Expectation," *Mind* 100, 265–267.

Calford, Evan M. (2021) "Mixed Strategies and Preference for Randomization in Games with Ambiguity Averse Agents," *Journal of Economic Theory* 197, 105326.

Cerreia-Vioglio, Simone, David Dillenberger, Pietro Ortoleva, & Gil Riella (2019) "Deliberately Stochastic," *American Economic Review* 109, 2425–2445.

Chandrasekher, Madhav, Mira Frick, Ryota Iijima, & Yves Le Yaouanq (2022) "Dual-Self Representations of Ambiguity Preferences," *Econometrica* 90, 1029–1061.

Cheung, Stephen L. (2015) "Risk Preferences Are not Time Preferences: On the Elicitation of Time Preference under Conditions of Risk: Comment (#11)," *American Economic Review* 105, 2242–2260.

Cohen, Michèle, Jean-Yves Jaffray, & Tanios Said (1987) "Experimental Comparisons of Individual Behavior under Risk and under Uncertainty for Gains and for Losses," *Organizational Behavior and Human Decision Processes* 39, 1–22.

Cox, James C., Vjollca Sadiraj, & Ulrich Schmidt (2014) "Asymmetrically Dominated Choice Problems, the Isolation Hypothesis and Random Incentive Mechanisms," *PLoS ONE* 9, e90742.

DeJarnette, Patrick, David Dillenberger, Daniel Gottlieb, & Pietro Ortoleva (2020) "Time Lotteries and Stochastic Impatience," *Econometrica* 88, 619–656.

Diamond, Peter A. (1967) "Cardinal Welfare, Individual Ethics, and Interpersonal Comparison of Utility: Comment," *Journal of Political Economy* 75, 765–766.

Diecidue, Enrico, Ulrich Schmidt, & Peter P. Wakker (2004) "The Utility of Gambling Reconsidered," *Journal of Risk and Uncertainty* 29, 241–259.

Eichberger, Jürgen & Illia Pasichnichenko (2021) "Decision-Making with Partial Information," *Journal of Economic Theory* 198, 105369.

Eichberger, Jürgen, Simon Grant, & David Kelsey (2016) "Randomization and Dynamic Consistency," *Economic Theory* 62, 547–566.

Epper, Thomas & Helga Fehr-Duda (2015) "Risk Preferences Are not Time Preferences: Balancing on a Budget Line: Comment (#12)" *American Economic Review* 105, 2261–2271.

Epper, Thomas & Helga Fehr-Duda (2021) "Risk in Time: The Intertwined Nature of Risk Taking and Time Discounting," working paper.

Epstein, Larry G. & Yoram Halevy (2019) "Ambiguous Correlation," *Review of Economic Studies* 86, 668–693.

Fleurbaey, Marc (2010) "Assessing Risky Social Situations?," *Journal of Political Economy* 118, 649–680.

Gorman, William M. (1968) "The Structure of Utility Functions," *Review of Economic Studies* 35, 367–390.

Grant, Simon, Atsushi Kajii, Ben Polak, & Zvi Safra (2010) "Generalized Utilitarianism and Harsanyi's Impartial Oberver Theorem," *Econometrica* 78, 1939–1971.

Gul, Faruk (1992) "Savage's Theorem with a Finite Number of States," *Journal of Economic Theory* 57, 99–110. ("Erratum," 1993, *Journal of Economic Theory* 61, 184.)

Halevy, Yoram (2008) "Strotz Meets Allais: Diminishing Impatience and the Certainty Effect," *American Economic Review* 98, 1145–1162.

Harsanyi, John C. (1955) "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63, 309–321.

Hartmann, Lorenz (2021) "Maxmin Expected Utility with Non-Unique Prior," working paper.

Hill, Brian (2019) "A Non-Bayesian Theory of State-Dependent Utility," *Econometrica* 87, 1341–1366.

Ida, Takanori & Rei Goto (2009) "Simultaneous Measurement of Time and Risk Preferences: Stated Preference Discrete Choice Modeling Analysis Depending On Smoking Behavior," *International Economic Review* 50, 1169–1182.

