A contrast between two decision rules for use with (convex) sets of probabilities:
Γ-Maximin versus *E-admissibilty*.
T.Seidenfeld
July 2002[1]

1. *Introduction*.

This paper offers a comparison between two decision rules for use when uncertainty is depicted by a non-trivial, convex[2] set of probability functions Γ. This setting for uncertainty is different from the canonical Bayesian decision theory of expected utility, which uses a singleton set, just one probability function to represent a decision maker's uncertainty. Justifications for using a non-trivial set of probabilities to depict uncertainty date back at least a half century (Good, 1952) and a foreshadowing of that idea can be found even in Keynes' (1921), where he allows that not all hypotheses may be comparable by qualitative probability – in accord with, e.g., the situation where the respective intervals of probabilities for two events merely overlap with no further (joint) constraints, so that neither of the two events is more, or less, or equally probable compared with the other.

Here, I will avail myself of the following simplifying assumption: Throughout, I will avoid the complexities that ensue when the decision maker's values for outcomes also are indeterminate and, in parallel with her or his uncertainty, are then depicted by a set of (cardinal) utilities. That is, for this discussion, I will contrast two decision rules when the decision maker's uncertainties, but not her/his values are indeterminate.

The more familiar decision rule of the pair under discussion, *Γ-Maximin*[3], requires that the decision maker ranks an option by its lower expected value, taken with respect to the convex set of probabilities, Γ, and then to choose an option whose lower expected value is maximum. This decision rule (as simplified by the two assumptions, above) was given a representation in terms of a binary preference relation over Anscombe-Aumann horse lotteries (Gilboa-Schmeidler, 1989), has been discussed by, e.g., Berger (1985, §4.7.6) and recently by (Grunwald and Dawid, 2002), who defend it as a form of *Robust Bayesian* decision theory.

The *Γ-Maximin* decision rule creates a preference ranking of options independent of the alternatives available: it is context independent in that sense. But *Γ-Maximin* corresponds to a preference ranking that fails the so-called (von Neumann-Morgenstern's) "Independence" or (Savage's) "Sure-thing" postulate of Bayesian Expected Utility theory.

The decision rule that I shall here contrast with *Γ-Maximin*, called *E-admissibility* ('*E*' for "*expectation*") I attribute to I. Levi (1974, 1980). *E-admissibility* constrains the decision maker's admissible choices to those feasible options that are Bayes for at least one probability $P \in \Gamma$. That is, given a choice set **S** of feasible options, the option $O \in S$ is *E*-

---

[2] The issue of convexity is also controversial. See Seidenfeld, Schervish, and Kadane (1995) for a representation of partially ordered strict preferences that does not require convexity, and for reasons why sometimes convexity is to be avoided. Rebuttal is presented by Levi (1999, section 7).
[3] When outcomes are cast in terms of a (statistical) loss function, the rule is then *Γ-Minimax*: rank options by their maximum expected risk and choose an option whose maximum expected risk is minimum.

admissible on the condition that, for at least one $P \in \Gamma$, O maximizes expected utility with respect to the options in $\boldsymbol{S}$.[4]

Savage (1954, §7.2[5]) defended a precursor to this decision rule in connection with cooperative group decision making. Variants to this rule have been the subject of representation theorems by, e.g., Giron and Rios (1980) and Se., Sc., and K. (1990, 1995).

E-admissibility does not support an ordering of options (real-valued or otherwise), so that it is inappropriate to characterize *E-admissibility* by a ranking of choices, independent of the set $\boldsymbol{S}$ of feasible options. However, the distinction between options that are and are not *E-admissible* does support the "Independence" postulate. For example, if neither option $\boldsymbol{A}$ or $\boldsymbol{B}$ is *E-admissible* in a given decision problem, then their convex combination, the mixed option $\alpha\boldsymbol{A} \oplus (1-\alpha)\boldsymbol{B}$ $(0 \le \alpha \le 1)$ likewise is *E-inadmissible* when added to the same set of feasible options, as is evident from the basic Expected Utility property that the utility of a convex combination is the convex combination of the separate utilities.

To foreshadow the principal theme of my remarks here, it is that "Independence" or "Sure-thing" reasoning, rather than the "Ordering" postulate which regulates coherent hypothetical preferences, relating to conditional probabilities. It is my contention, then, that *E-admissibility* has the normative advantage over Γ-*Maximin*, particularly with respect to decision problems that involve hypothetical preference.

In a general setting, the two rules are not equivalent. Here is a heuristic illustration of that difference.

> ***Example 1***: Consider a binary-state decision problem, $\Theta = \{E, E^{\mathbf{c}}\}$, with three feasible options. Option $\boldsymbol{A}$ yields an outcome worth 1 utile if state E obtains and an outcomes worth 0 utiles if E fails, i.e. if $E^{\mathbf{c}}$ obtains. Option $\boldsymbol{B}$ is the mirror image of $\boldsymbol{A}$ and yields an outcome worth 1 utile if $E^{\mathbf{c}}$ obtains and an outcomes worth 0 utiles otherwise, i.e. if E obtains. Option $\boldsymbol{C}$ is constant in value, yielding an outcome worth 0.4 utiles regardless whether E or $E^{\mathbf{c}}$ obtains.

Figure 1, below, graphs the expected utilities for these three acts. Let $\Gamma = \{\boldsymbol{P}: .25 \le \boldsymbol{P}(E) \le .75\}$. The surface of Bayes-solutions is highlighted in red.

The expected utility for options $\boldsymbol{A}$ and $\boldsymbol{B}$ each has the interval of values [.25, .75], whereas $\boldsymbol{C}$ of course has constant expected utility of .4. From the choice set of these three options $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}\}$ the Γ-*Maximin* decision rule determines that $\boldsymbol{C}$ is (uniquely) best, assigning it a value of 0.4, whereas $\boldsymbol{A}$ and $\boldsymbol{B}$ each has a value of 0.25. By contrast, under *E-admissibility*, only the option $\boldsymbol{C}$ is inadmissible from the trio. Either of $\boldsymbol{A}$ or $\boldsymbol{B}$ is *E-admissible*. The rules are diametrically opposed in what they allow in this decision problem.

---

[4] Levi's decision rule is lexicographic, with E-admissibilty the first tier. He allows a secondary "security" criterion to be employed in selecting from among the E-admissible options. For instance, one might maximize the minimum expected value among the E-admissible options. In Example1, that does not further narrow the admissible options, however. The evident symmetry between options $\boldsymbol{A}$ and $\boldsymbol{B}$ suggests that only a rule for "picking" rather than "choosing" (see Morgenbesser and Ullmann-Margalit, 1977) will resolve that problem.

