Axiomatic Coherent Choice Functions for IP Theory

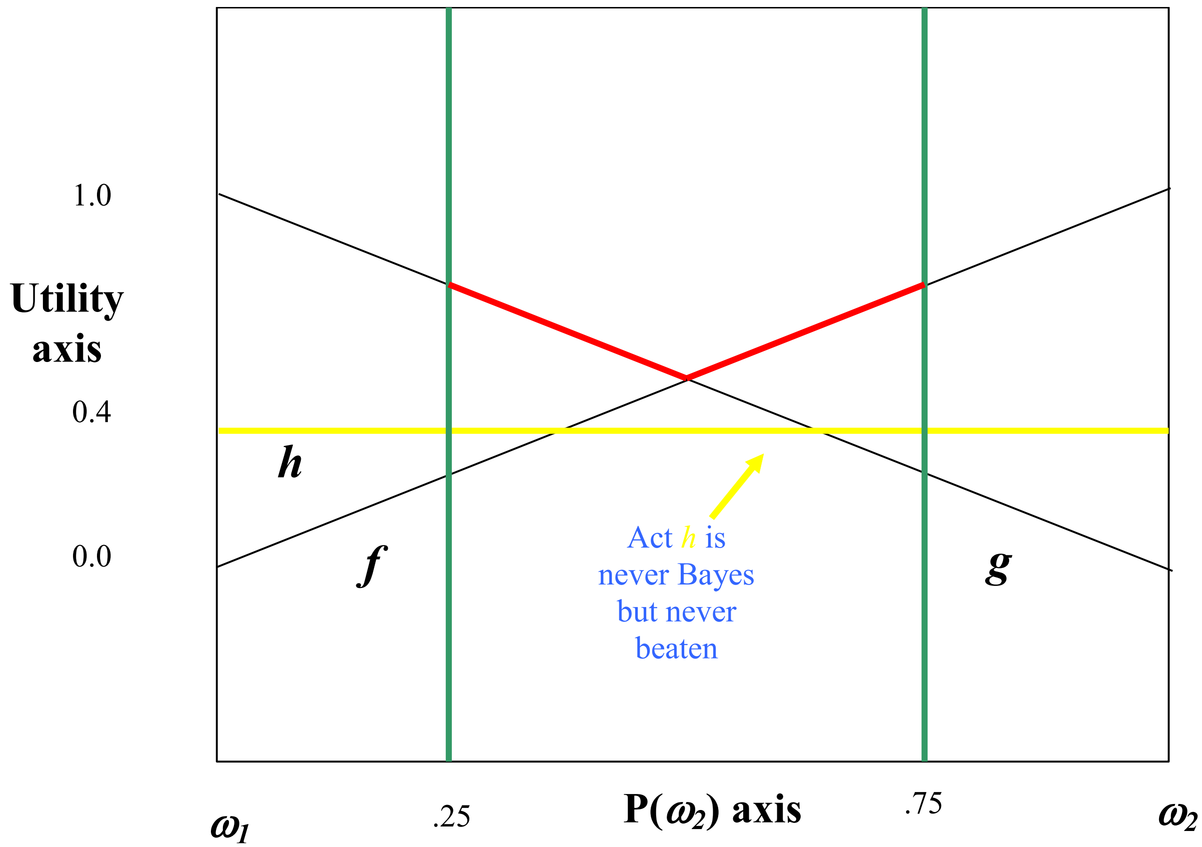
One way to improve IP-elicitation in accord with the central theme that

choice behavior reveals the agent's uncertainties

is to use *choice functions* rather than a (binary) *preference relation*.

As we'll see, then elicitation is not restricted to convex sets, IP-theory can be richer than the class of convex sets of probabilities.

Return to question the relation between IP-decision theory and dominance.



Only {*f*,*g*} are Bayes-admissible from the triple {*f*,*g*,*h*}; however, all pairs are Bayes-admissible in pairwise choices. Levi calls *h* _second worst**_ in the triple {*f*,*g*,*h*}.**

Contrast three *coherent* decision rules for extending Expected Utility [*EU*] theory when probability – but not cardinal utility – is indeterminate.