Jaffray, Jean-Yves (1989) "Linear Utility Theory for Belief Functions," *Operations Research Letters* 8, 107–112.

Johnson, Cathleen, Aurélien Baillon, Han Bleichrodt, Zhihua Li, Dennie van Dolder, & Peter P. Wakker (2021) "Prince: An Improved Method for Measuring Incentivized Preferences," *Journal of Risk and Uncertainty* 62, 1–28.

Kahneman, Daniel & Amos Tversky (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, 263–291.

Ke, Shaowei & Qi Zhang (2020) "Randomization and Ambiguity Aversion," *Econometrica* 88, 1159–1195.

Keeney, Ralph L. & Howard Raiffa (1976) "*Decisions with Multiple Objectives.*" Wiley, New York (2nd edn. 1993, Cambridge University Press, Cambridge).

Kochov, Asen (2015) "Time and No Lotteries: An Axiomatization of Maxmin Expected Utility," *Econometrica* 83, 239–262.

Machina, Mark J. (2014) "Ambiguity Aversion with Three or More Outcomes," *American Economic Review* 104, 3814–3840.

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

Martins-da-Rocha, Victor Filipe & Rafael M. Rosa (2021) "An Anscombe– Aumann Approach to Second-Order Expected Utility," working paper.

McCarthy, David, Kalle Mikkola, & Teruji Thomas (2020) "Utilitarianism with and without Expected Utility," *Journal of Mathematical Economics* 87, 77–113.

Miao, Bin & Songfa Zhong (2015) "Risk Preferences Are not Time Preferences: Separating Risk and Time Preference: Comment (#13)," *American Economic Review* 105, 2272–2286.

Miao, Bin & Songfa Zhong (2018) "Probabilistic Social Preference: How Machina's Mom Randomizes Her Choice," *Economic Theory* 65, 1–24.

Mongin, Philippe & Marcus Pivato (2015) "Ranking Multidimensional Alternatives and Uncertain Prospects," *Journal of Economic Theory* 157, 146–171.

Moscati, Ivan (2018) "*Measuring Utility: From the Marginal Revolution to Behavioral Economics.*" Oxford University Press, Oxford, UK.

Nataf, André (1948) "Sur la Possibilité de Construction de Certain Macromodèles," *Econometrica* 16, 232–244.

Oechssler, Jörg, Hannes Rau, & Alex Roomets (2019) "Hedging, Ambiguity, and the Reversal of Order Axiom," *Games and Economic Behavior* 117, 380–387.

Oechssler, Jörg & Alex Roomets (2021) "Savage vs. Anscombe-Aumann: An Experimental Investigation of Ambiguity Frameworks," *Theory and Decision* 90, 405–416.

Onay, Selçuk & Ayse Öncüler (2009) "How Do We Evaluate Future Gambles? Experimental Evidence on Path Dependency in Risky Intertemporal Choice," *Journal of Behavioral Decision Making* 22, 280–300.

Pivato, Marcus (2013) "Risky Social Choice with Incomplete or Noisy Interpersonal Comparisons of Well-being," *Social Choice and Welfare* 40, 123–139.

Rohde, Ingrid M. T. & Kirsten I. M. Rohde (2015) "Managing Social Risks – Tradeoffs between Risks and Inequalities," *Journal of Risk and Uncertainty* 51, 103–124.

Rohde, Kirsten I.M. & Xiao Yu (2022) "Intertemporal Correlation Aversion – A Model-Free Measurement," *Management Science*, forthcoming.

Saito, Kota (2013) "Social Preferences under Risk: Equality of Opportunity versus Equality of Outcome," *American Economic Review* 103, 3084–3101.

Saito, Kota (2015) "Preferences for Flexibility and Randomization under Uncertainty," *American Economic Review* 105, 1246–1271.

Schneider, Florian & Martin Schonger (2018) "An Experimental Test of the Anscombe-Aumann Monotonicity Axiom," *Management Science* 65, 1667–1677.

