SIPTA Summer 2004 lectures on

Some Decision Theoretic issues for *Imprecise Probability Theory*

Teddy Seidenfeld – using joint work with Jay Kadane and Mark Schervish
Carnegie Mellon University

OUTLINE

1. **Static (non-sequential) decision theory.**

   a. **The framework of horse-lotteries and the Anscombe-Aumann theory of Subjective Expected Utility [SEU].**

   b. **Motivating *imprecise probability* by with an Impossibility Result for Cooperative Groups who try to be Bayes-coherent**

   c. **Three decision theories that attempt to relax the Bayesian norms and their relationship to SEU.**

   d. **Some limitations using binary comparisons – and how to avoid them!**

2. **Sequential Decisions for Imprecise Probability Theory.**

   a. **Violations of the *independence* postulate only – *ordering* retained.**

   b. **The value of information and**

   c. ***Dilation* (based on work with Larry Wasserman)**

3. **Summary and some tentative conclusions.**

Part 1: Some preliminaries.

In this presentation, the framework of Anscombe-Aumann horse-lotteries is convenient for focusing on coherent choice rules generated by uncertainty in the decision maker's degrees of belief, while utility may be left determinate.

We will use *L* the set of von Neumann-Morgenstern lotteries on a finite set of rewards, $\{r_1, \ldots, r_m\}$

A von Neumann-Morgenstern lottery *l* is given as the distribution

$$< \alpha_1 , \ldots \alpha_m >$$

where $\alpha_i \geq 0$ and $\Sigma_i\, \alpha_i = 1$, and α_I is the chance of winning reward r_i.

The *convex combination* of two lotteries, $l_3 = \beta l_1 \oplus (1\text{-}\beta) l_2$ is merely the reward-by-reward β-mixture of their respective lotteries $\alpha_{3\iota} = \beta\alpha_{1\iota} + (1-\beta\alpha_{2\iota})$.

The fixed finite partition $\Omega = \{\omega_1, ..., \omega_n\}$ is the space of the agent's uncertainty.

An option, a horse-lottery h is a function from Ω to L, from states to lotteries.

	ω_1	ω_2	ω_3	...	ω_n
h	l_1	l_2	l_3	...	l_n

The mixture of two horse lotteries $h_3 = \beta h_1 \oplus (1\text{-}\beta)h_2$ is merely the state-by-state β-mixture of their respective lotteries

	ω_1	ω_2	...	ω_n
h_1	l_{11}	l_{12}	...	l_{1n}
h_2	l_{21}	l_{22}	...	l_{2n}
h_3	$\beta l_{11} \oplus (1\text{-}\beta)l_{21}$	$\beta l_{12} \oplus (1\text{-}\beta)l_{22}$	...	$\beta l_{1n} \oplus (1\text{-}\beta)l_{2n}$

For this discussion, the following two IMPORTANT restrictions will be in place:

- **Act-state independence: no cases of "moral hazards" are considered.**

- **State-independent utility: no cases where the value of a prize depends upon the state in which it is received.**

The Anscombe-Aumann SEU theory of preference on horse lotteries is given by these 4 axioms:

AA-1 **Strict preference $<$ is a *weak order* over pairs of horse lotteries.**

$<$ *is anti-symmetric, transitive, and non-preference* (~) is transitive indifference.

AA-2 ***Independence*** **For all h_1, h_2, h_3, and $0 < \alpha \leq 1$.**

$$h_1 < h_2 \text{ if and only if } \alpha h_1 \oplus (1\text{-}\alpha)h_3 < \alpha h_2 \oplus (1\text{-}\alpha)h_3.$$

AA-3 **An *Archimedean* condition**

For all $h_1 < h_2 < h_3$ **there exist** $0 < \alpha, \beta < 1$,

with $$\alpha h_1 \oplus (1\text{-}\alpha)h_3 < h_2 < \beta h_1 \oplus (1\text{-}\alpha)h_3.$$

AA-4 ***State-independent utility.*** **In words, whenever ω is non-null, between two horse lotteries that differ solely in that one yields *b* in ω where the other yields *w*, the agent prefers the former act.**

The central Anscombe-Aumann theorem is this:

***A-A Theorem*: A decision maker's binary preferences between horse lotteries satisfies these four axioms if and only if preference over horse lotteries is by Subjective Expected Utility maximization with respect to a single pair <*p*, *u*>,**

where *p* **is a subjective probability on the states**

and *u* **is a cardinal (von Neumann-Morgenstern) utility on the rewards.**

***Aside*: The first three axioms – *Ordering*, *Independence*, and *Archimedes*, constitute the von Neumann-Morgenstern theory of cardinal utility over lotteries.**

- **Motivating the need to relax SEU theory in the direction of imprecise probabilities. The case of cooperative Bayesian decision making.**

Consider two SEU decision makers, Dick and Jane, who wish to form a cooperative partnership that will make decisions, constrained by the following two principles.

- **The partnership must satisfy the A-A theory of SEU maximization.**
- **(Simple) Pareto coordination—if each of Dick and Jane strictly prefers one option h1 to a second h2, then so too does the partnership.**

For convenience, suppose that Dick and Jane each strictly prefers reward r^* over reward r_*.

Possibility/Impossibility Results for Cooperative SEU compromises.

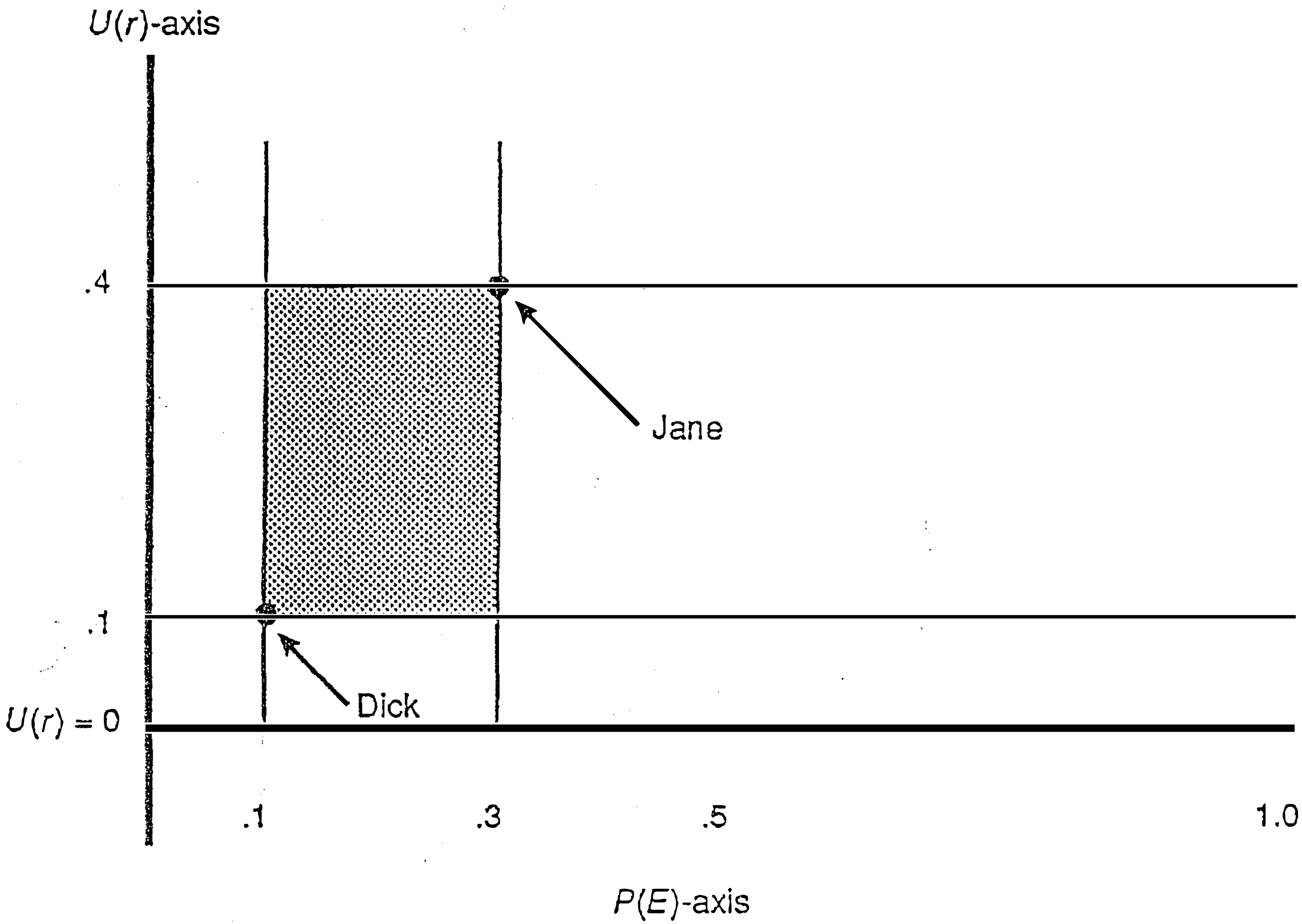
1. **If Dick and Jane share a common cardinal utility over the rewards, the candidate compromises for the group's preferences are given by an average of their two personal probabilities, and the common utility.**

