POSTERIOR PROBABILITIES OF THE INDEPENDENCE AXIOM WITH NON-EXPERIMENTAL DATA

Robert C. Marshall
Department of Economics, Duke University

Jean-Francois Richard
ISDS, Duke University

Gary A. Zarkin
Research Triangle Institute
DP #89-09

June 1989, Revised September 1989

Abstract

This paper addresses the issues associated with the construction of posterior probabilities for the independence axiom of expected utility from non-experimental data. To illustrate the methodology of analyzing non-experimental evidence we consider seat belt usage data. We find a posterior probability close to one of an Allais type paradox in this data set. In addition, the data is consistent with Machina's (1982) Hypothesis H but inconsistent with the "light" hypothesis of Chew and Waller (1986).

Acknowledgements: Marshall and Richard acknowledge support for this research from the National Science Foundation (SES-8708615) and the Pew Charitable Trusts. We are indebted to Phil Cook for helpful comments and to Leslie A. McFarland for skillful research assistance.
No analytic device has yielded greater insight into economic behavior under uncertainty than the expected utility paradigm of Von Neumann and Morgenstern (1944). Nevertheless, significant experimental evidence has developed over the past three decades which suggests that individuals often make choices that are inconsistent with the predictions of expected utility. The Allais paradox (1952) has come to be associated with an entire class of violations of expected utility. In fact, the choices which characterize these paradoxes can be shown to violate a particular axiom that serves as the underpinning for expected utility – the independence axiom.¹

The implications of the independence axiom and the nature of the paradoxes which constitute violations are most easily understood in a three event scenario where individuals select a lottery among each of two pairwise choices.² Within this context the issues and problems can be represented graphically in a simple two-dimensional diagram on the unit simplex where the worst event is graphed on the horizontal axis and the best event on the vertical axis, as in Figure 1. The independence axiom implies that indifference curves must be linear and parallel in this space. Consequently, when confronted with a choice between lottery W or X and then lottery Y or Z the only pairs of choices consistent with expected utility are W and Y or, alternatively, X and Z. Allais observed that the modal choice of individuals when confronting these choices was W and Z – choices that are inconsistent with the independence axiom since there exist no linear and parallel indifference curves that could lead to such selections.³

The violations of the independence axiom which have been found to date, such as the Allais paradox, are based entirely upon experimental evidence. Typically, subjects (almost always students) are asked to make choices between lotteries where the outcomes involve monetary awards. The main goal of this paper is to confront the independence axiom with non-experimental data – data which derives from the decisions of individuals in the marketplace.

We propose a Bayesian methodology for investigating the validity of the independence axiom of expected utility when non-experimental data is employed. The techniques are applied to 1983 data on time-of-day seat belt usage.⁴ The behavioral anomaly is that there are more drivers wearing belts in the morning than in the evening despite higher accident probabilities

¹ The Ellsberg paradox represents a different kind of violation which apparently stems from individuals' disaste for "ambiguity" in probabilities.
² Machina (1983a, 1983b) has classified known violations into three categories — common consequence, common ratio, and over (under) sensitivity to small probability events. The Allais paradox is a common consequence violation.
³ The Allais paradox involves four lotteries over three events. The events are payoffs of $0, $1, and $5. The two lottery pairs are (W: 0, 1, 0), (X: .01, .99, .00) and (Y: .89, .11, .00), (Z: .9, .0, .10), respectively.
⁴ Mandatory seat belt laws did not yet exist at this time.
in the latter period. We find that the posterior probabilities are largely inconsistent with the independence axiom.

The experimentally generated paradoxes have lead to theoretical efforts to propose alternatives to expected utility. Machina's (1982) generalized expected utility relaxes global linearity and requires only that preferences be "smooth". In terms of the two-dimensional diagram, Machina's indifference curves are upward sloping, continuous, and everywhere differentiable in the interior of the simplex. Furthermore, Machina's preferences are consistent with nearly all known experimental paradoxes if his indifference curves "fan out". The fanning out property has been formally conjectured by Machina in what is known as "Hypothesis II" -- if one lottery first order stochastically dominates (FOSD) another then an individual exhibits a higher degree of absolute risk aversion with the former.

In this paper we propose Bayesian methods for investigating Hypothesis II with non-experimental data. The aforementioned seat belt data is found to be consistent with the stochastic dominance condition of Hypothesis II and the observed choices of individuals are consistent with Hypothesis II with high posterior probability.

In addition to Machina's work, Chew and MacCrimmon (1979) and Fishburn (1983) have proposed a more restrictive generalization of expected utility called alpha-nu utility or weighted utility. Weighted utility employs a weaker form of the independence axiom which implies that indifference curves in Figure 1 would be straight lines that are not necessarily parallel. Non-parallel indifference curves have a very specific property -- they emanate from a single hub located in either the third quadrant -- implying fanning-out -- or the first quadrant northeast of the point (1,1) -- implying fanning-in. Chew and Waller (1986) refer to the fanning-out property of the indifference curves for weighted utility as the "light" hypothesis.\(^5\)

Along with Machina's Hypothesis II, the "light" hypothesis of weighted utility is investigated with our non-experimental seat belt usage data. The posterior probabilities for the "light" hypothesis are extremely low -- with few exceptions the hub is located in the second quadrant.

