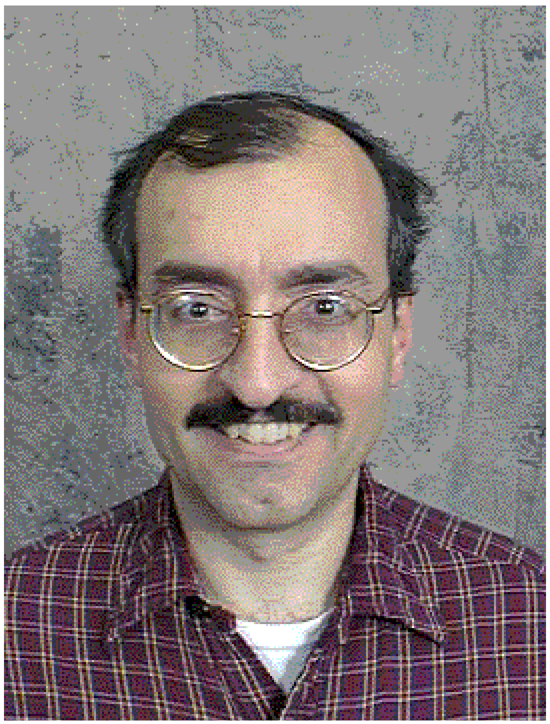
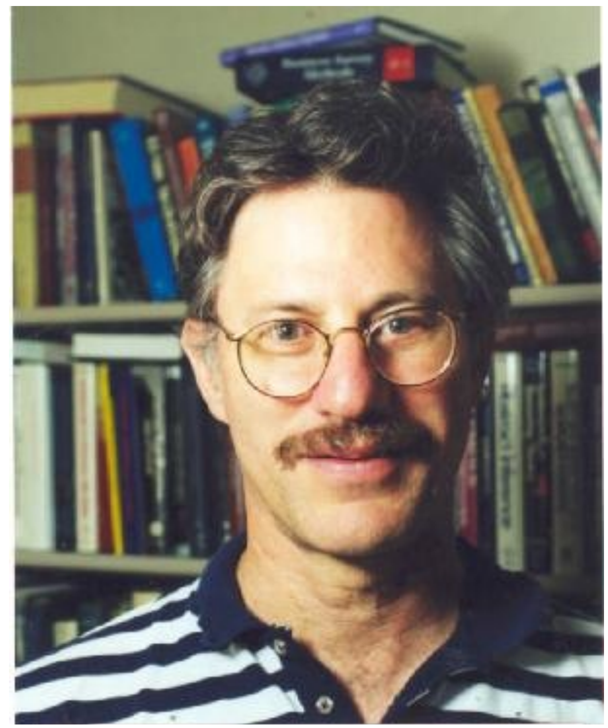
Extending Bayesian Theory to Cooperative Groups:

