A Bayesian Foundation for Classical Hypothesis Testing (formerly Provisional Beliefs and Paradigm Shifts)

Jonathan Weinstein*
Kellogg School of Management, Northwestern University

First draft: July 2010; This version: August 2012

Abstract

Bayesian inference has the advantage of dynamic consistency, but the drawback of rigidity. When a decision-maker's initial model fails a hypothesis test, he may wish to form a new model, violating Bayes' rule. We show that if such "paradigm shifts" are rare, he will be "approximately" dynamically consistent. More specifically, we show that in our setting dynamic consistency is equivalent to the non-existence of Dutch books, and that a decision-maker who is almost always Bayesian will suffer from only "small" Dutch books. This gives the decision-maker some latitude to revise his model while bounding the pain of inconsistency.

* I am grateful for conversations with Nabil Al-Najjar, Eddie Dekel, Tai-Wei Hu, Ehud Kalai, Eric Maskin, Allegra Petti, Phil Reny, Larry Samuelson, and Muhamet Yildiz, and seminar participants at Northwestern University, the Transatlantic Theory Conference, University of Montreal, and the Stony Brook Game Theory Festival.

"...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

If a decision-maker (DM) aspires to be rational according to Popper's criterion, he will try to avoid being trapped by a dogmatic belief in his model. He may therefore wish to update his model and adjust his beliefs when he sees a surprising pattern on the data. However, decision theory tells him that if he does so in any way other than according to Bayes' rule, he will be subject to a different sort of criticism: he will make dynamically inconsistent decisions, and be subject to arbitrage. He may feel that he is caught between a rock and a hard place: following Bayes' rule perfectly is the only way to avoid inconsistency, but it prevents him from responding to any pattern which was not present in his prior. An ideal Bayesian would respond by including every possible pattern in his prior, but such an approach is impractical and even provably uncomputable[1]. So, a practical Bayesian may wish to occasionally transgress against Bayes' rule, and may well wonder about the price of such transgressions in terms of inconsistency.

The message of this paper is: *Occasional* violations of Bayes' rule will make the DM only *slightly* inconsistent. An attentive reader will now wonder how we will formally define the italicized words, and indeed the merit of the paper will depend largely on how intelligently we do so. It is generally quite difficult to quantify inconsistency. In pure logic, it is impossible, as any inconsistency whatsoever implies that all propositions are both true and false, infecting the entire system of reasoning. In logic, then, it is useless to speak of a "minor" inconsistency. Here, though, we will be able to develop tools from decision theory into a method for measuring dynamic inconsistency. We will first show (Proposition 1) that, in our

[1] Solomonoff (1964) introduced the idea of a prior placing weight on every computable theory, with more weight on simpler theories according to Kolmogorov complexity. This is a fascinating formalization of Occam's razor but does not produce a practical method of reasoning.

framework, dynamic consistency (DC) is equivalent to Bayesian updating, and DC is also equivalent to the non-existence of arbitrage opportunities against the DM. An arbitrage, also called a Dutch book, is a set of gambles the DM accepts which, in aggregate, guarantee him a negative payoff. Then, it is natural to measure inconsistency by the magnitude of the potential Dutch book it creates, as measured by the amount of certain loss. Our two main results, Propositions 3 and 4, show that the magnitude of possible Dutch books against the DM is bounded by the product of (a) the subjective probability he initially attaches to non-Bayesian shifts in beliefs and (b) the maximum amount he wagers along any history. (The two propositions each refine this bound in distinct ways.) Notice that if the DM performs non-Bayesian updating only when his initial model fails a classical hypothesis test, the quantity (a) is precisely the significance level α of this test. Hence our results state that a DM who is primarily Bayesian but also tests, and sometimes replaces, his theory, will be "almost" as consistent as one who is purely Bayesian when α is small.

We hope that our results lessen the perceived paradigmatic conflict between Bayesian and classical statistics. A DM who forms an initial, provisional, model and updates its parameters using Bayes' rule, while also, in parallel, conducting a hypothesis test which may reject the model, is covered by our results. He can achieve almost the full consistency of Bayesian reasoning, while maintaining the flexibility to reject his theory when surprising[2] events occur.