[5] Savage's analysis of the decision problem depicted by Figure 1, p. 123, and his rejection of option b, p. 124 is the key point.
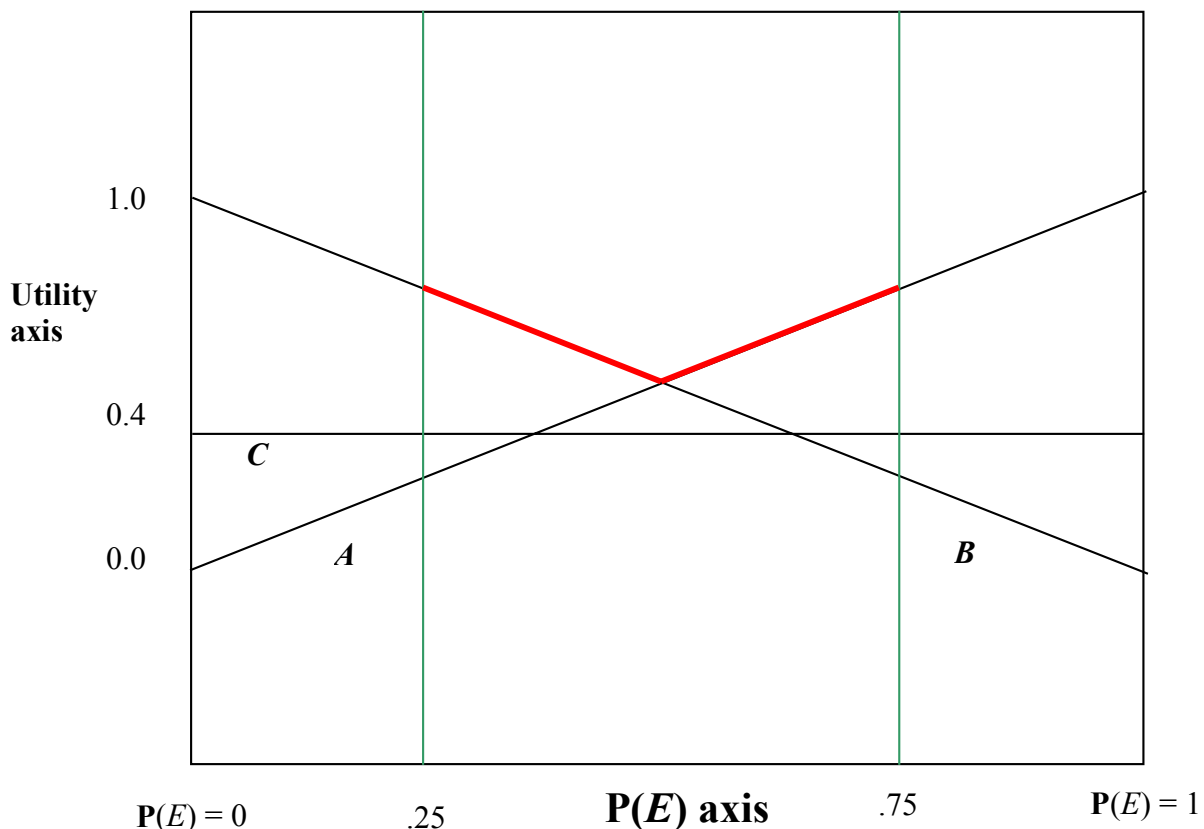
FIGURE 1

- *Aside on boundaries of the convex sets.*

I note in passing there are mathematical niceties in the delicate questions about the faces of the convex sets of probabilities, as to which are open and which are closed, and that these affect the comparison between Γ-Maximin and E-*admissibility*. To illustrate, consider first a coin that is thought to biased for heads, but by no known amount. That uncertainty is represented by the half closed interval of probabilities for the coin landing heads, (0.5, 1.0]. If the coin is thought only *not* to be biased for tails, that is represented by the closed interval of probabilities, [0.5, 1.0]. The difference is with respect to the closure of the convex set of probabilities. This makes a difference between the two decision rules under discussion here, as illustrated next.

Let money be linear in utility at modest stakes. Consider a choice between a fair money wager on heads (at modest stakes) and the status quo. The Γ-*Maximin* rule uses the infemum of expected utilities to rank order options. Both options (betting and abstaining) have the same lower expectation (0) under either set of probabilities. However, when uncertainty is represented by the half-open interval of probabilities, when the coin is thought to be biased for heads, the even money bet has higher expected value than does abstaining for each probability in the set. Hence, by E-*admissibility*, betting is strictly preferred in that case. When the closed interval of probabilities represents the decision maker's uncertainty, both options are E-*admissible*, as they have equal expected value for the case when P(Heads) = ½. Thus, between the two rules, only E-*admissibility* distinguishes the two different epistemic cases.

3

By similar reasoning, we see that *Γ-Maximin*, just as with ordinary Minimax theory, permits weakly dominated (so-called *inadmissible*) options, even when the set of dominating states has only positive probability for each element of *Γ*.

Example 1 *(modified)*: Increase the uncertainty for *E* so that *Γ* = {*P*: 0.0 < *P*(E) < 1.0}. Contrast the constant option *C* with the composite act [*A*+*C*] where [*A*+*C*](*E*) = 1.4 and [*A*+*C*](*E*^c) = 0.4. Then *C* and [*A*+*C*] have the same *Γ-Maximin* ranking (0.4), this despite that fact that *C* is, in the usual sense, *inadmissible* against the alternative [*A*+*C*], and is *E-inadmissible* against it as well. Note that the Bayes model for that supports the *Γ-Maximin* indifference between

Here is the same phenomenon formulated in terms of risk

Example 1 *(continued)*: Consider the expectation, given a parameter as the criterion to be maximized (or to be minimized, when formulated as *risk*, i.e., when utility is given by a loss function). Let the half open interval *Θ* = (0.5, 1.0] be the parameter space for the Bernoulli parameter θ, the chance for heads on a flip, and let *Γ* be the set of all probabilities over the parameter space. Then, for each possible value of the parameter, the even money bet on heads has higher expected value (better risk) than does the status quo. Each element of *Γ* is a convex combination of these extreme points. Nonetheless, *Γ-Maximin* ranks the bet on head indifferent with the status quo. The only Bayes models for this indifference require a finitely additive probability over the parameter space, agglutinated at the (missing) value 0.5. That is, *Nature*'s Minimax strategy here is a limit of proper (Bayes) "priors" over the parameter θ.

Last, note that under *Γ-Maximin* there is no value for this decision in the information of additional (*iid*) flips of the coin. That is, regardless the number of *iid* trials that may be observed, the even money bet on heads remains indifferent under *Γ-Maximin* to the status quo. This anomalous valuation for new data is examined in greater detail, below.

2. More about *Γ-Maximin*.

Because *Γ-Maximin* is a species of more general Minimax theory, it depends for its best results on the familiar assumption that the option space is convex, i.e., that mixed strategies are available. Let *M*(*S*) be the closure of the feasible option set *S* under mixtures. The following is well known and follows immediately from basic Minimax theory, e.g., Casella and Berger (1990, chapter 10).

Result **2.1**: Assume that *M*(*S*) is finitely generated, that *Γ* is closed and convex, and that its elements have a common, finite support set *Θ* including all the extreme (0-1) distributions on *Θ*. Then, a *Γ-Maximin* option exists and it is Bayes for some element of *Γ*.

