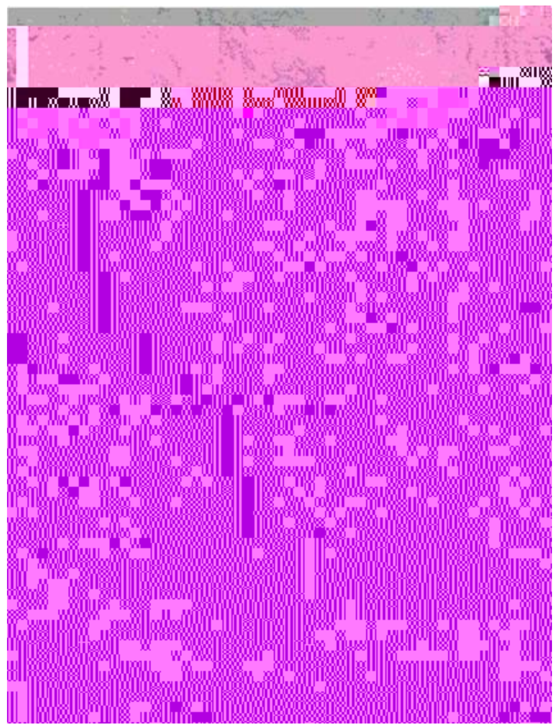
Extending Bayesian Theory to Cooperative Groups:

an introduction to Indeterminate/Imprecise Probability Theories [IP]
also see www.sipta.org