To illustrate why a DM may wish to violate Bayes' rule, some examples are in order. 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, with better foresight, the DM would have formed

[2] Some care is needed in interpreting "surprising." The DM may think that every specific sequence of 100 coin flips is individually unlikely, but if he thinks that every sequence justifies a new model, then he is "surprised" with probability 1 and our result will not help him. This pitfall is equivalent to that of uncorrected multiple hypothesis testing in statistics; it is the overall rate of rejection, conditional on the model being valid, which must be controlled.

a slightly different belief at time 0, one that acknowledges the possibility 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 a mixture somewhat concentrated around .5; this sounds like a sensible prior for an unknown coin. Now, on observing many consecutive heads, he will have beliefs which, as seems only reasonable, 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[3] distribution. Your belief was a mixture of (with high weight) an exchangeable distribution and (with much lower weight) a distribution including 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 that would cause him to believe the coin is not i.i.d., and mix these together into a grand belief. This is a burden he may find unmanageable. 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. Accordingly, the DM may value the latitude to form an initial set of beliefs without making a binding commitment as to his behavior at all future histories.

To formalize the notion of paradigm shifts, we will define a structure called **provisional beliefs**. Formally, a system of provisional beliefs is any mapping from histories to beliefs about the future. As a formal object, our system of provisional beliefs is similar to the "conditional probability systems" introduced by Myerson [7], but without the requirement that Bayes' rule always be used on

[3]A mixture of i.i.d. distributions. De Finetti famously showed, for infinitely repeated processes, that being such a mixture is equivalent to being invariant to permutations, hence the term "exchangeable."

positive-probability events. We are, as the foregoing discussion suggests, interested in those mappings where the updating is "usually" Bayesian. We will call the histories at which updating is non-Bayesian **paradigm shifts**. We should note that underlying the desire to revise one's beliefs is a conflict between the colloquial and formal meanings of belief. Colloquially, when we say someone "believes" a process to be exchangeable, we do not mean that he wouldn't change his mind when he sees HTHTHTH... In Bayesian language, of course, holding such a belief would mean that he never changes his mind but simply continues using Bayes' rule. Treating beliefs as provisional, therefore, may come closer to our natural understanding of the word. It does run the risk of dynamic inconsistency, but our results here provide for some control over this potential inconsistency.

Towards our goal of showing that DMs with occasional paradigm shifts are "approximately" consistent, we have already summarized our most important results, Propositions 3 and 4. One of the stepping-stones to this result, Proposition 2, is worth mentioning here in its own right. It says that for any DM with a system of provisional beliefs, there is a Bayesian DM who makes the same decisions on **shift-protected** bets. A bet fails to be shift-protected if it is made prior to a potential paradigm shift and may be affected by events which occur after that shift. This provides another sense in which the DM is close to dynamically consistent: If he avoids bets which are sensitive to events following a paradigm shift, he behaves just like a Bayesian and hence is dynamically consistent. This result should be unsurprising, but it is a useful step.

It is important to realize that the Bayesian beliefs, Q, constructed in Proposition 2 will generally be different, and more complex, than the initial provisional beliefs. The beliefs Q will include all models the DM will ever adopt under any circumstances, while the initial provisional beliefs may exclude models which are unlikely to be adopted. Therefore, even when the system of provisional beliefs rarely results in different decisions from those of a Bayesian, the non-Bayesian representation may be simpler and closer to the DM's natural thought processes. This is an additional motivation for introducing the formalism and results of this paper.

2 Literature

Versions of our preliminary result, Proposition 1, were proved by Freedman and Purves [6] and others; it is convenient for us to reprove it here to show how it fits our particular formalism. Also closely related to Proposition 1 is the work of Epstein and Le Breton [3]. They show that a dynamically consistent decision-maker whose static decisions are based on beliefs must in fact be Bayesian. Our assumption here of a very simple form for static decisions is mostly in order to focus attention on dynamic decision-making, but their result further justifies this choice.

The no-Dutch-book argument for Bayesian updating has been subject to critiques independent of the present paper. For instance, Border and Segal [1] showed that a bookie (who faces a problem similar to our DM) may wish to create odds that do not satisfy Bayes' rule, because strategic considerations involving the beliefs of his counterparties outweigh the issue of avoiding Dutch books. In general, more recent papers in decision theory are likely to focus on dynamic consistency alone rather than the accompanying issue of Dutch books. In the present context, the no-Dutch-book condition is an appealing equivalent formulation of DC because it lends itself naturally to measuring violations of DC.

