# Belief Movement, Uncertainty Reduction, & Rational Updating*

Ned Augenblick[†] and Matthew Rabin[‡]

November 2020

**Abstract**

When a Bayesian learns new information and changes her beliefs, she must on average become concomitantly more certain about the state of the world. Consequently, it is rare for a Bayesian to frequently shift beliefs substantially while remaining relatively uncertain, or, conversely, become very confident with relatively little belief movement. We formalize this intuition by developing specific measures of *movement* and *uncertainty reduction* given a Bayesian's changing beliefs over time, showing that these measures are equal in expectation, and creating consequent statistical tests for Bayesianess. We then show connections between these two core concepts and four common psychological biases, suggesting that the test might be particularly good at detecting these biases. We provide support for this conclusion by simulating the performance of our test and other martingale tests. Finally, we apply our test to datasets of individual, algorithmic, and market beliefs.

*We thank Tristan Gagnon-Bartsch and Kevin He for valuable research assistance, and Roland Benabou, Doug Bernheim, Aaron Bodoh-Creed, Stefano DellaVigna, Emir Kamenica, Michael Katz, David Laibson, Eben Lazarus, David Levine, Ulrike Malmendier, Muriel Niederle, Devin Pope, Josh Schwartzstein, Carl Shapiro, Avner Shlain, Charles Sprenger, Joel Sobel, Alex Steiny, Philipp Strack, Tomasz Strzalecki, Neil Shephard, Itai Sher, Dan Svirsky, and Eric Zitzewitz, as well as seminar audiences at UC-Berkeley, Dartmouth, Harvard, Princeton, University of Chicago Booth, BEAM, Stanford, Brown, Yale SOM, USC, University of Pittsburgh, Penn State, and UCSD, for helpful comments.

[†]Haas School of Business, University of California, Berkeley. ned@haas.berkeley.edu

[‡]Department of Economics and Business School, Harvard University. matthewrabin@fas.harvard.edu

# 1 Introduction

Suppose we observe Abigail's evolving beliefs about whether the Republicans will win the next Presidential election: her beliefs start at 35%, jump to 54% and then 78% after two shocking revelations, drop to 44% after the release of an economic forecast, and so on, until she finally learns the actual outcome. Is Abigail changing her beliefs in a rational way? Does she have a good model of how elections are won, based on reasonable priors? One way to judge Abigail's rationality would be to compare her assessment of announcements or of uneventfulness to expert judgment of the normatively right meaning of these events. However, even if we feel ill-informed about the "right" beliefs, we can still judge beliefs changes based solely on the intrinsic properties of Bayesian updating. This paper proposes an approach to this "agnostic" investigation of the rationality of beliefs built around a basic principle of rational information processing: changes in beliefs should on average correspond to a reduction of uncertainty.

This principle can be brought to mathematical life with simple measures of *belief movement* and *uncertainty reduction* whose equivalence on average follows directly from the martingale property of Bayesian updating. We use this relationship to create statistical tests to judge the likely rationality of belief changes. We then show why and how four categories of psychological biases are likely to lead to excess or insufficient movement of beliefs relative to uncertainty reduction, providing a novel way to categorize the biases. The connection between the biases and our core concepts suggests our approach might be particularly powerful in identifying these deviations, a conclusion we support by comparing in numerical simulations our approach to other potential tests. Finally, we illustrate the empirical potential for our approach, finding (1) excess movement on average in a panel of expert predictions, (2) evidence of small deviations from Bayesian beliefs in an algorithm used to predict baseball games, and (3) equality on average between movement and uncertainty reduction in large prediction markets for individuals sports events.

In Section 2 we present our framework and notation, formalize some universal basic properties of Bayesian updating, and use those properties to create statistical tests. Our setting is simple: a Bayesian uses signals in each period $t$ to update her beliefs about the likelihood of a binary state from $\pi_t$ to $\pi_{t+1}$, leading to a *stream* of beliefs over time. We then define our main objects of interest: (1) *uncertainty* at time $t$ as $\pi_t(1-\pi_t)$ and the *uncertainty reduction* between any periods $t_1$ and $t_2$ as $r_{t_1,t_2} \equiv \pi_{t_1}(1-\pi_{t_1}) - \pi_{t_2}(1-\pi_{t_2})$; and (2) the *movement* across those periods as $m_{t_1,t_2} \equiv \sum_{\tau=t_1}^{t_2-1}(\pi_{\tau+1}-\pi_\tau)^2$. Uncertainty reduction, which can be positive or negative, captures the gain in certainty about the state given the net effect of information. Movement, which is always positive, provides a simple measure of how much the person is changing her mind over time. We employ these particular definitions of movement and uncertainty because they are simple and familiar to economists, but we discuss how they could instead be captured using more-complex forms, such as Kullback–Leibler Divergence and Shannon Entropy.[1]