2. **If Dick and Jane share a common personal probability over the states, the candidate compromises for the group's preferences are given by an average of their two cardinal utilities, and the common probability.**

3. **If Dick and Jane have any difference in their personal probability and do not share the same cardinal utility over rewards there are only autocratic solutions. One of them makes *all* the decisions for the partnership!**

Table 1. *"Horse lotteries" used in fixing the upper and lower probabilities and utilities*

	E	$\neg E$
A_1	r^*	r_*
$A_{2\epsilon}$	$(.9+\epsilon)(r_*)+(.1-\epsilon)(r^*)$	$(.9+\epsilon)(r_*)+(.1-\epsilon)(r^*)$
$A_{3\epsilon}$	$(.7-\epsilon)(r_*)+(.3+\epsilon)(r^*)$	$(.7-\epsilon)(r_*)+(.3+\epsilon)(r^*)$
A_4	r	r
$A_{5\epsilon}$	$(.6-\epsilon)(r_*)+(.4+\epsilon)(r^*)$	$(.6-\epsilon)(r_*)+(.4+\epsilon)(r^*)$



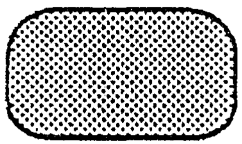
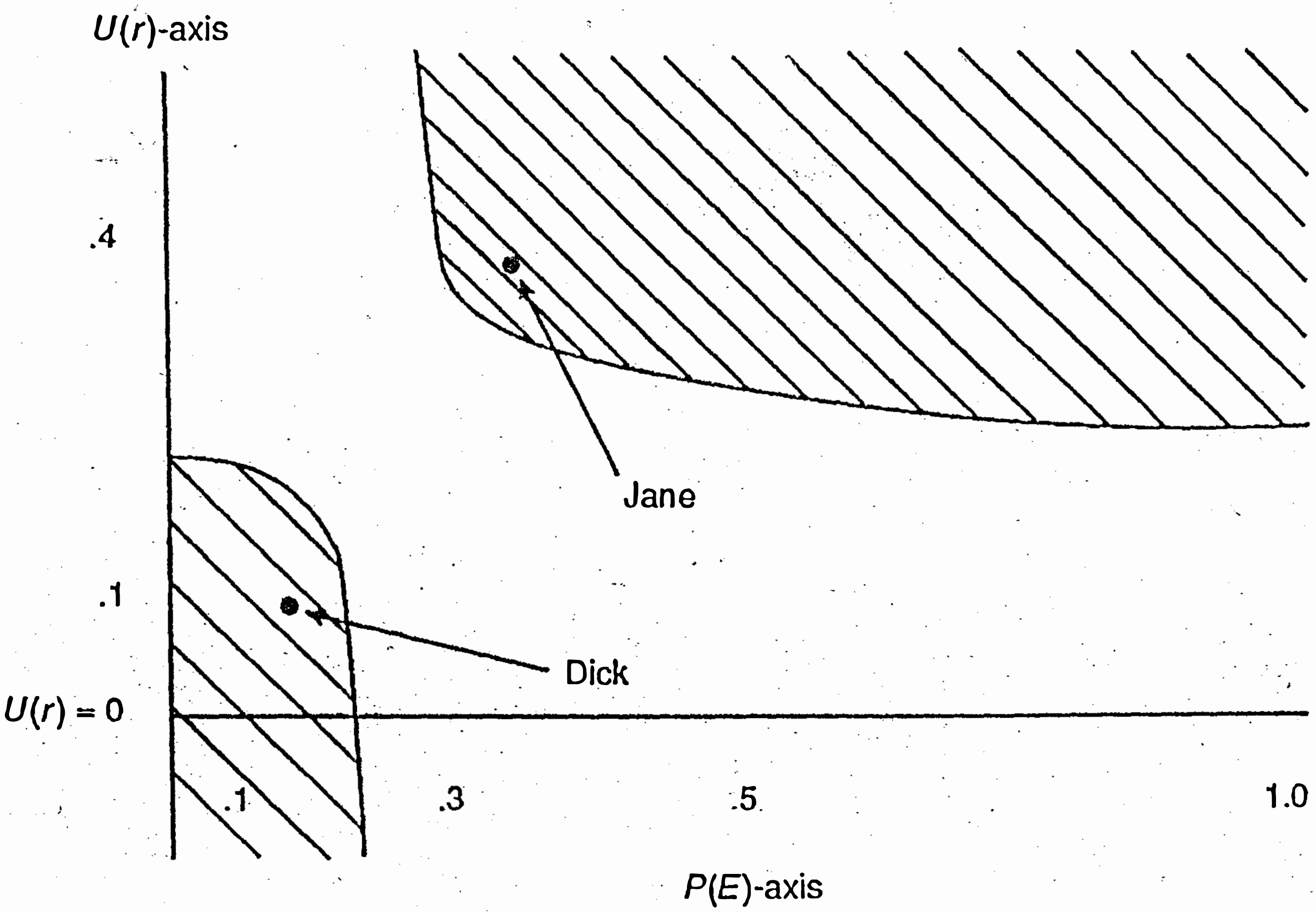
 designates the set of probability/utility pairs agreeing with the common preferences of Dick and Jane for the comparisons, above, in table 1.

Figure 2

Table 2. *"Horse lotteries" used in separating the set of compromises between Dick and Jane*

	E	$\neg E$
$A_{6\epsilon}$	$.785(r_*) + .215(r^*)$	$\epsilon(r_*) + .2(r) + (.8 - \epsilon)(r^*)$
$A_{7\epsilon}$	$(.2 - \epsilon)(r_*) + .8(r) + \epsilon(r^*)$	$.165(r_*) + .835(r^*)$



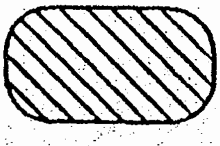

designates the set of probability/utility pairs agreeing with the common preferences of Dick and Jane for $A_{7\varepsilon}$ over $A_{6\varepsilon}$, $\varepsilon = .01$, in table 2.

Figure 3. Preferences which separate the family of agreeing probability/utility pairs

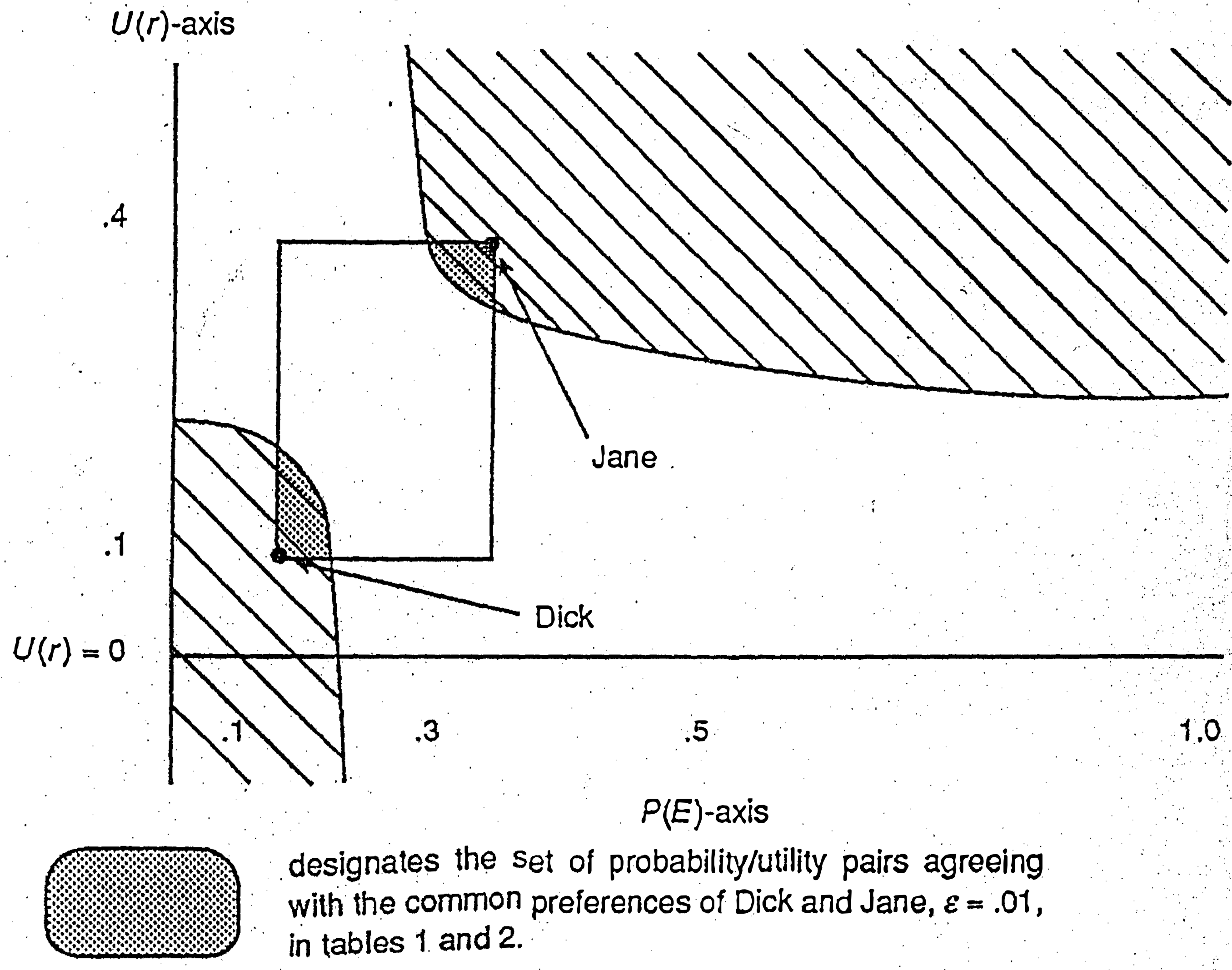


Figure 4. Preferences which separate the family of agreeing probability/utility pairs