Previous work on non-Bayesian updating includes the papers of Epstein [2] and Epstein, Noor and Sandroni [4],[5]. The decision-makers in their papers, unlike here, are sophisticated and anticipate their future non-Bayesian updating. This alternative modeling choice may reflect a difference in the motivation behind the non-Bayesian updating; in Epstein et. al. it is described as a temptation to overreact or underreact to news, whereas we are concerned with a DM who simply decides that his statistical model needs to be replaced. For us, if the DM could anticipate all possible eventualities, he would simply form an all-encompassing prior and update it. The fact that we are interested in limitations on foresight rather than on rationality has led us to analyze different issues than Epstein et. al.; they are not concerned with approximate dynamic consistency in our sense, and we do not address interesting issues in their work such as learning under imperfect Bayesian updating.

A recent paper by Ortoleva [8] also considers decision-makers who sometimes violate Bayesian updating. Our papers differ in focus; in [8] the main result is a representation theorem, while here we begin with a simple representation and analyze the impact of occasional non-Bayesian updating on our measure of dynamic consistency. Also, [8] assigns a very different meaning to approximate dynamic consistency. There, DC is relaxed by defining a very weak condition called "Dynamic Coherence[4]." The representation theorem in [8] shows that a DM who follows Dynamic Coherence (and other basic conditions) can be represented as performing Bayesian updating after all events of probability greater than some ϵ, and arbitrary[5], potentially non-Bayesian updating after events with probability less than ϵ. Notice that even for small ϵ, this does not imply our notion of approximate DC, since it may be certain that *some* event of probability ϵ occurs – the familiar problem of multiple hypothesis testing. Furthermore, the representation theorem makes no conclusion regarding ϵ (i.e. ϵ may be arbitrarily close to 1), so it implies virtually no restriction on how the DM forms his beliefs[6]. The main theorem in [8] includes a converse, so that no stronger conclusion is available from the assumptions.

[4]Dynamic Coherence states: For any cyclical sequence of events $A_1, A_2, \ldots, A_{n+1} = A_1$ such that for each i, A^c_{i+1} is a null event when conditioned on A_i, preferences conditional on A_1 and A_n are identical. To understand this better, consider first the case that, conditional on any A, each state $\omega \in A$ is non-null, our focus in this paper. Then if B^c is null when conditioned on A, $A \subseteq B$. Then the antecedent in Dynamic Coherence implies that $A_1 \subseteq \cdots \subseteq A_{n+1} = A_1$, and therefore that all the A_i are equal, so that without null states, Dynamic Coherence is vacuous. More generally, when some states are null, the condition says roughly that events that differ only on null states lead to the same preferences, or in other words that the DM's preferences are unaffected by events he was certain would occur.

[5]In the representation in [8], non-Bayesian updating is performed by applying maximum likelihood to a prior over priors ρ, which may make the updating seem non-arbitrary. However, the proof that a representation exists proceeds by showing that there is sufficient freedom in choosing ρ to fit any updating at all, except for the small restriction involving null states we discussed in footnote 4. Because of the unrestrictive nature of Dynamic Coherence, this freedom is needed for the result to hold.

[6]The only restriction is that, since ϵ is strictly less than 1, the DM does not revise his beliefs after an event he was certain would occur. This conclusion is closely related to the assumption of Dynamic Coherence; see footnote 4.

3 Model and Results

3.1 Definitions and Notation

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 an initial segment 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 $\omega \in \Omega$), we denote its truncation to k periods by h^k, and its truncation to $|h| - k$ periods by h^{-k}. The key object of study will be:

Definition 1. *A* ***system of provisional beliefs****,* $P_h : h \in H$*, is a collection of distributions on* Ω*, one for each history, such that* $P_h \in \Delta(\Omega), P_h(\{\omega : h \leq \omega\}) = 1$ *for each history* h*.*

This collection represents the DM's beliefs over future events contingent on each history. In particular, $P_\emptyset$ is his initial belief. As a convenient shorthand we write $P_{h_1}(h_2) = P_{h_1}(\{\omega : h_2 \leq \omega\})$ for the probability of reaching h_2 conditional on reaching h_1, according to the subjective belief held at h_1. Also as shorthand, we will write $P_h(A)$, where $A \subset H$, to mean the probability at h of reaching any history in A, that is, $P_h(A) \equiv P_h(\cup_{h' \in A}\{\omega : h' \leq \omega\})$.