The paper proceeds as follows. In Section I a number of issues associated with the relevant hypotheses and the use of non-experimental data to investigate them are discussed. In Section

---

\(^5\) Machina (1983b) and Fishburn (1984) provide excellent surveys of the alternative axioms that have been proposed to replace the independence axiom.

\(^6\) The word "light" refers to the weights, relative to those which generate parallel indifference curves, that produce "fanned-out" indifference curves.
II we define the lotteries, events, and probabilities of interest for the seat belt usage example. In Section III the main results from our Bayesian inference approach are discussed. Sensitivity analysis is also presented to demonstrate the robustness of our findings to the selection of priors. Extensions of this line of research are discussed in Section IV along with a discussion of both the limitations and advantages of bringing non-experimental data to bear on the validity of the independence axiom and substitute hypotheses.

I. The Relevant Hypotheses and Bayesian Inference

Although generalization to more than three events is not problematic, we restrict attention to lotteries over three events for two reasons. First, all hypotheses and cases of interest can be portrayed graphically in two dimensions. Second, our empirical work is based upon lotteries over three events.

Consider the two pairs of lotteries discussed in the introduction as portrayed now in Figure 2. We will refer to the segments connecting lotteries W and X as WX, connecting Y and Z as YZ, etc. It is assumed that these segments are positively sloped (otherwise, the choice between, say, W and X is uninteresting since one lottery FOSD's the other). All subsequent discussion in this Section focuses on Figure 2.

If an individual chooses X over W and Z over Y or, alternatively, W over X and Y over Z then there exist linear parallel indifference curves which are consistent with these selections. Consequently, the only cases that can lead to a refutation of the independence axiom are the choices X over W and Y over Z (denoted as X&Y) or, alternatively, W over X and Z over Y (denoted as W&Z). If the slope of WX is not less than the slope of YZ then the selection W&Z is inconsistent with the independence axiom. In addition, if the slope of WX is not greater than the slope of YZ then the selection X&Y is inconsistent with the independence axiom.

Machina’s Hypothesis II has a precondition that both WY and XZ have non-positive slopes. In other words, W must FOSD Y and X must FOSD Z. If this condition holds and if the slope of WX is greater then the slope of YZ then the only selection incompatible with Machina’s Hypothesis II is X&Y. In particular and unlike the independence axiom, W&Z are compatible with Machina’s Hypothesis II that the indifference curves "fan-out".

The "light" hypothesis also implies that indifference curves fan-out but the indifference curves are straight lines that emanate from a single hub (the limiting case being parallel indifference
curves). If the slope of WX is greater than the slope of YZ then the only selection incompatible with the "light" hypothesis is X&Y. Again, W&Z are compatible. However, one further condition must be checked to corroborate the "light" hypothesis which involves the location of the hub. For the "light" hypothesis the hub must be located in the third quadrant. Otherwise, indifference curves can be negatively sloped implying that an individual is indifferent between two lotteries where one FOSD's another.

We can now state the relevant hypotheses in terms of our lotteries W, X, Y, and Z assuming that both WX and YZ have positive slopes. \( H_{1a} \), \( H_{m2} \), and \( H_l \) denote the hypotheses corresponding to the independence axiom, Machina's Hypothesis II, and the "light" hypothesis, respectively.  

1. \( H_{1a} \): If the slope of WX is greater than (less than) the slope of YZ then an individual will not select W&Z (X&Y). If WX and YZ are parallel then neither W&Z nor X&Y will be selected.

2. \( H_{m2} \): Provided that the slopes of WY and XZ are non-positive and if the slope of WX is not less than the slope of YZ then an individual will not select X&Y.

3. \( H_l \): If the slope of WX is not less than the slope of YZ then an individual will not select X&Y and, furthermore, the lines coincident with WX and YZ must intersect in the third quadrant.

Each of the above hypotheses has been tested with experimental data. Specifically, individuals have been confronted with two lottery pairs and modal responses have been computed and reported. A key advantage of the experimental approach is that the lottery probabilities are determined by the experimental design.

On the other hand, with non-experimental data the event probabilities must be estimated. The estimation of the basic event probabilities raises an entire set of issues which are not relevant to experimental studies. Specifically, a measure of reliability must be attached to any conclusion reached regarding the validity of a hypothesis. A classical approach requires construction of test statistics for each hypothesis and evaluation of their sampling distributions under these hypotheses and also, whenever possible, under (local) alternatives. However, this approach can raise a number of problems.