The decision problems involve (bounded) sets of lotteries, where the outcomes have well-defined cardinal utility but where the (act-independent) states are uncertain, represented by a *convex* set of probabilities $\mathcal{P}$.

- ***Γ-Maximin* (Gilboa-Schmeidler, 1989) – maximize min. expectations over $\mathcal{P}$.**
- ***Maximality* (Walley 1990) – admissible choices are undominated in expectations over $\mathcal{P}$ by any single alternative choice.**
- ***E-admissibility* (Levi/Savage) – admissible choices have Bayes' models, i.e., they maximize *EU* for some probability in the (convex) set $\mathcal{P}$.**

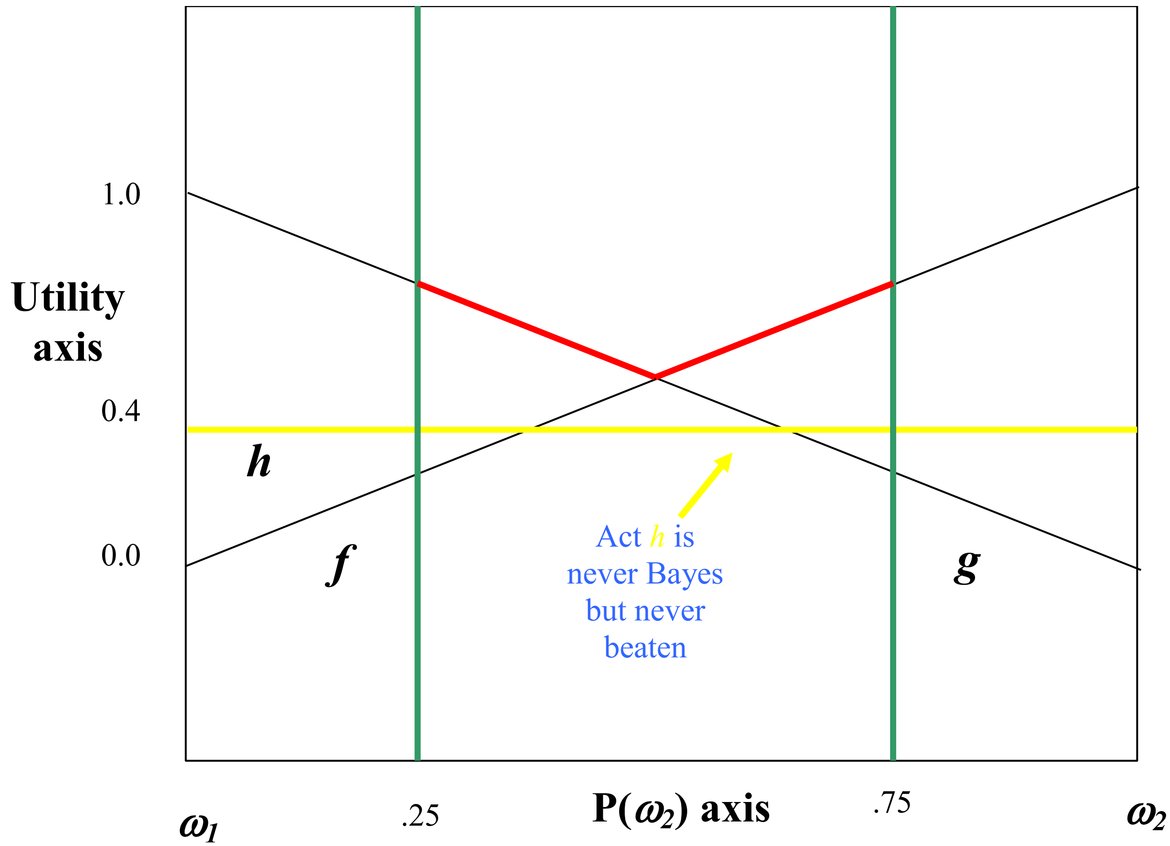
Each rule has *EU* Theory as a special case when probability is determinate, i.e., when $\mathcal{P}$ is comprised by a single probability distribution.