Fixing a system of beliefs f, a history h is said to be **normal** if P_h is formed by a Bayesian updatefrom $P_{h^{-1}}$, and otherwise is said to be a **paradigm shift**. Write $S \subseteq H$ for the set of all paradigm shifts. Let $\bar{S}$ be the set of terminal histories with a shift somewhere along their path, i.e $\bar{S} = \bigcup_{s \in S}\{\omega \in \Omega : s \leq \omega\}$. Also, let $\hat{S} = \{s \in S : \nexists s' \in S : s' < s\}$ be the set of "initial" paradigm shifts, those without a prior shift. We call the DM **Bayesian** if $S = \emptyset$, i.e. he is normal at all histories.

We use elements of $V = \mathbb{R}^\Omega$ to describe state-contingent payoffs. We write $V_h = \{v \in V : v_\omega \neq 0 \rightarrow h \leq \omega\}$ for the set of vectors which have non-zero value only at states consistent with history h. A **bet** is a pair (h, v) where h denotes the history at which the bet is offered and $v \in V_h$ denotes the net gain or loss for the DM at each terminal history. The interpretation is that the bet is offered *after* history h is observed; if this history is not reached, it is never offered. The restriction to V_h imposes a conventional requirement[7] that a bet has non-zero value only at states consistent with the current history. A bet (h, v) is **accepted** by the DM if the expectation of v according to the measure P_h is non-negative, i.e. if $P_h \cdot v \geq 0$, where P_h is viewed in the natural way as a vector in $\mathbb{R}^\Omega$. A finite set $D = \{(h, v)\}$ of bets is a **weak (dynamic) Dutch book** if all elements of D are accepted and $\hat{v} \equiv \sum_D v < 0$, i.e. $\hat{v}_\omega$ is nowhere positive and is negative for sone ω. It is a **strong Dutch book** if $\hat{v} << 0$, i.e. v_ω is negative for every ω. We will use the sup norm for vectors, denoted $||v|| = \max_\omega |v_\omega|$.

Note that we have assumed the simplest possible form for static decisions, expected utility with risk neutrality, in order to focus attention on issues of dynamic consistency. The axioms which lead to such a representation for static decision-making are well-known, and we will not review them here.

3.2 Equivalence of Bayesian inference, dynamic consistency, and absence of Dutch books

It will be convenient to prove the following proposition in our specific context, but as mentioned earlier, the core of the result is certainly not new. For this result and through the rest of the paper, we assume that $P_\emptyset$ has full support, and indeed that for each h, P_h has support $\omega : h \leq \omega$. The motivation for this assumption is that among his theories about the data, the DM considers it possible that the sequence is generated by a fair coin. Issues involving zero-probability histories are certainly vital in studies of dynamic games, but are orthogonal to our concerns here.

Proposition 1. *Suppose the DM's beliefs at each history have full support, i.e. $P_h(\omega) > 0$ whenever $h \leq \omega$. 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.*

[7] This convention departs slightly from the usual setup in which a bet may specify non-zero payoffs at impossible states. These payoffs will always be simply ignored by the DM. Since here we assume that all relevant parties know the history h, it is natural to assume that they do not bother specifying non-zero payoffs at impossible states. This convention loses no substantive freedom and simplifies the statements of our results.

3. *The DM is not Bayesian.*

4. *There exist v, h_1, h_2 with $v \in V_{h_1} \cap V_{h_2}$ such that (h_1, v) is accepted and (h_2, v) is rejected.*

5. *There exist h and $v \in V_h$ such that $(\emptyset, v)$ is rejected but (h, v) is accepted.*

Proof. $5 \Rightarrow 4$: Trivial.

$4 \Rightarrow 3$: Clearly $v \neq 0$. For $V_{h_1} \cap V_{h_2}$ to be non-trivial, it is necessary that one history is an initial segment of the other, say $h_1 \leq h_2$. If the DM were Bayesian, then for each $\omega \geq h_2$ we would have $f(h_1)(\omega) = f(h_1)(h_2) * f(h_2)(\omega)$. Since these are the only histories where v is non-zero, $f(h_1) \cdot v = f(h_1)(h_2) * f(h_2) \cdot v$, so different decisions are impossible. The full-support assumption on $f(h_1)$ is needed here.