[1]In each, the movement measure is a convex function, such that an immediate jump from one belief to another

Our paper centers around a strong and simple equivalence relationship between the two measures: for *any* data-generating process (DGP) determining the distribution of signals across *any* time periods, a Bayesian's expected movement must equal her expected uncertainty reduction. Intuitively, if a person's change in beliefs is rational, it should be because she is—on average—learning a corresponding amount about the state of the world. The relationship implies that we should be skeptical of a person's "Bayesianess" if we either consistently observe that she repeatedly changes her mind a relatively large amount without growing more confident, or, conversely, consistently ends up very confident but with relatively little fluctuation in beliefs. Mathematically, the equivalence is an almost-immediate implication of the fact that Bayesian beliefs are a martingale combined with the law of iterated expectations: the one-period excess movement statistic $M_{t,t+1} - R_{t,t+1}$ can be simplified as $(2\pi_t - 1)(\pi_t - \pi_{t+1})$, which must have an expectation of equal zero because $\mathbb{E}[f(\pi_0, \pi_1, ..\pi_t) \cdot (\pi_t - \pi_{t+1})] = 0$ for any *instrument* $f$ given that $\mathbb{E}[\pi_t - \pi_{t+1} | \pi_0, \pi_1, ..\pi_t] = 0$ for a martingale. We return to this formulation below. While the connection between movement and uncertainty is simple, it has not to our knowledge been used as the basis to test the rationality of belief updating.

We start by presenting a set of simple tests that demonstrate the principles of the relationship. First, a simple corollary shows that for *resolving DGPs*—where the person eventually ends up certain about the state of the world—the expected movement of the belief stream over all periods must equal the initial uncertainty $\pi_0(1 - \pi_0)$. For example, consider someone who observes two sequentially tossed fair coins and has beliefs about the likelihood of observing two heads. With probability $\frac{1}{4}$, she observes a "HT" leading to a belief stream of 25% → 50% → 0%, which has movement of $\frac{5}{16}$. She also might observe "TT","HH" or "TH," each with probability $\frac{1}{4}$, leading to movement statistics of $\frac{1}{16}$,$\frac{5}{16}$, and $\frac{1}{16}$, respectively. The expectation of these four movement statistics is equal to the initial uncertainty of $\frac{3}{16}$. The corollary implies that all other DGPs with the same prior $\frac{1}{4}$ can have different likelihoods of different signal realizations (implying different beliefs streams with different movement statistics), but the expectation of the movement statistic must always equal $\frac{3}{16}$.

Given this equality and the fact that movement must be positive, it is straightforward to statistically bound the movement in one belief stream. For example, DGPs with a prior of $\frac{1}{2}$ (and therefore initial uncertainty of $\frac{1}{4}$) cannot lead to a larger than 5% chance of a movement statistic greater than 5, because this would generate expected movement greater than $\frac{1}{4}$. Consequently, one can reject Bayesianess with 95% confidence after observing one stream with movement greater than the cutoff of 5. While it is not surprising that some belief movements are rare for a Bayesian—after all, a Bayesian can only move from 5% to 100% with 5% probability—our test allows rejection given a combination of a set of smaller belief changes regardless of the person's initial and ending beliefs. We then bound the variance of the movement distribution, which allows for a reduction

has a larger measure than a gradual change between these beliefs in small monotonic steps.

of these cutoffs (for example from 5 to 2.69 when the prior is $\frac{1}{2}$).[2] We also demonstrate an interesting asymmetry: it is not possible to statistically reject Bayesianess given a single stream of very small movements. While we can provide upper and lower statistical cutoffs when observing multiple resolving streams, the tests are very conservative because they are constructed using layered worst-case scenarios.

The test can be significantly strengthened by relying on the central limit theorem (CLT) approximation to construct test statistics. That is, given that the one-period belief movement minus the uncertainty reduction must have an expectation of zero at any period for any DGP, it is possible to perform a simple means test given any collection of belief movements from potentially different DGPs. The CLT approximation can fail: we provide examples of arbitrarily large datasets of very small belief movements that cause the approximation to be very inaccurate and lead to the common false rejection of Bayesian behavior. This failure demonstrates an important point to which we return below: because the likelihood of a belief movement under rationality is dependent on its size, the *scale* of movements is an important feature of any test. To address this issue, we introduce a simple rule of thumb, derived from numerical simulations, that this test should only be employed with datasets where the total amount of belief movement is at least 3. We also discuss the effect of measurement error and rounding: while reasonably small amounts of noise and rounding have little effect on the statistic, very large levels of error can substantially bias the statistic upward, although we discuss why this size of random error might better be considered a deviation to be detected rather than a nuisance to remove.