Example 1 *(continued)*: Let *D* = .5*A* ⊕ .5*B* be the mixed strategy, which with probability ½ is *A* and with probability ½ is *B*, say on the flip of a fair coin: *Heads* yields *A* and *Tails* yields *B*. *D* is an *equalizer* strategy, with a constant expected value 0.5. It is the *Γ-Maximin* solution from *M*(*S*) with respect to the set of all probabilities *Γ** on the two-state partition *Θ*, and it is Bayes against the minimax

"prior" $P(E) = P(E^c) = .5$. Since this uniform "prior" belongs to the set $\Gamma = \{P: .25 \leq P(E) \leq .75\}$, **D** is the Γ-*Maximin* solution to the decision problem with respect to $M(S)$ as well. (Let $\boldsymbol{D}_r$ be the reverse mixture, *Heads* yields **B** and *Tails* yields A, using the same flip of a fair coin. $\boldsymbol{D}_r$ will be useful for a sequential decision problem, below.)

Of course, though the Γ-*Maximin* solution from $M(S)$ is Bayes (and hence is *E-admissible*) the Γ-Maximin *ranking* of options from $M(S)$ does not have a Bayes model. This is evident as Γ-*Maximin* (just as Minimax theory, generally) does not satisfy the *Mixture Dominance* (so-called "Betweeness") principle, which is a logical consequence of *Independence* and *Ordering*.

Expressed in terms of a preference ordering, $\leq$, the *Mixture Dominance* postulate requires:
> *Mixture Dominance*: For each pair of options $\{A, B\}$, and for each $0 \leq \alpha \leq 1$,
$$\text{Min}_{\{A, B\}} \leq \alpha A \oplus (1-\alpha)B \leq \text{Max}_{\{A, B\}}.$$
Where '$\oplus$' denotes convex combination, e.g., the lottery mixture of two options.

In *Example 1*, we have that under Γ-*Maximin* $0.25 \approx A$, $B < D \approx 0.5$, and *Mixture Dominance* fails as **D** is the ($\alpha = .5$) convex combination of A and B.

It is sometimes alleged that preference orderings which are not Bayesian are not *coherent*, in the technical sense that they admit a *Dutch Book*. That is incorrect. Γ-*Maximin* preferences are immune to a Book. They do not lead to a sure loss under the rules of Dutch Book play.

Call an option Γ-*favorable* if it is ranked by Γ-Maximin over the status-quo, 0 point.
> ***Result* 2.2**: No finite combination of Γ-favorable options can result in a Dutch Book.
> *Proof*: Reason indirectly. Suppose that the sum of a finite set of Γ-*favorable* gambles is negative in each state of a finite partition. Choose an element **P** from Γ. The probability **P** is a convex combination of the extreme (0-1) probabilities, corresponding to a convex combination of the finite partition by states. The expectation of a convex combination of probabilities is the convex combination of the individual expectations. This makes the **P**-expectation of the sum of the finite set of Γ-favorable options negative. But the **P**-expectation of the sum cannot be negative unless at least one of the finitely many gambles has a negative **P**-expectation. Then that gamble cannot be Γ-*favorable*, since **P** is an element of Γ.

Thus, (unconditional) Γ-*Maximin* preferences are not subject to sure loss.

However, the same cannot be said about *hypothetical* preferences, particularly as those relate to sequential choice problems. For the current investigation of Γ-*Maximin* and its comparison with *E-admissibilty*, I want to highlight how Γ-*Maximin* addresses *hypothetical* preferences and, through *hypothetical* preferences, how it values new information. By an *hypothetical* preference I mean a preference relation with respect to an augmented body of evidence, a preference given some new information, $Y = y$, not already known.

Hypothetical preference, in this sense, is not to be confused with unconditional, *called-off preference, given $Y = y$*, which, for example, is how Savage defines *conditional* preference.

For Savage (or deFinetti) *conditional preference, given Y = y*, is a comparison using unconditional preference, among acts all of which agree on outcomes for states $Y \neq y$.

Instead, *hypothetical preference, given Y = y*, compares acts that are defined with respect to a quotient algebra obtained by equating $Y \neq y$ with the impossible event. In Savage's theory, what I am calling *hypothetical preference*, given $Y = y$, corresponds to a comparison between partially defined (unconditional) acts, partial functions defined only for states consistent with the new information, $Y = y$.

Elsewhere (Seidenfeld, 1988), I have shown that each (real-valued) decision rule that satisfies the *Ordering* postulate but which fails *Independence* is not *sequentially coherent*. That is, in extensive form (i.e., in sequential) decisions, choices are not invariant over which preferentially *indifferent* outcomes appear at the terminal choice points. From what is shown above (where *Γ-Maximin* fails *Mixture Dominance*), this result applies to *Γ-Maximin*. Moreover, by contrast, *E-admissibility* is sequentially coherent.
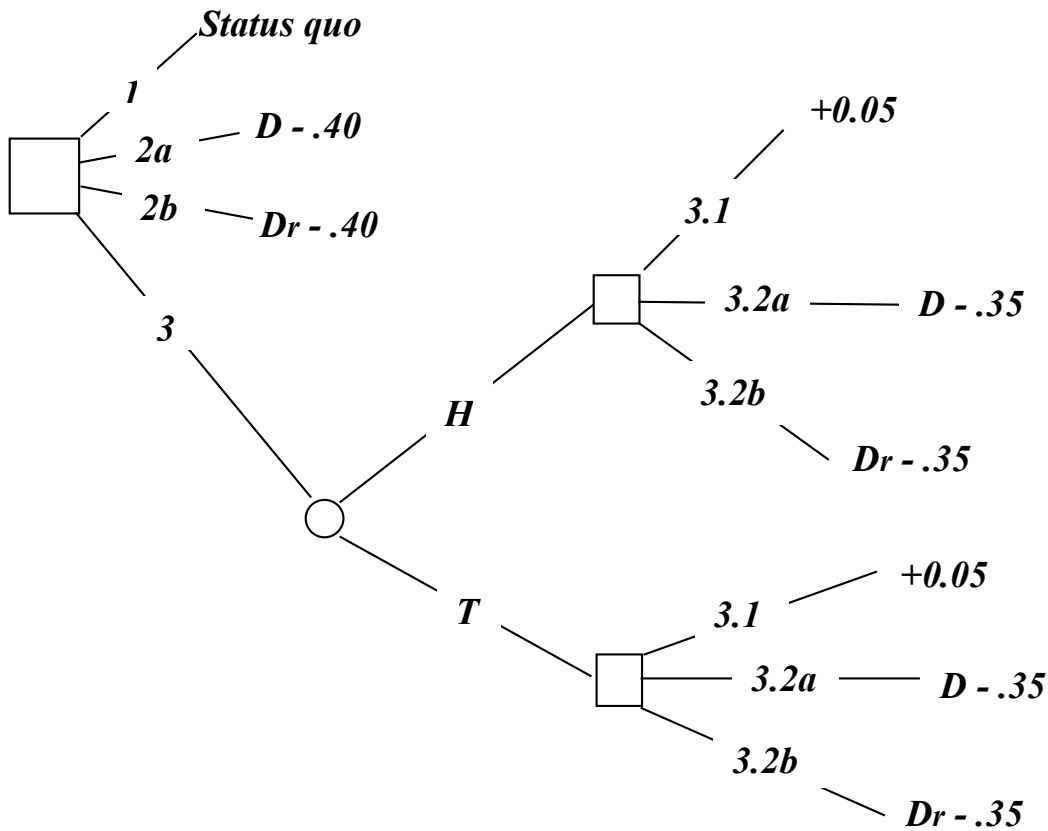
Choice rules that obey the *Ordering* assumption are supposedly context insensitive in that it is sometimes argued that *coherent* orderings (those free from Dutch Book) equate *extensive* and *normal* form decisions. But this is not true of *Γ-Maximin* because of the way it addresses hypothetical preference.