$3 \Rightarrow 2$: If the DM is not Bayesian, let h be a paradigm shift. There must be two states compatible with h whose likelihood ratio shifts[8] 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 payoff $s - r < 0$ at state ω_2 and zero elsewhere.

$2 \Rightarrow 1$: Let ω have negative payoff in the weak Dutch book. Append to the book a bet $(\emptyset, v)$ with $v_\omega = \epsilon$ and $v_{\omega'} = -\epsilon f(\emptyset, \omega)$ for all $\omega' \neq \omega$. This bet will be accepted and gives a strong Dutch book for sufficiently small $\epsilon > 0$. The full support of $P_\emptyset$ is needed here.

$1 \Rightarrow 5$: If this implication failed, the DM would accept all of the bets in the Dutch book at time 0. Then there would also be a strong static Dutch book at time 0; by adding all of the bets involved we would get a strictly negative vector with non-negative expectation according to measure $P_\emptyset$.

∎

Note that in the absence of the full-support assumption, it is easy to find counterexamples for the implications $4 \Rightarrow 3$ and $2 \Rightarrow 1$. Full support of $P_\emptyset$ alone would suffice to show $5 \Rightarrow 3$ and $2 \Rightarrow 1$, and hence that 1, 2, 3, and 5 are

[8] In fact, Bayesian updating is equivalent to the likelihood ratio of all pairs of events consistent with h being unchanged by the update.

equivalent. We tacitly assume full support in the remainder of the paper for ease of interpretation, though it is not used directly in later results.

3.3 Shift-protected bets

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 $W_h \subseteq V_h$ be the set of v such that (h, v) is shift-protected; note that W_h is a vector subspace of V_h, since it is defined by equality constraints. A bet that is not shift-protected is called **shift-exposed**.

Proposition 2. *Given any DM with a system of provisional beliefs P, there is a Bayesian DM with prior $Q_\emptyset$ who, on all shift-protected bets, makes the same decisions as the original DM. More specifically, if $Q_h : h \in H$ is the Bayesian system of beliefs with $Q_\emptyset = Q$, then $P_h \cdot v = f'(h) \cdot v$ for all h and $v \in W_h$.*

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

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

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

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

That is, Q is precisely the prior under which all the "myopic" forecasts $f(\omega^i, \omega^{i+1})$ (of the next observation) are identical to those of f. A Bayesian who begins with prior Q will have, at every history, the same opinion as f about the next observation, but will have different predictions in the longer term when f has paradigm shifts.

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

To prove the claim: the definition of a shift-protected bet (h, v) can be restated by saying that v assigns the same outcome to any states which are equivalent under the relation

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

It then suffices to show that P_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 if there are no shifts between h and h', then

$$P_h(h') = \prod_{i=|h|}^{|h'|-1} f(\omega^i, \omega^{i+1}) = f'(h)(h')$$

because on the relevant histories both systems of beliefs are Bayesian with the same myopic forecasts. ∎

Along with Proposition 1, this implies:

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 $h_1 < h' \leq h_2$.

The following lemma relies on the fact that Q is identical to $P_\emptyset$ in predicting events leading up to a shift.

Lemma 1. *The probability of ever reaching a shift is identical under Q and $P_\emptyset$, i.e. $Q(\bar{S}) = P_\emptyset(\bar{S})$*

Proof. Recall that $\hat{S} \equiv \{s \in S : \nexists s' \in S : s' < s\}$. By construction it is clear that for each $s \in \hat{S}$, $Q(\{\omega : s \leq \omega\}) = P_\emptyset(\{\omega : s \leq \omega\})$. But $\bar{S}$ is the disjoint union of such sets, implying the result. ∎

Define an inner product on V_h 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 P_h. Note that for any h we can write V_h as a direct sum $V_h = W_h + W_h^\perp$ where $W_h^\perp$ is the orthogonal complement to W_h with respect to this inner product. That is, we can write any $v \in V_h$ as $v' + v''$ where (h, v') is a shift-protected bet and $v'' \in W_h^\perp$.

To better understand the space $W_h^\perp$, note that a basis for W_h is given by indicator functions for the sets $\{\omega : h' < \omega\}$ for each paradigm shift $h' > h$ with no shift in between, i.e. no $h'' \in S$ with $h' > h'' > h$, together with indicator functions for singleton states $\omega \geq h$ with no prior shift $h'' \in S$, $\omega > h'' > h$. Then $W_h^\perp$ is the set of vectors orthogonal to each basis element. These are the vectors with zero expectation (according to the measure P_h) at each paradigm shift $h' > h$, as well as zero payoff at each ω with no prior shift.