We then relax the binary-state assumption, considering a person who holds beliefs over many possible states. We first show that the expected sum of movement statistics for beliefs in all states must equal the expected sum of uncertainty reduction statistics. We then turn to the presumably-more-empirically-relevant situation in which the person has beliefs over possible values of some variable—like the expected value of a new good or the expected inflation rate next quarter—but only reports her changing expectation of that value over time rather than her full belief histogram. A natural conjecture would be that, given that the movement of beliefs are statistically bounded, there is some bound on the movement of expectations. In fact, a simple but strong negative result shows this is clearly false: for any observed stream of expected value predictions, there is some DGP in which that exact stream occurs with arbitrarily high probability. To understand this simple result, consider an expectations stream in which the person moves from 0.01 to 0.99 and back again many times. While this stream is clearly rare when the possible outcomes are 0 and 1 (as with beliefs), it can be extremely common for some DGP when outcomes range from -1,000 and 1,000.

This negative result reflects an important theme in the paper and crucial to understanding our results with binary beliefs: if there is no way to judge the scale of movements, there is no way to

[2] These cutoffs are not tight: given a prior of $\frac{1}{2}$, a more-complicated proof technique shows a 95%-confidence cutoff of 2.07, and given numerical simulations we believe that the actual cutoff is likely closer to 1.31.

conclude that streams are rare. In our results on binary beliefs, we were implicitly exploiting the fact that beliefs have a defined scale in the unit interval, which allowed us to determine when collections of movements were "too large" or "too small." The exact nature of the scaling is apparent given our next positive result about expectation streams: the expected movement in expectations must equal the expected reduction in variance about the value, suggesting that the second moment encodes the necessary scaling information. Crucially, because binary beliefs are defined with two outcomes, the variance is encoded in beliefs, allowing us to make statements without any further information. This fact makes beliefs particularly amenable to tests of movement.

In Section 3, we explore the connection between our binary-belief test and four psychological biases. The connections indicate why our approach may be particularly well-suited to detect (and possibly differentiate among) common departures from Bayesian updating, a theme continued in Section 4, where we compare other tests of the martingale property and their statistical performance in detecting the four biases. We start by developing a simple dynamic updating model in which the person can misweight her prior or her signal. We focus on four simple biases embedded in the model: underweighting the prior ("base-rate neglect"), overweighting the prior ("confirmation bias"), underweighting the signal ("underreaction"), and overweighting the signal ("overreaction").[3] We analyze the effect of these biases in two simple environments: (1), where the person starts with the correct prior and observes a single symmetric signal; and (2), where the person starts with the correct prior of $\frac{1}{2}$ and observes an infinite sequence of symmetric identical conditionally-independent signals.

We show how the psychological biases connect to the relationship between expected movement $\mathbb{E}M$ and uncertainty reduction $\mathbb{E}R$. In the one-shot situation, regardless of the prior or the level of bias, both base-rate neglect and overreaction always lead to $\mathbb{E}M > \mathbb{E}R$, while confirmation bias and underreaction always lead to $\mathbb{E}M < \mathbb{E}R$. While the sequential-signal environment is more complex because agents update incorrectly using priors that have accumulated past errors, the results on overreaction and underreaction hold, and the results on base-rate neglect and confirmation bias hold as long as the biases are strong enough.

Results about the sign of the difference $\mathbb{E}M - \mathbb{E}R$ obviously cannot categorize these or other biases into more than two types. To further differentiate the biases, Section 3 also explores how the biases affect movement and uncertainty reduction *separately*, by comparing $\mathbb{E}M$ and $\mathbb{E}R$ to those of a Bayesian observing the same signals, which we label $\mathbb{E}M^B$ and $\mathbb{E}R^B$. This analysis requires the departure from our driving assumption—held throughout the rest of the paper—that the normatively-correct beliefs are unobserved. The four biases have four different implications for these comparisons in our simple dynamic environment. The most intuitive case is base-rate neglect

[3] We model confirmatory bias as the opposite of base-rate neglect, and (more obviously) under-reaction to signals as the opposite of over-reaction. As we discuss in Section 3, however, we do not view the varied parameter values in our setting as reflecting different assumptions about the underlying psychological biases any individual succumbs to. Rather, they reflect different informational environments where each of the four biases will dominate a person's updating.

where, as a result of "forgetting" information, the base-rate neglector shifts beliefs more than a Bayesian but ends up being less certain: $\mathbb{E}M > \mathbb{E}M^B, \mathbb{E}R < \mathbb{E}R^B$. Conversely, confirmation bias—when strong enough—leads the person to predictably move in the direction of her original signal, leading to too little movement but over-confident beliefs in the long run: $\mathbb{E}M < \mathbb{E}M^B$ and $\mathbb{E}R > \mathbb{E}R^B$. Meanwhile, overreaction leads to $\mathbb{E}M > \mathbb{E}M^B$ and $\mathbb{E}R > \mathbb{E}R^B$ and underreaction leads to $\mathbb{E}M < \mathbb{E}M^B$ and $\mathbb{E}R < \mathbb{E}R^B$ even in the short run: over-interpreting information leads to both too much movement and causes the person to be too certain, while under-interpretation leads to too little movement and not enough certainty.