1. The data collection process often includes a complex calibration of the raw data (i.e. national weighting schemes) and can not be "objectively" characterized in full.

---

7 If the observed choice is W&Z as above, then the FOSD conditions are sufficient for compatibility with \( H_{m2} \).
2. The transformation from the “baseline” probabilities (i.e. those on which the data bear direct evidence) to the lottery probabilities is non-linear and the derivation of sampling distributions for the lottery probabilities would, therefore, typically require the use of large sample approximations. However, if certain events occur with very low probability (such as death in an auto accident when wearing a seat belt), and therefore sample evidence is very limited, such approximations are quite inappropriate.

3. None of the hypotheses of interest are point hypotheses. They are all composite hypotheses that are characterized by sets of inequalities on the lottery probabilities. The construction of appropriate test statistics for such hypotheses is non-trivial and raises a number of conceptual problems.

Consequently, we adopt instead a Bayesian approach which permits a mix of “objective” and “subjective” evidence and allows more flexibility in the characterization of the sampling process.

II. Seat-Belt Usage: Events, Lotteries, “Baseline” Probabilities, and Data Sources

The first issue addressed in this section concerns the event space and the lotteries. We then define the probabilities of interest. Finally, the data sources are identified.

a. The Events

For any given automobile trip there is a risk of an accident. In terms of physical injury to the driver an accident can result in a continuum of impact-related maladies ranging from no injury to death. Law enforcement officers use the Abbreviated Injury Scale (AIS) classification system to provide a simple scalar measure of an accident’s severity in terms of injury to the driver. In our opinion the detailed disaggregation is somewhat ambiguous and probably subject to non-negligible classification errors by the reporting agents. Consequently, we restrict ourselves to three events:

\[ A_0: \text{no accident} \]
\[ A_1: \text{non-critical accident (AIS codes 0-4)} \]
\[ A_2: \text{critical or fatal accident (AIS codes 5-6)} \]

---

The seven possible AIS codings are defined as: no injured (AIS0), "minor" injury (AIS1), "moderate" injury (AIS2), "severe" injury (AIS3), "serious" injury (AIS4), "critical" injury (AIS5), and fatal injury (AIS6).
b. The Lotteries

In the era prior to mandatory seat belt laws survey information indicated that individuals who wore seat belts on long distance trips often did not wear seat belts on local trips. This behavior appeared paradoxical since accidents of all severities are more common for local trips than for long distance trips. The two lottery pairs of interest in this context are (i) whether or not to wear a seat belt on a local trip and (ii) whether or not to wear a seat belt on a long distance trip. Two major problems exist with this example. First, the data on trip length is very poor. Second, and more importantly, the costs of "buckling-up" per mile driven are much smaller for long distance trips than for local trips. Discovery of an Allais type paradox in the data would be subject to the criticism that the expected payoffs of the three basic events were not invariant to an individual's trip length. In other words, if the cost per mile of buckling-up were included explicitly in the definition of the events it is possible that an Allais type paradox would not be found. These two problems have redirected our attention toward commuter seat belt usage.

The cost per mile of buckling-up is equal for trips of equal length. The most common trips of equal length are commutes. Surprisingly, seat belt usage is lower for the late afternoon or early evening commute than for the morning commute despite the fact that accident probabilities are higher in the evening. Furthermore, relatively good data exists for inference purposes. Therefore, the two lottery pairs on which we focus for this study are:

- Lottery X: Unbelted (B₀) for the Morning Commute (T₀)
- Lottery W: Belted (B₁) for the Morning Commute

and

- Lottery Z: Unbelted for the Evening Commute (T₁)
- Lottery Y: Belted for the Evening Commute

c. The Probabilities of Interest

In order to examine the validity of the independence axiom and related hypotheses the eight basic probabilities (for drivers only) below must be estimated

\[ P_{i,j,k} = Pr \left( A_i \mid B_j, T_k \right) \quad i=1,2 ; j,k=0,1 \]  

\[ 9 \] See Waller and Barry (1969). These authors noted a significant difference between reported and observed usage. In fact, Robertson et. al. (1972) reported "no significant difference in the percentage of visually observed use in and out of town" with out of town use slightly higher than in town use (27% versus 24%).
Our data does not permit evaluation of these probabilities. Hence, our first task is to relate the $p_{i,j,k}$'s to a coherent set of probabilities for which observations are available. Note first that since the $A_i$'s are subsets of the space of accidents, $\Lambda$, the following identity holds

$$Pr(A_i, B_j, T_k, A) = Pr(A_i, B_j, T_k) \quad . \quad (2)$$

This identity validates the use of the following factorization which constitutes the cornerstone of our analysis

$$p_{i,j,k} = \frac{p_{i,j,k,a} \cdot p_{j,k|a} \cdot p_a}{p_{j,k} \cdot p_k} \quad i=1,2; \quad j, k = 0,1 \quad . \quad (3)$$

The thirteen "baseline" probabilities in the right hand term (rht) are the following.