Siniscalchi, Marciano (2022) "Structural Rationality in Dynamic Games," *Econometrica*, 90, 2437–2469.

Starmer, Chris (2000) "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature* 38, 332–382.

Strzalecki, Tomasz (2013) "Temporal Resolution of Uncertainty and Recursive Models of Ambiguity Aversion," *Econometrica* 81, 1039–1074.

Tsetlin, Ilia & Robert L. Winkler (2009) "Multiattribute Utility Satisfying a Preference for Combining Good with Bad," *Management Science* 55, 1942–1952.

Tversky, Amos & Daniel Kahneman (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297–323.

Wakker, Peter P. (2010) "*Prospect Theory: For Risk and Ambiguity.*" Cambridge University Press, Cambridge, UK.

Zimper, Alexander (2008) "Revisiting Independence and Stochastic Dominance for Compound Lotteries," *B.E. Journal of Theoretical Economics* (MS #1444).

Online Appendix of

"The Deceptive Beauty of Monotonicity, and the Million-Dollar Question: Row-First or Column-First Aggregation?"

Chen Li, Kirsten I.M. Rohde, & Peter P. Wakker

January, 2023

ONLINE APPENDIX A. PRECEDING MATHEMATICAL RESULT

The mathematics underlying our results has been known longtime. We do not bring mathematical novelties. This appendix discusses preceding literature. Theorem 3 has been known since Nataf (1948). There, rows described producers and columns described production inputs. Nataf presented[8] Theorem 3 to show when macro (column-first, or column-monotonic, defined below) aggregation of production inputs can be equivalent to micro (row-first, or row-monotonic) aggregation: only if there is not any interaction[9]. van Daal & Merkies (1988) provided an early historical account. This production example further illustrates the wide applicability of our framework.

Mongin & Pivato (2015 Proposition 1) provided the mathematically most general versions of Nataf's (1948) result, implying our Theorem 3. (Hence, we gave no proof of it.) In the mathematical theory of functional equations, these results have been known as generalized bisymmetry equations.[10] See Maksa (1999), who also pointed out their relatedness to economic aggregation. The special case of proportional representations in our Theorem 1 is equivalent to mathematical theorems on multisymmetry functional equations, explained by Münnich, Maksa, & Mokken (2000). Mongin & Pivato (2015 Theorem 1) is the most general result of this kind. See Zuber (2016) for related results and literature with Anscombe-Aumann outcome sets.

---

[8] He heavily used differentiability and his proof is not easily accessible.

[9] Formally, we use the suggestive term interaction to indicate preference relations that violate complete separability. In general, not only rows and columns, but every subset of cells can be nonseparable, i.e., be impacted by (interacting with) any other subset of cells.

[10] They search for functions allowing identity of Eqs. 5 and 6 im the main text.

Thus, the mathematics underlying our results has been known longtime. As for Theorem 1 on DEU, its novelty is not in mathematical generality but in simplicity and appeal. The preference conditions there can be stated verbally and are accessible to nonspecialists more than any preeding axiomatization of DEU. Although many authors, several cited later, used advanced implications of the preceding results in various decision theories, their basic impact for empirical and theoretical work, specified in Section 4 and applied in the rest of this paper, has not been presented before.

ONLINE APPENDIX B. EXTENSION TO INFINITE-DIMENSIONAL MATRICES

Extensions to matrices with infinitely many rows and/or columns are often of interest. This holds mainly for Theorem 1. Infinite-dimensional extensions of Theorem 3 are less common because they involve nonstandard functionals. Wakker & Zank (1999) examined them. We focus on Theorem 1 henceforth, interpreting rows as states and columns as timepoints, but using the general notation of Figure 2.

Equal-likely states in Figure 2 can capture all simple lotteries with rational probabilities (McCarthy, Mikkola, & Thomas 2020). Mixture-closedness or continuous distributions require a continuum of r^i. Such extensions can be obtained by standard techniques from mathematical measure theory. Theorem 9 provides a typical example. It is explained next.

