Provisional Probabilities and Paradigm Shifts

Preliminary Version *

Jonathan Weinstein

Kellogg School of Management, Northwestern University

January 23, 2011

Abstract

In this paper we show that procedures in which models are replaced, allowing for "paradigm shifts," can be partially reconciled with the principles of decision theory which lead to Bayesian updating. Under certain conditions, a belief-revising decision-maker is identical to or closely resembles a Bayesian. In these cases we have two alternative ways to represent the same decision-maker: with model revision, in which case we refer to his beliefs as "provisional beliefs," or as an ordinary Bayesian who has a different "belief" that covers all of the possible revisions. This is an attempt to bridge the gulf between decision theory and statistical practice.

* I am grateful for conversations with Eddie Dekel, Tai-Wei Hu, Ehud Kalai, Eric Maskin, and Muhamet Yildiz, and seminar participants at Northwestern University and the Transatlantic Theory Conference. I am especially grateful for many conversations over the years with Nabil Al-Najjar.

Contents

"...I equate the rational attitude and the critical attitude. The point is that, whenever we propose a solution to a problem, we ought to try as hard as we can to overthrow our solution, rather than defend it. Few of us, unfortunately, practice this precept..."

Karl Popper, *The Logic of Scientific Discovery*

1 Introduction

As theorists we work with many definitions of rationality, all far from perfect. The statement by Popper, above, is related to all of these notions. Axiomatic definitions of rationality have the flavor of presenting certain possible avenues of self-criticism to a decision-maker, and calling him rational if he is immune to these. This attitude is typified by the famous story of Savage's reaction to the Allais paradox: he initially made the "normal" pair of choices, which contradict the substitution axiom. When this was pointed out, he considered his decisions a mistake. This is the essence of the normative point of view: Savage apparently did not think that he or anyone else followed the axioms perfectly, but did think it was an appropriate goal. One theme of this paper will be to ask if we can assign a useful meaning to following a certain set of principles approximately, since true purity is an unreasonable expectation.

Descartes defined *homo sapiens* as a "rational animal," emphasizing our dual nature. Indeed, being a "fully rational" decision-maker makes no sense, since tastes must come from somewhere, and these would be considered animal. (Indeed, brain-damaged patients who lose their emotions are much less functional than those who lose their reason.) A DM who is "fully animal" is possible, but is not a good subject for theorists. There may well be useful patterns in his behavior, but not of a kind best captured by axioms and deductive reasoning. The object of our study is a true "rational animal" where I take rational in the sense of Popper, as engaging one's critical faculty in an effort to improve one's decisions.

This sense of rational does not carry any expectation of perfection. Indeed, any set of formal rules will leave out some possible criticisms, i.e. be insufficient for rationality. For instance, no abstract theory of decisions prevents me from believing that cars use the left side of the road everywhere (at least until I am flattened.) Conversely there is almost always an argument that a principle, however compelling, is not necessary for rationality. Furthermore, each principle takes a certain amount of mental energy to check, and the art of good decision-making must involve making wise decisions as to which principles to prioritize.

Decision theorists can, in principle aid this process by proving equivalence results, i.e. "If you want to follow Savage's axioms, use a subjective probability distribution." Unfortunately, it is difficult to conceive that there will ever be a Grand Unified decision theory which aids the decision-maker in avoiding every possible criticism. Actually, such a theory would be tantamount to "strong AI," the problem of building a machine which mimics or exceeds human capacities, which is considered at least decades away. Decision theory is not so ambitious, but merely tries to help people avoid selected mistakes in well-defined areas.

In briefest terms, this paper will explore the tension between the desires of the decision-maker (DM) for internal and external consistency. More specifically: We will assume throughout this paper that the DM indeed uses a subjective probability for all *static* decision problems. For dynamic decision problems, the classic procedure is to use Bayesian updates of one's initial subjective distribution. Indeed, as we will show, using Bayesian updating is necessary and sufficient for the DM to avoid "diachronic" Dutch books – combinations of bets made at different times which in total guarantee a negative total payoff.