- $p_a = Pr(A)$ denotes the probability per trip that a driver will have an accident;
- $p_k = Pr(T_k)$ denotes the probability of a vehicle trip in the morning period (7–10 AM) for $k=0$ or in the afternoon/evening period (4–7 PM) for $k=1$;
- $p_{1|k} = Pr(B_1 \mid T_k)$ denotes the probability that a driver will wear a seat belt given a trip in period $k$;
- $p_{j|k|a} = Pr(B_j, T_k \mid A)$ denotes the probability that a driver was driving in period $k$ and either wearing ($j=1$) or not wearing ($j=0$) a seat belt given that he had an accident;
- $p_{2|j,k,a} = Pr(A_2 \mid B_j, T_k, A)$ denotes the probability that a driver suffers a "critical" injury given that he was involved in an accident in period $k$ with seat belt status $j$. Also, $p_{1|j,k,a} = 1 - p_{2|j,k,a}$.

Data exist which allow "direct" evaluation of these thirteen probabilities and, therefore, inference on the eight lottery probabilities of interest via equation (3).

d. The Data Sources

We analyze seat belt usage in 1983. At that time there were no states with mandatory seat belt laws so usage was entirely a matter of unconstrained individual choice. As already alluded

\[ \text{10} \text{ Since } B_0 \text{ and } B_1 \text{ are mutually exclusive and complementary we also have } p_{0|k} = Pr(B_0 \mid T_k) = 1 - p_{1|k}. \]

\[ \text{11} \text{ Note that the four events } B_j \times T_k \text{ are not mutually exclusive since } A \text{ includes accidents at all periods of the day.} \]
to, the morning period of observation runs from 7 to 10 AM and the afternoon period from 4 to 7 PM. Four main data sources were employed to construct the necessary probabilities. They are referred to subsequently by the acronyms below.


III. Seat Belt Usage by Commuters: Bayesian Inference

Posterior distributions are constructed for the thirteen baseline probabilities. As noted in Section I, these distributions are based upon a mix of “objective” and “subjective” evidence depending on the quality of the available sample evidence and on the characterization of the sampling process. Monte Carlo simulation procedures are employed to derive estimates of the exact posterior means and standard deviations of the probabilities of interest — summarized in Table 1 for ease of reference — and of the exact posterior probabilities for each hypothesis.\(^{12}\)

\(^{12}\) Contains summary and descriptive information for the data set.

\(^{13}\) More specifically, assume we draw from the posterior distribution \(N\) independent realizations of the vector consisting of the thirteen baseline probabilities. For any given hypothesis \(H\) — however complicated its characterization — we can easily count the number \(N_H\) of “successes”, i.e. of validations of \(H\) by an individual drawing. The ratio \(N_H/N\) yields an estimate of the posterior probability \(p_H\) attached to \(H\). The standard deviation of the estimate is given by \(\sigma_H = \left(\frac{1}{N_H} \cdot PH \cdot (1 - PH)\right)^{1/2}\). We can select \(N\) in such a way that the ratio \(\sigma_H/p_H\) — or its estimate — be as small as desirable, notwithstanding numerical rounding errors. In the context of our application where some of the key \(p_H\)'s are of the order of .9 by selecting \(N=10^9\) we obtain a ratio approximately equal to \(10^{-1}\).
The probability $p_A$ is defined by the ratio $N_A/N_V$ and its posterior distribution is that of a ratio of Normally distributed random variables.\(^{14}\)

\textit{b. Probabilities of Travel by Time of Day}

The NPTS includes a table for the “Distribution of Vehicle Trips for Work and All Other Trip Purposes by Time of Day” (Table E-50, page E-52). From a total of 126,871 million vehicle trips, 21.78% of them or 27,631 million are taken in the afternoon (4-7 PM). No standard deviation is reported for the latter number. The sample consists of 6,348 households who were asked to report all trips during a specific 24 hour travel day. The average number of day trips reported by these households is 4.07. Consequently, as a first approximation, we assign to $p_1=Pr(T_1)$ a beta posterior distribution corresponding to a non-informative prior and a sample of 5,709 (= .2178 x 26,213) successes out of 26,213 (= 4.07 x 6,438) trials.\(^{15}\)

The determination of a posterior distribution for $p_0=Pr(T_0)$ raises additional problems. In accordance with the NHTSA study our morning period of reference is 7–10 AM. The NPTS uses instead the periods 6–9AM and 9AM-1PM (together with other 3 or 4 hour periods). NPTS also draws a distinction between “home to work” trips and trips for “other purposes”. These two categories have very different time patterns across the day. The morning “home to work” trips are largely concentrated in the 6–9AM period for a total of 11.6 million. We take as a working assumption that the total over the 7–10AM period is slightly lower, approximately 10.0 million, with a substantial margin of error. Except for the beginning of the morning period the “other purposes” trips seem to be spread quite uniformly over the day maintaining a rate of 6.5 million per hour. A rough estimate for the period 7–10AM is 13.0 million. Overall, we consider 23.0 million trips with a standard deviation of 2.0 million to constitute a reasonable basis for the selection of a posterior density for $p_0$.