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



Mark Schervish

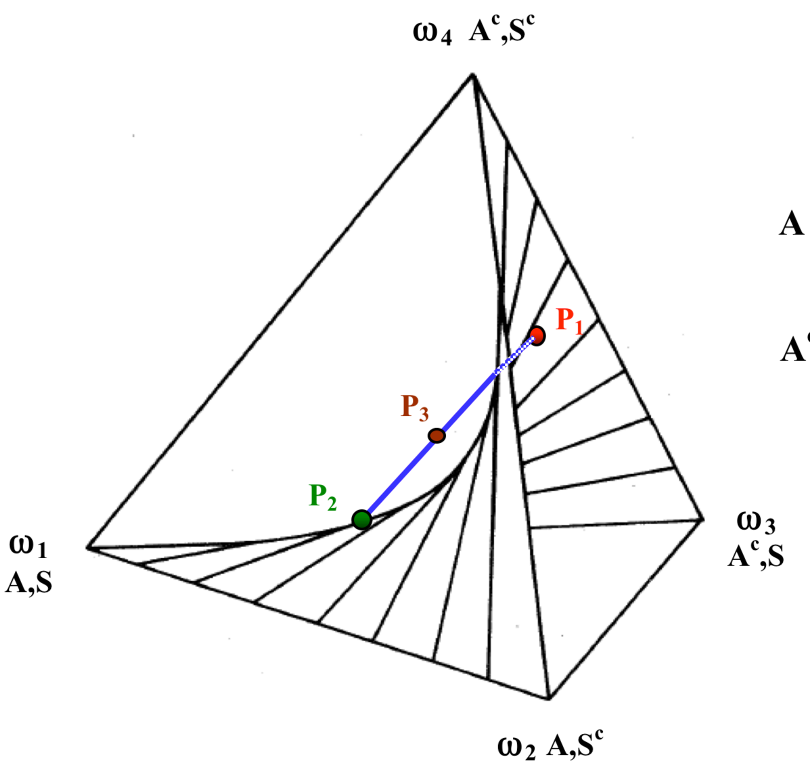


Joseph ("Jay") Kadane

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!**



	S	S^c
A	ω_1	ω_2
A^c	ω_3	ω_4

We used this fact to give a simple decision problem with these features. Choose among three treatment plans, $\{T_1, T_2, T_3\}$, for a patient whose allergic state $\{A, A^c\}$ has different probabilities for the two experts. The state of weather in China is denoted by the partition $\{S, S^c\}$. Each expert judges the patient's allergic state and the weather in China independent. But the Linear Pool makes them positively dependent.

	$\omega_1 = A\&S$	$\omega_2 = A\&S^c$	$\omega_3 = A^c\&S$	$\omega_4 = A^c\&S^c$
T_1	0.00	0.00	1.00	1.00
T_2	1.00	1.00	0.00	0.00
T_3	0.99	-0.01	-0.01	0.99

Distribution P_3 is the .50-.50 (convex) mixture of P_1 and P_2.

	ω_1	ω_2	ω_3	ω_4
P_1	0.08	0.32	0.12	0.48
P_2	0.48	0.12	0.32	0.08
P_3	0.28	0.22	0.22	0.28

T_3 is the SEU maximizer under P_3, which is objectionable to the experts as it spends resources (-0.01) to learn the weather in China in order to determine which drug to give the patient.

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.

2.2 About experiments that are sure to increase uncertainty – some things that you rather not know!

Consider a cooperative group of Bayesian decision makers who have common goals – a common utility function.

What features of their shared beliefs will be reflected in their determination of acceptable options when they function as a team?

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

Note: With binary choice problems, this is equivalent to the familiar *Pareto rule* –

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

However, as we have seen, in a decision with more than two options it may be that an option (T_3) is unanimously unacceptable without another option Pareto dominating it.

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 $\mathcal{P}$ such that acceptable options are $\mathcal{P}$ –Bayes.**

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

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 $o \in O$ *has a local Bayes model* P if**

o maximizes the P-expected utility over the options in O.

- ***Theorem 1*** **(Pearce, 1984 for finite state spaces): If an option $o \in O$ fails to have a local Bayes model then it is uniformly, strictly dominated by a finite mixture of options already available from O.**

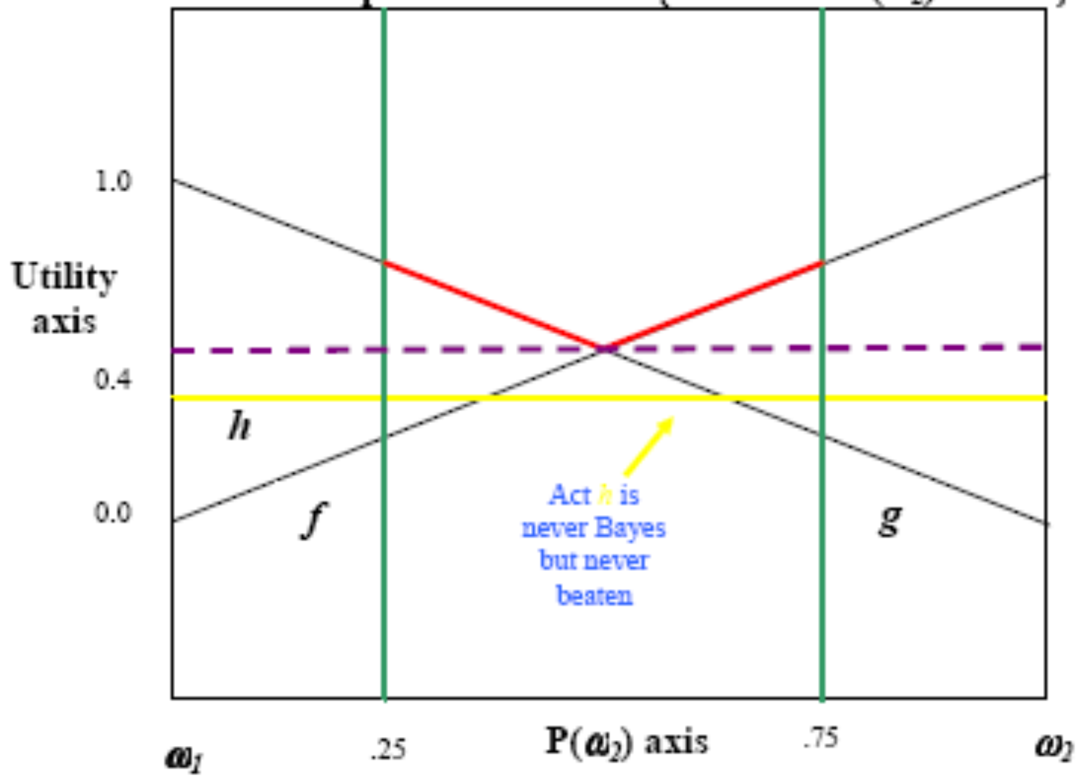
So – at least when the option space is closed under (finite) mixtures –