But, while Bayesian updating guarantees a pleasing internal consistency, it may lead the DM to beliefs that stimulate a different kind of self-criticism. Suppose the DM believes he is observing repeated flips of a fair coin, i.e. his initial distribution is uniform on sequences in $\{H, T\}^N$. According to Bayesian updating, if he observes 100 consecutive heads he must continue to believe that the next flip is 50-50. This will doubtless put him in mind of Emerson's dictum that "A foolish consistency is the hobgoblin of little minds," and he will wish to change his belief in a non-Bayesian way.

The natural response is that, foreseeing this desire to revise, the DM should have formed a slightly different belief at time 0, one which was a mixture of the uniform distribution with a small probability that the coin is unfair. Suppose then that the DM's initial belief is a mixture of i.i.d distributions with full support on the frequency parameter, perhaps not uniform but somewhat concentrated around .5; this seems like a sensible prior for an unknown coin. Now, on observing many heads, he will have reasonable beliefs, which converge to the belief that the coin is double-headed. But another problem arises: if he observes a long alternating sequence HTHTHTHT..., his posterior will converge to the belief that the flips are i.i.d. 50-50. Clearly he will not consider this reasonable, and will instead want to conclude that the coin is alternating (perhaps due to prestidigitation.)

We can suggest more sophisticated beliefs as before, telling the DM: "Aha! Your actual belief was not fully represented by the exchangeable distribution. Your belief was a mixture of (with high weight) an exchangeable distribution and (with much lower weight) a distribution which includes many finer patterns such as the alternation. Sufficient data can swamp your prior and cause your posterior to be concentrated on a non-exchangeable belief."

The DM can attempt to construct his initial belief according to this advice, but this places a rather large onus on him. Apparently, he must anticipate *a priori* every possible pattern which would cause him to believe the coin is not i.i.d, and mix these together into a grand belief. This is a rather large burden. Even if an exponentially small fraction of the possible paths lead to a "paradigm shift," the number may still be exponentially large. A DM being constrained to follow Bayesianism with full purity is analogous to a chess player being *forced* to decide on his entire strategy (in the formal sense) in advance.

Relatedly, there is a conflict between the colloquial and formal meanings of belief. Colloquially, when we say someone believes a process to be i.i.d. with unknown parameter (exchangeable), we do not mean that he wouldn't change his mind when he sees HTHTHTH... Formally, of course, holding such a belief implies that you never change your mind but simply continue using Bayes' rule. To formalize the colloquial notion of revisable beliefs, we will define a structure called **provisional beliefs**.

Let us reexamine the assertion that Bayesian updating is necessary and sufficient for avoiding diachronic Dutch books. Indeed, avoiding Dutch books is equivalent to the *existence* of a representation of the DM via Bayesian updating. A core observation of this paper is that the Bayesian representation is not necessarily the *only* representation, or the most useful. Instead we conceive of a "flexible" DM who usually updates his probabilities according to Bayes' rule but at some histories decides that his belief is "wrong" and changes it according to some other principle. Accordingly we call his beliefs **provisional beliefs**, and the periods at which he is non-Bayesian **paradigm shifts**. We will show two fundamental results:

- For any flexible DM, there is a Bayesian DM who makes the same decisions on **shift-protected** bets: bets that do not depend on events subsequent to any paradigm shift.
- The magnitude of possible Dutch books against the DM depends on how sensitive the bets he makes are to post-shift events.

When applying the first principle, it is important to realize that the Bayesian beliefs associated with the DM are different from the provisional beliefs. Our purpose in introducing the representation via provisional beliefs is that it will often be more convenient for the DM to use, and will correspond more closely to our intuitive notion of beliefs. Indeed, as discussed earlier, a "belief" as a complete catalogue of all things that I might believe in the future is clearly at odds with the everyday meaning of the word.