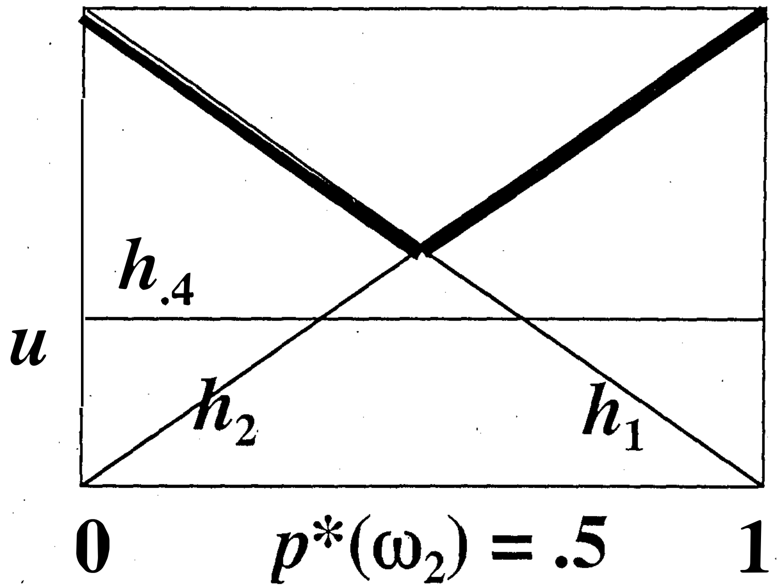
QUESTION: How to relax the SEU axioms in order to avoid such results?
Here are three decision theories that use *Imprecise Probabilities* in order to do it.

- **Γ-*Maximin*: Assign to each horse lottery its minimum of SEU values with respect to a (closed) set of probabilities and utilities. Choose from a (closed) feasible set of options any horse lottery whose value so assigned is maximum in that choice set.**

- ***Maximality* (Sen-Walley): For each pair of horse lotteries compare them to see whether one has greater SEU value than the other with respect to each probability (and utility) allowed. If so, the first is strictly preferred to the second. Choose from a (closed) feasible set of options any horse lottery that does not have another strictly preferred to it.**

- ***Coherence* (Levi's rule of E-admissibility): Choose from a (closed) feasible set of horse lotteries any one that maximizes SEU with respect to one of the allowed probability/utility pairs.**

Here is a useful example illustrating how these 3 rules differ.



Axiomatics for these three rules.

Γ-*Maximin* has been represented in the structure of horse-lotteries by Gilboa-Schmeidler (1989) in the following way:

They consider the setting where the decision maker has a determinate cardinal utility for rewards – so A-A theory applies to constant horse lotteries.

They retain A-A axioms 1, 3, and 4 – Ordering, Archimedes, and State-Independent Utility.

They do not require A-A axiom 2 – Independence – except for constant acts.

***Maximality* has, in effect, been axiomatized for an arbitrary set *S* of probability/utility pairs by SSK (1995) in the following way.**

Relax A-A axiom 1 to require only a *strict partial order* over horse lotteries.
Retain *Independence* and *State-independent Utility* AA-axioms 2 and 4.
A modified Archimedean axiom 3, as motivated by the following example.

- **Choose from a feasible set of options those that are not strictly preferred by any others.**

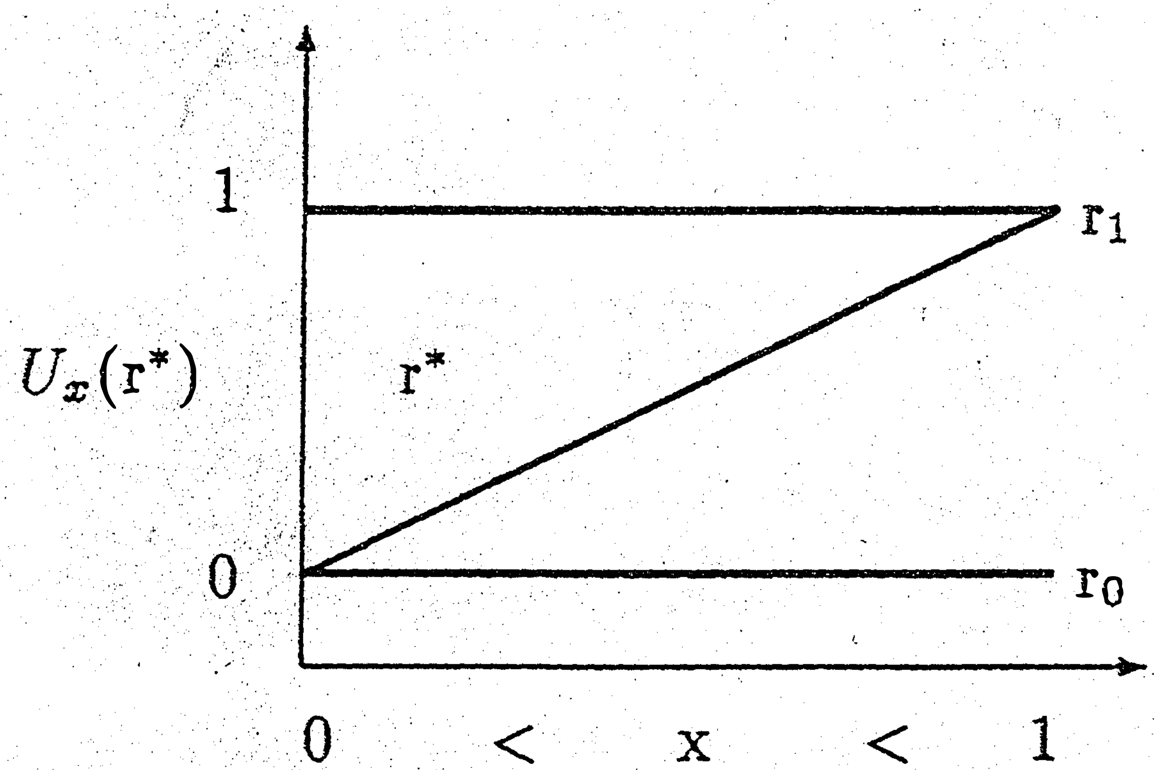


Figure 7. Example of restrictions of the usual Archimedean axiom

Coherence is the ongoing subject of current work, which I now summarize.

A coherent choice rule identifies the Bayes-solutions to a decision problem – the Bayes solutions taken with respect to the set *S*.

That is, given any (closed) set *O* of options, a choice rule *C* identifies the admissible options $C[O] = B$, for a non-empty subset $B \subseteq O$.

***Definition*: The choice rule is *<u>coherent</u>* with respect to a set *S* of prob/util pairs**

if for each $b \in B$ there is a pair $<p, u>$ in *S* such that *b* maximizes the

***p*-expected *u*-utility of options from *O*,**

and

***B* constitutes the set of all such solutions.**

***B* is the set of Bayes-solutions from *O*: solutions that are Bayes with respect to *S*.**

Alternatively, we can examine the *rejection* rule $R[O] = O \setminus C[O]$, which identifies the *inadmissible* options in O.

When C is coherent, the inadmissible options are those that fail to maximize p-expected u-utility for all pairs $\langle p, u \rangle$ in S: there is unanimity over S in the rejection of inadmissible options.

Three important cases, in increasing order of generality:

- S is a singleton pair, we have traditional *Subjective Expected Utility* theory.
- S is of the form $P \otimes u$ for a single utility u and a closed, convex set of probabilities P, and when all option sets are closed under mixtures, we have Walley-Sen's principle of *Maximality*.
- When S is a cross product of two convex sets $P \otimes U$, we have I. Levi's rule of *E-admissibility*.

There are several themes that lead to a set *S* that is not a singleton pair.

Here are 3 examples.

1. **Robust Bayesian Analysis – where *S* reflects a (typically) convex set of distributions obtained by varying either/or the *prior* or the *statistical model*, together with a single loss function.**

For example, the ε-*Contamination Model* is obtained by using a set of distns:

$\{(1\text{-}\varepsilon)\mu + \varepsilon q$: with μ a fixed distn, and $q \in Q$, where Q is a set of distributions$\}$

The idea is that with probability 1-ε the intended model μ generates the data. And with probability ε any distribution from *Q* might generate the data.