Clearly $p_0$ and $p_1$ should not be treated as independent random variables. In particular, our data suggest that $p_0$ is less than $p_1$ with high probability. An obvious way of handling positive dependence between $p_0$ and $p_1$ consists of reasoning in terms of a marginal probability for $p_1$ and a conditional probability for $p_0$ given $p_1$, with an upward sloping regression function. It proves convenient to assume that the conditional posterior distribution for $p_0$ | $p_1$ is uniform on

\(^{14}\) Theoretically, the distributions of $N_A$ and $N_V$ ought to be restricted to the positive real line. In practice, however, truncation is irrelevant since the posterior means are approximately 80 standard deviations from the origin.

\(^{15}\) Random drawings from such a beta distribution are obtained via the transformation of random variables $p_1 = X_1 \cdot (X_1 + X_2)^{-1}$ where $X_1$ and $X_2$ are draws from standardized gamma distributions with parameters 5,709 and 26,213, respectively.
the interval (.7p_1, p_1). As shown in Table 1, the corresponding marginal posterior moments for p_1 then take the intended values and our specification induces a moderate positive correlation between p_0 and p_1 (approximately .11).\textsuperscript{16}

Note that we would have preferred to limit our attention to "home to work" trips on the grounds that we would then be considering the same drivers in the morning and evening periods. Unfortunately our data on seat belt usage do not discriminate between trip purposes. We believe, however, that the addition of the trips for "other purposes" during the periods of interest consist of a fairly homogeneous population (school commuters, shoppers, etc.) and, hence, is most unlikely to severely bias our conclusions.

c. Probability of Wearing a Seat Belt by Period of the Day

The NHTSA survey consists of a total of 146,305 observations on drivers at designated intersections and freeway exit locations. The driver seat belt usages for the two periods of interest are\textsuperscript{17}:

7–10AM: 15.4% of 30,013 observations.

4–7PM: 13.9% of 22,944 observations.

Under the working assumption that the sampling process consists of a pair of (independent) binomial processes and under the non-informative beta prior densities, the posterior distributions for p_{1|0} = Pr(B_1 | T_0) and p_{1|1} = Pr(B_1 | T_1) are independent beta distributions. They are very closely approximated by Normal distributions and, therefore, we use independent Normal posterior distributions. Their moments are reported in Table 1.\textsuperscript{18}

d. Probability of Driver Belt Use by Time of Day Conditional on an Accident

The data are from the 1983 NASS survey and they consist of the time of each accident, seat belt use by drivers, and the injuries sustained by drivers. The raw data of interest consists of 14,554 accident observations which can be regrouped as follows:

\textsuperscript{16} Our assessment of the posterior distribution for (p_0, p_1) is on the conservative side and, as discussed later, our conclusions are quite robust against substantial variations in the specification of this distribution.

\textsuperscript{17} For the entire day 14.0% of drivers from a total of 146,305 were belted.

\textsuperscript{18} An additional reason for considering the Normal distribution is that we can easily allow for correlation between p_0 and p_1 in the subsequent sensitivity analysis.
\[ B_0, T_0: 1,881 \]
\[ B_1, T_0: 305 \]
\[ B_0, T_1: 3,139 \]
\[ B_1, T_1: 509 \]
\[ \text{Other times of the day: 8,720} \]

The NASS weights are then used to transform the data into national estimates. A complete analysis of the process whereby the national estimates are generated and a precise quantification of all possible sources of error goes beyond the objectives of our analysis and, furthermore, our subsequent sensitivity analysis indicates it is largely irrelevant anyway. Hence, we simply scaled down the national estimates in such a way that their total coincides with the "effective" sample size of 14,554 which is largely the relevant sample size for the purpose of constructing the posterior distributions and standard deviations. The corresponding figures which serve as the basis for our inference are as follows:

\[ B_0, T_0: 1,678 \]
\[ B_1, T_0: 451 \]
\[ B_0, T_1: 3,362 \]
\[ B_1, T_1: 604 \]
\[ \text{Other times of the day: 8,459} \]

We assume that the sampling process can be approximated by a multinomial process. Under a non-informative Dirichlet prior the posterior distribution for the \( p_{j,k} \)'s is Dirichlet and its parameters are given by the sample figures above. The corresponding posterior means and standard deviations are regrouped in Table 1.

e. **Conditional Probabilities of an Accident by Severity**

The NASS data also contains a (subjective) classification of the severity of driver injuries. The raw data are as follows.