strict dominance assures that admissible options are locally coherent.

That is, then the choice function needs to be *locally coherent* at least.

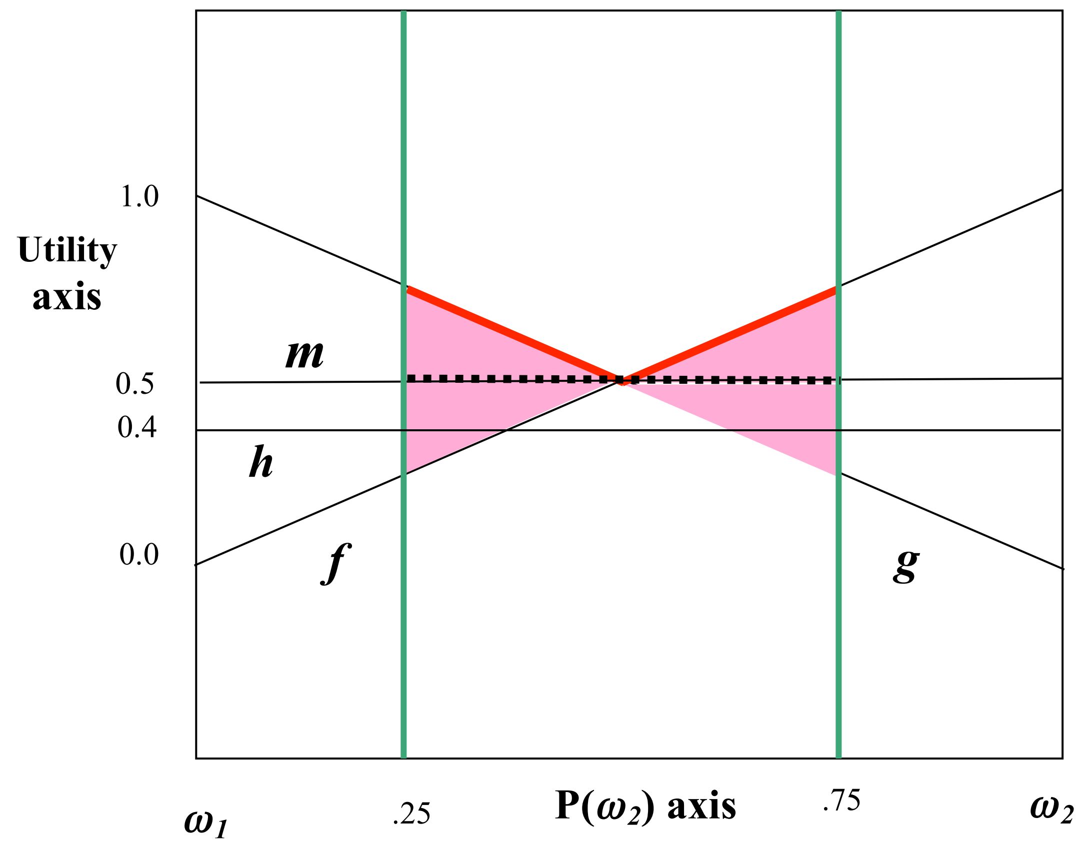
- **This result is a generalization of de Finetti's "Book" argument for incoherent betting.**
- **It strengthens Wald's "Complete Class" theorem as the standard for inadmissibility is *strict dominance*, not *weak dominance*.**

Consider a decision problem on 2 states $\{\omega_1, \omega_2\}$, with 3 options $\{f, g, h\}$, and a convex set of probabilities $\mathcal{P} = \{ P: .25 \leq P(\omega_2) \leq .75 \}$.