Seqential Example 1 (continued): Construct a sequential decision problem involving these options from Example 1, as follows. At the initial choice node, there are four options, the first three of which are terminal and the third is sequential:

(1) the status quo – a terminal option
(2a) the opportunity to purchase **D** at a cost of 0.4 units – a terminal option.
(2b) the opportunity to purchase D_r at a cost of 0.4 units – a terminal option.
(3) the opportunity to learn the outcome of the mixing variable {*Heads, Tails*}, and then either to purchase **D** or D_r at a cost of 0.35 units or to receive 0.05 units outright – a sequential option involving hypothetical preferences.

The figure, below, provides a schematic illustration of this choice problem.

Recall that, as evaluated by *Γ-Maximin* (or by *E-admissibility*), at the initial node **D** (or **D**$_r$) has a constant value of 0.5 units. Thus, either option (2a) or (2b) is a *Γ-Maximin favorable* (and *E-admissibility favorable*) *option*, with a positive value of 0.1 unit, as assessed at the initial choice point. Hence, option (1) is *Γ*-inadmissible (and *E-inadmissible*), being strictly dispreferred to options (2a) and (2b).

What is the decision maker's ranking (at the initial node) of the sequential option (3)? Given the outcome that the mixing coin lands *Heads*, act **D** *is* the act *A* and act **D**$_r$ *is* the act **B**, as acts are here function from states to (utility) outcomes.

Given that the mixing coin lands *Heads*, the *hypothetical Γ-Maximin* ranking of **D** (and of **D**$_r$) drops to 0.25 units, so that the prospective trade of 0.35 units to purchase **D** (or **D**$_r$) has a *Γ-Maximin* hypothetical value of −0.10 units, which is less than the (hypothetical) value of the *Γ-Maximin favorable* alternative to receive 0.05 units outright. That is, the option to purchase **D** (or **D**$_r$) at a cost of 0.35 units is *hypothetically unfavorable* given that the coin lands *Heads*, and the *Γ-Maximin* decision maker will opt, instead for the gain of 0.05 units at that terminal node of the sequential decision. By symmetry, the option to purchase **D** (or **D**$_r$) at a cost of 0.35 units is hypothetically *Γ-Maximin unfavorable* given that the coin lands *Tails*, and the *Γ-Maxinim* decision maker will opt, instead for the sure gain of 0.05.

Thus, for the *Γ-Maximin'er*, sequential option (3) reduces to a sure gain of 0.05 units. But that amount is dispreferred to option (2a) or (2b) which is each valued (indifferently) at +0.1 units. Therefore, faced with this sequential decision problem the *Γ-Maximin* decision maker will choose option (2).

What is surprising about the *Γ-Maximin* ranking of these three options is that the sequential option (3) is itself equivalent to *paying* the decision maker 0.5 units at the initial node in order first to learn the outcome of the mixing variable {*Heads*, *Tails*}, and then to present him with same three terminal choices as among options (1), (2a), and (2b). That is, in mandating a higher value for option (2a) or (2b) over option (3), *Γ-Maximin* puts a *negative* value on the new data of the mixing variable. The *hypothetical* preferences associated with *Γ-Maximin* disvalues cost-free (relevant) information![6]

E-admissibility recommends differently. Under the sequential option (3), the hypothetical preferences (given either *Heads* or *Tails*) follows the unconditional preferences for the choices in Example 1. That is, the triple of the three options {*+.05*, ***D**-0.35*, ***D**ᵣ-0.35* } has a hypothetical *E*-preference worth no less than +0.15. Moreover, the first of these (3.1) is *E-inadmissible*. Hence, using *E-admissibility* (with or without the security condition), we have that the sequential option (3) is uniquely *E-admissible*.

In short, because *Γ-Maximin* undervalues the triple of options in Example 1 (as judged by E-admissibility), that decision rule produces a negative value for the (ancillary) mixing datum when that information collapses risk back to uncertainty.

Moreover, this sequential decision problem does not receive the same *Γ-Maximin* solution when put into normal form. Then, sequential option (3) expands to form 9 new terminal options. It is obvious that the *Γ-Maximin* solution to this normal form problem, involving 6 options, is the *Γ-favorable* option to purchase ***D*** (or ***D***ᵣ) at a cost of only 0.35 units, with a net gain of 0.15 units. Of course, in the extensive form of the decision problem, this is not a *Γ-Maximin* feasible alternative, viewed from the initial node.

In *Example* 1, ***A*** [or respectively, ***B***] is a called-off option relative to the status quo, called off in case state E^c [respectively, *E*] fails. The failure of the *Independence* postulate, through the failure of *Mixture Dominance*, results in a *sequentially incoherent* treatment of *hypothetical* preferences, as shown above. Next, I pursue this theme in connection with failures of the law of iterated expectations for bounded random variables:

$$\mathbf{E} [\mathbf{E} [X | Y]] = \mathbf{E} [X].$$

In a recent, and very interesting paper, Grunwald and Dawid (2002, §6) show that for a variety of proper scores (so that the decision problem is cast in terms of a loss function) *Γ-Minimax* solutions are equivalent to using Bayes' decisions against a *Maximum Entropy* distribution (as a worst-case "prior") and, generally, against minimum *Kullbach-Lieber* [*K-L*] shifts. But we have just seen that *Γ-Maximin* induces hypothetical preferences in violation of Mixture Dominance principle. This can be linked, via the Grunwald and Dawid results, to counterpart incoherence in the *MaxEnt* principle, and in the principle of making *minimum K-L* shifts. The following two examples illustrate the phenomenon.

---

[6] See (Herron, Seidenfeld, and Wasserman, 1997) for analysis of such cases of "dilation" of uncertainty.

Example 2 (Shimony, 1973). Let X be a simple random variable with possible values $\{x_i: i = 1, \ldots, n\}$ such that one of them is the average value.

That is, for one k, $1 \le k \le n$, $x_k = \bar{x} = \sum_i x_i/n$.

- The "apriori" MaxEnt distribution P_0 for X is uniform $P_0(X = x_i) = 1/n$. So the "apriori" expectation for X is $\mathbf{E}_0[X] = \bar{x}$.

- Using the "linear" constraint $C = \mathbf{E}_1[X] = c$, with $min_X \le c \le max_X$ the *MaxEnt* distribution P_1 for X that satisfies constraint C, also satisfies

  - $P_1(X = x_k) < 1/n$, if $c \ne \bar{x}$
  - $P_1(X = x_k) = 1/n$, if $c = \bar{x}$.

Understanding $P_1(X)$ as a conditional probability $P_0(X \mid C)$, then in order to satisfy the law of iterated expectations $\mathbf{E}_0[\mathbf{E}_0[X \mid C]] = \mathbf{E}_0[X]$, we must have that

  - $P_0(C = \bar{x}) = 1$,

which is a degenerate Bayes model that requires the constraint to be constant, almost surely.