2 Model and Results

The decision-maker (DM) observes in each of N periods an element of a finite set A. The set of possible sequences is $\Omega = A^N$. A history of length k is an element $h \in A^k$, and we then write $|h| = k$. We use $\emptyset$ for the empty history, and write $h_1 \leq h_2$ when h_1 is a subhistory of h_2. The set of all histories is H. We denote the set of distributions over Ω by $\Delta(\Omega)$. Given a history $h \in H$ (including terminal histories ω), we denote its truncation to k periods by h^k, and its truncation to $|h| - k$ periods by h^{-k}. A **system of provisional beliefs** is any function $f : H \to \Delta(\Omega)$ such that $f(h)(\{\omega : h \leq \omega\}) = 1$ for all histories h. A history h is said to be **normal** if $f(h)$ is formed by a Bayesian update[1] from $f(h^{-1})$, and otherwise is said to be a **paradigm shift**. Write $S \subseteq H$ for the set of all paradigm shifts. Also, let $\bar{S}$ be the set of terminal histories with a shift somewhere along their path, i.e $\bar{S} = \bigcup_{s \in S}\{\omega \in \Omega : s \leq \omega\}$. We call the DM **Bayesian** if $S = \emptyset$, i.e. he is normal at all histories. As a convenient shorthand we write $f(h_1, h_2) = f(h_1)(\{\omega : h_2 \leq \omega\})$ for the probability of reaching h_2 conditional on reaching h_1, according to the subjective belief held at h_1.

A **bet** is a pair (h, v) where h denotes the history at which the bet is offered and $v \in V = \mathbb{R}^\Omega$ denotes the net gain or loss for the DM at each terminal history, with the requirement that $v_\omega = 0$ when ω is incompatible with h. A bet (h, v) is accepted by the DM if $f(h) \cdot v \geq 0$, where $f(h)$ is viewed in the natural way as a vector. (The interpretation is that the bet is accepted contingent on reaching history h.) A finite set $D = \{(h_i, v_i)\}$ of bets is a **weak (dynamic) Dutch book** if all elements of D are accepted and $\sum v_i < 0$. It is a **strong Dutch book** if $\sum v_i << 0$. We will use the sup norm for vectors, denoted $||v|| = \max_\omega |v_\omega|$.

It will be convenient to prove the following proposition, but the core of the result is certainly not new, and in some form may be as old as Bayes' rule.

Proposition 1. *Suppose the DM's initial belief has full support, i.e. $f(\emptyset)(\omega) > 0$ for all ω. Then the following are equivalent:*

1. *There is a strong Dutch book against the DM.*
2. *There is a weak Dutch book against the DM.*
3. *The DM is not Bayesian.*
4. *There exist legal bets (h_1, v) and (h_2, v) on which the DM makes different decisions.*
5. *There exist legal bets $(\emptyset, v)$ and (h_1, v) on which the DM makes different decisions.*

[1] This includes the case that h has zero probability under $f(h^{-1})$, although we will usually work with full-support distributions.

Proof. 5 ⇒ 4: Trivial.

4 ⇒ 3: Clearly $v \neq 0$. For both bets to be legal, $v_\omega \neq 0 \to h_1, h_2 \leq \omega$. This means one is a subhistory of the other, say $h_1 \leq h_2$. If the DM were Bayesian, the weights assigned by $f(h_1)$ and $f(h_2)$ to terminal histories compatible with h_2 would be proportional. Since these are the only histories where v is non-zero, different decisions would be impossible.

3 ⇒ 2: If the DM is not Bayesian, let h be a paradigm shift. There must be two states compatible with h whose likelihood ratio shifts between h^{-1} and h, say $r = f(h^{-1}, \omega_1)/f(h^{-1}, \omega_2)$ and $s = f(h, \omega_1)/f(h, \omega_2)$ with $r > s$. Then, restricting payoff vectors to (ω_1, ω_2), $D = \{(h^{-1}, (1, -r)), (h, (-1, s))\}$ is a weak Dutch book, giving negative payoff at state ω_2 and zero elsewhere.

2 ⇒ 1: Let ω have negative payoff in the weak Dutch book. Add to the book a bet $(\emptyset, v)$ with $v_\omega = \epsilon$, $v_{\omega'} = -\epsilon f(\emptyset, \omega)$ for all $\omega' \neq \omega$. This bet will be accepted and gives a strong Dutch book for sufficiently small $\epsilon > 0$. This is the only implication where the full-support assumption is used.