With local coherence, only $\{f,g\}$ are admissible from the triple $\{f,g,h\}$; however, all pairs are admissible in pairwise choices.

Isaac Levi calls *h* <u>*second worst*</u> in the triple $\{f, g, h\}$.



Convexify the option set.

The locally Bayes mixed strategies $\alpha f \oplus (1-\alpha) g$ are pink

By Pearce's Theorem, e.g., the mixed act $m = .5f \oplus .5g$ 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 **$P(\omega_2) = .5$ belongs to the IP set $\mathcal{P}$.**

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

<u>SEU Coherent Preference</u> $<$

Axiom*₁** $<$ ***is a weak order.

Axiom*₂** $<$ ***obeys Independence $\quad \mathbf{o_1 < o_2}$ ***iff*** $\mathbf{x o_1 \oplus (1-x) o_3 < x o_2 \oplus (1-x) o_3}$

Axiom*₃ *Archimedes

If $\mathbf{o_1 < o_2 < o_3}$**, then** $\exists\ \mathbf{0 < x, y < 1}$

$$\mathbf{x o_1 \oplus (1-x) o_3 < o_2 < y o_1 \oplus (1-y) o_3}$$

***Axiom*₄ State-independent Utilities**

$<$ **over constant acts reproduces within each non-null state.**

Coherent Choice Functions

(SSK 2010)

We adapt our presentation to match these four axioms.

For ease of exposition some conditions are formulated in terms of the *rejection function*, $R(\cdot)$ which identifies the C-inadmissible options from a feasible set O.

***Definition*:** $$R(O) = O - C(O)$$

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.

With Axioms 1a and 1b, define a strict partial order $\langle$ on sets of options as follows.

Let O_1 and O_2 be two option sets.

Definition: $O_1 \langle O_2$ ***if and only if*** $O_1 \subseteq R[O_1 \cup O_2]$.

So $O_1 \langle O_2$ obtains when O_1 contains only inadmissible options in a choice among the options in both sets, $O_1 \cup O_2$.

> **Aside: This maneuver allows us to engage our (1995) work on strict partial orders.**

The role of mixtures between options is captured in the following pair of axioms for $\langle$.

With O_1 an option set and o an option, the notation $\alpha O_1 \oplus (1\text{-}\alpha)o$ denotes the set of pointwise mixtures, $\alpha o_1 \oplus (1\text{-}\alpha)o$ for $o_1 \in O_1$.

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 $\langle$ over sets of options. Specifically, let o be an option and $0 < \alpha \leq 1$.**

$$O_1 \langle\; O_2 \quad \textit{if and only if} \quad \alpha O_1 \oplus (1\text{-}\alpha)o \;\langle\; \alpha O_2 \oplus (1\text{-}\alpha)o.$$

Axiom 2b – Mixtures **If $o \in O$ and $o \in R[H(O)]$, then $o \in R[O]$.**

Axiom **2b asserts that inadmissible options from a mixed set remain so even before mixing.**

***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 *Γ-Maximin* theory.**

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

Note: *Γ-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 Archimedean condition for coherent choice functions requires a technical adjustment from the canonical form used by, e.g. von Neumann-Morgenstern theory or Anscombe-Aumann theory. The canonical form is too restrictive in this setting. (See section II.4 of our 1995.) The reformulated version of the Archimedean condition is as a continuity principle compatible with strict preference as a strict partial order. It reads as follows.

Let A_n and B_n (n = 1, …) be sets of options converging pointwise, respectively, to the option sets A and B. Let N be an option set.

Axiom 3a **If, for each n, $B_n \langle A_n$ and $A \langle N$, then $B \langle N$.**

Axiom 3b **If, for each n, $B_n \langle A_n$ and $N \langle B$, then $N \langle A$.**

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 $h_i(\omega_j) = \beta_{ij}1 \oplus (1-\beta_{ij})0$; $i = 1, 2$ $j = 1, \ldots, n$.