And each rule is *coherent* in the sense that sure loss (*Book*) is not possible. The three rules are chosen to reflect the following progression, where each rule relaxes more of the ordering assumption than does its predecessors:

- ***Γ-Maximin* produces a (real-valued) ordering of options; hence, defined by binary comparisons – but it fails *Independence*.**

- ***Maximality* does not generate an ordering of options; however, it is given by binary comparisons.**

- ***E-admissibility* does not generate an ordering, nor is it given by binary comparisons.**



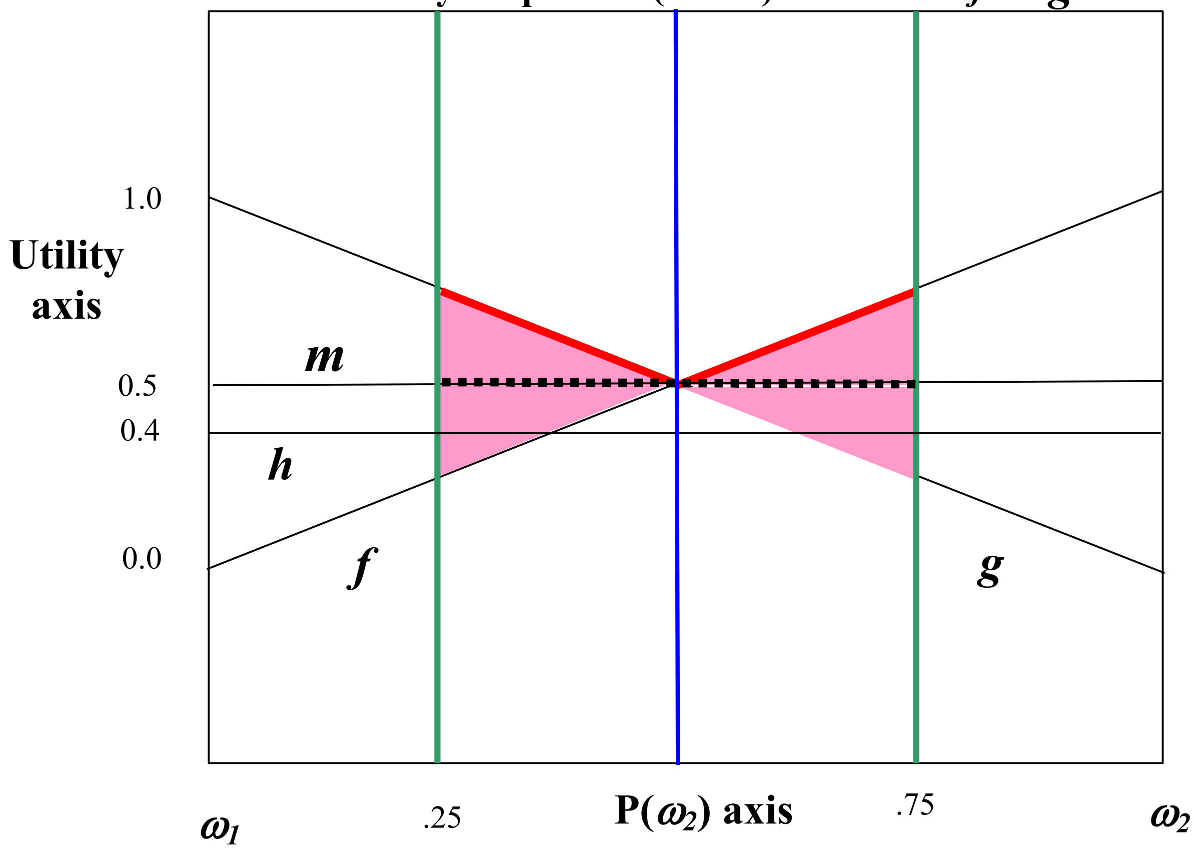
- ***The Γ-Maximin* solution is {*h*}.**
- **The *E-admissible* solution set is {*f*, *g*}.**
- **And *Maximality* finds all three options admissible, {*f*, *g*, *h*}.**

Thus, each rule gives a different set of admissible options in this problem.