Figure 2, below, illustrates this result for the case of a three sided die, $X = \{1, 2, 3\}$, using Barycentric coordinates. The concave function is the graph of the *MaxEnt* solutions P_1 for the outcome of the die, given the constraint C of the new P_1-average for X. It is evident that the unique Bayes' model for this application of *MaxEnt* is the point-mass probability P_0 that assigns probability 1 to the die being fair.

Also shown, for comparison, is the Bayes solution using a uniform "apriori" probability P_0 over the simplex of all possible die loadings, for which P_0 of X is, of course, the uniform distribution: $P_0(X = i) = 1/3$, $i = 1, 2, 3$.   Then (marginal) conditional probability $P_0(X \mid C = c)$ is the inverted-V function.
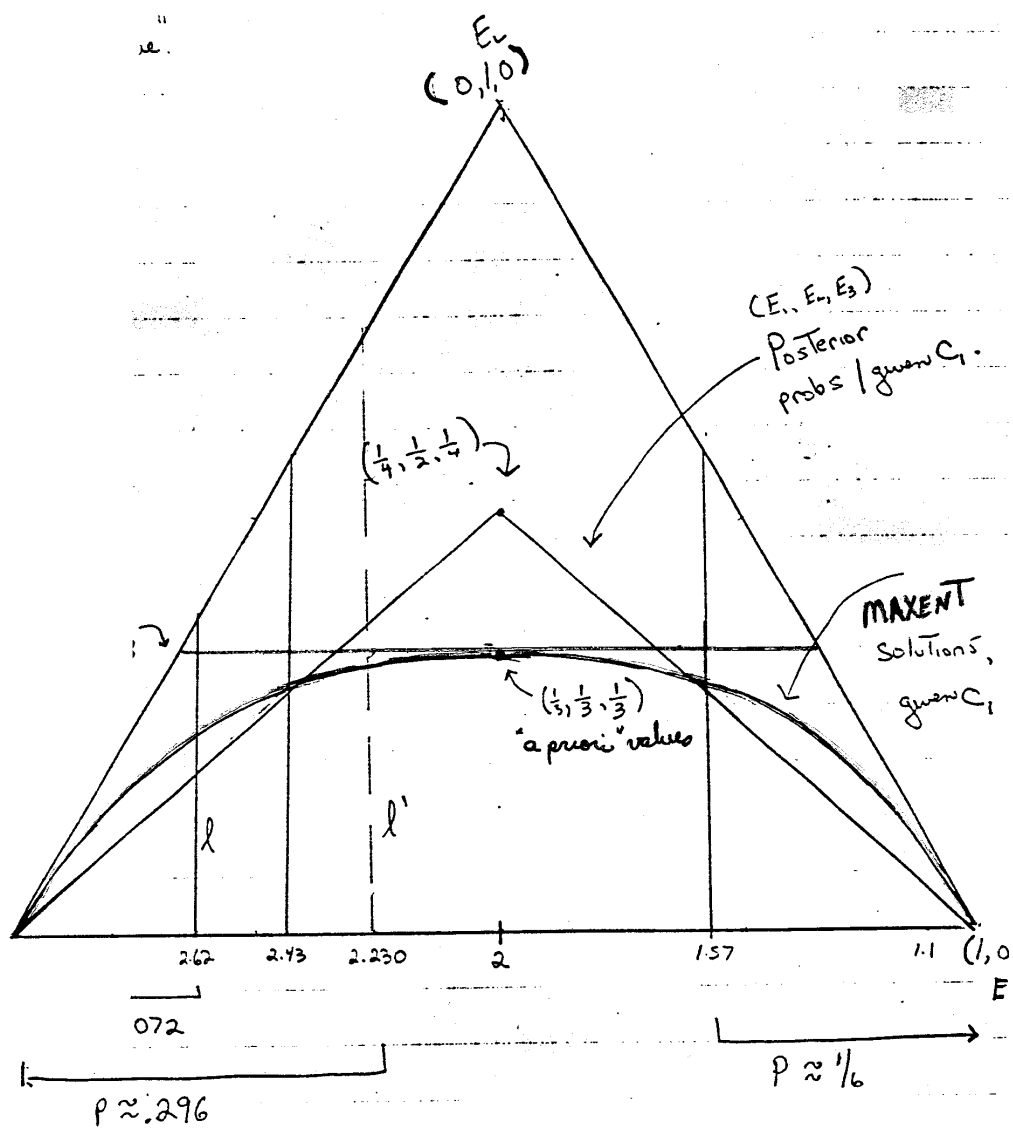
FIGURE 2

Next, I give an illustration of the same anomaly that applies to all instances of the principle of *Minimum K-L shifts*.

> ***Example 3*** (Seidenfeld, 1987): Partition the space into three disjoint, nonempty events $\{E_1, E_2, E_3\}$ and let P_0 be any probability distribution over these. Consider the "linear" constraint of a new odds ratio for E_1 to E_2, $P_1(E_1) / P_1(E_2) = \alpha/(1-\alpha)$.
> This takes the form of a linear constraint, using the deFinetti's called-off wager on E_1 to E_2, as follows.
> Define the random variable Y_α by,
> $$Y_\alpha(E_i) = -(1-\alpha) \qquad \text{for } i = 1$$

10

$$= \alpha \quad \text{for } i = 2$$
$$= 0 \quad \text{for } i = 3.$$

Then set $C = c_\alpha$ as the condition that $\mathbf{E}_1[Y_\alpha] = 0$.

It follows that the minimum K-L shift from P_0 to P_1 satisfies:

- $P_1(E_3) > P_0(E_3)$ if $P_0(E_1) / P_0(E_2) \neq \alpha/(1\text{-}\alpha)$
- $P_1(E_3) = P_0(E_3)$ if $P_0(E_1) / P_0(E_2) = \alpha/(1\text{-}\alpha) = c_{\alpha_0}$

Then, just as in Example 2, in order to satisfy the law of iterated expectations $\mathbf{E}_0[\mathbf{E}_0[E_3 \mid C]] = \mathbf{E}_0[E_3]$, we must have that

- $P_0(C = c_{\alpha_0}) = 1$,

which is a degenerate Bayes model that requires the constraint to be constant, almost surely.

Figure 3, below, illustrates this result for the case of a three sided die, $X = \{1, 2, 3\}$, again using Barycentric coordinates. P_0 is taken to be the uniform distribution over the three events. The concave function is the graph of the *minimum K-L shift* to P_1 for the constraint C of the new odds ratio $P_1(E_1) / P_1(E_2) = \alpha/(1\text{-}\alpha)$. It is evident that the unique Bayes' model for this application of *MaxEnt* is the point-mass probability P_0 that assigns probability 1 to the die being fair.