***Definition*: h_2 *weakly dominates* h_1 if $\beta_{2j} \geq \beta_{1j}$ 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 $o_2 \in O$ and $a \in R(\{o_1\} \cup O)$ then $a \in R(O)$.**

Axiom 4b **If $o_1 \in O$ and $a \in R(O)$ then $a \in R(\{o_2\} \cup [O - \{o_1\}])$.**

In words, Axiom *4a* says that when a weakly dominated option is removed from the set of options, other inadmissible options remain inadmissible. So, by Axiom 1, when an option is replaced in the option set by one that it weakly dominates, other admissible options remain admissible.

Axiom *4b* says that when an option is replaced by one that weakly dominates it, (other) inadmissible options remain inadmissible.

Trivially by Axiom 1, merely adding a weakly dominating option cannot promote an inadmissible option into one that is admissible.

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:

$P_T = \cup_i P_i$.

2 - Experimental design within this model of consensus for a team.

2.1 Outline of an adaptive clinical trial – Kadane (1996).

The trial investigates the efficacy of two treatments for controlling blood pressure during open-heart surgery: verapamil vs. nitroprusside.

***Team Utility*: The agreed goal for each patienbt is regulation of the 30-minute deviation of mean arterial systolic pressure from the target of 80 mmHg.**

***Probabilities*: 5 experts were interviewed and their opinions elicited to identify their (prior) opinions about the relevant medical factors for predicting patient outcomes (for qualifying patients) under the rival treatments.**

This yielded a model with 16 patient types based on 4 binary classifiers.

Table 12.1. Patient types as a function of patient characteristics

Patient Type	Beta Blockers	Calcium Antagonists	Wall Movement Abnormality	Previous Hypertension
1	1	1	1	1
2	1	1	1	0
3	1	1	0	1
4	1	1	0	0
5	1	0	1	1
6	1	0	1	0
7	1	0	0	1
8	1	0	0	0
9	0	1	1	1
10	0	1	1	0
11	0	1	0	1
12	0	1	0	0
13	0	0	1	1
14	0	0	1	0
15	0	0	0	1
16	0	0	0	0

Note: A "1" indicates the presence of the condition, a "0" its absence.

***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 based only on LADEV

Patient Type	Expert *A*	*B*	*C*	*D*	*E*
1	N	N	V	N	V
2	N	N	V	N	V
3	N	N	V	N	V
4	N	N	N	N	V
5	N	N	V	N	V
6	N	N	N	N	V
7	N	N	V	N	V
8	N	N	N	N	V
9	N	N	V	N	V
10	N	N	V	N	V
11	N	N	V	N	V
12	N	N	V	N	V
13	N	N	V	N	V
14	N	N	V	N	V
15	N	N	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

Patient Type	Expert *A*	*B*	*C*	*D*	*E*
1	V	V	V	V	V
2	V	N	N	V	V
3	V	V	V	V	V
4	V	V	N	V	V
5	V	V	V	V	V
6	V	N	N	V	V
7	V	V	V	V	V
8	V	V	N	V	V
9	N	N	V	V	V
10	N	N	N	N	V
11	V	V	V	V	V
12	V	N	N	N	V
13	N	N	V	V	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.