3.4 Measuring deviations from dynamic consistency

Because the defining property of a Dutch book is a *certain* loss, we define the **magnitude** of a Dutch book as the smallest absolute loss the DM experiences in any state.

Definition 2. *The magnitude of a Dutch book D is $||D|| = \min_{\omega \in \Omega} |\sum_{(h,v) \in D} v_\omega|$.*

By this measure, a Dutch book with moderate equal losses in all states is worse than one with widely varying losses. While very large losses in selected states may certainly be imprudent, there is nothing *inconsistent* about tolerating such losses; one may simply hold a strong belief that those states are very unlikely[9]. If the reader considers small certain losses less important than large losses caused by erroneous beliefs, this paper is designed to be sympathetic. As mentioned in the introduction, the results here seek to free the DM from a fruitless quest for

[9]This is not only true for a DM who bases static decisions on a single belief, as in our model. Even an ambiguity-averse DM may assign some states small mass according to all distributions he considers possible, and hence tolerate large losses in those states.

perfect internal consistency, so that he can pay attention to the potentially more important task of refining his beliefs when unusual events occur.

Our stated program is that given a system of beliefs, we will measure its inconsistency by the magnitude of potential Dutch books. In this risk-neutral environment, though, Dutch books can be multiplied by an arbitrary scalar. Clearly, then, we cannot simply use the largest absolute magnitude of any Dutch book as in Definition 2. We will have to divide the magnitude by a measurement proportional to the scale of the bets made. There are two useful ways to define this scale factor, leading to the distinct bounds in the two main results below. In Proposition 3, the scale factor is $||\sum_{(h,v)\in P} v||$, the largest possible gain or loss from our *post-shift* bets along any history[10]. Notice that it is easy for the DM to control this quantity, since even under the imperfect introspection we have in mind, he knows when he changes paradigm after it happens. For instance, if he knows there will be a shift (subjectively) only 5% of the time according to his prior $P_\emptyset$, and he can commit to losing at most \$100 after any shift, the magnitude of the largest possible Dutch book is \$5.

Proposition 3. *Let D be a Dutch book. Let $\alpha = P_\emptyset(\bar{S})$ be the $P_\emptyset$-probability of ever reaching any paradigm shift. Let $C \subseteq D$ be the bets which are subsequent to some paradigm shift, i.e. $C = \{(h,v) \in D : \exists h' \in S : h' \leq h\}$. Then*

$$\frac{||D||}{||\sum_{(h,v)\in C} v||} \leq \alpha$$

More specifically, the $P_\emptyset$-expectation of $\sum_{(h,v)\in D} v$ is at worst $-\alpha||\sum_{(h,v)\in C} v||$.

Proof. We proceed by evaluating the $P_\emptyset$-expectation of each bet. We divide bets into two groups:

1. For bets $(h,v) \in D - C$, which are made before any shift, we know that (by convention) $v_\omega \neq 0 \rightarrow h \leq \omega$. For each such ω, $P_\emptyset(\omega) = P_\emptyset(h) * P_h(\omega)$, because there is Bayesian updating between $\emptyset$ and h. It follows that the $P_\emptyset$-expectation is proportional to the P_h-expectation and so is non-negative, since the bet is accepted.

[10] See formal definition of P in the proposition. It is important to note that this quantity is in general much smaller than $\sum_{(h,v)\in P} ||v||$, the total magnitude of all bets made at all post-shift histories.

2. Bets in C have non-zero outcomes only in $\bar{S}$, making it immediate that the $P_\emptyset$-expectation of their sum is at worst $-\alpha||\sum_{(h,v)\in C} v||$.

The desired result follows; the certain loss from a Dutch book cannot be worse than its expectation under a given measure. ■

The scale factor in Proposition 3 depended only on magnitudes of bets that are made *after* a shift. The next bound, in Proposition 4, is complementary: It depends only on the shift-sensitive components of bets, i.e. the projection on $W_h^\perp$. These are bets that are made *before* a shift but depend on events that occur after the shift. Neither bound is necessarily weaker or stronger than the other. The bound in Proposition 3 is probably more useful, since under the limited introspection we have in mind, it may be difficult for the DM to monitor the shift-sensitive components of his bets; he would have to know when a shift was coming. Still, we find Proposition 4 worthwhile for a complete understanding of the requirements for Dutch books. They occur only when the DM makes shift-sensitive bets prior to a shift, and bets again after the shift; if either kind of bet is bounded, so proportionally will be the Dutch book.