Also shown, for comparison, is the Bayes solution using a uniform "prior" probability P_0 over the simplex of all possible die loadings, for which P_0 of E_i is, of course, the uniform distribution: $P_0(E_1) = 1/3$, $i = 1, 2, 3$. Then the (marginal) conditional probability $P_0(E_e \mid C = c) = 1/3$ is constant, as the odds ratio, C, is irrelevant to E_3.
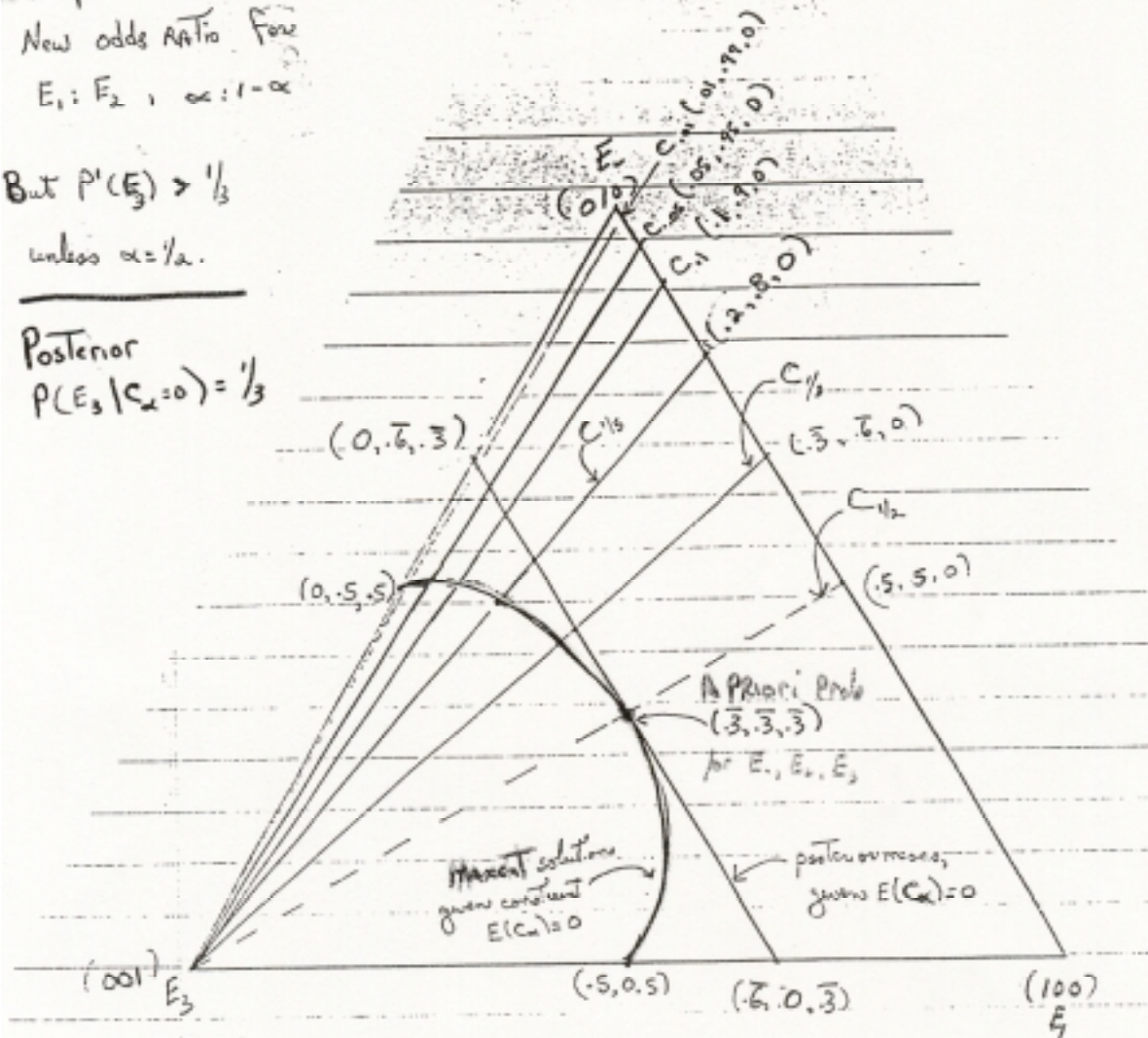
Example: Uniform prior on (E_1, E_2, E_3)

New odds ratio for

$E_1 : E_2$, $\alpha : 1-\alpha$

But $P'(E_3) > 1/3$

unless $\alpha = 1/2$.

___

Posterior

$P(E_3 \mid C_\alpha = 0) = 1/3$

(Figure: triangle with vertices $E_3 \ (001)$, $E_1 \ (100)$, $E_2 \ (010)$; labeled points $(0,.\bar{6},.3)$, $(0,.5,.5)$, $(.3,.\bar{6},0)$, $(.5,.5,0)$, $(.2,.8,0)$, $C_{.01}(.01,.99,0)$, $C_{.05}(.05,.95,0)$, $C_{.1}(.1,.9,0)$, $C_{1/5}$, $C_{1/3}$, $C_{1/2}$, A Priori prob $(.\bar{3},.\bar{3},.\bar{3})$ for E_1, E_2, E_3; "Present solutions given constant $E(C_\alpha)=0$"; "posterior means given $E(C_\alpha)=0$"; $(.5,0,.5)$, $(.\bar{6},0,.\bar{3})$.)

The prior is uniform over all three-point distributions (p_1, p_2, p_3)

C_α: the "constant" is an expectation of 0 for

$$F(E_1) = -(1-\alpha) \qquad F(E_2) = \alpha \qquad F(E_3) = 0$$

fixing an $\alpha : 1-\alpha$ odds ratio for $E_1 : E_2$.

The distributions satisfying $C_\alpha = 0$ is a ray from $(0,0,1)$

FIGURE 3

12

Based on the results of Grunwald and Dawid (2002), *Examples 2* and *3* illustrate how Γ-*Minimax* provides assessments of hypothetical preferences that violate the law of total probability -- unless the principle is made into a degenerate Bayes model – a model in which (almost surely) learning is precluded.

Next, I offer additional contrast with *E-admissibility*.

3. More about E-admissibility.
3.1 Context sensitivity. *E-admissibility* liberalizes the canonical Bayesian Expected Utility (EU) decision theory by sacrificing "Ordering" in order to preserve "Independence."   That move presents an operationally distinct liberalization of canonical Bayesian theory than does *Γ-Maximin*, as we saw in *Example 1*.   However, just as with a set of *Γ-Maximin favorable* options, and for the same reasons, a finite set of *E-admissible favorable* options cannot result in a Book. (An option is *E-admissible favorable* if it is strictly preferred according to *E-admissibility* over the status quo, that is, if the status quo is *E-inadmissible* in a take-it-or-leave-it pairwise choice.)

The brief analysis that follows examines more closely the extent to which "Ordering" fails when *E-admissibility* is applied and how that leads to choices in sequential choices that differ from *Γ-Maximin*. For this purpose, I rely on Sen's (1977) classification of choice rules, which follow next.

Let $C(S)$ be a choice rule applied to a feasible set of options S. Assume that the choice function is well defined, i.e., if $S \neq \varnothing$ then $\varnothing \neq C(S) \subseteq S$.

Given a choice function $C(\bullet)$ define the *(weak) preference relation* $\leq_C$ over pairs of options, by the condition:

$A \leq_C B$ if and only if there exists a feasible set S, with $B \in C(S)$ and $A \in S$.