<table>
<thead>
<tr>
<th></th>
<th>( B_0, T_0 )</th>
<th>( B_0, T_1 )</th>
<th>( B_1, T_0 )</th>
<th>( B_1, T_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>1,870</td>
<td>3,118</td>
<td>305</td>
<td>508</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>11</td>
<td>21</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>1,881</td>
<td>3,139</td>
<td>305</td>
<td>509</td>
</tr>
</tbody>
</table>

11
As in Section III.d the weighted NASS figures are used and then scaled down in order to keep track of effective sample sizes. The national average figures for $A_2$-type accidents are substantially lower than the raw figures above and the rescaling generates non-integer numbers. Since beta distributions can accommodate non-integer parameters we use them to preserve the (weak) informational content of the sample.\textsuperscript{19} The figures are:

<table>
<thead>
<tr>
<th></th>
<th>$B_0,T_0$</th>
<th>$B_0,T_1$</th>
<th>$B_1,T_0$</th>
<th>$B_1,T_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>1,675.35</td>
<td>3,354.25</td>
<td>451</td>
<td>603.84</td>
</tr>
<tr>
<td>$A_2$</td>
<td>2.65</td>
<td>7.75</td>
<td>0</td>
<td>0.16</td>
</tr>
<tr>
<td>Total</td>
<td>1,678</td>
<td>3,362</td>
<td>451</td>
<td>604</td>
</tr>
</tbody>
</table>

Except for the case $(B_1,T_0)$ these figures are used to characterize independent beta posterior distributions derived with non-informative beta priors. Assigning a zero posterior probability distribution to $A_2$ conditionally on $(B_1,T_0)$ would only strengthen our conclusion. Hence, we adopt a conservative approach and assume that conditionally on $p_{1,1_0} = Pr(B_1,T_1 \mid A)$ the probability $p_{1,0_0} = Pr(B_1,T_0 \mid A)$ is uniformly distributed on the range $(0, p_{1,1_0})$. In doing so we introduce a positive correlation of about .85 between these two probabilities which is certainly quite sensible. Alternative specifications are considered in the sensitivity analysis.

\textbf{f. Results under the Baseline Posterior Distribution}

The posterior means and standard deviations for the thirteen baseline probabilities under the posterior distributions described above are presented in Table 1. The corresponding moments for the eight lottery probabilities of interest are found in Table 2. Figure 3 illustrates the lotteries when evaluated at their posterior means.\textsuperscript{20} Posterior probabilities for the hypotheses of interest are presented in Table 3. The estimates are based upon 100,000 drawings and are correct up to the digits which are reported in our tables.

The three most interesting results are:

1. The posterior probability against the independence axiom is .936 which strongly supports the inconsistency of the choices as depicted in Figure 3.

\textsuperscript{19} The use of integer parameters requires multiplying effective sample sizes by a factor of six and dividing standard deviations by the same amount. This need not be unreasonable since the weighting procedure used by NASS obviously relies upon additional information that just that contained in their report. We nevertheless opt for the less informative version on the grounds that results under the latter would be even more precise under tighter posterior distributions.

\textsuperscript{20} We presume that exogenous factors cause the segments WX and YZ to be positively sloped.
2. There is a posterior probability of .896 that the choices are consistent with Machina’s Hypothesis II.

3. The approximate zero probability for the “light” hypothesis stems entirely from the implied location of the hub — it does not lie in the third quadrant.

<table>
<thead>
<tr>
<th>Pa</th>
<th>$0.242 \times 10^{-3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{Pa}$</td>
<td>(0.005)</td>
</tr>
</tbody>
</table>

| $P_0$ | 0.185 |
| $\sigma_{P_0}$ | (0.019) |

| $P_1$ | 0.218 |
| $\sigma_{P_1}$ | (0.003) |

| $P_{0,0a}$ | 0.115 |
| $\sigma_{P_{0,0a}}$ | (0.003) |

| $P_{0,1a}$ | 0.231 |
| $\sigma_{P_{0,1a}}$ | (0.004) |

| $P_{1,0a}$ | $0.310 \times 10^{-1}$ |
| $\sigma_{P_{1,0a}}$ | (0.014) |

| $P_{1,1a}$ | $0.415 \times 10^{-1}$ |
| $\sigma_{P_{1,1a}}$ | (0.017) |

| $P_{10}$ | 0.154 |
| $\sigma_{P_{10}}$ | (0.002) |

| $P_{11}$ | 0.139 |
| $\sigma_{P_{11}}$ | (0.002) |

| $P_{20,0,a}$ | $0.158 \times 10^{-2}$ |
| $\sigma_{P_{20,0,a}}$ | (0.096) |

| $P_{20,1,a}$ | $0.231 \times 10^{-2}$ |
| $\sigma_{P_{20,1,a}}$ | (0.083) |

| $P_{21,0,1}$ | $0.138 \times 10^{-3}$ |
| $\sigma_{P_{21,0,1}}$ | (0.040) |

| $P_{21,1,a}$ | $0.275 \times 10^{-3}$ |
| $\sigma_{P_{21,1,a}}$ | (0.067) |

**Note:** These figures are estimates based on 100,000 Monte Carlo drawings from the posteriors. They coincide — up to the digits reported — with the exact moments which are known analytically.
Table 2: Posterior Mean and Standard Deviations for Lottery Probabilities