We continue to assume n columns $c_1, \dots, c_n$ with $n \geq 2$ fixed. A row continues to be an element of $\mathbb{R}^n$. In the main text, we considered the special case of risk where each r^i had probability $1/m$, so that matrices could be identified with some simple probability distributions over columns. We now consider more general probability distributions over columns, such as the space of all simple probability distributions or all bounded ones. To this effect, instead of $R = \{r^1, \dots, r^m\}$, we now assume $R = [0,1)$, endowed with the uniform distribution P and instead of finite-dimensional matrices as before, we now consider functions from $R \times \{c_1, \dots, c_n\}$ to the reals. We continue to call such functions matrices. Preferences will be over matrices. We make the assumption characteristic of decision under risk: functions on $R \times \{c_1, \dots, c_n\}$ that generate the same probability distribution over columns are indifferent.

Using obvious notation, a simple probability distribution over rows can be denoted $(p^1: r^1, \dots, p^k: r^k)$, with k variable, and all probabilities positive. We identify it with a matrix that assigns row r^i to each set R^i, where $(R^1, \dots, R^k)$ partitions $[0,1)$ and $P(R^i) = p^i$ for each i. It, thus, is like the matrix in Figure 2, with R^i for r^i for each i, and $m = k$. It will be sufficient to impose our intuitive axioms only on such simple finite-dimensional matrices. *Row* and *column monotonicity* are now defined to hold for all simple matrices.[11] For each fixed $(R^1, \dots, R^k)$, Theorem 1 then gives a DEU representation. Normalizing $U(0) = 0$, $U(1) = 1$, these DEU representations agree on common domain by standard uniqueness results, giving a probability measure P' on $[0,1)$ that at this stage might be thought to possibly differ from P and even be only finitely additive. However, partitions $(R^1, \dots, R^k)$ with $P(R^i) = 1/k$, by symmetry, imply $P'(R^i) = 1/k = P(R^i)$. The unions of such R^i show that P' agrees with P on all $R \subset [0,1)$ with rational P probability. By monotonicity w.r.t. set inclusion, P' and P are identical. We have obtained a DEU representation for all simple matrices.

The extension of our theorems to all bounded matrices now follows using standard techniques from mathematical measure theory. Monotonicity with respect to rows and columns, but also with respect to outcomes, is imposed only on simple matrices. Thus, null events are avoided and strict preferences are properly implied. We reinforce outcome monotonicity to infinite dimensions by adding *pointwise monotonicity*: a matrix is weakly preferred if all its cells weakly dominate. This condition is as unobjectionable for infinite dimensions as it is for finitely many. Every bounded matrix is now “sandwiched” more and more tightly by pointwise dominating and dominated simple matrices. This determines a unique DEU value, such that strict inequality of DEU values implies strict preference (using transitivity). Next, we reinforce continuity into supnorm continuity, ensuring existence of constant equivalents. Then equality of DEU values, again using transitivity, implies indifference and, hence, we have a DEU representation. We have shown the following result.

[11] Bear in mind that we assume strictly positive probabilities, avoiding null events as required for outcome monotonicity.

THEOREM 9. Assume that: (a) matrices map $[0,1) \times \{c_1, \ldots, c_n\}$ to the reals and are measurable; (b) preferences are over matrices; (c) decision under risk holds with respect to the uniform distribution on $[0,1)$. That is, our domain of matrices is equivalent to probability distributions over "rows" in $\mathbb{R}^n$. On the domain of simple matrices/distributions, and also on the domain of all bounded matrices/distributions, discounted expected utility holds if and only if weak ordering, supnorm continuity, pointwise monotonicity, and monotonicity with respect to outcomes, rows, and columns hold. □

Extension to unbounded matrices and connected topological outcome spaces (including all convex sets of commodity bundles) can be obtained by Wakker's (1993) truncation continuity. The total subjective weight of space R is still assumed bounded here. Unbounded subjective weight of R may occur, for instance, if R reflects time rather than uncertainty, or populations of variable size. Then further continuity conditions have to be invoked, discussed for instance by Asheim et al. (2010), Banerjee & Mitra (2007), Christensen (2022), Drugeon & Huy (2022), Marinacci (1998), and Pivato (2022). For extensions to infinitely many columns, besides infinitely many rows, our extension techniques are similarly aplied to columns.