Create a convex option space by allowing mixed strategies.

Expected Utility for the (Bayes) mixed options $\alpha f \oplus (1-\alpha)g$ is in pink; they maximize EU at $p(\omega_2) = .5$ (blue)

The Bayes equalizer (mixed) act is $m = .5f \oplus .5g$



- ***The Γ-Maximin* solution is the EU-*equalizer* {*m*}.**
- **The *E-admissible* and *Maximality admissible* options are the same set of *Bayes* solutions (pink).**

Agreement of the 3 decision rules on Bayes solutions then is no accident:

- ***Walley*** **(Theorem 3.9.5, 1990) establishes that when the option set is** ***convex*** **and the (convex) set of probabilities $\mathcal{P}$ is** ***closed*****,**

or (SSKL) if **the set $\mathcal{P}$ is open and the option set is finitely generated,**

then ***E-admissibility*** **and** ***Maximality*** **give the same solution sets.**

Their admissible sets are precisely the Bayes-admissible options.

- **And then it also follows that the** ***Γ-Maximin*** **admissible acts are a (proper) subset of the Bayes-admissible options.**
- ***Under these conditions, pairwise comparisons of acts suffice to determine the set of Bayes-admissible choices and from them we can elicit the IP-model.***
- ***However, otherwise options that are admissible by Maximality may be E-inadmissible.*** **(SSKL, 2003).**

D.Pearce (1984), reports a related result which is important for understanding the underlying connection between *dominance* and *Bayes-admissibility*.

***Theorem* (Pearce, 1984): In a decision problem under uncertainty,**

- **with finitely many states and finitely generated option set *O*,**
- **with utility of outcomes determinate – cardinal utilities,**

if an option $o \in O$ fails to be Bayes-admissible,

then *o* is (uniformly) dominated by a finite mixture from *O*.

Aside: This result can be extended to infinite decision problems.

In this sense, incoherent choices suffer deFinetti's penalty – being (uniformly) dominated by a mixed option – within the decision at hand – and not merely for the prevision game, which is a specialized decision.

In accord with Pearce's Theorem, in the example above,
the mixed act $m = .5f \oplus .5g$ strictly dominates *h*.