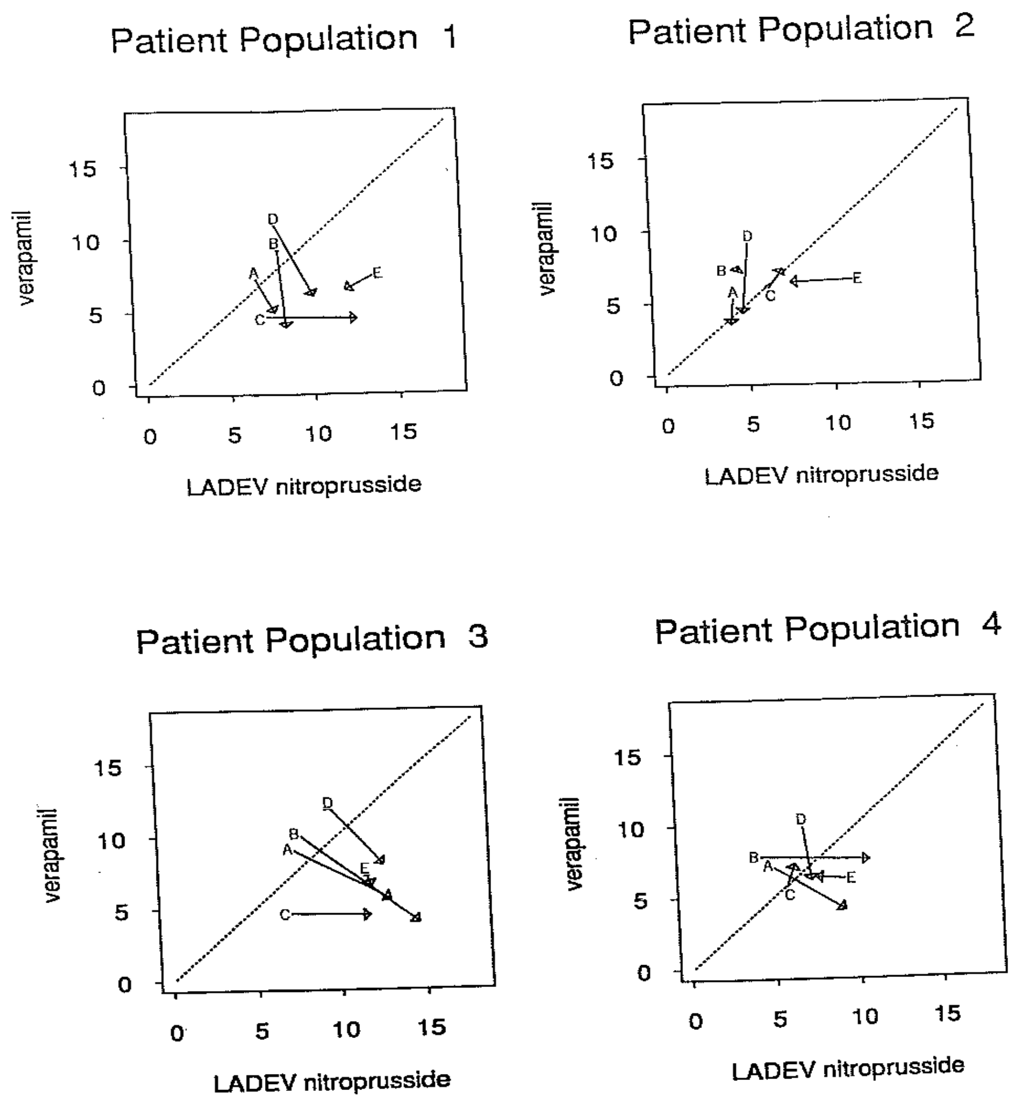


Figure 12.1. Patient populations 1–16.

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 $\mathcal{P}$ be a (convex) set of probabilities on an algebra $\mathcal{A}$.
For an event E,

denote by $P_*(E)$ the "lower" probability of E: $\inf_{\mathcal{P}} \{P(E)\}$

and denote by $P^*(E)$ the "upper" probability of E: $\sup_{\mathcal{P}} \{P(E)\}$.

Let $B = \langle B_1, \ldots, B_n \rangle$ be a (finite) partition.

Think of B as an experiment to determine which B_i obtains.

Defn. The set of conditional probabilities $\{P(E \mid B_i)\}$ dilate if

$$P_*(E \mid B_i) < P_*(E) \leq P^*(E) \leq P^*(E \mid B_i) \qquad (i = 1, \ldots, n).$$

Dilation occurs provided that, for each event (B_i) in a partition, the conditional probabilities for an event E, given B_i, properly include the unconditional probabilities for E.

With dilation, IP-uncertainty – the spread between the lower and upper probability – is sure to increase: anti-convergence of posterior probabilities.

Heuristic Example

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

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

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 $P \in \mathcal{P}$. 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 $P \in \mathcal{P}$.

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

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

Dilation and Independence.

Independence is sufficient for dilation.

Let Q be a convex set of probabilities on algebra A and suppose we have access to a fair coin which may be flipped repeatedly: algebra C for the coin flips.

Assume the coin flips are mutually independent and, with respect to Q, also independent of events in A.
Let P be the resulting convex set of probabilities on $A \times C$

(This condition is similar to, e.g., DeGroot's assumption of an extraneous continuous r.v., and is similar to the "fineness" assumptions in the theories of Savage, Ramsey, Jeffrey, etc.)