Teddy Seidenfeld – Carnegie Mellon University

based on joint work with Jay Kadane and

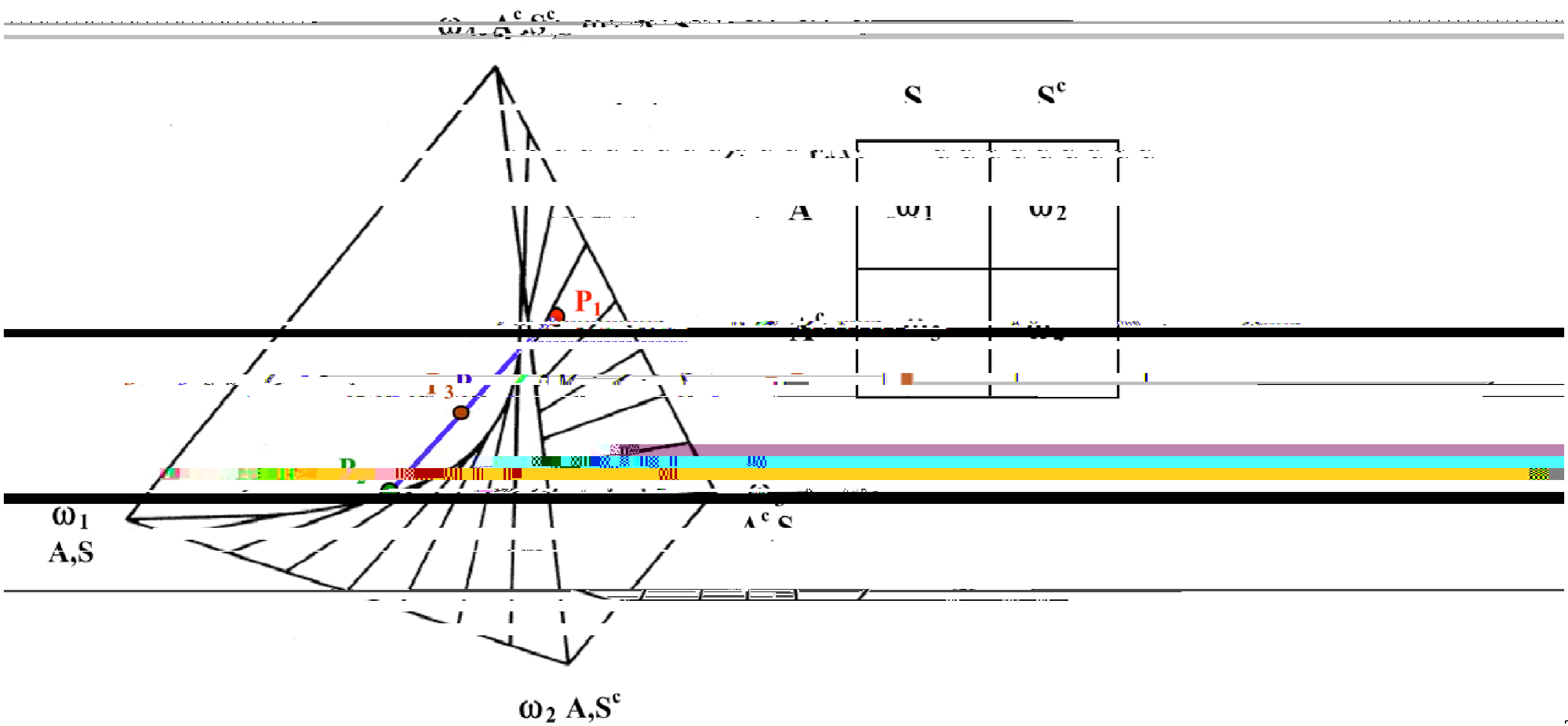


Mark Schervish

Review from the earlier presentation.

In our examination of the Linear Pool – combining probabilistic opinions into a convex combination of those distributions – we illustrated its failure to be "Externally Bayesian." There two experts judged events A and S independent, $P_i(AS) = P_i(A)P_i(S)$ for $i = 1, 2$. But the Linear Pool created a group opinion P_3 with positive dependence. $P_3(A|S) > P_3(A)$.

- **Pooling and conditioning do not commute!**



Extending Bayesian Theory to Cooperative Groups

The experts agree that T_3 is inadmissible in this three-way choice, because of their unanimity about the irrelevance of $\{S, S^c\}$.

However, neither T_1 nor T_2 is Pareto superior to T_3.

There is no one alternative to T_3 that the experts agree is better.

A lesson to be learned is that:

- **Pairwise comparisons between options is insufficient for determining a consensus among Bayesian agents.**

The decision theory for Bayesian consensus does not begin with a binary relation of preference.

What follows in this presentation

Part 1: Outline of a theory of coherent choice for use with IP models of consensus for a *team*.

Axiomatic representation of a coherent choice function.

Part 2: Some issues of experimental design within this model of consensus

2.1 Summary of an adaptive clinical trial following this model.

Consider a cooperative group of Bayesian decision makers who have

Given a (closed) set *O* of feasible options, a *choice function C* identifies the set *A* of *<u>acceptable options</u>* *C*[*O*] = *A*, for a non-empty subset *A* ! *O*.

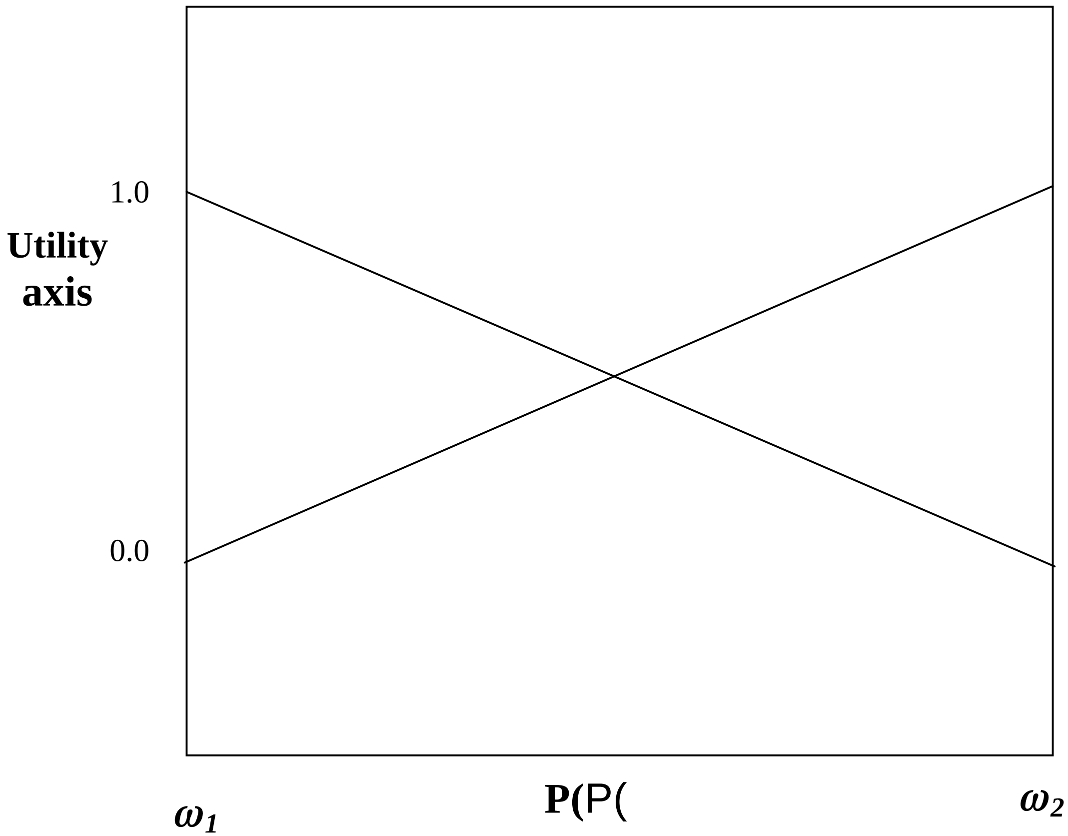
- **A choice function *C* is *<u>coherent</u>* across a class of problems if there is a set of probabilities P such that acceptable options are P –Bayes.**