1 ⇒ 5: If this implication failed, there would also be a strong static Dutch book at time 0, i.e. a strictly negative vector with non-negative expected value. ■

Note that in the absence of the full-support assumption, any of 3, 4 and 5 implies 2 and is implied by 1.

Call a bet (h, v) **shift-protected** (with respect to a fixed system f) if whenever $h < h' < \omega_1, \omega_2$ for a paradigm shift h', $v(\omega_1) = v(\omega_2)$. That is, a shift-protected bet is not sensitive to events subsequent to any future paradigm shift – note that the definition depends on h as well as v. Let $W_h \subseteq V$ be the set of v such that (h, v) is shift-protected; note that W_h is a vector subspace of V. A bet that is not shift-protected is called **shift-exposed**.

Proposition 2. *Given any DM with a system of provisional beliefs f, there is a Bayesian DM with prior P_f who makes identical decisions on all shift-protected bets.*

Proof. Given a terminal history ω, let $h_1 < h_2 < \ldots < h_n$ be the paradigm shifts which are subhistories of ω. Define a prior P_f by

$$P_f(\omega) = \prod_{i=0}^{n} f(h_i, h_{i+1})$$

where $h_0 = \emptyset, h_{n+1} = \omega$. Equivalently, P_f could be defined by a product of one-period-ahead probabilities:

$$P_f(\omega) = \prod_{i=0}^{N-1} f(\omega^i, \omega^{i+1})$$

That is, P_f is precisely the prior under which all the "myopic" forecasts (of the next observation) are identical to those of f.

Let f' be the system of provisional beliefs formed by Bayesian updating from P_f. Our claim is that f and f' lead to the same decisions on all bets that are shift-protected (with respect to f). Then the result will follow from Proposition 1.

To prove the claim: the definition of a shift-protected bet (h, v) can be restated as saying v is measurable with respect to the equivalence relation defined by

$$\omega_1 \equiv_h \omega_2 \Leftrightarrow \exists h' \in S : h < h', h' < \omega_1, h' < \omega_2$$

It then suffices to show that $f(h)$ and $f'(h)$ assign the same weight to each equivalence class. Indeed, an equivalence class consists either of a single state ω with no shifts between h and ω, or a set $\{\omega : h' < \omega\}$ where h' is a shift following h with no intermediate shifts. In either case the result follows from the fact that the one-step-ahead predictions are the same for h and h' whenever there is no shift. ∎

Corollary 1. *Any Dutch book must contain a shift-exposed bet.*

More specifically, a Dutch book must include bets (h_1, v_1) and (h_2, v_2) where $f(h_2)$ is not a Bayesian update of $f(h_1)$ and $v_1 \notin W_{h_1}$. That is, (h_1, v_1) is exposed to some shift h' with $h_1 < h' \leq h_2$.

One way of measuring the degree of deviation by the shifting DM from Bayesian updating is to look at the distance between his initial belief $f(\emptyset)$ and the Bayesian prior constructed in Proposition 2. We measure this distance with the total variation metric:

$$\delta(p, q) = \frac{1}{2} \sum_{\omega \in \Omega} |p_\omega - q_\omega| = \sum_{\omega \in \Omega} (p_\omega - q_\omega)^+ = - \sum_{\omega \in \Omega} (p_\omega - q_\omega)^-$$

The next proposition tells us that if a DM has a small (subjective) probability of reaching a paradigm shift, his initial belief $f(\emptyset)$ is close to the Bayesian belief constructed in Proposition 2.