Given a (complete and reflexive) weak preference relation $\leq$ over pairs of options, define the *binary choice function* $C_\leq(S)$ based on $\leq$ by the condition:

$B \in C_\leq(S)$ if and only if $B \in S$ and for each $A \in S, A \leq B$.

Third, call a choice function $C(\bullet)$ *normal* if its weak preference relation $\leq_C$ regenerates itself through $C_{\leq_C}$. That is, $C_{\leq_C}(S) = C(S)$, applied to all feasible sets, S.

Sen shows (1977, p. 65) that a choice function is normal and generates an ordering if and only if it satisfies the following two of properties.

> ***Property α***: If $A \in S \subseteq S'$ and if $A \in C(S')$, then $A \in C(S)$.

> ***Property β***: Let $\{A, B\} \subseteq S \subseteq S'$. If $\{A,B\} \subseteq C(S)$ and $\{A,B\} \subseteq S'$, then $A \in C(S')$ if and only if $B \in C(S')$.

Moreover, he shows (1977, p. 64) that a choice function is normal if it satisfies Property α and the following

> ***Property γ***: Let $M = \{S\}$ be a class of feasible sets. If $A \in C(S)$ for each $S \in M$, then $A \in C(\cup M)$.

13

Consider two variations of the *E-admissibility* principle: first without and second with a supplemental consideration for choosing among *E-admissible* options. In order to simplify the contrast with *Γ-Maximin*, let the supplemental consideration (what Levi (1980) calls a "security" rule) be to choose from among those *E-admissible* options any that maximizes minimum expectations. The following are straightforward:

Results:

> **3.1** *E-admissibilty* with no supplemental "security" consideration satisfies Property α but fails Properties β and γ.
>
> **3.2**. *E-admissibility* supplemented by a *Γ-maximin* "security" rule fails all three Properties.

Proof:

3.1 That *E-admissibility* (without security considerations) satisfies Property α is immediate. Specifically, if $A \in C(S')$ then with respect to some $P \in \Gamma$, A maximizes expected utility over all options in S'. Then with respect to the same $P \in \Gamma$, A maximizes expected utility over all options in $S \subseteq S'$ on the assumption that $A \in S$.

For the remainder of the proof, use Example 1 to reason as follows.

- To see that Properties β fails, observe that both options A and C are *E-admissible* in a pairwise choice between them. However, upon adding option B, option C is *E-inadmissible* from the triple $\{A,B,C\}$, while A remains admissible..
- To see that Property γ fails note, from what was just shown, that option C is *E-admissible* from each of the pairs of feasible sets $\{A,C\}$ and $\{B,C\}$; however it is *E-inadmissible* from the feasible set of their union $\{A,B,C\}$.

3.2 When *E-admissibility* is supplemented by a *Γ-maximin* "security" rule.

- Property α fails by noting that A is admissible from the triple of options $\{A,B,C\}$, but C alone is admissible from the pair $\{A,C\}$.
- Properties β and γ fail in the same settings illustrated in 2.1.

By contrast, the *Γ-Maximin* choice rule satisfies all three properties, since it assigns each option a real-valued rank, independent of any other feasible options, and the reals are simply ordered by magnitude.

Thus, the price for adopting *E-admissibility* (in either version) as the decision rule is to introduce context sensitivity into non-sequential decision analysis that is not present when binary choices suffice to determine the choice function, as is the case with a normal decision rule (such as *Γ-Maxmiin*) that satisfies Ordering. This context sensitivity will also display itself in a general, non-equivalence of extensive form (sequential) decisions and their normal form (non-sequential) versions, just as with *Γ-Maximin*.

However, the context sensitivity of the *E-admissibility* choice rule, even in sequential decisions, does not lead to failures of the law of total probability, as happens in *MaxEnt* and *Minimum K-L shifts* (as in *Examples* 2 and 3), which are the decision theoretic duals to the *Γ-Maximin* rule. The reason that *E-admissibility* does not lead to such failures is, simply, because its hypothetical decisions always apply Bayes' decision rule to the (convex) set of *conditional probabilities* associated with the set Γ. *Γ-Maximin* fails to choose Bayes' solutions to hypothetical decisions, with the consequence that its hypothetical preferences do not link together in agreement with the law of total probability. This allows for a difference between *E-admissibility* and *Γ-Maximin* over the value of (cost-free) new information, as I discuss next.

3.2 *E-admissibility* and *Value of New Information.* In *Sequential Example 1*, we found that
Γ-Maximin **requires** a *negative* value for new information – so that a decision maker using
that decision rule will even pay to avoid learning data prior to making a terminal decision.
Moreover, in that same example, we saw that a decision maker using *E-admissibility* may
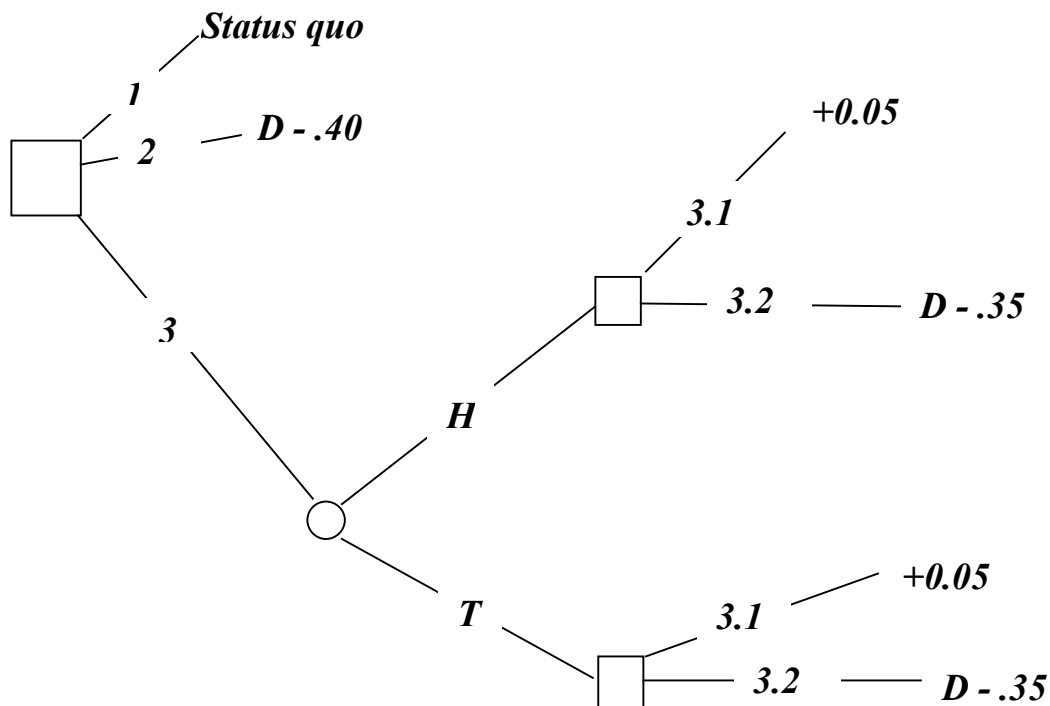respect the value of new (cost free) information, prior to making a terminal decision.