An option is acceptable, *o* A (= C[*O*]), just in case *o* is a Bayes solution to problem O for some P P.

Aside: There may be no acceptable option if the option set is not closed, e.g., there is no "best" option from the continuum of utility values in [0, 1). We use closed sets of options in decision problems.

Definition: **Option** ***o*** ***O*** ***has a local Bayes model P***

With local coherence, only {*f*,*g*



Convexify the option set.

The locally Bayes mixed strategies *f* # (1-)*g* are pink

By Pearce's Theorem, e.g., the mixed act *m* = .5*f* # .5*g* strictly dominates *h*.

<u>Note well</u> that for a coherent choice function

act *m* is among the acceptable acts from {*f*, *g*, *m*}

if and only if **P(! $_2$) = .5 belongs to the IP set P•**

This observation about the acceptability of a mixed option generalizes.

- **Each (arbitrary) IP set of probabilities has its own distinct coherent choice function.**
- **For each two different sets of distributions there is a (finite) decision problem where they have distinct coherent choices.**

***Application*:**

We can represent the IP set of probability distributions that make two events independent, since convexity of the IP set is not required in our approach.

Coherent choice functions may be characterized by axioms on acceptable sets that parallel familiar axioms for SEU theory

SEU Coherent Preference $<$

Axiom$_1$ $<$ ***is a weak order.***

Axiom$_2$ $<$ ***obeys Independence*** $\mathbf{o}_1 < \mathbf{o}_2$ ***iff*** $x\mathbf{o}_1 \oplus (1-x)\mathbf{o}_3 < x\mathbf{o}_2 \oplus (1-x)\mathbf{o}_3$

Axiom$_3$ ***Archimedes***

If $\mathbf{o}_1 < \mathbf{o}_2 < \mathbf{o}_3$**, then** $\exists\ 0 < x, y < 1$

$$x\mathbf{o}_1 \oplus (1-x)\mathbf{o}_3 < \mathbf{o}_2 < y\mathbf{o}_1 \oplus (1-y)\mathbf{o}_3$$

Coherent Choice Functions

(SSK 2010)

In place of the ordering axiom, we require the following two conditions:

Axiom 1a **– Sen's (1977) property *alpha***

If $O_2 \subseteq R(O_1)$ and $O_1 \subseteq O_3$, then $O_2 \subseteq R(O_3)$.

You cannot promote an unacceptable option into an acceptable option by adding options to the feasible set.

Axiom 1b **– a variant of Aizerman's 1985 condition**

If $O_2 \subseteq R(O_1)$ and $O_3 \subseteq O_2$, then $O_2 - O_3 \subseteq R(closure[O_1 - O_3])$.

You cannot promote an unacceptable option into an acceptable option by deleting unacceptable options from the option set.

Note: We require *closure of* $[O_1 - O_3]$ since $O_1 - O_3$ may not be a closed set, despite the fact that O_1 and O_3 are closed.

The role of mixtures between options is captured in the following pair of axioms for &.

With O_1 an option set and *o* an option, the notation # O_1 % (1-#)*o* denotes the set of pointwise mixtures, # o_1 % (1-#)*o* for o_1 ! O_1.

Denote by H(*O*) the closed, convex hull of the option set *O*, to include mixed options.

***Axiom 2a – Independence* is formulated for the relation & over sets of options. Specifically, let *o* be an option and 0 < # 1.**

***O_1 & O_2 if and only if* # O_1 % (1-**

Aside: **Two rival decision theories that have been proposed within IP theory each violate a different part of Axiom 2.**

- ***Independence* (Axiom 2a) fails in -*Maximin* theory.**

    -*Maximin*: Maximize minimum expected utility with respect to the distributions P in $\mathcal{P}$. (See Berger, 1985)

Note: -*Maximin* uses only binary comparisons, since it generates a (real-valued) ordering of options.