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

> **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).**

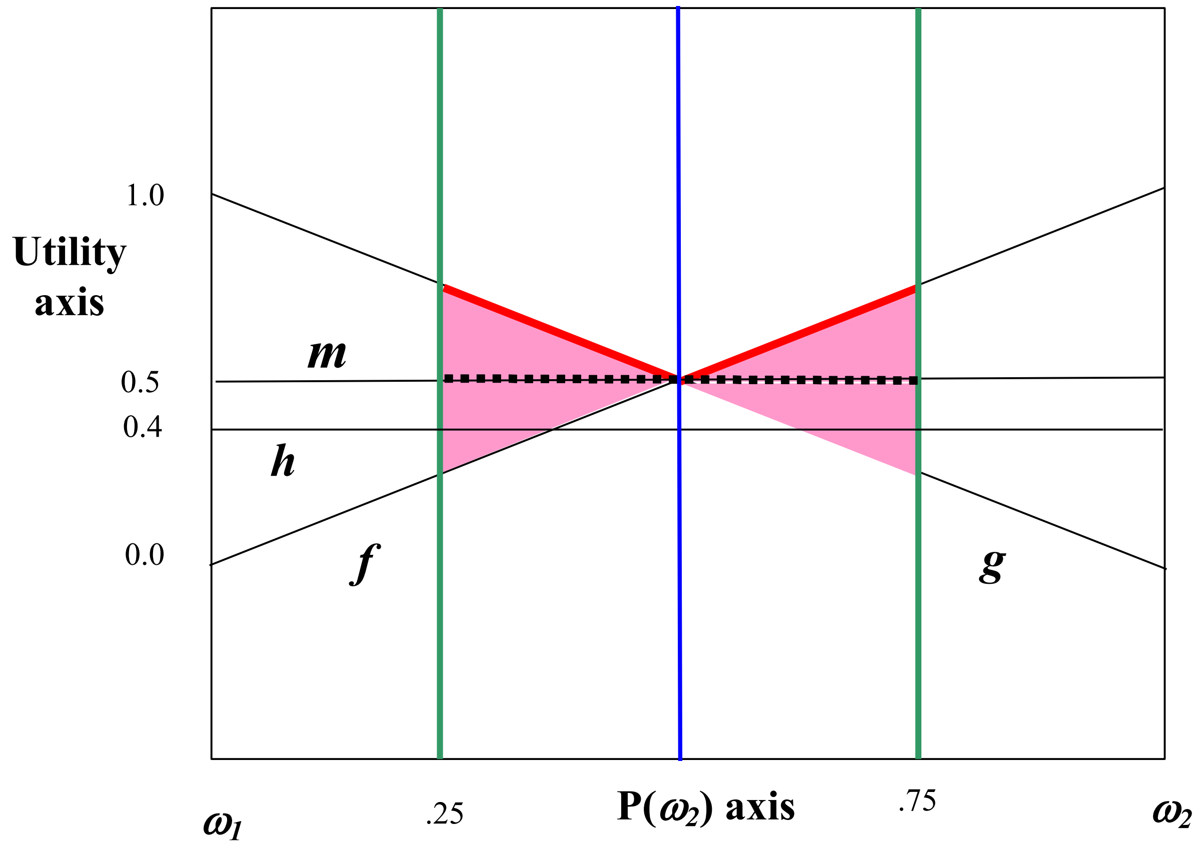
***Definition:* Option *o* ∈ *O* <u>*has a local Bayes model*</u> *P* if**
**   *o* maximizes the *P*-expected utility over the options in *O*.**

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

So – at least when the option space is closed under (finite) mixtures –
(uniform) dominance assures that admissible options are locally coherent.
That is, then the choice function needs to be *locally coherent* at least.

Definition**: A choice function C is *coherent* if there exists an IP-set $\mathcal{P}$ of probabilities such that the acceptable options under C are precisely those that maximize expected P-utility for some P $\in \mathcal{P}$.**

Reconsider this decision problem:



- **Note well that option *m* has a local Bayes model from the choice set {*f*, *g*, *m*}**
  if and only if **P(ω_2) = .5 belongs to the IP set *𝓟*.**

This observation about the admissibility of a mixed option generalizes to allow very fine IP elicitations using coherent choice functions.

- **Each (arbitrary) IP set 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 with coherent choice functions.

This leads to different admissible options in a simple (normal form) decision problem than results when the IP set is, instead, the convex hull formed with extreme points satisfying independence between two events.

Combining Expert Bayesian Opinions. Can it be done?

The challenge is to determine whether there are defensible rules for combining a set of *n*-many "expert" probability distributions into one common probability distribution.

By contrast, and IP-model might use the set of expert opinions, without pooling.

We suppose that each of our *n*-many experts has an opinion about some common domain of interest, represented by the partition into relevant states:

$$\Omega = \{\omega_1, \ldots, \omega_k\}.$$

Expert_i's opinion is probability distribution $P_i = \langle p_{i1}, \ldots, p_{ik} \rangle$ over Ω, $i = 1, \ldots, n$.

- **Can we combine these *n*-many probabilities, P_i, into a single probability P_G that reflects the group's combined wisdom?**

<u>Linear Pooling</u>:

Assign each expert a non-negative *weight* $w_i \geq 0$ to reflect her/his relative expertise in the group, and standardize these so that $\sum_i w_i = 1$

Form $P_G = \sum_i w_i P_i$, the w_i-weighted average of their separate opinions.

- **P_G is called a *Linear Pool* of the expert opinions.**
- **The Linear Pool puts P_G inside the *hull* (= closed, convex set) of the n-many points P_i ($i = 1, \ldots, n$).**

What are some of the nice features of a Linear Pool?