2. **Lower Probability with a (closed) convex set of distributions P and a single u.**
   - **CAB Smith, P. Walley (and many others), where S is based on one-sided previsions.**
   - ***Γ-Maximin* applied to convex sets of bets: Gilboa-Schmeidler theory.**
   - **Dempster-Shafer theory, with a decision rule suggested by Dempster.**

3. **Consensus in cooperative group decision making:**
   - **Levi – where $S = P \otimes U$. P is the convex set representing the group's uncertain degrees of belief, and U the convex set of its uncertain values.**
   - **SSK – S is a set of prob/util pairs that represent the unrestricted Pareto (strict) preferences in a group of coherent decision makers.**

It is convenient for focusing on coherent choice rules generated by uncertainty in the decision maker's degrees of belief, while utility is left determinate.

We will use *L* the set of von Neumann-Morgenstern lotteries on two prizes,

b ("better") and w ("worse"),

which serve respectively as the 1 and 0 of the cardinal utility function.

Thus, a von Neumann-Morgenstern lottery *l* is given as the mixture

$$\alpha b \oplus (1-\alpha)w,$$

with determinate utility α.

The fixed finite partition $\Omega = \{\omega_1, \ldots, \omega_n\}$ is the space of the agent's uncertainty.

An option, a horse-lottery h is a function from Ω to L, from states to lotteries.

	ω_1	ω_2	ω_3	…	ω_n
h	l_1	l_2	l_3	…	l_n

The mixture of two horse lotteries $h_3 = \beta h_1 \oplus (1\text{-}\beta)h_2$ is merely the state-by-state β-mixture of their respective lotteries

	ω_1	ω_2	…	ω_n
h_1	l_{11}	l_{12}	…	l_{1n}
h_2	l_{21}	l_{22}	…	l_{2n}
h_3	$\beta l_{11} \oplus (1\text{-}\beta) l_{21}$	$\beta l_{12} \oplus (1\text{-}\beta) l_{22}$	…	$\beta l_{1n} \oplus (1\text{-}\beta) l_{2n}$

- **The final preliminary point is to understand the importance of using a *strict preference* ($<$) relation, rather than a *weak preference* ($\leq$) relation when dealing with uncertainty represented by coherent choice rules.**

Here is a simple example using a two-state partition, $\Omega = \{\omega_1, \omega_2\}$, corresponding to the outcome of a toss of a coin landing *tails* (ω_1) or *heads* (ω_2).

Let $h_{.5}$ be the constant horse lottery that pays off $\alpha = .5$ in either state.

Let h be the horse lottery that pays *w* if the *tails* and *b* if *heads*.

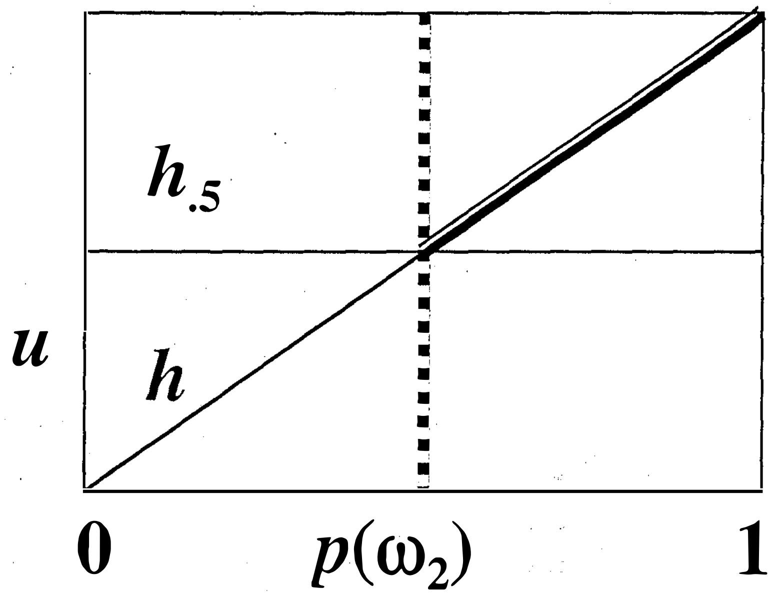
Contrast these two states of uncertainty.

$$S_1 = \{p\colon p(\omega_1) < p(\omega_2);\ \text{the coin is biased for } \textit{heads}\}$$

$$S_2 = \{p\colon p(\omega_1) \leq p(\omega_2);\ \text{the coin is } \textit{not} \text{ biased for } \textit{tails}\}$$

In a binary choice from the pair of option $O = \{h, h_{.5}\}$

- under S_1 only option h maximizes expected utility
- under S_2 both options are coherently admissible.



Strict preference (<) but not *weak preference* (≤) captures this distinction (SSK 95).

Limitations using strict preference: a binary comparison between acts.

Distinct convex sets of distributions that generate the same strict partial order over horse lotteries.

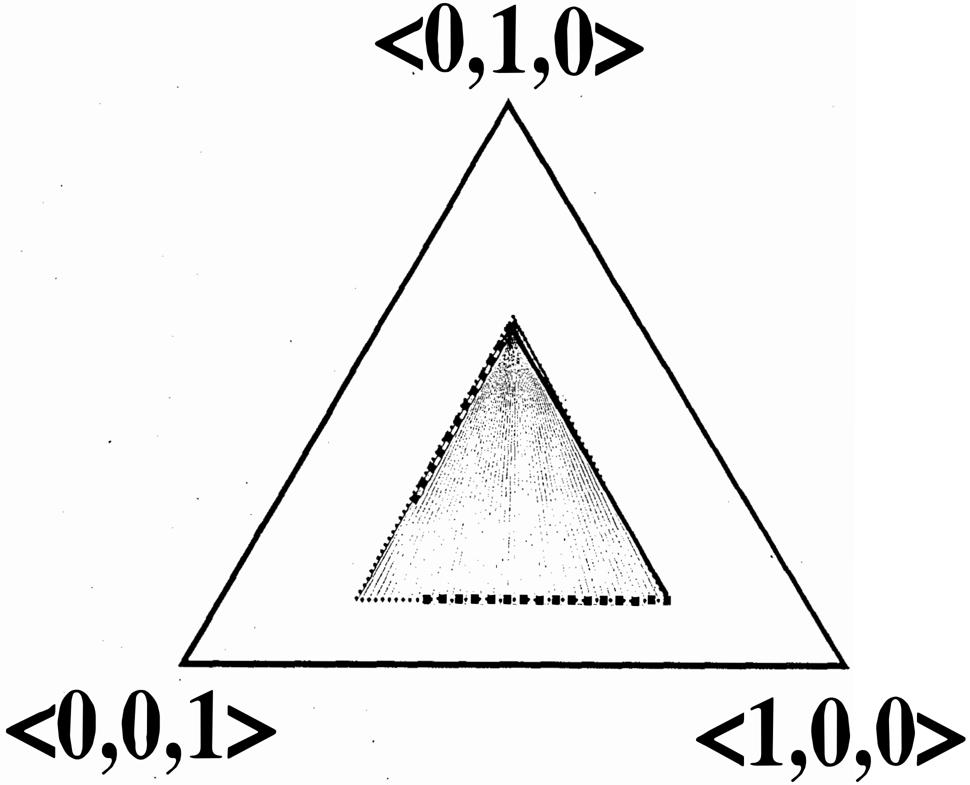
The general problem is that a strict preference between two of our horse lotteries

$$h_1 < h_2$$

defines a hyperplane of separation between the (convex) set of those distributions that give h_2 greater expected utility than h_1, and those distributions that do not.

But many distinct (convex) sets of distributions all may agree on these binary comparisons. They differ with regard to their boundaries.

- **Here is an example using three states: $\Omega = \{\omega_1, \omega_2, \omega_3\}$.**



A related problem is that choice rules that use only binary comparisons between options to determine what is admissible from a choice set, even when choice sets include all mixtures are:

- **coherent rules for *closed* (convex) sets of probabilities (Walley, 1990)**
- **and are generally *not* coherent rules otherwise (SSK&L, 2003).**

That is, such choice rules that reduce to binary comparisons conflate all the different sets of probabilities that meet the same set of supporting hyperplanes. These sets of probabilities are different at their boundaries only.

- **As we see next, these different (convex) sets of probabilities – though they share all the same supporting hyperplanes – are distinguished one from another by what they make admissible in non-binary choices.**

Using a coherent choice rules to distinguish between sets of probabilities

The following toy-example illustrates how to use a coherent choice rule to "test" for the presence/absence of a specific distribution $p^* \in P$ when $S = P \otimes u$, regardless the nature of the set P.