Theorem: If Q is not a singleton, there is a 2×2 table of the form $(E,E^c) \times (H,T)$ where both:

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

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

That is, dilation occurs.

Independence is necessary for dilation.

Let P be a convex set of probabilities on algebra A. The next result is formulated for subalgebras of 4 atoms: (p_1, p_2, p_3, p_4)

The case of 2×2 tables.

	B_1	B_2
A_1	p_1	p_2
A_2	p_3	p_4

Define the quantity
$\mathbf{S_P}(A_1,B_1) = p_1/(p_1+p_2)(p_1+p_3) = P(A_1,B_1) / P(A_1)P(B_1)$.

Thus, $\mathbf{S_P}(A_1,B_1) = 1$ iff A and B are independent under P.

Lemma: If P displays dilation in this sub-algebra, then
$\inf_P\{\mathbf{S_P}(A_1,B_1)\} < 1 < \sup_P\{\mathbf{S_P}(A_1,B_1)\}$.

Theorem: If P displays dilation in this sub-algebra, then there exists $P^{\#} \in$ P such that $\mathbf{S_{P\#}}(A_1,B_1) = 1$.

Dilation and the ε-contaminated model.

Given probability P and $1 > \varepsilon > 0$, define the convex set

$\mathcal{P}_{\varepsilon}(P) = \{(1-\varepsilon)P + \varepsilon Q$: Q an arbitrary probability$\}$.

This model is popular in studies of Bayesian Robustness.

(See Huber, 1973, 1981; Berger, 1984.)

Also, it is equivalent to the model formed by fixing effective lower probabilities for the atoms of an algebra.

Lemma In the ε-contaminated model, dilation occurs in an algebra $\mathcal{A}$ iff it occurs in some 2×2 subalgebra of $\mathcal{A}$.

So, the next result is formulated for 2x2 tables.

$\mathcal{P}_\varepsilon(P)$ experiences dilation *if and only if*

case 1: $S_P(A_1,B_1) > 1$

$\varepsilon > [S_P(A_1,B_1) - 1] \times \max\{ P(A_1)/P(A_2) \; ; \; P(B_1)/P(B_2) \}$

case 2: $S_P(A_1,B_1) < 1$

$\varepsilon > [1 - S_P(A_1,B_1)] \times \max\{ 1 \; ; \; P(A_1)P(B_1)/P(A_2)P(B_2) \}$

and case 3: $S_P(A_1,B_1) = 1$

P is internal to the simplex of all distributions.

Thus, dilation occurs in the ε-contaminated model if and only if the focal distribution, P, is close enough (in the tetrahedron of distributions on four atoms) to the saddle-shaped surface of distributions which make A and B independent.

Here, S_P provides one relevant index of the proximity of the focal distribution P to the surface of independence.

Dilation creates a new challenge for the design of experiments.

- Design experiments to avoid dilation!

The significance of this challenge is heightened by the following result.

A neighborhood model (with focal distribution P) is called *symmetric* if, when P is the uniform distribution, a neighborhood is closed under permuation of the atoms.

- **The only symmetric neighborhood model that is dilation immune is the *Density Ratio* model!**

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. Dept. of Statistics. Carnegie Mellon Univ.

Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.

Kadane, J.B. ed. (1996) *Bayesian Methods and Ethics in a Clinical Trial Design* Wiley, N.Y.

Levi, I. (1974) "On indeterminate probabilities" *J.Phil*. 71: 391-418.

Levi, I. (1980) *The Enterprise of Knowledge*. MIT Press.

Pearce, D. (1984) "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica* 52: 1029-1050.

Seidenfeld, T., Kadane, J.B., and Schervish, M. (1989) "On the shared preferences of two Bayesian decision makers," *J.Phil.* 89: 225-244.

Seidenfeld, T., Schervish, M.J., and Kadane, J.B. (1995) "A representation of partially ordered preferences," *Ann Stat*. 23: 2168-2217.

Seidenfeld, T., Schervish, M.J., and Kadane, J.B. (2010) "Coherent Choice Functions," *Synthese* 172: 157-176, 2010.

Seidenfeld, T. and Wasserman, L. (1993) "Dilation for Sets of Probabilities" *Ann. Statistics* 21: 1139-1154.

Sen, A. (1977) "Social choice theory: a re-examination," *Econometrica* 45: 53-89.

Walley, P. (1990) *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.