- ***Mixing* (Axiom 2b) fails for *Maximality*.**

    ***Maximality*: An option *o* is *Maximal* if there is no option o where $E_P U(o\ ) > E_P U(o)$ for each P in $\mathcal{P}$. The admissible options are those that are Maximal. (See Walley, 1990.)**

Note: *Maximality* uses only binary comparisons, also, though it does not generate an ordering.

The counterpart to Axiom 4 for state-neutrality is captured by the following dominance relations. Introduce two rewards, {1, 0}.

Consider Anscombe-Aumann (1963) horse lotteries h_1 and h_2, with $h_i((_j)$ = $\&_{ij}1$) $(1-\&_{ij})0$; $i = 1, 2$ $j = 1, \ldots, n$.

***Definition*: h_2 *weakly dominates* h_1 if $\&_{2j}$ $\&_{1j}$ for $j = 1, \ldots, n$.**

Assume that o_2 weakly dominates o_1, and that a is an option different from each of these two.

Axiom 4a **If o_2 # O and a # $R(\{o_1\}$ % $O)$ then a # $R(O)$.**

Axiom 4b **If o_1 # O and a # $R(O)$ then a # $R(\{o_2\}$ % $[O$! $\{o_1\}])$.**

Main Result on Representation

A choice function C is coherent *if and only if* it satisfies these (4-pairs of) axioms.

A choice function satisfies these axioms *if and only if* it is given by a non-empty set *P* of global Bayes probability models.

The axioms suffice for representing a choice function with the coherence rule for admissibility applied to a (unique) set of *Probability/Almost-state-independent utility* pairs.

Different sets *P* are identified with different coherent choice functions.

We offer a sufficient condition for representation using a single, state-independent utility on rewards.

Return to the principal question about consensus.

What features of their shared beliefs and values will be reflected in their determination of acceptable options as a team?

Proposal**: Preserve unanimity of *unacceptable* options.**

Note: With binary choice problems, this is equivalent to an unrestricted *Pareto rule* –

If everyone strictly prefers o_1 over o_2, then so does the team.

This proposal results in taking the team's coherent choice function to be the one given by a set of global probabilities, P_T, formed by taking the *union* of the experts' individual sets P_i (i = 1, .., n) of global probabilities:

$P_T = \quad {}_i P_i$.

***Allocation rule*: Patients were admitted sequentially. In each case, based on an updated IP model for the 5 experts – updated by the data acquired to date in the trial – it was determined whether the group of 5 was unanimous: Was one of the two treatments T* Pareto superior for that patient.**

If so, that treatment T* was used.

If not, so that relative to the set of 5 updated expert opinions each treatment was acceptable with respect to the goal of regulating the patient's mean blood pressure deviation, then the patient was assigned in order to make the outcome most informative, e.g., by balancing the legs of the trial.

The comparison of prior and posterior favored treatments (after 49 patients) is reported in the next slide.