Proposition 4. *Let D be a Dutch book, and for each $(h, v) \in D$ let $v = v' + v''$ be the decomposition of v into $W_h + W_h^\perp$. Let $\alpha = P_\emptyset(\bar{S})$ as in Proposition 3. Then*

$$\frac{||D||}{||\sum_D v''||} \leq \alpha$$

More specifically, the Q-expectation of $\sum_D v$ is at worst $-\alpha||\sum_D v''||$.

Proof. Let $C = \{(h, v) \in D : \exists h' \in S : h' \leq h\}$ as in Proposition 3. Let Q be as in Proposition 2. This proof will proceed somewhat similarly to that of Proposition 3, but by calculating the Q-expectation of each bet rather than the $P_\emptyset$-expectation.

For any h, let Q_h be the Bayesian update of Q at history h. Recall that by

Proposition 2, $Q_h \cdot v' = P_h \cdot v'$ for any shift-protected bet v'. Then

$$\begin{aligned} Q_h \cdot v &\geq (Q_h - P_h) \cdot v \\ &= (Q_h - P_h) \cdot (v' + v'') \\ &= (Q_h - P_h) \cdot v'' \\ &= Q_h \cdot v'' \end{aligned}$$

where the first inequality holds because the bet (h, v) is accepted, and last equality holds because $P_h \cdot v'' = \langle \lambda_h, v'' \rangle = 0$ where λ_h is the characteristic vector for $\{\omega : h \leq \omega\}$, since $\lambda_h \in W_h$.

Because $v \in V_h$, we can write

$$Q \cdot v = Q(h) * (Q_h \cdot v) \geq Q(h) * (Q_h \cdot v'') = Q \cdot v''$$

It now follows that

$$Q \cdot \sum_D v = \sum_D Q \cdot v \geq \sum_D Q \cdot v'' = Q \cdot \sum_D v''$$

As per the prior discussion of the space $W_h^\perp$, each vector $v'' \in W_h^\perp$ is non-zero only at terminal states in $\bar{S}$. It follows that

$$Q \cdot \sum_D v \geq Q \cdot \sum_D v'' \geq -Q(\bar{S}) || \sum_D v'' ||$$

The final claim in the stated result now follows from Lemma 1, and the primary claim follows since, again, the certain loss from a Dutch book cannot be worse than its expectation under a given measure.

■

4 Bounding rejection probability with limited introspection

Our interpretation of the model is that while the DM's beliefs depend deterministically on the data, via the function f, the DM does not know all the details of the function f at time 0. The astute reader should then wonder how he could apply our main results, which require knowledge of $\alpha = P_\emptyset(\bar{S})$, the overall probability of rejection. As in classical hypothesis testing, a DM who rejects the initial theory only when some unlikely event occurs is not sufficiently disciplined, because all sufficiently long histories may be unlikely. Then, the DM may go through a paradigm shift on every history ($\alpha = 1$), and he may not even know that he has this property, because he only lives once.

The dilemma is: How, if the DM does not even know at time 0 what the set S looks like, can he know that its $P_\emptyset$ probability is low? Fortunately, there is a way. While he need not know the entire set S at time 0, we will require that he plan his paradigm shifts *somewhat* in advance. Our idea is that at some histories he generates new alternative hypotheses, and specifies later histories at which he will consider these alternative hypotheses to be verified and undergo a paradigm shift. He must do so far enough in advance that the shifts still have a low conditional probability at the time they are planned. We can show that any DM who disciplines himself in this way will have rejection probability at most α. To be much more precise:

Fix as before a system of provisional beliefs P. We call a function $g : H \to 2^H$ an α-**rejection plan** for P if it has these properties:

1. For every h, $g(h) \subset \{h' : h' \geq h\}$

2. $S = \cup_{h \in H} g(h)$

3. For every $\omega \in \Omega$,
$$\sum_{h \leq \omega} \frac{P_\emptyset(g(h))}{P_\emptyset(h)} \leq \alpha \tag{1}$$

The intended interpretation of $g(h)$ is as follows: at history h, the DM comes to suspect a pattern in the data which contradicts his initial theory, and decides that he will reject the theory and undergo a paradigm shift at histories in $g(h)$. The probability of set $g(h)$ conditional on history h must be small enough that the total probability of all rejection sets he constructs is at most α. The virtue of the final condition (1) is that it can be confirmed along the history which actually transpires. That is, it can be checked by a DM *without* the unrealistic ability to introspect about his behavior at *every* feasible history. Please note that in the applications we have in mind, the DM only periodically generates new theories, so that $g(h) = \emptyset$ for the vast majority of histories h, but this need not be so for the result to go through.