In Section 4, we compare and contrast our binary-belief method with the most common tests of Bayesian updating that also do not require knowledge of the signal DGP. To frame the discussion, recall from above that the martingale property of Bayesian beliefs implies that $\mathbb{E}[f(\pi_0, \pi_1, ..\pi_t) \cdot (\pi_t - \pi_{t+1})] = 0$ for any *instrument* $f$. Different tests of the martingale property arise from different instruments. Our test boils down to the simple instrument $(2\pi_t - 1)$, such that excess movement can interpreted as a summary statistic of the combined propensity to revert to a belief of $\frac{1}{2}$, the point of highest uncertainty.[4] But, why use this particular instrument? We are not claiming that our instrument is universally more powerful to detect all deviations from rationality: different tests based on different instruments will be better at detecting different deviations.[5] However, we present evidence that it is particularly powerful to detect deviations associated with the common psychological biases in Section 3, as well as having certain desirable characteristics, such as time-invariance, interpretability, and scale-variance.

It is instructive to first compare our test with a simple one-period *autocorrelation* test, which uses the instrument $(\pi_{t-1} - \pi_t)$ to determine if the current belief movement is correlated with the most recent belief movement. While negative autocorrelation implies that a person's next-period belief change will be downward in expectation if the previous belief change was upward, $\mathbb{E}M > \mathbb{E}R$ implies an expected downward belief change if the current belief is above $\frac{1}{2}$. Importantly, the biases from Section 3 do not predict any autocorrelation *except* the mechanical autocorrelation that comes from the fact that the relation of current beliefs to $\frac{1}{2}$ will on average reflect recent movement away from $\frac{1}{2}$ rather than towards it. Therefore, we expect our test to have more power to detect these biases, which we confirm with numerical simulations.[6] That said, there are other

[4] If $\pi_t > \frac{1}{2}$, then $(2\pi_t - 1) > 0$, implying that $(2\pi_t - 1)(\pi_t - \mathbb{E}[\pi_{t+1}|\pi_t]) > 0$ if and only if $\mathbb{E}[\pi_{t+1}|\pi_t] < \pi_t$, which occurs if the expected direction of movement is toward $\frac{1}{2}$.

[5] As a simple example, given that our test detects reversion to $\frac{1}{2}$, $\mathbb{E}M = \mathbb{E}R$ for all movements starting from a belief of $\frac{1}{2}$. Therefore, for example, our test will not reject a dataset of movements from $\frac{1}{2}$ to $\frac{3}{4}$, a clear violation of the martingale property.

[6] There are also instruments, created from different definitions of movement and uncertainty, that similarly detect reversion to $\frac{1}{2}$. For example, defining movement as the Kullback–Leibler divergence and uncertainty as Shannon Entropy yields the instrument $(-\log[1 - \pi_t] + \log[\pi_t]) = 2\tanh(2\pi_t - 1)$. As with our instrument $(2\pi_t - 1)$, this instrument tests reversion to $\frac{1}{2}$, but it is non-linear and places more weight on beliefs with more certainty. We do not claim nor suspect that our choice is optimal for all deviations or DGPs, but this of course doesn't invalidate our test. Similarly, while we show it is easy to construct alternative variance-ratio tests given our family of movement definitions, this does not invalidate the common form.

potential biases—such as overreaction to a signal that is corrected in the following period—for which autocorrelation is likely a more powerful test, just as there are other biases—such as drift toward one particular state—for which completely different tests will be more powerful.

Beyond being suited to detect particular common biases, our test does have additional benefits. First, note that autocorrelation tests require an additional stance on the definition of time-period and the amount of time required before a reversion, while our test is *time-invariant* because it detects reversion to a fixed point.[7] Second, while the interpretation of the autocorrelation statistic likely varies across situations and datasets with different timing, the (normalized) excess movement has a *consistent and portable interpretation*: a statistic of 1.2 implies that the person's movement was 20% more than expected, regardless of the timing or dataset. That is, even though instrument is simple, the statistic is endowed with meaning from our more complicated core concepts. Third, autocorrelation tests share a characteristic (common in many tests) of being invariant to scale transformations of the movement data, thereby ignoring important scale information. All forms of our test intentionally *vary with scale*—either in the test-statistic formulas or in the sample-size criteria—so that the probability of rejecting a set of beliefs increases when the belief differences are scaled up.