Proposition 3. *Suppose* $\sum_{h \in S} f(\emptyset, h) = \epsilon$. *Then* $\delta(f(\emptyset), P_f) \leq \epsilon$, *where* P_f *is as defined in proposition 2.*

Proof. It is immediate from the construction of P that for any $\omega \in \Omega - \bar{S}$, $P(\omega) = f(\emptyset, \omega)$. The result follows. ∎

Note that for any h we can write V as a direct sum $V = W_h + W_h^\perp$ where W_h' is the orthogonal space to W_h with respect to the inner product defined by

$$\langle v, w \rangle_h = \sum_{\omega \in \Omega} f(h, \omega) v_\omega w_\omega$$

That is, $\langle v, w \rangle_h$ is the expected value of the product of the two payoffs with respect to the measure $f(h)$. What does the space $W_h^\perp$ look like? Well, a spanning set for W_h is given by indicator functions for the sets $\{\omega : h' < \omega\}$ for each paradigm shift $h' > h$. This means that $W_h^\perp$ is the set of bets with zero expectation (according to the measure $f(h)$) at each paradigm shift $h' > h$.

By the magnitude of a Dutch book we shall mean the smallest absolute loss the DM experiences in any state. The magnitude of a Dutch book can be bounded if we can estimate the subjective probably of reaching a shift, and the magnitudes of the shift-sensitive components of our bets.

Proposition 4. *Let* $D = \{(h_i, v_i)\}$ *be a Dutch book, and let* $v_i = w_i + w_i'$ *be the decomposition of* v_i *into* $W_\emptyset + W_\emptyset^\perp$. *Let* ϵ *be the* $f(\emptyset)$*-probability of ever reaching any paradigm shift. Then the magnitude of* D *is at most* $\epsilon \sum_i ||w_i'||$.

Proof. It is without loss to let the DM be indifferent to each bet, i.e. $f(h_i) \cdot v_i = 0$. The idea of the proof is the same as the previous derivation: we evaluate each bet according to the measure P_f, and show that its expectation is at worst $-\epsilon||w'_i||$. Indeed, let q be the Bayesian update of P_f at h_i, and recall that this measure agrees with $f(h_i)$ on all shift-protected bets. Then

$$\begin{aligned} q \cdot v_i &= (q - f(h_i)) \cdot v_i \\ &= (q - f(h_i)) \cdot w'_i \\ &\geq -\delta(q, f(h_i))||w'_i|| \\ &\geq -\sum_{h \in S, h_i < h} f(h_i, h)||w'_i|| \end{aligned}$$

Recalling that v_i is required to be 0 at states incompatible with h_i, we can say

$$P_f \cdot v_i = f(\emptyset, h_i)(q \cdot v_i) \geq -f(\emptyset, h_i) \sum_{h \in S, h_i < h} f(h_i, h)||w'|| \geq$$

Now we consider two cases: if there is a shift $h < h_i$, then $f(\emptyset, h_i) \leq \epsilon$. If not, then $f(h_i)$ is a Bayesian update of $f(\emptyset)$ and therefore

$$f(\emptyset, h_i) \sum_{h \in S, h_i < h} f(h_i, h) = \sum_{h \in S, h_i < h} f(\emptyset, h) \leq \epsilon$$

Either way, $P_f \cdot v_i \geq -\epsilon||w'_i||$. Summing over all bets, our P_f-expectation is at worst $-\epsilon \sum_i ||w'_i||$, which implies the desired result.

■

Please note that while the bound in this result is based on the DM's *subjective* probability of reaching a paradigm shift, the conclusion measures an *objective* quantity. That is, it provides an objective measure of the internal inconsistency of the DM's decision process.

3 Example

It is important to realize that a bet may be shift-sensitive even if does not specifically pertain to post-shift events. For instance, consider the following situation:

4 Further Comments

The point of this exercise is to draw a moral for statistical practitioners who, for practical reasons, use a procedure which is not fully Bayesian, but find the internal consistency of Bayesian inference appealing. How concerned should they be about the inconsistency represented by occasional non-Bayesian updating, e.g. switching statistical models? Proposition 4 gives an answer. It tells the DM that if his initial *subjective* probability of reaching a non-Bayesian shift is small, and if the dependence of his bets on "post-shift" events is bounded, the degree to which he is subject to an *objective* Dutch book is small.

Recall that the idea of a DM who uses non-Bayesian beliefs is motivated by a lack of full introspection; examining one's potential beliefs after every possible sequence is too costly. The application of Proposition 4 requires only *limited* introspection; the DM need not know every possible future belief to put a bound on the probability of a shift.