Proposition 5. *If there exists an α-rejection plan for P, then $P_\emptyset(S) \leq \alpha$.*

Proof.

$$
\begin{aligned}
P_\emptyset(S) &= P_\emptyset(\cup_{h\in H} g(h)) \\
&\leq \sum_{h\in H} P_\emptyset(g(h)) \\
&= \sum_{h\in H} \left[P_\emptyset(g(h)) \frac{\sum_{\omega \geq h} P_\emptyset(\omega)}{P_\emptyset(h)} \right] \\
&= \sum_{h\in H} \sum_{\omega\geq h} \frac{P_\emptyset(g(h))}{P_\emptyset(h)} P_\emptyset(\omega) \\
&= \sum_{\omega\in\Omega} P_\emptyset(\omega) \sum_{h\leq\omega} \frac{P_\emptyset(g(h))}{P_\emptyset(h)} \\
&\leq \sum_{\omega\in\Omega} P_\emptyset(\omega) * \alpha \\
&\leq \alpha
\end{aligned}
$$

■

So, we can conclude that if the DM only undergoes paradigm shifts according to an α-rejection plan, then the conclusions of Propositions 3 and 4 apply. Furthermore, such a plan need not be formed in detail *ex ante*, but rather its properties can easily be checked along the realized history.

5 Further Comments

Please note that while the bounds in our main results are based on the DM's *subjective* probability of reaching a paradigm shift, the conclusions measure an *objective* quantity which is defined without reference to any distribution. That is, we provide an objective, external measure of the internal inconsistency of the DM's decision process. We thus provide theoretical support to statistical practitioners who find the internal consistency of Bayesian inference appealing, but, for practical reasons, use a procedure that is not fully Bayesian. If a DM's criterion for rejecting his model has a strictly defined level as defined in classical hypothesis testing, the degree to which he is subject to an *objective* Dutch book is small.

Recall that our DM's use of non-Bayesian updating is motivated by a lack of full introspection; specifying *ex ante* one's potential beliefs after every possible sequence is very costly. The application of Propositions 3 and 4 requires only *limited* introspection; the DM need not know every possible future belief to put a bound on the probability of a shift. If he can guarantee that the possible paradigm shifts are confined within a set of small subjective probability, he does not need to know what his beliefs will be when these histories are reached.

References

[1] **Border, Kim and Uzi Segal.** 1994. "Dutch Books and Conditional Probability." *Economic Journal*, 104: 71-75.

[2] **Epstein, Larry.** 2006. "An Axiomatic Model of Non-Bayesian Updating." *Review of Economic Studies*, 73: 413-436.

[3] **Epstein, Larry and Michel Le Breton.** 1993. "Dynamically Consistent Beliefs Must be Bayesian." *Journal of Economic Theory*, 61: 1-22.

[4] **Epstein, Larry, Jawwad Noor and Alvaro Sandroni.** 2008. "Non-Bayesian Updating: A Theoretical Framework." *Theoretical Economics*, 3: 193-229.

[5] **Epstein, Larry, Jawwad Noor and Alvaro Sandroni.** 2010. "Non-Bayesian Learning." The B.E. Journal of Theoretical Economics (Advances), 10(1).

[6] **Freedman, David and Roger Purves.** 1969. "Bayes' Method for Bookies." *The Annals of Mathematical Statistics*, 40: 1177-1186.

[7] **Myerson, Roger.** 1986. "Multistage Games with Communication." *Econometrica*, 54: 323-358.

[8] **Ortoleva, Pietro.** 2011. "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News." *American Economic Review*, forthcoming.

[9] **Solomonoff, Ray.** 1964. "A formal theory of inductive inference, Part 1 and Part 2." *Information and Control*, 7:1-22 and 224-254.