In Section 5, we apply our tests in three empirical contexts with three main findings. We start with a dataset borrowed from Mellers *et al* (2014) and Moore et al. (2017) where beliefs about nearly 300 geopolitical events throughout 2011-2014 were elicited continuously from over three thousand forecasters. Our first finding is that, on average, forecasters exhibit 20% more movement than resolution. Furthermore, forecasters with a relatively high level of movement for one event (compared to other forecasters) are likely to have a high level of movement on other events, suggesting that the statistic is capturing a stable individual trait. We then apply our test to the dynamic probabilistic assessments by algorithms. Specifically, we analyze the dynamic predictions of all baseball games in 2006-2010 from *Fangraphs*, which uses play-by-play information to estimate the probability of the home team winning as baseball games progress. Our second finding is that the algorithm exhibits slight excess uncertainty reduction and we speculate on how seemingly well-designed algorithms can lead to predictions that violate the movement-uncertainty relationship. Finally, we go beyond the theoretical model of the belief movements of individuals by studying sporting-event data from the large prediction market *Betfair* under the assumption that market prices map directly to Bayesian "market beliefs." This analysis leads to the third finding that, after eliminating high-frequency changes suggestive of transaction noise, movement and uncertainty reduction are close to equal on average.

We conclude in Section 6 with a broader discussion of our approach, suggesting multiple future

[7] It is easy to construct other time-invariant tests using the instrument $(\pi_t - \pi^*)$ that detect reversion to a fixed point $\pi^*$ rather than $\frac{1}{2}$. This instrument likely has more power to detect certain deviations from rationality, such as motivated beliefs toward a known $\pi^*$. However, this instrument requires knowledge of the normative meaning of states and does not occur in the biases we highlight in Section 3.

directions for research. We discuss past related research throughout the paper in the relevant sections, with the exception of two papers that employ a similar theoretical result and multiple papers on excess volatility in the finance literature, which we discuss now.

First, as our main result is simplistic and not difficult to prove, it is not surprising that related results have appeared in other papers. In a paper devoted to exploring the bounds of the *absolute* deviations of a bounded martingale, Mertens and Zamir (1977) note the equality in our Corollary 1 without comment in a footnote. Ely, Frankel, and Kamenica (2015) use the same relationship in Corollary 1 to examine different functions (including the squared deviation) of belief movements to capture preferences over different DGPs as a result of utility from "suspense" and "surprise." Finally, a continuous-time version of the movement and uncertainty relationship for semimartingales also appears in various finance papers focused on estimating real-time volatility using high-frequency trading data, such Andersen, et al. (2010) and Barndorff-Nielsen and Shephard (2001).

The closest work on excess volatility conducts empirical tests for excess movement in asset prices (Shiller (1979, 1981), LeRoy and Porter (1981), De Bondt and Thaler (1985), Campbell and Shiller (1987), and Stein (1989)). As a representative example, Shiller (1981) studies the connection between the movement of prices and the movement of dividends, finding that the variation in prices cannot be explained by the variation in dividends. Our method differs from these by testing the internal consistency of Bayesian beliefs rather than relying on relationships that must hold in a model of financial markets. Augenblick & Lazarus (2018), which applies our methodology to changes in risk-neutral beliefs inferred from asset prices, is closer to the finance literature and discusses the related work in much greater detail.

## 2 Bayesian Updating: Properties and Tests

In this section, we examine the properties of the beliefs of a Bayesian who observes information about the state of the world over time. For simplicity and exposition, we focus the discussion on belief movements in the simple binary-state case. We outline the theoretical setup in Section 2.1, discuss a simple equivalence relationship involving belief movement in 2.2, and examine how the equivalence can be leveraged to create statistical tests in Sections 2.3 and 2.4, with the former mainly demonstrating important principles and the latter providing our primary empirical test. Then, we generalize these results in two related ways. First, in Section 2.6, we allow for the existence of many states. Second, in Section 2.7, we discuss the movement of a person's expectation of a variable given changes in her beliefs about its possible realizations. The latter analysis highlights a crucial characteristic of beliefs—that they are scaled within the unit interval—which implicitly drives the previous results. Finally, Section 2.8 concludes with a discussion of some notable properties, interpretations, and caveats of the tests.

## 2.1 Setup: Data Generating Processes and Belief Streams

Consider a person observing signals across periods indexed by $t \in \{0, 1, 2, ...T\}$. At the beginning of each period $t \geq 0$, the person observes a signal realization $s_t \in S$ regarding *two* mutually exclusive and exhaustive events, which we will call *states* $x \in \{0, 1\}$. We define a history of signal realizations at period $t$ as $H_t$, with typical element $H_t = [s_1, s_2, ...s_t]$. The signals each period are drawn from a discrete signal distribution, $DGP(s_t|x, H_{t-1})$, that depends on the state, and potentially the signal history. This function defines the *data-generating process* (DGP). Our setup puts no restrictions on the DGP: signals can be non-informative, be informative at later periods, provide information about the likelihood of future signals, or be complementary with other signals. We define the probability of observing history $H_t$ induced by the DGP as $P(H_t)$.