However, in general with *E-admissibility*, the value of new information in sequential
decisions depends upon values reflected in secondary ("security") considerations. For
example, consider the following modified version of Sequential Example 1 obtained by
deleting option D_r.

       Sequential Example 1 (modified): Construct a sequential decision problem
involving these options from Example 1, as follows. At the initial choice node, there are
three options, the first two of which are terminal and the third is sequential:
- (1)   the status quo – a terminal option
- (2)   the opportunity to purchase D at a cost of 0.4 units – a terminal option.
- (3)   the opportunity to learn the outcome of the mixing variable {*Heads*, *Tails*},
     and then either to purchase D at a cost of 0.35 units or to receive 0.05 units
     outright – a sequential option involving hypothetical preferences.

The figure, below, provides a schematic illustration of this choice problem.

A second Sequential Decision Involving the Value of New Information

As with *Γ-Maximin*, terminal option 1 (*status quo*) is *E-inadmissible*, as terminal option 2
(*D*-.40 units) has a determinate expected utility of .10 units. However, given the datum of

15

the mixing variable, that is given outcome *Heads* (or given *Tails*), the hypothetical preference between 3.1 and 3.2 makes each *E-admissible*. Using *Γ-Maximin* as a secondary crieterion in this problem will result in choices identical to what *Γ-Maximin* alone recommends; namely, then 3.1 alone will be admissible and, by backward induction, sequential option 3 is *E-inadmissible* compared with terminal option 2. Thus, the primary criterion of *E-admissibility* supplemented by the secondary criterion of *Γ-Maximin* results in a decision rule that is not guaranteed to respect the value of new, cost-free information.

However, as shown below, there do exist secondary criteria that may be used along with *E-admissibility* to produce a non-negative value for cost-free information. Let **S** be a sequential decision problem that, as in the sequential examples above, offers the decision maker the opportunity to postpone a terminal decision in order instead first to conduct an experiment and then choose among the same terminal options. Let $P \in \Gamma$ be a probability that assigns to the experiment available in **S** *the least added value*, relative to all the probability distributions in Γ.
Definition: Call such a probability *P least informative* for the experiment provided by **S**.

Illustration: In the modified sequential decision problem, above, each *P* satisfying $.4 \leq P(E) \leq .6$ is a least informative probability from the set $\Gamma = \{P: .25 \leq P(E) \leq .75\}$. The expected utility under such a *P* of first learning the outcome of the mixing variable before choosing between options 3.1 and 3.2 is .15 units. (The optimal sequential decision under such a distribution *P* is to select 3.2 regardless the outcome of the mixing variable, with an expected value of .15.) This a gain of only .05 units over the value of the terminal option 2, i.e., the value of **D** - .40, which is .10. For example, with $P(E) = .25$, the optimal sequential decision is to choose 3.1 if *Heads* and 3.2 if *Tails*. This has an expected value of .25 units, making the experiment worth .15 units – so that this distribution is not a least informative probability from the set Γ.

> *Result 3.3*: When *E-admissibility* is supplemented with the secondary criterion to maximize expectations with respect to a *least informative* distribution, then this lexicographic decision rule respects the value of (cost-free) information.
> *Proof*: The result is immediate from the elementary fact that, with respect to *each* P $\in \Gamma$ the optimal (sequential) decision respects the value of cost-free information. Hence, optimizing among *E-admissible* decisions with respect to a least informative distribution respects the value of cost free information.

4. Summary
The discussion here contrasts two decision rules that apply when *uncertainty* is represented by a (convex) set of probabilities, Γ, rather than as in the canonical Bayes case when uncertainty is represented by a single probability distribution. The *Γ-Maximin* decision rule, which ranks options by the infemum of expected utility with respect to the set Γ, satisfies the canonical Bayesian (EU) theory's *Ordering* postulate, but at the expense of violating its *Independence* postulate. A rival decision rule, *E-admissibility*, reverses this trade and abandons *Ordering* but not *Independence*. Neither decision rule is in jeopardy of having Book made against it in normal form decisions. Each rule distinguishes extensive and normal forms of a decision problem. But, since *Γ-Maximin* alone fails the *Independence* postulate, the rankings associated with its hypothetical preferences do not knit together in

accord with the law of total probabilities. This is in contrast with the hypothetical *E-admissible* preferences, which do.

The upshot is that in sequential decision problems *E-admissibility* permits a decision maker (through the appropriate choice of secondary criteria) to respect the value of new information. By contrast, in some sequential decision problems Γ-Maximin mandates a negative value for new information. Is it not a normative failing of a decision theory that it mandates a negative value for cost-free learning?

References
Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). Springer-Verlag: New York.

Casella, G. and Berger, R.L. (1990) *Statistical Inference*. Wadsworth Publishing: Belmont, California.

Gilboa, I. and Schmeidler, D. (1989) "Maxmin Expected Utility with Non-Unique Prior," *J. Math. Econ.* **18**, 141-153.

Giron, F.J. and Rios, S. (1980) "Quasi-Bayesian behavior: A more realistic approach to decision making?" In *Bayesian Statistics* (J.M.Bernardo et al. eds.), 17-38. Univ. of Valencia Press.

Good, I.J. (1952) "Rational Decisions," *J. Royal Statistical Society* **B 14**, 117-114.

Grunwald, P.D. and Dawid, A.P. (2002), "Game Theory, Maximum Entropy, Minimum Discrepancy, and Robust Bayesian Decision Theory," *Report #223*, Dept. of Statistical Science, University College London.

Herron, T, Seidenfeld, T., andWasserman, L. (1997) "Divisive Conditioning: further results on dilation," *Philosophy of Science*. **64**, 411-444.

Keynes, J.M. (1921) *A Treatise on Probability*. MacMillan and Co.: London.

Levi, I. (1974) "On Indeterminate Probabilities," *J.Philosophy* **71**, 391-418.

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

Levi, I. (1999) "Value Commitments, Value Conflict, and the Separability of Belief and Value," *Philosophy of Science* **66**, 509-533.

Morgenbesser, S. and Ullmann-Margalit, E. (1977) "Picking and Choosing," *Social Research* **44**, #4, 757-785.

Savage, L.J. (1954) *The Foundations of Statistics*. Wiley: New York.

Seidenfeld (1987) "Entropy and Uncertainty," In *Foundations of Statistical Inference*, I.B. MacNeil and G.J. Umphrey (eds), 259- 287. Reidel: Dordrecht.

Seidenfeld, T. (1988) "Decision Theory without *Independence* or without *Ordering*, what is the difference?" with discussion, *Economics and Philosophy* 4, 267-315.

Seidenfeld, T., Schervish, M.J., and Kadane, J. (1990) "Decisions Without Ordering" In *Acting and Reflecting* (W. Sieg, ed.), 143-170. Kluwer Academic Publishers: Dordrecht.

Seidenfeld, T., Schervish, M.J., and Kadane, J. (1995) "A Representation of Partially Ordered Preferences," *Annals of Statistics* **23**, 2168-2217.

Sen, A.K. (1977) "Social Choice Theory: a re-examination." *Econometrica* **45**, 53-89.

Shimony, A. (1973) "Comment on the Interpretation of Inductive Probabilities," *J. Statistical Physics* **9**, 187-191.