- **Preservation of unanimity of (unconditional) probabilistic opinions**

$$\text{If } c_1 \le P_i(E) \le c_2 \ (i = 1, \ldots, n) \text{ then } c_1 \le P_G(E) \le c_2.$$

Suppose there is a common utility U for outcomes across the group, that is, suppose the group is a *Team*.

- **If each expert judges that Act_1 is better than Act_2 by the standards of SEU, then so too the Team will make the same Pareto judgment – using the shared utility U and pooled opinion P_G.**

- **The Linear Pool is computationally convenient in the following sense of being a *local* computation.**

Once the w_i ($i = 1, \ldots, n$) are fixed $P_G(E)$ depends solely on the *n*-values $P_i(E)$.

In other words, $P_G(E)$ does not depend upon how the n-many experts divide up their probabilities on E^c.

- **But is there a problem with the Linear Pool? What more might we want of a Bayesian consensus than is required by the Pareto condition for pairwise comparisons?**

Learning and Pooling.

Let us use conditional probability as the rule for updating new information.

- **$P_i(\bullet \mid F)$ is the revised opinion for P_i when new information F is added.**

(1) Consider allowing the experts all to learn the same new information F before pooling their opinions with weights w_i.

So, by this method of first updating and then pooling we obtain

$$P^1_G(\bullet \mid F) = \sum_i w_i P_i(\bullet \mid F).$$

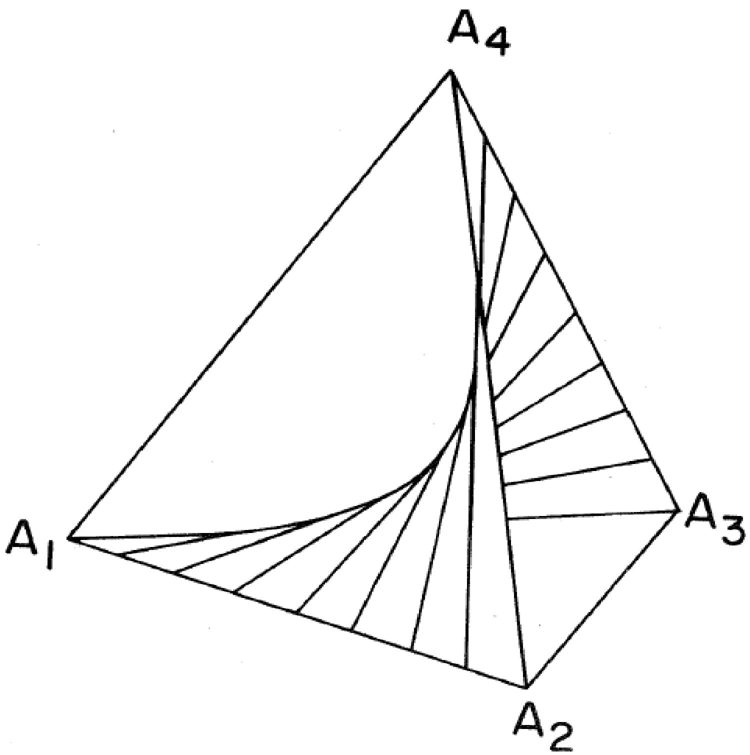
(2) However, we might first pool the expert opinions and then update P_G with the same information F, to yield

$$P^2_G(\bullet \mid F) = P_G(\bullet \cap F) \div P_G(F)$$
$$= \sum_i w_i P_i(\bullet \cap F) \div \sum_i w_i P_i(F).$$