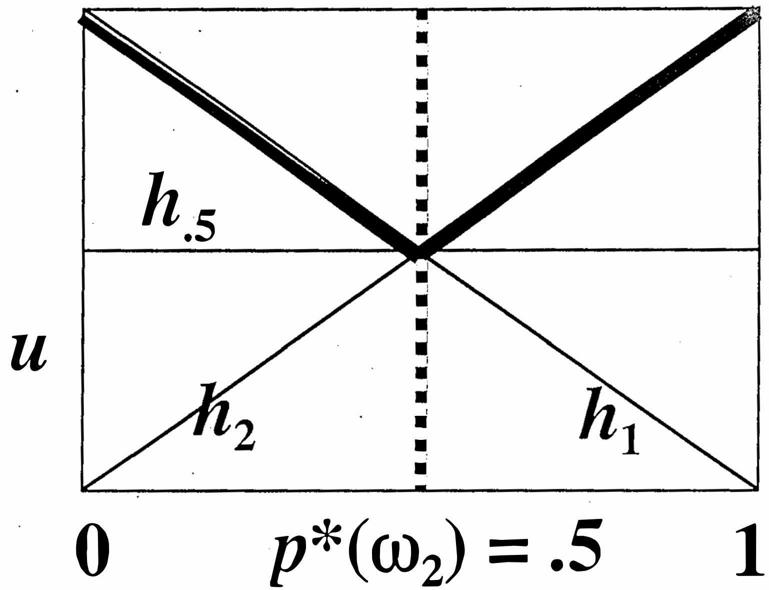
Return to the example of the coin with an unknown bias for landing *heads* (ω_2). Consider now the choice problem with these three options

Let $h_{.5}$ be the constant horse lottery that pays off $\alpha = .5$ in either state.

Let h_1 be the horse lottery that pays w if the *heads* and b if *tails*.

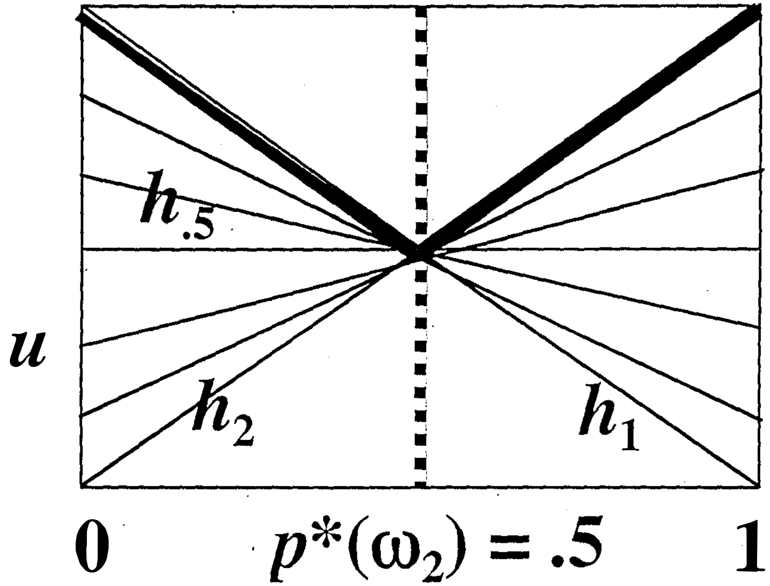
Let h_2 be the horse lottery that pays w if the *tails* and b if *heads*.

We use the choice problem $O = \{h_{.5}, h_1, h_2\}$ to "test" for $p^* = (1/2, 1/2)$



- $C[O] = \{h_{.5}, h_1, h_2\}$, **all three options are admissible, if and only if** $p^* \in P$.

- **And if we close the option set under convex combinations, the same holds.**



The entire (convex) set of mixed options generated by h_1 and h_2 is admissible
if and only if $p^* \in P$.

- **This technique generalizes to permit a "test" for an arbitrary distribution p on an arbitrary (finite) Ω against an arbitrary set P.**

3. *A representation of coherent choice rules.*

The Anscombe-Aumann SEU theory of preference on horse lotteries is given by these 4 axioms:

AA-1 **Strict preference $<$ is a *weak order* over pairs of horse lotteries.**

$<$ *is anti-symmetric, transitive, and non-preference* (~) is transitive indifference.

AA-2 ***Independence*** **For all h_1, h_2, h_3, and $0 < \alpha \leq 1$.**

$$h_1 < h_2 \text{ if and only if } \alpha h_1 \oplus (1\text{-}\alpha)\mathrm{h}_3 < \alpha h_2 \oplus (1\text{-}\alpha)\mathrm{h}_3.$$

AA-3 **An *Archimedean* condition**

For all $h_1 < h_2 < h_3$ there exist $0 < \alpha, \beta < 1$,

with $$\alpha h_1 \oplus (1\text{-}\alpha)\mathrm{h}_3 < h_2 < \beta h_1 \oplus (1\text{-}\alpha)\mathrm{h}_3.$$

AA-4 ***State-independent utility*.** **In words, whenever ω is non-null, between two horse lotteries that differ solely in that one yields *b* in ω where the other yields *w*, the agent prefers the former act.**

In the language of this presentation, the Anscombe-Aumann *SEU* Theorem is that such a preference over pairs of horse lotteries is represented by a coherent choice rule, with *P* a singleton set comprised by one probability distribution *p* over Ω.

Since the first axiom AA-1 is that preference is a weak order, we can use Sen's result that the corresponding choice rule satisfies *Properties alpha* and *beta*:

- **Property *alpha*: You cannot promote an inadmissible option into an admissible option by adding to the choice set of options.**
- **Property *beta*: If two options are both admissible from some choice set, then whenever both are available, either both are admissible or neither one is.**

With these two properties, define the (strict) preference relation between two options $h_1 < h_2$ to mean that $C[\{h_1, h_2\}] = \{h_2\}$, i.e., only h_2 is admissible from the pair,

Then the rest of the axioms are easily expressed in terms of choice rules.

In our setting, we can generalize Anscombe-Aumann theory to accommodate all coherent choice rules, as follows. I will express the axiom in words. Each is evidently necessary for a choice rule to be coherent.

Structural Assumption:

Each choice set is O is *closed*, to insure admissible options exist.

- **<u>Axiom 1a</u>:** ***Property alpha*** **– you can't promote an inadmissible option into an admissible option by adding to the option set.**
- **<u>Axiom 1b</u>: You cannot promote an inadmissible option into an admissible option by deleting inadmissible options from the choice set.**

From these two, define a strict partial order $\lessdot$ on sets of options as follows. Let O_1 and O_2 be an option sets. Recall that $R[\bullet]$ is the rejection rule associated with the choice rule $C[\bullet]$.

Defn: $O_1 \lessdot O_2$ *if and only if* $O_1 \subseteq R[O_1 \cup O_2]$.

That is, it follows from Axioms 1a & 1b that $\lessdot$ is a strict partial order.

Axiom 2a ***Independence*** **is expressed for $\prec$ over sets of options just as before.**

$$O_1 \prec O_2 \textit{ if and only if } \alpha O_1 \oplus (1\text{-}\alpha)h \prec \alpha O_2 \oplus (1\text{-}\alpha)h.$$

Let $H(O)$ be the result of taking the (closed) convex hull of the choice set O. That is, H augments O with all its mixed options – and then close up the set.

Axiom 2b ***Convexity*** **If $h \in O$ and $h \in R[H(O)]$, then $h \in R[O]$.**
Inadmissible options from a mixed set remain inadmissible even before mixing.

Note: Axiom 2b is needed to eliminate the choice rules Walley-Sen *Maximality*, and *Γ-Maximin*, which are not generally coherent.

Axiom-3: The Archimedean condition requires a technical adjustment, as the canonical form used by von Neumann-Morgenstern and Anscombe-Aumann theory is too restrictive in this setting.

The reformulated version expresses the Archimedean condition as a continuity principle compatible with strict preference as a strict partial order.

Axiom 4 is unchanged from the Anscombe-Aumann theory.

Main Result: A choice rule is coherent *if and only if* it satisfies these 4 axioms.

Corollary: The coherent choice rule associated with the set *P* of probabilities on Ω is unique to *P*. Different sets *P* yield different choice coherent choice rules.

The emphasis here is on the fact that the set *P* of probabilities used to represent the coherent choice rule is entirely arbitrary. There is no assumption that P is closed, or convex, or even that it is connected.

Summary and principal conclusions regarding coherent choice rules when the decision maker's degrees of belief are given by a set of probabilities.

- **We saw that some choice rules, *Maximality*, are coherent only in special circumstances, as when the option set *O* is convex and the set *P* is closed.**
- **Coherent choice rules based on binary comparisons between pairs of horse lotteries fail to distinguish among different convex sets of probabilities, all of which share the same supporting hyperplanes.** SSK-95
- **This is not the case for decision rules like Levi's *E-admissibility*.**
- **A fully general representation exists for coherent choice rules, modeled on the A-A theorem. It applies with any set P of probabilities.**
- **Coherent choice rules are capable of distinguishing between any two different sets of probabilities, regardless their structure.**

3. Sequential Decisions for Imprecise Probability Theory.

a. Violations of the *independence* postulate only – *ordering* retained.
applied, e.g., to Γ-*Maximin*.

b. The value of information
Applied to all three rules: Γ-*Maximin*; *Maximality*; *E-admissibility*.

c. *Dilation* (based on work with Larry Wasserman)

First, however, we review static coherence – deFinetti's "Book" criterion and that *each of our rules is coherent in this sense*.

Begin with a short review of deFinetti's *Book* argument for coherent wagering.

A Zero-Sum (sequential) game is played between a *Bookie* and a *Gambler*, with a *Moderator* supervising.