<table>
<thead>
<tr>
<th>Lotteries</th>
<th>$A_1$</th>
<th>$A_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W: B_1, T_0$</td>
<td>$0.266 \times 10^{-3}$</td>
<td>$0.037 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.106)</td>
</tr>
<tr>
<td>$X: B_0, T_0$</td>
<td>$0.180 \times 10^{-3}$</td>
<td>$0.285 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.177)</td>
</tr>
<tr>
<td>$Y: B_1, T_1$</td>
<td>$0.332 \times 10^{-3}$</td>
<td>$0.092 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.225)</td>
</tr>
<tr>
<td>$Z: B_0, T_1$</td>
<td>$0.298 \times 10^{-3}$</td>
<td>$0.687 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.246)</td>
</tr>
</tbody>
</table>

Note: Estimates based upon 100,000 Monte Carlo drawings.

Table 3: Posterior Probabilities for Hypotheses of Interest

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Pr(H_{1a})$</td>
<td>0.064</td>
</tr>
<tr>
<td>$\Pr(H_{m2})$</td>
<td>0.896</td>
</tr>
<tr>
<td>$\Pr(H_0)$</td>
<td>-0.000</td>
</tr>
</tbody>
</table>


g. Sensitivity Analysis

It is crucial that the sensitivity of the results to modifications of the prior and/or posterior distributions be analyzed since some of the features of the posterior distributions are based upon thin sample evidence. In order to understand the sensitivity of our results to assumptions that are implicit or explicit in the construction of the posterior distributions the simulations are rerun under a number of variants. These changes affect posterior standard deviations (not reported) and, more importantly, the posterior probabilities of the hypotheses of interest. Only the latter are reported. We report in Table 4 changes in the posterior probabilities of the hypotheses for the following eleven variations:

Row 1. The standard deviation of $N_A$ (which is the numerator of $p_A$) is multiplied by 4, resulting in a threefold increase of the standard deviation of $p_A$.

Row 2. The distribution of $p_0 = Pr(T_0)$ is made tighter. We now assume that the distribution of $p_0 | p_1$ is uniform on the range $(.85p_1, p_1)$ instead of $(.7p_1, p_1)$. 

14
Row 3. The standard deviation of $p_1 = Pr(T_1)$ is multiplied by two.\footnote{Multiplying the standard deviation of a beta (or Dirichlet) distribution by a factor $\lambda$ only requires dividing its parameters by $\lambda^2$ and leaves the means unaffected.}

Row 4. The standard deviations of $p_{1|0} = Pr(B_1 | T_0)$ and $p_{1|1} = Pr(B_1 | T_1)$ are both multiplied by two.

Row 5. The standard deviations of $p_{j,k|a} = Pr(B_j, T_k | A)$ are both multiplied by two.

Row 6. The standard deviations of $p_{2|0,k,a} = Pr(A_2 | B_0, T_k , A)$ are both multiplied by two. In contrast with the previous change we note a bigger impact on both $Pr(H_{ia})$ and $Pr(H_{m2})$. The following alternative scenario yields some insight into this finding.

Row 6a. The number of $A_2$ accidents conditionally on $B_0$ are modified as follows: 11 and 21 are replaced by 13 and 18. As a result, the posterior means of the corresponding probabilities are drawn closer together — 0.00187 and 0.00198 instead of 0.00158 and 0.00231 — but their standard deviations are not affected in any major way.

Row 7. The range of the uniform conditional distribution for $Pr(A_2 | B_1, T_0, A)$ given $Pr(A_2 | B_1, T_1, A)$ is doubled causing the posterior means of these two probabilities to be equal.

Row 8. The standard deviation of $Pr(A_2 | B_1, T_1, A)$ is doubled.

Row 9. All changes from 1 through 8, excluding 6a, are jointly implemented.

We finally consider two alternative scenarios which incorporate additional variations.

Row 10. The posterior distribution of $p_0$ and $p_1$ is derived under a non-informative prior. Consider now the scenario where a statistician wants to incorporate the informative prior that, in his opinion, $p_0$ and $p_1$ are equal. More specifically, assume he sets both prior means at 0.140, the survey fraction for all periods of the day, both standard deviations at 0.0020 (an order of magnitude comparable to sample information) and the prior correlation to .9 (so that the prior standard deviation of the difference $p_0 - p_1$ is equal to 0.00087). Multiplying the baseline posterior distribution by this informative prior generates an overall posterior
with means 0.145 and 0.135, standard deviations both equal to 0.0013, and correlation equal to -0.78. The results in row 10 are derived under this “overall” posterior distribution.