Alas, generally, $\quad P^1_G(\bullet \mid F) \neq P^2_G(\bullet \mid F)$ **!!**

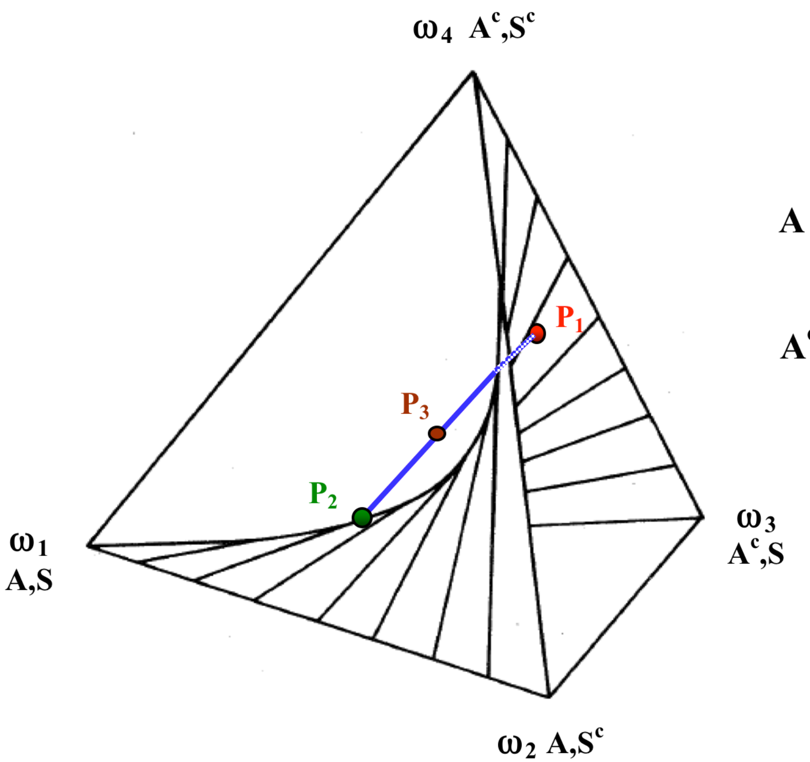
- **The Linear Pool is not "Externally Bayesian"!**

- **Consider the 3-dimensional simplex of probabilities on two events.**



	F	F^c
E	A_1	A_2
E^c	A_3	A_4

We see that, generally, linear pooling two probability distributions that make the events E and F independent will make them dependent!



	S	S^c
A	ω_1	ω_2
A^c	ω_3	ω_4

This method of pooling creates some strange decisions for the group.

If *n* = 2 and both experts think that E and F are independent events, then each will refuse to pay anything to learn about F before betting on E.

However, if a linear opinion pool is formed first, that opinion may make E and F dependent events, and under the pooled-opinion, there will be value in first learning F before wagering on E.

Example 5 (sketch): Consider two doctors who are unsure both about your allergic state and about the weather in China, but who agree these are independent events. Do you mind if, instead of checking your medical history for information about your drug allergies, instead they spend the insurance money learning about the weather in China and using that information to decide on your treatment?

Here is the normal form version of that sequential problem.

***Example 5*: Consider a decision problem among three options – three treatment plans $\{T_1, T_2, T_3\}$ defined over 4 states $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ with determinate utility outcomes given in the following table. That is, the numbers in the table are the utility outcomes for the options (rows) in the respective states (columns).**

	ω_1	ω_2	ω_3	ω_4
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

Consider the convex set of probabilities be generated by two extreme points, distributions P_1 and P_2. 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**

- **Note well that (for i = 1, 2, 3) under probability P_i, only option T_i is Bayes-admissible from the option set of $\{T_1, T_2, T_3\}$.**

Now, interpret these states as the cross product of two binary partitions:

- **a binary medical event – A (patient allegric), A^c (patient not-allergic),**
- **a binary meteorological partition – S (sunny) and S^c (cloudy).**

Specifically: $\omega_1 = A\&S$ $\omega_2 = A\&S^c$ $\omega_3 = A^c\&S$ $\omega_4 = A^c\&S^c$

- **Under P_1, the two partitions are independent events with $P_1(A) = .4$ and $P_1(S) = .2$.**
- **Likewise, under P_2, the events are independent, $P_2(A) = .6$ and $P_2(S) = .8$.**
- **But under linear pooling P_3, A and S are positively correlated:**

$$.56 = P_3(A \mid S) > P_3(A) = .5,$$

as happens with each distribution that is a non-trivial mixture of P_1 and P_2.