Table 12.2 Expert elicited prior preferences

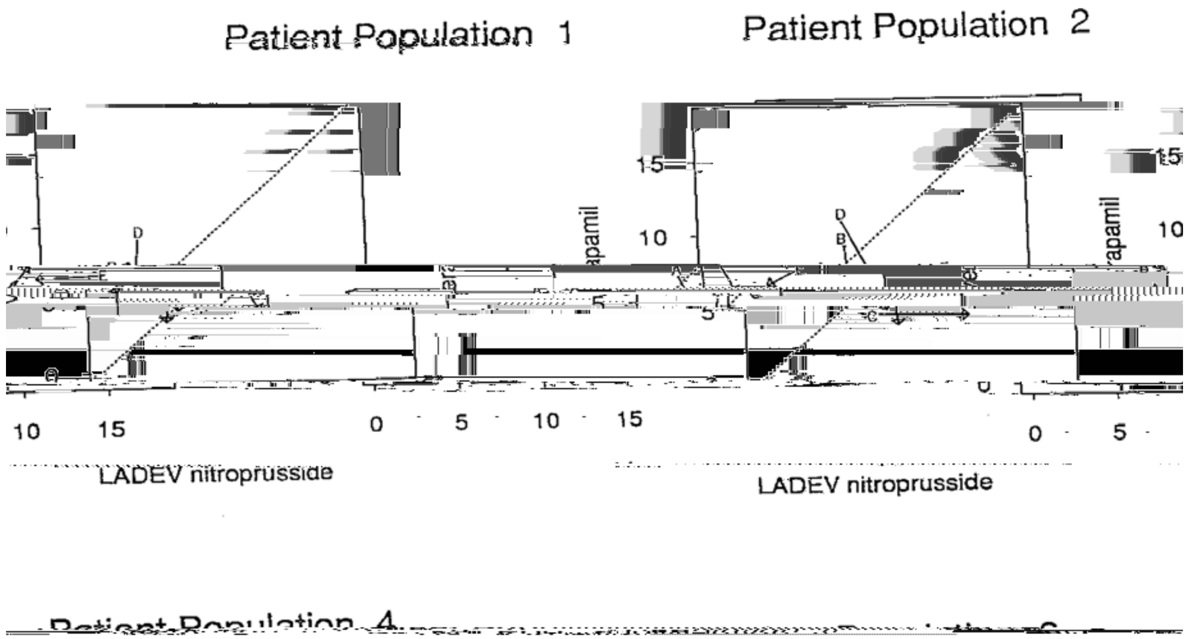
Patient Type	A	B	C	D	E
1	N	N	V	N	V
[illegible]	V	V	V	V	V
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	N	N	N	N	V
5	N	N	V	N	V
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7	N	N	V	N	V
8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12	V	N	N	N	V
13	N	N	V	N	V
14	N	N	N	N	V
15	V	V	V	V	V
16	N	N	N	N	V

Note: An "N" indicates a preference for nitroprusside, a "V" for verapamil.

Table 12.3 Expert computed posterior preferences based only on LADEV

[illegible]

Note: An "N" indicates a preference for nitroprusside, a "V" for verapamil.



More informative are shifts from prior to posterior predictive means.
Note: The allocation rule does not require randomization!

2.2 *Dilation* for IP sets of probabilities – some things you rather not know! [S & W, 1993]

Let P

Heuristic Example

Suppose A is a highly *uncertain* event in the added sense of "uncertainty" that comes with a set of probabilities P.

That is $P^*(A)$ - $P_*(A)$ 1.

Let {H,T} indicate the flip of a fair coin whose outcomes are independent of A. That is, $P(A,H) = P(A)/2$ for each P ! P. Define event E by, $E = \{(A,H), (A^c,T)\}$.

	H	T
A	E	E^c
A^c	E^c	E

It follows, simply, that P(E) = .5 for each P ! P.

Then 0 $P_*(E \mid H) < P_*(E) = P^*(E) < P^*(E \mid H)$ 1

and 0 $P_*(E \mid T) < P_*(E) = P^*(E) < P^*(E \mid T)$ 1.

Dilation and independence [illegible]

[illegible]

[illegible] convex set [illegible]

[illegible]

the coin flips are mutually independent with [illegible] Assume [illegible] respect to Q, also independent of events in A. [illegible]

Let P be the resulting convex set of probabilities on A × C. [illegible]

[illegible]

[illegible] extraneous [illegible] and is similar to the "fine [illegible] of Savage, Ramsey, Jeffrey, etc.) [illegible]

[illegible] singleton, there is a 2 × 2 table of the [illegible] Theorem: If Q is not a [illegible]

[illegible]

$$P_*(E \mid H) < P_*(E) = .5 = P^*(E) < P^*(E \mid H)$$

[illegible]

That is, dilation occurs.

Dilation creates a new challenge for the design of experiments.

Summary of our IP-model of consensus for a team

- **Coherent choice does not reduce to binary comparisons between the options available.**

- **Each two IP sets of probabilities yield different coherent choices.**

- **Coherent choice is axiomatized by constraints on choice functions that parallel the familiar axioms for coherent (binary) *preferences*.**

- **Experimental design with respect to an IP-set may permit:**
  - **The shared data to induce a (familiar) merging of posterior probabilities and a resulting concentration of the posterior IP-set.**

  OR

  - **Dilating the set of IP probabilities, resulting in added uncertainty for sure.**

Experimental design for an IP set is not Fisher's *Design of Experiments*!

Selected References

Anscombe, F.J. and Aumann, R.J. (1963) "A definition of subjective probability" *Ann Math. Statist*. 34: 199-205.

Berger, J. (1985). *Statistical Decision Theory*, 2nd ed. Springer, New York.

Goodman, J. (1988) *Existence of Compromises in Simple Group Decisions*. Ph.D. thesis.