Is Ignorance Bliss with Sets of Probabilities?

Or, some more things you might rather not know!

Teddy Seidenfeld – CMU

A Basic Theorem of (Bayesian) Expected Utility Theory

If you can postpone a terminal decision in order to observe, cost free, an experiment whose outcome might change your terminal decision, then it is strictly better to postpone the terminal decision in order to acquire the new evidence. [I.J.Good; F.P.Ramsey]

The analysis also provides a value for the new evidence, to answer:

How much are you willing to pay for the new information?
An agent faces a current decision:
• with \( k \) terminal options \( D = \{d_1, \ldots, d^*, \ldots, d_k\} \) \( (d^* \) is the best of these)
• and one sequential option: first conduct experiment \( X \), with outcomes \{\( x_1, \ldots, x_m \)\} that are observed, then choose from \( D \).
Terminal decisions (acts) as functions from states to outcomes

The canonical decision matrix: **decisions × states**

<table>
<thead>
<tr>
<th></th>
<th>s_1</th>
<th>s_2</th>
<th>s_j</th>
<th>s_n</th>
</tr>
</thead>
<tbody>
<tr>
<td>d_1</td>
<td>( 0_{11} )</td>
<td>( 0_{12} )</td>
<td>( 0_{1j} )</td>
<td>( 0_{1n} )</td>
</tr>
<tr>
<td>d_2</td>
<td>( 0_{21} )</td>
<td>( 0_{22} )</td>
<td>( 0_{2j} )</td>
<td>( 0_{2n} )</td>
</tr>
<tr>
<td>d_k</td>
<td>( 0_{k1} )</td>
<td>( 0_{k2} )</td>
<td>( 0_{kj} )</td>
<td>( 0_{kn} )</td>
</tr>
</tbody>
</table>

\( d_i(s_j) = \text{outcome } o_{ij} \).

What are “outcomes”? That depends upon which version of expected utility you consider. We will allow arbitrary outcomes, providing that they admit a von Neumann-Morgenstern cardinal utility \( U(\bullet) \).
A central theme of Subjective Expected Utility \([SEU]\) is this:

- axiomatize (weak) preference \(_<\) over decisions so that
  \[
d_1 \_<\_ d_2 \quad \text{iff} \quad \sum_j P(s_j)U(o_{1j}) \leq \sum_j P(s_j)U(o_{2j}),
\]
  for \textbf{one} subjective (personal) probability \(P(\bullet)\) defined over states
  and \textbf{one} cardinal utility \(U(\bullet)\) defined over outcomes.

- Then the decision rule is to choose that (an) option that \textit{maximizes} \(SEU\).
Note: In this version of SEU, which is the one that we will use here:

1. Decisions and states are probabilistically independent, $P(s_j) = P(s_j \mid d_i)$.

Reminder: This is sufficient for a fully general dominance principle.

2. Utility is state-independent, $U_j(o_{ij}) = U_h(o_{gh})$, if $o_{ij} = o_{gh}$.

Here, $U_j(o_{\cdot j})$ is the conditional utility for outcomes, given state $s_j$.

3. (Cardinal) Utility is defined up to positive linear transformations, $U'(\bullet) = aU(\bullet) + b$ ($a > 0$) is also the same utility function for purposes of SEU.

Note: More accurately, under these circumstances with act/state prob. independence, utility is defined up to a similarity transformation: $U_j'(\bullet) = aU_j(\bullet) + b_j$.

So, maximizing SEU and Maximizing Subjective Expected Regret-Utility are equivalent decision rules.
Reconsider the value of cost-free evidence when decisions conform to maximizing \( SEU \). Recall, the decision maker faces a choice now between \( k \)-many terminal options \( D = \{d_1, ..., d^*, ..., d_k\} \) (\( d^* \) maximizes \( SEU \) among these \( k \) options). There is one sequential option: first conduct experiment \( X \), with sample space \( \{x_1, ..., x_m\} \), and then choose from \( D \) having observed \( X \). Options in \textcolor{red}{red} maximize \( SEU \) at the choice nodes, using \( P(s_j \mid X = x_j) \).
By the law of conditional expectations: $E(Y) = E(E[Y | X])$.