We are interested in the beliefs about state 1 (vs. state 0) by a person observing these signals. We denote the person's initial beliefs that the state is 1 by $\pi_0$, and her belief in state 1 given history $H_t$ as $\pi_t(H_t)$. Given a history $H_t$, we define the *belief stream* $\pi(H_t) = [\pi_0, \pi_1(s_1), \pi_2(\{s_1, s_2\}), ...]$ the collection of a person's beliefs at the end of each period given this history, which we will commonly write as $\pi$ or $[\pi_0, \pi_1, \pi_2, ..., \pi_t]$ to simplify notation. We define the belief stream as *resolving* if the person always achieves certainty about the state by the final period: for all $H_T$, $\pi_T(H_T) \in \{0, 1\}$. Throughout we are mainly interested, of course, in the standard case where the person is Bayesian with a prior $\pi_0$ that matches the objective ex-ante probability of state 1, such that $\pi_t(H_t) = \frac{\pi_0 P(H_t|x=1)}{\pi_0 P(H_t|x=1)+(1-\pi_0)P(H_t|x=0)}$.

Our test focuses on a combination of two particular summary statistics about these beliefs, designed to capture the amount of belief movement induced by pieces of information and the overall effect of these signals on the person's certainty about the state of the world. The first, which we call the *movement* of belief stream $\pi$ from period $t_1$ to period $t_2 > t_1$, is defined as the sum of squared-deviations of beliefs over these periods:

$$m_{t_1,t_2}(\pi) \equiv \sum_{\tau=t_1}^{t_2-1} (\pi_{\tau+1} - \pi_\tau)^2.$$

Then, defining the *uncertainty* of belief stream $\pi$ at period $t$ as $u_t(\pi) \equiv (1 - \pi_t)\pi_t$, the second statistic is called the *uncertainty reduction* of belief stream $\pi$ from period $t_1$ to period $t_2 > t_1$, and is defined as

$$r_{t_1,t_2}(\pi) \equiv \sum_{\tau=t_1}^{t_2-1} (u_\tau(\pi) - u_{\tau+1}(\pi)) = u_{t_1}(\pi) - u_{t_2}(\pi).$$

As shorthand, we drop time subscripts to denote uncertainty reduction and movement *over the full stream*: $r(\pi) \equiv r_{0,T}(\pi)$ and $m(\pi) \equiv m_{0,T}(\pi)$. For all of these variables, we define the concomitant random variables in capital letters (such as $M_{t_1,t_2}$) with distribution determined by the likelihood of different streams induced by the DGP. Finally, we sometimes suppress the expectations bracket to simplify notation (such as $\mathbb{E}M_{t_1,t_2}$ rather than $\mathbb{E}[M_{t_1,t_2}]$).

These are not the only measures that could be used to encapsulate movement and uncertainty.

Below, we discuss a family of other formulations that could be used to similar effect, as well as showing how other natural ways of measuring movement, such as the sum of Euclidean distances, cannot provide similar results.

While nearly all of our results are agnostic to the form of the data-generating process, we will commonly illustrate concepts and provide intuitions in reference to three examples. In each, we label states 1 and 0 as "$H$" and "$L$," respectively. The first example process is the familiar symmetric binary *noisy-signal* process, in which a person observes an independent signal $s_t \in \{h, l\}$ in each period, with a conditional distribution $\Pr(h|H) = \Pr(l|L) = \theta$, where we call $\theta$ *precision.* The second process captures a politician Mr. $H$ who has $T$ days left in an election and has a probability $\phi$ of making a *gaffe* each day, such that he wins (the state is $H$) if he lasts until election without a gaffe, but loses (the state is $L$) if he commits any gaffes. In this case, the observer's beliefs with $T$ days left are $(1-\phi)^T$, which change to 0 given a gaffe on that day and $(1-\phi)^{T-1}$ given no gaffe. Finally, the third *double gaffe* example builds on the second process by adding another politician, Mrs. $L$, who takes turns with Mr. $H$. In this model, if Mr. $H$ makes a gaffe in a period, he loses the election, but if he does not make a gaffe, Mrs. $L$ has a similar chance of making a gaffe in the following period and losing the election, and so on. In this case, if the horizon is long enough ($T$ is very large), the observer's beliefs are approximately $\frac{1-\phi}{2-\phi}$ if it is Mr. $H$'s turn, which change to 0 given a gaffe and approximately $\frac{1}{2-\phi}$ given no gaffe.