Let X be a random variable defined on a space Ω of possibilities, a space that is well defined for all three players by the Moderator.

The *Bookie*'s prevision *p(X)* on the r.v. *X* has the operational content that,

when the *Gambler* fixes a real-valued quantity $\alpha_{X,\,p(X)}$

then the resulting payoff to the *Bookie* is

$$\alpha_{X,\,p(X)}\,[\,X - p(X)\,],$$

with the opposite payoff to the Gambler.

A simple version of deFinetti's *Book* game proceeds as follows:

1. **The *Moderator* identifies a (possibly infinite) set of random variables $\{X_i\}$**

2. **The *Bookie* announces a prevision $p_i = p(X_i)$ for each r.v. in the set.**

3. **The *Gambler* then chooses (*finitely many*) non-zero terms $\alpha_i = \alpha_{X_i, p(X_i)}$.**

4. **The *Moderator* settles up and awards Bookie (Gambler) the respective SUM of his/her payoffs: *Total payoff to Bookie* = $\sum_{i=1}^{n} \alpha_i [X_i - p_i]$.**

$$\textit{Total payoff to Gambler} = -\sum_{i=1}^{n} \alpha_i [X_i - p_i].$$

***Definition*:**

The *Bookie*'s previsions are <u>*incoherent*</u> if the *Gambler* can choose terms α_i that assures her/him a (*uniformly*) positive payoff, regardless which state in Ω obtains – so then the *Bookie* loses for sure.

A set of previsions is <u>*coherent*</u>, if not incoherent.

Theorem **(deFinetti):**

A set of previsions is coherent ***if and only if***

each prevision $p(X)$ **is the expectation for** X **under a common (finitely additive) probability** P**.**

That is, $$p(X) = E_{P(\bullet)}[X] = \int_{\Omega} X \, dP(\bullet)$$

Two Corollaries**:**

Corollary 1: **When the random variables are** ***indicator functions*** **for events** $\{E_i\}$**, so that the gambles are simple bets – with the** α**'s then the stakes in a winner-take-all scheme – then the previsions** p_i **are coherent** ***if and only if*** **each prevision is the probability** $p_i = P(E_i)$**, for some (f.a.) probability** P**.**

Aside on conditional probability:

Definition: A *called-off prevision* $p(X \,||\, E)$ for X,

made by the *Bookie* on the condition that event E obtains,

has a payoff scheme to the Bookie: $\alpha_{X||E}\ E[\ X - p(X \,||\, E)\ ]$.

Corollary 2: Then a called-off prevision $p(X \,||\, E)$ is coherent alongside the

(coherent) previsions $p(X)$ for X, and $p(E)$ and E *if and only if*

$p(X \,||\, E)$ is the *conditional expectation* under P for X, given E.

That is, $p(X \,||\, E) = E_{P(\bullet|E)}[X] = \int_\Omega X\, dP(\bullet|E)$ and is $P(X \mid E)$ if X is an event.

- In this sense, the *Bookie*'s conditional probability distribution $P(\bullet|E)$ is the norm for her/his *static called-off* bets.
- Coherence of *called-off* previsions is not to be confused with the norm for a *dynamic learning rule*, e.g., when the Bookie learns that E obtains.

There are two aspects of deFinetti's coherence criterion that we relax.

1. Previsions may be *one-sided*, to reflect a difference between *buy* and *sell* prices for the *Bookie*, which depends upon whether the *Gambler* chooses a *positive* or *negative* α-term in the payoff $\alpha_{X,\,p(X)}\,[\,X - p(X)\,]$ to the *Bookie*.

For positive values of α, allow the *Bookie* to fix a maximum *buy*-price.

- **Betting on event *E*, this gives the *Bookie*'s lower probability $p_*(E)$,**

$$\alpha^{+}\,[\,E - p_*(E)\,].$$

For negative values of α, allow the *Bookie* to fix a minimum *sell*-price.

- **Betting against event *E*, this gives the *Bookie*'s upper probability $p^*(E)$,**

$$\alpha^{-}\,[\,E - p^*(E)\,].$$

At odds between the lower and upper probabilities, the *Bookie* rather not wager!

This approach has been explored for more than 40 years!

(See http://www.sipta.org/ the *Society for Imprecise Probabilities, Theories and Practices*)

For example, when dealing with upper and lower probabilities:

Theorem **[C.A.B. Smith, 1961]**

- **If the *Bookie*'s one-sided betting odds $p_*(\bullet)$ and $p^*(\bullet)$ correspond,**

  respectively, to the infemum and supremum of probability values from a *convex* set of (coherent) probabilities, then the Bookie's wagers are coherent: then the *Gambler* can make no *Book* against the *Bookie*.

- **Likewise, if the *Bookie*'s one-sided *called-off* odds $p_*(\bullet \| E)$ and $p^*(\bullet \| \mathrm{E})$**

  correspond to the infemum and supremum of conditional probability values, given E, from a *convex* set of (coherent) probabilities, then they are coherent.

Say that a one-sided *previson* is (strictly) *favorable* to the bookie just in case it is strictly preferred to abstaining in a pairwise choice.

Static coherence result, for each of our 3 decision rules:

No (finite) set of strictly favorable previsions leads to a sure loss.

Proof: **Use the representation theorem for the respective decision rule and apply the one-sided version of deFinetti's result, i.e.,**

If a one-sided prevision if strictly favorable, it has positive expected value. The sum of finitely many random variables with strictly positive expected value cannot have negative value in each state of Ω.

We use sequential decisions to explore the dynamical properties of our rival decision rules.

The (standard) notation for a sequential decision is to represent

a choice node with a box □

and a chance node with a circle ○

Lines radiating from a box represent options.

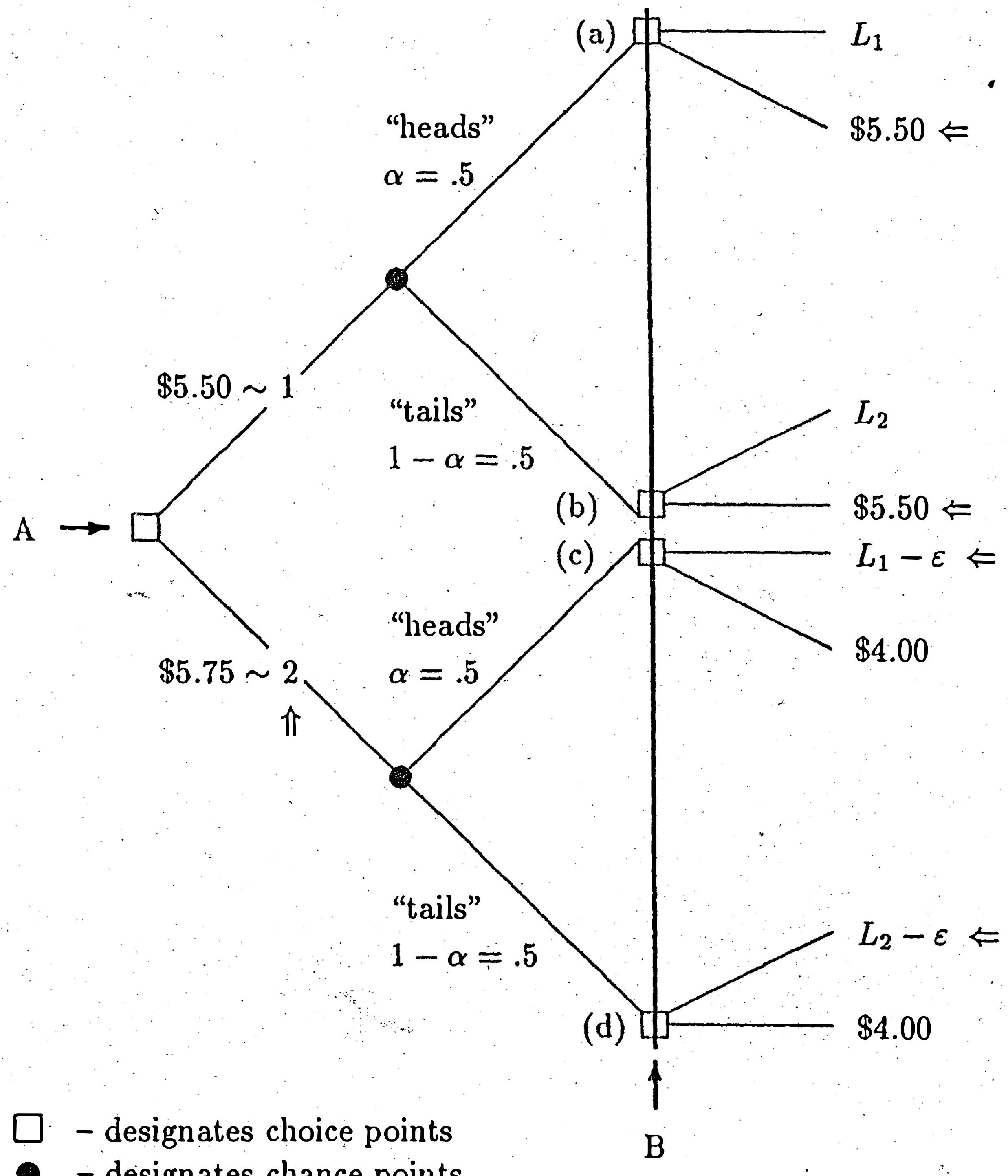
Lines radiation from a circle represent outcomes of a random variable.