With $Y$ the Utility of an option $U(d)$, and $X$ the outcome of the experiment,

$$\text{Max}_{d \in D} E(U(d)) = E(U(d*))$$

$$= E(E(U(d*) | X)) \text{ (“ignoring } X \text{” when choosing)}$$

$$\leq E(\text{Max}_{d \in D} E(U(d) | X))$$

$$= U(\text{sequential option}).$$

• Hence, the academician’s first-principle:
  Never decide today what you might postpone until tomorrow in order to learn something new.

• $U(d*) = U(\text{sequential option})$ if and only if the new evidence $X$ never leads you to a different terminal option.

• $U(\text{sequential option}) - E(U(d*))$ is the value of the experiment: what you will pay (at most) in order to conduct the experiment prior to making a terminal decision.
When the law of conditional expectations fails, then “cost free” evidence may have negative value, in violation of the basic result.

For a review of cases, see our *Is Ignorance Bliss? J.Phil.* 2008.

Several years ago, at the Info-Metrics workshop honoring Arnold Zellner, I illustrated how this failure may arise when there is act/state dependence (*moral hazard*) in a two person “Bayesian” game. It may be Pareto superior for both players to remain ignorant of some information $X$ rather than having it be common knowledge that Player-1 learns $X$.

Here, I focus on the case where the decision rule is not Expected Utility Maximization with a single probability distribution.


Let the decision rule be [Berger, 1985; Gilboa & Schmeidler, 1989]

- $\Gamma$-Maximin – Maximize minimum expected utility w.r.t. set $\mathcal{P}$.

Then the value of (cost free) information may be negative.

I will illustrate how this arises with *Dilation* (S&W, 1993) of IP sets.
Dilation for sets of probabilities.

Let \( \mathcal{P} \) be a (convex) set of probabilities on algebra \( \mathcal{A} \).

For an event \( E \), denote by

\[
P_*(E) \text{ the lower probability of } E: \inf_{P \in \mathcal{P}} \{P(E)\}
\]

and

\[
P^*(E) \text{ the upper probability of } E: \sup_{P \in \mathcal{P}} \{P(E)\} \{P(E)\}.
\]

Let \( X = (x_1, \ldots, x_n) \) be a partition, here taken to be finite for simplicity.

The set of conditional probabilities \( \{P(E \mid x_i)\} \) (strictly) dilate if

\[
P_*(E \mid x_i) < P_*(E) \leq P^*(E) < P^*(E \mid x_i)
\]

for each \( i = 1, \ldots, n \).

That is, dilation occurs provided that, for each event \( (X = x_i) \) in a partition,

the set of conditional probabilities for an event \( E \), given \( x_i \), properly include

the unconditional probabilities for \( E \).

Dilation of conditional probabilities is the opposite phenomenon to the more familiar “shrinking” of sets of opinions with increasing shared evidence.
Heuristic Example of Dilation

Suppose \( A \) is a highly uncertain event in the added sense of “uncertainty” that comes with a set of probabilities \( \mathcal{P} \).

That is

\[
P^*(A) - P^*(\bar{A}) \approx 1.
\]

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, \( E = \{(A,H), (A^c,T)\} \).

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

\[
0 \approx P^*(E \mid T) < P^*(E) = P^*(E) < P^*(E \mid T) \approx 1.
\]
Thus, regardless how the coin lands, $X$, conditional probability for the event $E$ dilates to a large interval, from a determinate value .5.

This example mimics Ellsberg’s (1961) *paradox*, where the mixture of two *uncertain* events has a determinate probability.

The law of conditional expectations fails for lower expectations in an IP setting:

$$0.5 = P_\star(E) > E[P_\star(E \mid X) \approx 0$$

So, with $\Gamma$-Maximin and a non-trivial IP model, the information from the experiment \{H,T\} dilates the Imprecise Probabilities for $E$, and the experiment has negative value in deciding a bet on $E$.  