To clarify our notation and example DGPs, Table 1 notes the possible signal realizations and belief streams given specific parameters of the three processes. The top panel of Table 1 considers the symmetric noisy signal DGP with $\pi_0 = \frac{1}{2}$ and $\theta = \Pr(h|H) = \Pr(l|L) = \frac{3}{4}$ for 2 periods; the middle panel the gaffe DGP with a gaffe probability $\phi$ of $\frac{1}{4}$ and three periods; and the bottom panel the long-horizon double-gaffe DGP with gaffe probability $\phi$ of $\frac{1}{4}$ for the first three periods.

## 2.2 Relating Movement and Uncertainty Reduction

Our paper centers around the surprisingly strong relationship between movement and uncertainty reduction in any DGP:

**Proposition 1** ***(Implication of Martingale Property for Beliefs)*** *For any DGP and for any periods $t_1$ and $t_2$, the expected movement from period $t_1$ to period $t_2$ must equal the expected uncertainty reduction from period $t_1$ to period $t_2$: $\mathbb{E}M_{t_1,t_2} = \mathbb{E}R_{t_1,t_2}$.*

The proof is straightforward and instructive (and, as with all proofs, is in the Appendix). First, it is possible to rewrite the one-period difference as:

$$\mathbb{E}M_{t,t+1} - \mathbb{E}R_{t,t+1} = \mathbb{E}[(2\pi_t - 1) \cdot (\pi_t - \pi_{t+1})].$$

Table 1: Example DGPs, Belief Streams, and Movement and Uncertainty Resolution Statistics

| Signals $[s_1, s_2, ...]$ | Prob $P(H_t)$ | Bayesian Stream $\pi([s_1, s_2, ...])$ | Movement / Uncertainty Reduction $m_{0,1}$ | $r_{0,1}$ | $m_{1,2}$ | $r_{1,2}$ | $m$ | $r$ | $m-r$ |
|---|---|---|---|---|---|---|---|---|---|
| (1) Symmetric noisy-signal DGP with $\theta = \frac{3}{4}$ for two periods | | | | | | | | | |
| $[l, l]$ | 5/16 | [1/2 1/4 1/10] | 1/16 | 1/16 | 9/400 | 39/400 | 17/200 | 4/25 | −3/40 |
| $[l, h]$ | 3/16 | [1/2 1/4 1/2] | 1/16 | 1/16 | 1/16 | −1/16 | 1/8 | 0 | 1/8 |
| $[h, l]$ | 3/16 | [1/2 3/4 1/2] | 1/16 | 1/16 | 1/16 | −1/16 | 1/8 | 0 | 1/8 |
| $[h, h]$ | 5/16 | [1/2 3/4 9/10] | 1/16 | 1/16 | 9/400 | 39/400 | 17/200 | 4/25 | −3/40 |
| Average | | [1/2 1/2 1/2] | 1/16 | 1/16 | 3/80 | 3/80 | 1/10 | 1/10 | 0 |
| (2) Gaffe DGP with $\phi = \frac{1}{4}$ and $T = 3$ (resolving) | | | | | | | | | |
| No gaffe | 1/8 | [1/8 1/4 1/2 1] | 1/64 | −5/64 | 1/16 | −1/16 | 21/64 | 7/64 | 7/32 |
| Week-1 gaffe | 1/2 | [1/8 0 0 0] | 1/64 | 7/64 | 0 | 0 | 1/64 | 7/64 | −3/32 |
| Week-2 gaffe | 1/4 | [1/8 1/4 0 0] | 1/64 | −5/64 | 1/16 | 3/16 | 5/64 | 7/64 | −1/32 |
| Week-3 gaffe | 1/8 | [1/8 1/4 1/2 0] | 1/64 | −5/64 | 1/16 | −1/16 | 21/64 | 7/64 | 7/32 |
| Average | | [1/8 1/8 1/8 1/8] | 1/64 | 1/64 | 1/32 | 1/32 | 7/64 | 7/64 | 0 |
| (3) Double-Gaffe DGP with $\phi = \frac{1}{4}$ and large $T$ for three periods | | | | | | | | | |
| No gaffe | 27/64 | [3/7 4/7 3/7 4/7] | 1/49 | 0 | 1/49 | 0 | 3/49 | 0 | 3/49 |
| Week-1 gaffe | 1/4 | [3/7 0 0 0] | 9/49 | 12/49 | 0 | 0 | 9/49 | 12/49 | −3/49 |
| Week-2 gaffe | 3/16 | [3/7 4/7 1 1] | 1/49 | 0 | 9/49 | 12/49 | 10/49 | 12/49 | −2/49 |
| Week-3 gaffe | 9/64 | [3/7 4/7 3/7 0] | 1/49 | 0 | 1/49 | 0 | 11/49 | 12/49 | −1/49 |
| Average | | [3/7 3/7 3/7 3/7] | 3/49 | 3/49 | 9/196 | 9/196 | 111/784 | 111/784 | 0 |

Note: This table provides bayesian beliefs given different signal realizations for three DGPs discussed in the main text, along with the corresponding movement and uncertainty reduction statistics defined in the main text. As the third DGP has a large number of periods, this statistic represents $m_{0,3}$ rather than $m$.