Theorem 9 can be used for all interpretations of columns. If they refer to ambiguous events (horses), versions of the AA framework result. Here it is usually assumed that only marginal distributions conditional on horses matter, which can be added as a preference condition. Then our structure becomes isomorphic to the set of maps from $\{c_1, \ldots, c_n\}$ to probability distributions over $\mathbb{R}$. Correlations between different c_j then play no role.

ONLINE APPENDIX C. THEORETICAL APPLICATIONS OF NATAF'S AGGREGATION RESULT TO PREFEFENCE AXIOMATIZATIONS

We briefly sketch some further theoretical applications to preference axiomatizations, in addition to Theorem 1 in the main text. We first assume that both rows and columns refer to events. Thus, $\{r^1, \ldots, r^m\}$ and $\{c_1, \ldots, c_n\}$ are two partitions of the universal event. In Figure 1, the intersection event $r^i \cap c_j$ gives outcome x_j^i. Outcome monotonicity implies that none of those intersections is empty or null. Uniform row and column monotonicity can be interpreted as versions of stochastic independence: being informed about one partition does not affect preferences over the other. Theorem 1 then gives an appealing axiomatization of subjective expected utility, alternative to Savage (1954). Pfanzagl (1968; Section 12.5) presented this result using the stochastic independence interpretation for $m = n = 2$. Mongin (2020) and Ceron & Vergopoulos (2021) independently generalized it to general m, n.

We next continue to assume that rows and columns refer to events, but we further assume decision under risk for the r^i, with probability $1/m$ for each r^i. We first consider the case where the c_js may have unknown probabilities. Theorem 1 gives expected utility for risk (evaluating each column). Our equally-likely case can cover all simple rational-probability distributions. Online Appendix B shows how more general probability distributions can be incorporated, and that subjective probabilities over rows must be equal to the objective probabilities over rows. Theorem 1 also gives expected utility for the horse events c_j and, thus, provides an alternative axiomatization of the original expected utility model of AA, using the two-stage framework that has become standard today. AA referred to standard mixture independence to axiomatize expected utility for risk, and also assumed horse monotonicity. In our approach, their mixture independence is weakened to risk monotonicity. For our monotonicities the event, say row, to be conditioned on always only involves one outcome per column, whereas for von Neumann-Morgenstern mixture independence (or Savage's sure-thing principle) such events to be conditioned on must be allowed to involve any number of rows, i.e. any number of outcomes per column. The symmetry of our two monotonicity conditions and, thus, of the treatment of risk and uncertainty, adds to the appeal of our alternative theorem. As a price to pay, we need continuous utility whereas AA allowed for complete generality in this regard.

If we interpret the c_js as persons rather than events, Theorem 1 becomes an alternative to Harsanyi's (1955) welfare result based on the veil of ignorance. His Pareto optimality is column monotonicity. Like AA, he refers to mixture independence to obtain EU, and we similarly generalize here. In Theorem 1 there is no middle ground: if the social welfare function is ordinal in the individual utilities then it must be cardinal, leading to a linear sum. This is the essence of Harsanyi's result. Grant et al. (2010) provided generalizations that relaxed the independence and monotonicity conditions in Harsanyi's result.