Is Ignorance Bliss with Sets of Probabilities?  Info-Metrics Inst. April 2013
Consider a choice between

Two terminal options:

Terminal option $d_1$ — Lose, -$1.00$ if $E$ and Win, +$1.00$ if $E^c$.

Terminal option $d_2$ — Pay a fine of -$0.50$.

And one Sequential option:

*Sequential* – Observe the outcome $X = \{H, T\}$, then choose between $d_1$ and $d_2$.

Because $X$ dilates the conditional probability for $E$,

$$P_\ast(E \mid H) = P_\ast(E \mid T) \approx 0,$$

and the lower (conditional) expected value of $d_1$ is approximately -$1.00$.

Then, in the *Sequential* option, $\Gamma$-Maximin requires choosing $d_2$, regardless how the coin lands, with a sure loss of (only) -$0.50$.

Terminal decision $d_1$ has (ex-ante) value $0$, which is strictly better than $d_2$.

Hence, using $\Gamma$-Maximin, in this sequential decision, the experiment $X$ has a *negative value*, -$0.50$.

You’d even pay, e.g. $0.25$, to choose between $d_1$ and $d_2$ without first learning $X$. 
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 × 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:

\[
\begin{align*}
P_\ast (E \mid H) &< P_\ast (E) = .5 = P^\ast (E) < P^\ast (E \mid H) \\
P_\ast (E \mid T) &< P_\ast (E) = .5 = P^\ast (E) < P^\ast (E \mid T).
\end{align*}
\]

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 $2 \times 2$ tables.

\[
\begin{array}{cc}
  & B_1 & B_2 \\
A_1 & p_1 & p_2 \\
A_2 & p_3 & p_4 \\
\end{array}
\]

Define the quantity

\[S_P(A_1,B_1) = \frac{p_1}{(p_1+p_2)(p_1+p_3)} = \frac{P(A_1,B_1)}{P(A_1)P(B_1)}.\]

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

**Lemma:** If $P$ displays dilation in this sub-algebra, then

\[\inf_P \{S_P(A_1,B_1)\} < 1 < \sup_P \{S_P(A_1,B_1)\}.\]

**Theorem:** If $P$ displays dilation in this sub-algebra, then there exists $P^\# \in P$ such that $S_P^#(A_1,B_1) = 1$. 
Dilation and the $\varepsilon$-contaminated model.

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

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

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 $\varepsilon$-contaminated model, dilation occurs in an algebra $\mathcal{A}$ iff it occurs in some $2 \times 2$ subalgebra of $\mathcal{A}$.

So, the next result is formulated for $2 \times 2$ tables.
\( P_\varepsilon(P) \) experiences dilation \emph{if and only if}

\begin{align*}
\text{case 1: } & \quad S_P(A_1,B_1) > 1 \\
& \quad \varepsilon > [S_P(A_1,B_1) - 1] \times \max \{ P(A_1)/P(A_2) ; \ P(B_1)/P(B_2) \} \\
\text{case 2: } & \quad S_P(A_1,B_1) < 1 \\
& \quad \varepsilon > [1 - S_P(A_1,B_1)] \times \max \{ 1 ; \ P(A_1)P(B_1)/P(A_2)P(B_2) \} \\
\text{and case 3: } & \quad S_P(A_1,B_1) = 1 \\
& \quad P \text{ is internal to the simplex of all distributions.}
\end{align*}

Thus, dilation occurs in the \( \varepsilon \)-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.
• *An aside on another IP-theory*

Consider, Dempster-Shafer Belief Theory with Dempster’s update rule, rather than updating an IP set by Bayes-conditioning of each of its elements.

However, the $\varepsilon$-contamination model is a Dempster-Shafer Belief model and there updating by Bayes’ conditioning is equivalent to updating by Dempster’s rule.

Hence, the conditions for Dilation in the $\varepsilon$-contamination model are exactly the same for Dempster-Shafer theory.
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 permutation of the atoms.

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

---

**Summary**

- The Basic SEU Result – that cost-free information has non-negative value – does **not** extend to IP-decision theory with, e.g., Γ-Maximin as the decision rule.
- Dilation of sets of probabilities is a recipe for creating cost-free data that have negative value.
Selected References