Row 11. We include AIS severity code 4 in the definition of an $A_2$-type accident in order to test the robustness of our conclusion against potential ambiguities in the severity classification. The last set of figures from Section III.e are replaced by those below.

<table>
<thead>
<tr>
<th></th>
<th>$B_0,T_0$</th>
<th>$B_0,T_1$</th>
<th>$B_1,T_0$</th>
<th>$B_1,T_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>1,674.40</td>
<td>3,347.40</td>
<td>450.87</td>
<td>603.84</td>
</tr>
<tr>
<td>$A_2$</td>
<td>3.60</td>
<td>14.60</td>
<td>0.13</td>
<td>0.16</td>
</tr>
<tr>
<td>Total</td>
<td>1,678</td>
<td>3,362</td>
<td>451</td>
<td>604</td>
</tr>
</tbody>
</table>

The sensitivity analysis produces three main conclusions. First, our earlier evidence against the independence axiom is very robust to a number of variations. Second, our evidence in support of Machina’s Hypothesis II is less robust to variations in assumptions regarding the posterior distributions. Finally, the “light” hypothesis fails under all variations because of the invalid location of the hub.

Table 4: Sensitivity Analysis

<table>
<thead>
<tr>
<th>Variation</th>
<th>Pr($H_{la}$)</th>
<th>Pr($H_{m2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>0.064</td>
<td>0.896</td>
</tr>
<tr>
<td>1</td>
<td>0.064</td>
<td>0.896</td>
</tr>
<tr>
<td>2</td>
<td>0.064</td>
<td>0.939</td>
</tr>
<tr>
<td>3</td>
<td>0.063</td>
<td>0.895</td>
</tr>
<tr>
<td>4</td>
<td>0.067</td>
<td>0.888</td>
</tr>
<tr>
<td>5</td>
<td>0.085</td>
<td>0.843</td>
</tr>
<tr>
<td>6</td>
<td>0.188</td>
<td>0.777</td>
</tr>
<tr>
<td>6a</td>
<td>0.076</td>
<td>0.796</td>
</tr>
<tr>
<td>7</td>
<td>0.097</td>
<td>0.572</td>
</tr>
<tr>
<td>8</td>
<td>0.049</td>
<td>0.894</td>
</tr>
<tr>
<td>9</td>
<td>0.158</td>
<td>0.573</td>
</tr>
<tr>
<td>10</td>
<td>0.067</td>
<td>0.874</td>
</tr>
<tr>
<td>11</td>
<td>0.081</td>
<td>0.531</td>
</tr>
</tbody>
</table>

Note: Pr($H_{l}$) remains arbitrarily close to zero for all variations
IV. Extensions and Discussion

It is important to note that the very nature of non-experimental data implies a lack of control on the sampling process. This lack of control typically implies that a number of explanations are available for the anomalies in the data. Our analysis of the seat belt evidence requires the use of four different data sets and a number of aggregate statistics. None of the data are panel or longitudinal in nature, an obvious problem when attempting to draw inferences regarding intertemporal individual decision making. However, as with any non-experimental study, this single piece of research is not meant to constitute definitive evidence against the independence axiom. Only a large number of non-experimental results like the one presented herein would, in our opinion, raise serious concerns regarding the use of the expected utility paradigm — much more serious concerns than those raised by experimental evidence.

A specific potential problem with our non-experimental data is the aggregation of events. Although the events "no accident" and "death" are identical whether one is wearing a seat belt or not, the event "non-fatal" accident may not have the same payoff for belted versus unbelted drivers. It seems reasonable that belted drivers would incur less serious injuries in a non-fatal accident than would unbelted ones. Disaggregation of the sole intermediate event into several events of ranked severity would partially address this issue but the problem would only be eliminated with total disaggregation — an impossibility given existing data.

Outside of typical data availability problems, a difficult issue in constructing non-experimental tests of the independence axiom is the identification of lotteries with the same event space. For example, smokers who always wear seat belts are making rather curious choices but since lung cancer is not a possibility from an automobile accident and becoming a quadriplegic is not a smoking risk it is difficult to test the independence axiom with these apparently anomalous choices. However, one possibility for future research involves smokers whose homes are potentially exposed to radon gas. The event space could be construed as "healthy", "non-fatal lung disease", and "death". Even in this example there are difficulties since the intermediate events for smokers are, to the best of our knowledge, more heterogeneous than for those exposed to radon (i.e. heart disease is not caused by radon). Again, this is the nature of non-experimental evidence. Nevertheless, this should not detract from efforts to augment the body of experimental evidence against the expected utility paradigm with evidence from the marketplace.
Bibliography