We, finally, present an implication where only one component is available at the outset, but we construct a second kind for auxiliary purposes. Gul (1992) considered a finite state space $\{r^1, \ldots, r^m\}$. *Acts* $(x^1, \ldots, x^m)$ map states to $\mathbb{R}$. Gul's preference relation on acts, denoted $\succcurlyeq'$ here, satisfies weak ordering, continuity, and outcome monotonicity, implying that all states are nonnull. One fixed event A (nontrivial subset of the state space) plays a special role explained later (reminiscent of Ramsey's (1931) ethically neutral event). We define the function $\overline{C}$ on acts as the certainty equivalent ("constant equivalent") function, and $R^1(y_1, y_2) = \cdots = R^m(y_1, y_2)$ as the certainty equivalent function of acts $(A: y_1, A^c: y_2)$, using obvious notation.

We take matrices as in Figure 2 with $n = 2$, $c_1 = A, c_2 = A^c$. We define our preference relation $\succcurlyeq$ over matrices as represented by Eq. 5. Thus, row monotonicity holds (Observation 4) and it is uniform because all R^j's are the same. The act $\left(R^1(x_1^1, x_2^1),\ \ldots\ldots, R^m(x_1^m, x_2^m)\right)$ can be identified with the equivalence class of corresponding matrices with entries $x_1^{j\prime}$ and $x_2^{j\prime}$ such that $R^j\left(x_1^{j\prime}, x_2^{j\prime}\right) = R^j\left(x_1^j, x_2^j\right)$ for all j. Uniform column monotonicity for $\succcurlyeq$ over matrices in Figure 2 is equivalent to Gul's Assumption 2 for $\succcurlyeq'$ on acts, a condition called *act independence* nowadays (Chew & Karni 1994). Thus, we obtain as a corollary of Theorem 1:

THEOREM 10. Under the assumptions of this subsection, the following four statements are equivalent:

(i) Expected utility holds for $\succcurlyeq'$ over acts.

(ii) Discounted expected utility holds for $\succcurlyeq$ over matrices.

(iii) Uniform column monotonicity holds for $\succcurlyeq$ over matrices.

(iv) Act independence holds for $\succcurlyeq'$ over acts.

□

In the above result, standard uniqueness results for DEU imply that the "discount weight" d_1 of the left column, after normalization, is the probability of event A resulting from the row probabilities. The conditions in Statements (iii) and (iv) are appealing because they mimic mixture independence for risk to the context of uncertainty.

Gul's axiomatization of subjective expected utility through act independence thus follows as a corollary of our Theorem 1. Our result is more general because Gul required the event A to satisfy a symmetry condition implying that it has subjective probability 0.5, which we do not need. Chew & Karni (1994) also provided this generalization. Our verbal proof, involving the Appendix in the main text and the preceding paragraphs, is considerably shorter and more accessible than that in Gul (1992 pp. 104-109) or Chew & Karni (1994). It is remarkable that Gul (1992) can be obtained as, essentially, a corollary of Nataf (1948).

Some other axiomatizations of expected utility used generalizations of bisymmetry axioms that are all more restrictive than Gul's Assumption 2: they also consider more than two columns and many events A (Köbberling & Wakker 2003 Theorem 16). Hence, they also follow as corollaries of our Theorems 1 and 10. Such results include Krantz et al. (1971, Theorem 6.9.10 which assumes $m = n =$2) and Münnich, Maksa, & Mokken (2000 Theorem 2).

REFERENCES FOR ONLINE APPENDIX

Asheim, Geir B., Walter Bossert, Yves Sprumont & Kotaro Suzumura (2010) "Infinite-Horizon Choice Functions," *Economic Theory* 43, 1–21.

Banerjee, Kuntal & Tapan Mitra (2007) "On the Continuity of Ethical Social Welfare Orders on Infinite Utility Streams," *Social Choice and Welfare* 30, 1–12.

Ceron, Federica & Vassili Vergopoulos (2021) "On Stochastic Independence under Ambiguity," *Economic Theory* 71, 925–960.

Chew, Soo Hong & Edi Karni (1994) "Choquet Expected Utility with a Finite State Space: Commutativity and Act-Independence," *Journal of Economic Theory* 62, 469–479.

Christensen, Timothy M. (2022) "Existence and Uniqueness of Recursive Utilities without Boundedness," *Journal of Economic Theory* 200, 105413.