The three options have the following interpretations:

T_1 and T_2 are ordinary medical options, with outcomes that depend solely upon the patient's allergic state.

T_3 is an option that makes the allocation of medical treatment a function of the meteorological state, with a "fee" of 0.01 utile assessed for that input.

T_3 is the option "T_1 if cloudy and T_2 if sunny, while paying a fee of 0.01."

Suppose P_1 represents the opinion of medical expert$_1$, and P_2 represents the opinion of medical expert$_2$.

Without linear pooling, T_3 is inadmissible for each expert.

This captures the shared agreement between the two medical experts that T_3 is unacceptable from the choice of three $\{T_1, T_2, T_3\}$, and it captures the pre-systematic understanding that under T_3 you pay to use medically irrelevant inputs about the weather in order to determine the medical treatment.

However, with linear pooling of the pair P_1 and P_2, then T_3 (or a variant of T_3) becomes uniquely admissible for the group.

- **There is no violation of the (binary) Pareto condition under the group opinion formed by the linear pool since the experts disagree about which option, T_1 or T_2, is better than T_3, though they agree that T_3 is not best.**

***Aside*: *Consensus* is not bargaining!**

From a *bargaining* point-of-view, it makes good sense for each expert to accept option T_3.

Option T_3 allows each party in a bargaining problem to think that, with probability .8, his/her medical view will decide the treatment allocation for the patient.

Externally Bayesian Pooling Rules.

There is a family of pooling rules that is invariant over the order of pooling and updating by conditioning. These are called *Externally Bayesian* Pooling rule.

It is a *Logarithmic Pool*: $P_G \propto \prod_i P_i^{w_i}$

It is a linear pool in the logarithms of the expert opinions.

- **What is problematic about this pooling rule?**

Example 3: Using three states and two experts.

$\Omega = \{\omega_1, \omega_2, \omega_3\}$ $P_1 = <.3, .5, .2>$, $P_2 = <.3, .2, .5>$, and $w_1 = w_2$.

***Exercise*: Show that using the logarithmic pooling rule,** $P_G(\omega_1) \neq .3$**, which is a violation of unanimity for pooling of the unconditional probabilities.**

Coherent choice functions may be characterized by axioms on admissible sets that parallel familiar axioms for coherent preferences over horse-lotteries.

Coherent Preference **$<$**

$Axiom_1$ **$<$ *is a weak order***

$Axiom_2$ **$<$ *obeys Independence***

$o_1 < o_2$ *iff*

$xo_1 \oplus (1\text{-}x)o_3 < xo_1 \oplus (1\text{-}x)o_3$

$Axiom_3$ Archimedes

If $o_1 < o_2 < o_3$, then $\exists\ 0 < x, y < 1$

$xo_1 \oplus (1\text{-}x)o_3 < o_2 < yo_1 \oplus (1\text{-}y)o_3$

$Axiom_4$ **State-independent Utilities Preference over constant acts reproduces within each non-null state.**

Coherent Choice Functions

$Axiom_{1a}$ **Sen's Property α:**

An inadmissible option remains so upon addition of other options.

$Axiom_{1b}$ **Aizerman condition, *almost*:**

Deleting inadmissible options does not promote other inadmissible options.

These two axioms determine a strict partial order O_1 « O_2 on option *sets*: O_1 contains no admissible options from among the choice of $O_1 \cup O_2$.

The remaining 3 pairs of axioms are expressed in terms of the partial order « and parallel the axioms for coherent (binary) preference.

Representation Theorem (SSK 2007)

- **The 4-pairs of axioms are necessary for a choice function to be coherent.**

- **The axioms suffice for representing a choice function with the Bayes-styled decision rule (the option is admissible if Bayes for some probability in set $\mathcal{P}$) applied to a set of *Probability/Almost-state-independent utility* pairs.**

- ***Each*** **two different sets of probabilities on the same state space generate different coherent choice functions.**

  Finite decision problems suffice for the identification.
  Choice functions provide a "behavioral" characterization of IP sets.

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