Let us consider a decision rule, such as *Γ-Maximin*, that violates the Independence axiom but retains Ordering.

One such violation is where two von Neumann-Morgenstern lotteries are indifferent, but where their simple mixture is strictly preferred to either.

$$5.00 \approx L_1 \approx L_2 < .5L_1 \oplus .5L_2 \approx 6.00.$$

Consider the following sequential two decision problems.



□ – designates choice points

● – designates chance points

⇒ – designates chosen alternative

Figure 5. First version of the sequential decision: an illustration of sequential incoherence for a failure of mixture dominance ("betweenness"). At choice node A option 2 is preferred to option 1. At each choice node B this preference is reversed.

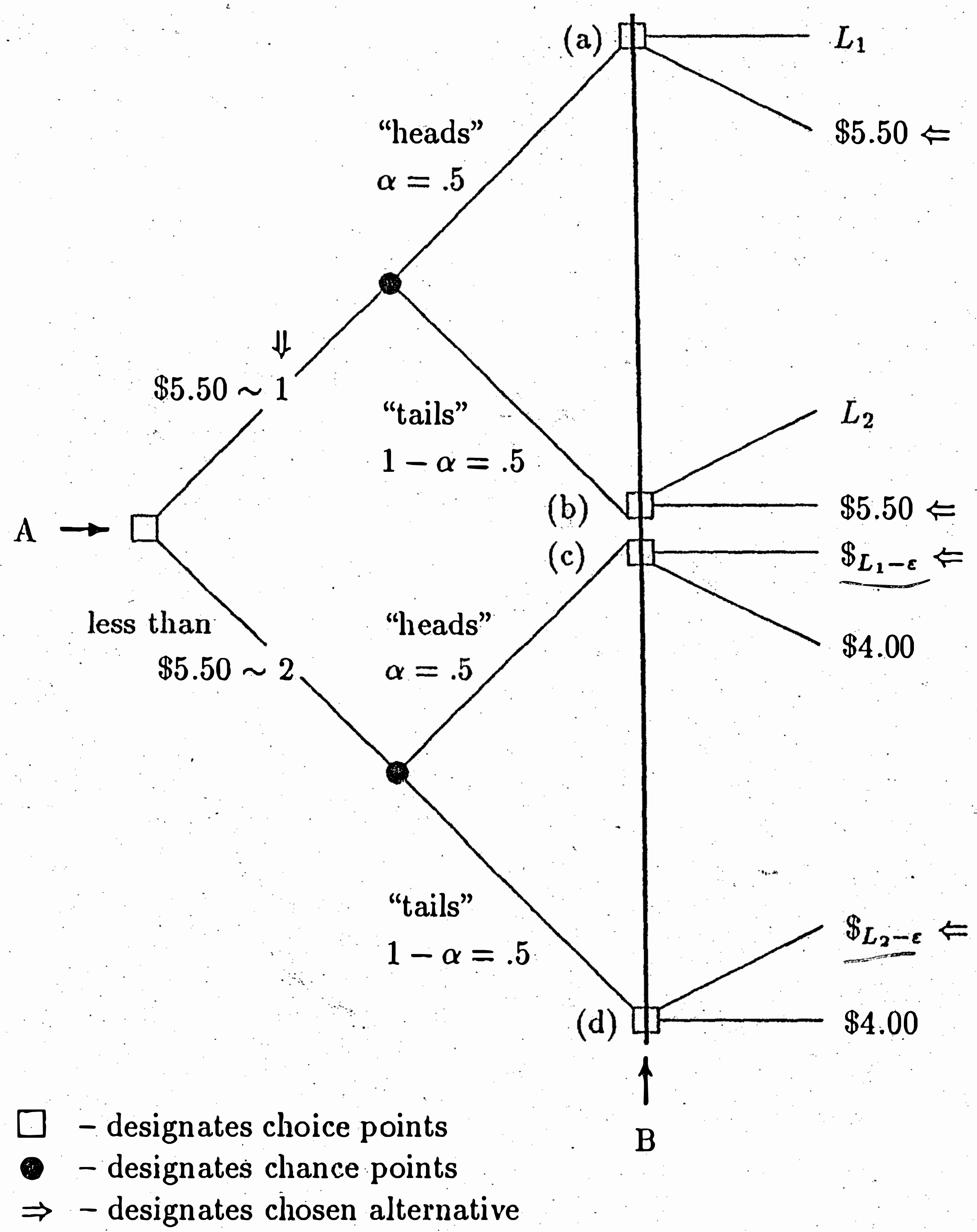


Figure 6. Second version of the sequential decision: an illustration of sequential incoherence for a failure of mixture dominance ("betweenness"). At choice node A option 1 is preferred to option 2. The tree results by replacing $L_i - \varepsilon$ $(i = 1, 2)$ from Figure 6.5 with \$-equivalents under $\preceq$.

What we see here is that

the decision maker's choices in these sequential decisions are NOT invariant over substitution of indifferent outcomes at terminal nodes in the tree.

***General Result*:**

- **The decision maker who is coherent in the static sense but who violates only the Independence postulate will suffer such *sequential incoherence*.**

- **The decision maker using *E-admissibility* will be sequentially coherent**

Next, we examine how our three decision rules work to assess the value of new information.

For this, we begin with an old result about SEU theory:

- **By the standards of SEU-theory, if we may postpone a terminal decision without costs to a acquire new (cost-free) information, we will do so, and have a strict preference for doing so in case the new information might alter our choice.**

Outline

1. Introduction
2. The expected value of sample information
   a. Countably additive, proper Bayesian
   b. Finitely but not countably additive Bayesian
   c. Improper proper Bayesian
3. Generalized Bayesian Decision Theory with Sets of Probabilities
   a. Γ- Maximin
   b. E-admissibility
   c. Maximality
4. Is Ignorance Bliss?

2. The Expected Value of Sample Information

* Intuitive argument

* Precise argument

Let $U(d, \theta)$ be your utility function, which depends on both your decision d and on $\theta \epsilon \Omega$, the unknown state of the world. You have a distribution $p(x, \theta)$ that jointly describes your probabilities for the data $x \epsilon \mathcal{X}$ and for θ. Without the data x, you would choose d to maximize

$$\int_{\mathcal{X}} \int_{\Omega} U(d, \theta) p(x, \theta) d\theta dx \tag{1}$$

If you were to learn the data x, you would maximize your utility with respect to your conditional distribution $p(\theta \mid x)$ i.e. maximize

$$\int_{\Omega} U(d, \theta) p(\theta \mid x) d\theta, \tag{2}$$

which has expectation, with respect to the unseen value of x,

$$\int_{\mathcal{X}} \left[\max_{d} \int_{\Omega} U(d, \theta) p(\theta \mid x) d\theta \right] p(x) dx. \tag{3}$$

The intuitive argument above suggests that (3) is no smaller than (1).

To show this, let d^* be a maximizer of (1). [The argument works just as well, if such a d^* does not exist, for d^* to be an ϵ-maximizer of (1)]. Then for each x in (2),

$$\max_{d} \int_{\Omega} U(d, \theta) p(\theta \mid x) d\theta \geq \int_{\Omega} U(d, \theta) p(\theta \mid x) d\theta. \tag{4}$$

Integrating both sides of this inequality with respect to x, yields

$$
\begin{aligned}
\int_{\mathcal{X}} \left[\max_d \int_\Omega U(d,\theta)p(\theta \mid x)d\theta\right] p(x)dx &\geq \int_{\mathcal{X}} \int_\Omega U(d,\theta)p(\theta \mid x)d\theta p(x)dx \\
&= \int_{\mathcal{X}} \int_\Omega U(d,\theta)p(\theta, x)d\theta dx \\
&= \max_d \int_{\mathcal{X}} \int_\Omega U(d,\theta)p(\theta, x)d\theta dx,
\end{aligned} \tag{5}
$$

as claimed.

Example 1 (finitely, but not countably additive distributions)

Suppose that there are two states of the world, A and B, each of which has probability .5 in your current opinion. Imagine that you can observe a positive integer n. If A is true, the integer n has a geometric distribution, as follows:

$$P(n|A) = (1/2)^{(n+1)} \tag{6}$$

for each integer $n > 0$. However, if B is the case, the integer n is uniformly distributed on the integers in your opinion.

If a particular integer, say $n = 3$, is observed, an easy application of Bayes Theorem shows that

$$P(A|n = 3) = 1 \tag{7}$$

and in fact, this is true whatever value of n is observed. Hence you are in the peculiar state of belief that although your prior is even between A and B, you know that conditional on the observation of N, regardless the value of N, you believe with certainty that A is true. Which then is your prior, what you believe now, or what you know you would believe if you could observe N?

Suppose that you currently hold a ticket that pays \$1 if B is true, and \$-1 if A is true. Currently your expected winnings are \$0. However, you know that if you were to obtain the integer n, your expected winnings would be \$-1. Would you pay \$.50 not to receive the data? It seems that you would have to.