Drugeon, Jean-Pierre & Thai Ha Huy (2022) "A not so Myopic Axiomatization of Discounting," *Economic Theory* 73:349–376.

Grant, Simon, Atsushi Kajii, Ben Polak, & Zvi Safra (2010) "Generalized Utilitarianism and Harsanyi's Impartial Oberver Theorem," *Econometrica* 78, 1939–1971.

Gul, Faruk (1992) "Savage's Theorem with a Finite Number of States," *Journal of Economic Theory* 57, 99–110. ("Erratum," 1993, *Journal of Economic Theory* 61, 184.)

Harsanyi, John C. (1955) "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63, 309–321.

Köbberling, Veronika & Peter P. Wakker (2003) "Preference Foundations for Nonexpected Utility: A Generalized and Simplified Technique," *Mathematics of Operations Research* 28, 395–423.

Krantz, David H., R. Duncan Luce, Patrick Suppes, & Amos Tversky (1971), "*Foundations of Measurement, Vol. I (Additive and Polynomial Representations).*" Academic Press, New York. (2nd edn. 2007, Dover Publications, New York.)

Maksa, Gyula (1999) "Solution of Generalized Bisymmetry Type Equation without Surjectivity Assumptions," *Aequationes Mathematicae* 57, 50–74.

Marinacci, Massimo (1998) "An Axiomatic Approach to Complete Patience and Time Invariance," *Journal of Economic Theory* 83, 105–144.

McCarthy, David, Kalle Mikkola, & Teruji Thomas (2020) "Utilitarianism with and without Expected Utility," *Journal of Mathematical Economics* 87, 77–113.

Mongin, Philippe (2020) "Bayesian Decision Theory and Stochastic Independence," *Philosophy of Science* 87, 152–178.

Mongin, Philippe & Marcus Pivato (2015) "Ranking Multidimensional Alternatives and Uncertain Prospects," *Journal of Economic Theory* 157, 146–171.

Münnich, Àkos, Gyula Maksa, & Robert J. Mokken (2000) "n-Variable Bisection," *Journal of Mathematical Psychology* 44, 569–581.

Nataf, André (1948) "Sur la Possibilité de Construction de Certain Macromodèles," *Econometrica* 16, 232–244.

Pfanzagl, Johann (1968) "*Theory of Measurement.*" Physica-Verlag, Vienna.

Pivato, Marcus (2022) "A Characterization of Cesaro Average Utility," *Journal of Economic Theory* 201 105440.

Ramsey, Frank P. (1931) "Truth and Probability." *In* Richard B. Braithwaite (ed.), *The Foundations of Mathematics and other Logical Essays,* 156–198, Routledge and Kegan Paul, London.

Reprinted in Henry E. Kyburg Jr. & Howard E. Smokler (1964, eds.) *Studies in Subjective Probability,* 61–92, Wiley, New York. (2nd edn. 1980, Krieger, New York.)

Savage, Leonard J. (1954) "*The Foundations of Statistics."* Wiley, New York. (2nd edn. 1972, Dover Publications, New York.)

van Daal, Jan & Arnold H.Q.M. Merkies (1988) "The Problem of Aggregation of Individual Economic Relations; Consistency and Representativity in a Historical Perspective." *In* Wolfgang Eichhorn (ed.) *Measurement in Economics* (Theory and Applications of Economic Indices), 607–637, Physica-Verlag, Heidelberg.

Wakker, Peter P. (1993) "Unbounded Utility for Savage's "Foundations of Statistics," and other Models," *Mathematics of Operations Research* 18, 446–485.

Wakker, Peter P. & Horst Zank (1999) "State Dependent Expected Utility for Savage's State Space; Or: Bayesian Statistics without Prior Probabilities," *Mathematics of Operations Research* 24, 8–34.

Zuber, Stéphane (2016) "Harsanyi's Theorem without the Sure-Thing Principle: On the Consistent Aggregation of Monotonic Bernoullian and Archimedean Preferences," *Journal of Mathematical Economics* 63, 78–83.