Example 3 Consider a binary-state decision problem, $\Omega = \{\omega_1, \omega_2\}$, with three feasible options. Option f yields an outcome worth 1 utile if state ω_1 obtains and an outcome worth 0 untiles if ω_2 obtains. Option g is the mirror image of f and yields an outcome worth 1 utile if ω_2 obtains and an outcomes worth 0 utiles if ω_1 obtains. Option h is constant in value, yielding an outcome worth 0.4 utiles regardless whether ω_1 or ω_2 obtains. Figure 1 graphs the expected utilities for these three acts.

Let the mixing variable α equal 1 or 0 as a *fair* coin lands Heads up or Tails up on a toss, so that $P(\alpha = 1) = P(\alpha = 0) = .5$. Assume, also, that α is independent of the states, Ω, over which the *pure* options are defined, for that each $P \in \mathcal{P}$, $P(\alpha, \omega) = .5P(\omega)$.

As a modification of Example 3, consider the mixed options m, and n, defined as follows.

$$m = \alpha f \ \oplus (1 - \alpha)g$$

$$n = \alpha g \ \oplus (1 - \alpha)f.$$

Thus, m is the mixed act that uses the fair coin to bet on ω_1 if Heads and to bet on ω_2 if Tails. Likewise, n is the dual mixed act that uses the same fair coin to bet on ω_2 if Heads and to bet on ω_1 if Tails.

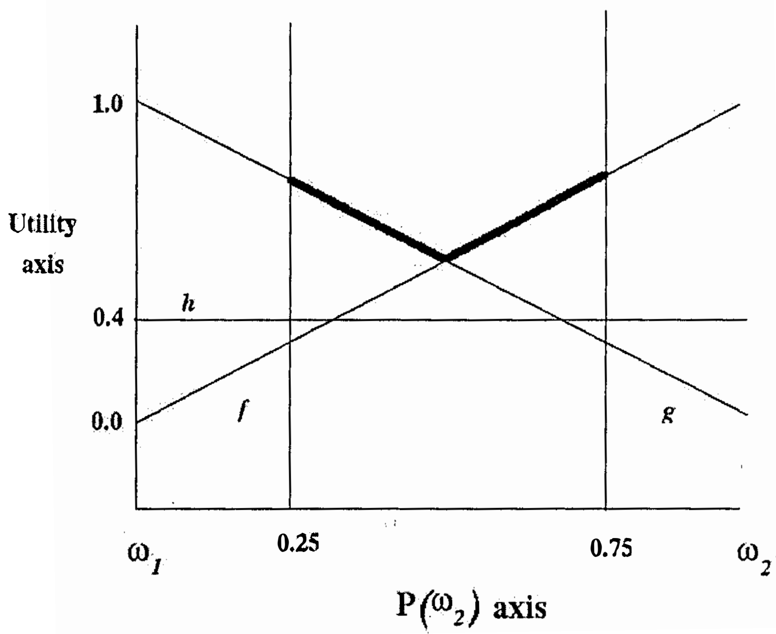


Figure 1: Expected utilities for three acts in Example 3. The thicker line indicates the surface of Bayes solutions.

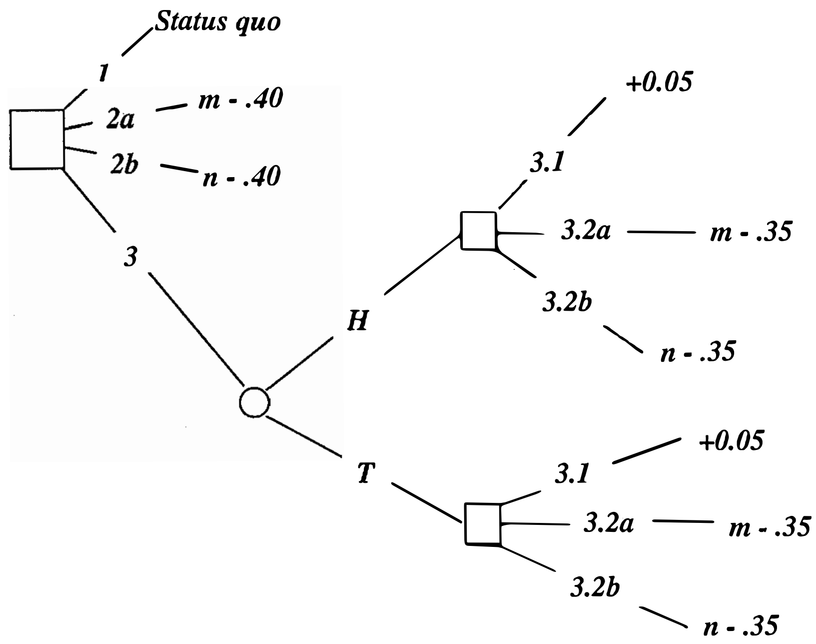


Figure 2: Sequential Decision Problem 1

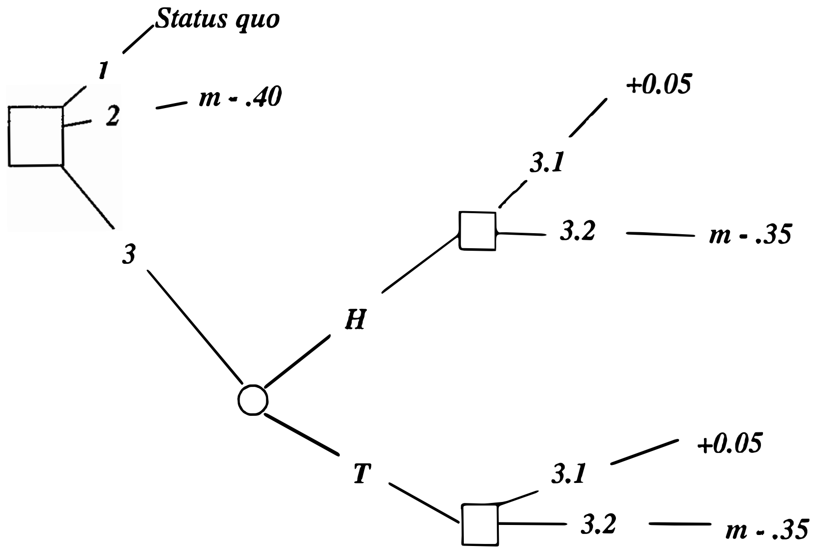


Figure 3: Sequential Decision Problem 2

Sequential Decision Problem 2: Here, at each of the two terminal choice nodes that might be reached under the sequential option, both options are E-admissible. Then, it is permissible for the decision maker who decides by E-admissibility to choose the constant (worth .05) if the sequential option is taken at the initial choice node. But .05 is an inadmissible option at the initial choice node, since $m - .40$ is strictly preferred to .05. Hence, it appears that in this sequential decision problem, E-admissibility does not require a decision maker to assign a non-negative value to potential cost-free information. Of course, since, E-admissible options are always a subset of the options permissible by Maximality, to the extent that this phenomenon is a problem with E-admissibility as the decision rule, it is only more so of a problem for decision makers using Maximality.

Dilation of sets of probabilities (work with Larry Wasserman)

The familiar Bayesian result is that coherent agents who share information, use Bayes' rule for updating, and who do not have extreme "prior" views will come to agree in their "posterior" opinions.

Savage (1954) argued this for shared *iid* data from a statistical model

Blackwell-Dubins (1962) showed that the result holds even without a common statistical model, providing that the coherent agents have mutually absolutely continuous joint opinions – they agree on all "null" events.

In contrast with such merging of opinions when opinions are given each by a single coherent probability, say that when uncertainty is represented by a set P of probabilities,

the random variable B dilates the uncertainty about the event A when, for all possible values $B = b$

$$\inf_P p(A \mid b) < \inf_P p(A) \leq \sup_P p(A) < \sup_P p(A \mid b).$$

That is, when B dilates A, the uncertainty in A is sure to increase by learning B, at least when updating is by Bayes' rule.

The previous examples showing when Maximality and E-admissibility allow that new cost-free information may be given a negative value, are cases of dilation.

Basic Dilation results:

- **If *B* dilates *A*, with respect to the set *P* of probabilities, then *A* and *B* are independent for some probability p in P.**
- **If a randomizer is available, and *P* is non-trivial, dilation may be created.**
- **If *P* is a symmetric-neighborhood model, then the only case where *P* is immune to dilation for some pair of random variables is if *P* is the density ratio model.**
- **In a case like the ε-contamination model, which is also a lower-probability model, the conditions for dilation have a simple geometric flavor.**
- **Moreover, the ε-contamination model is also a *D-S* Belief model, and in such cases Bayes-updating is the same as Dempster-updating. Hence, dilation applies equally to the D-S theory.**

Conclusion:

IMPRECISE PROBABILITY THEORY NEEDS A VIABLE THEORY OF SEQUENTIAL DECISION MAKING.

Or, at least,

WE DO NOT YET HAVE AN IMPRECISE PROBABILITY THEORY OF EXPERIMENTAL DESIGN.