The only requirement on Bayesian beliefs is that they are bounded in the unit interval and satisfy the martingale property: $\mathbb{E}[\pi_{t+1}|\pi_0, \pi_1, ..\pi_t] = \pi_t$. This property implies that, for any *instrument* $f$:

$$\mathbb{E}[f(\pi_0, \pi_1, ..\pi_t) \cdot (\pi_t - \pi_{t+1})] = 0.$$

$\mathbb{E}M_{t,t+1} - \mathbb{E}R_{t,t+1}$ is equivalent to the use of the instrument $(2\pi_t - 1)$ and therefore must equal zero in expectation. Consequently, the sum must also equal zero: $\mathbb{E}M_{t_1,t_2} - \mathbb{E}R_{t_1,t_2} = \sum_{\tau=t_1}^{t_2-1}(\mathbb{E}M_{\tau,\tau+1} - \mathbb{E}R_{\tau,\tau+1}) = 0$. In Section 4, we discuss the properties and advantages of this type of instrument. This formulation implies that excess movement occurs when the expectation of posteriors deviates from the prior in the direction of a belief of $\frac{1}{2}$. For example, the statistic will be positive when $\pi_t = \frac{3}{4}$ and $\mathbb{E}[\pi_{t+1}] < \frac{3}{4}$, or if $\pi_t = \frac{1}{4}$ and $\mathbb{E}[\pi_{t+1}] > \frac{1}{4}$. Conversely, the statistic is negative when the expectation of posteriors deviates from the prior away from $\frac{1}{2}$. Under this interpretation, the statistic summarizes the tendency to revert to the point of highest uncertainty.[8]

[8] Note that the equation then also implies that the statistic is always zero if $\pi_t = \frac{1}{2}$, *regardless* of the person's posterior beliefs $\pi_{t+1}$, implying that our test will never detect errors in updating from a uniform prior.

As an example of the equivalence, note that in the top example from Table 1 above, the uncertainty reduction from period 1 to period 2 $R_{1,2}$ takes the value of $-\frac{1}{16}$ or $\frac{39}{400}$ and movement $M_{1,2}$ $\frac{1}{16}$ or $\frac{9}{400}$ with probabilities $\frac{3}{8}$ and $\frac{5}{8}$, respectively. Therefore, the expected uncertainty reduction $\mathbb{E}R_{1,2}$ and expected movement $\mathbb{E}M_{1,2}$ between these periods are equal at $\frac{3}{80}$, as implied by the Proposition. Similarly, $\mathbb{E}M_{0,1} = \mathbb{E}R_{0,1}$ and $\mathbb{E}M = \mathbb{E}R$. The Proposition additionally implies that $\mathbb{E}[M_{t_1,t_2}|H_{t_0}] = \mathbb{E}[R_{t_1,t_2}|H_{t_0}]$ for any history with $t_0 \leq t_1$, because everything that follows history $H_{t_0}$ is equivalent to a new DGP with prior $\pi(H_{t_0})$. Therefore, in the table, $\mathbb{E}[M_{1,2}|(s_1 = l)] = \mathbb{E}[R_{1,2}|(s_1 = l)]$ and $\mathbb{E}[M_{1,2}|(s_1 = h)] = \mathbb{E}[R_{1,2}|(s_1 = h)]$. Note that all of the analogous relationships hold in the second and third examples in Table 1.

Aggregating these statistics for all time periods between period 0 and $T$ for resolving streams (in which the uncertainty at the final period must equal zero) leads to a simple corollary:

**Corollary 1** ***(Resolving Streams)****: For any resolving DGP, the expected movement over the entire stream must equal the expected uncertainty reduction:* $\mathbb{E}M = u_0 = \pi_0(1-\pi_0)$.

Given that the uncertainty at the end of a resolving stream is zero, the expected uncertainty reduction over the entire stream must be equal to the initial uncertainty. Consequently, for example, the Bayesian beliefs following *any* resolving DGP with $\pi_0 = \frac{1}{2}$ must have $\mathbb{E}M = \frac{1}{4}$. While every DGP with the same prior induces a potentially different distribution of movement statistics from different belief streams associated with different signal realizations, the expected value of these distributions must be the same for any DGP.

Analogous relationships are also true for other measurements of movement and uncertainty reduction. Specifically, for any decreasing *surprisal* function $\gamma(\pi_t)$ with $\gamma(1) = 0$, there is a corresponding definition of uncertainty $u_t(\pi) \equiv \pi_t\gamma(\pi_t) + (1-\pi_t)\gamma(1-\pi_t)$, uncertainty reduction $r_{t_1,t_2}(\pi) \equiv u_{t_1}(\pi) - u_{t_2}(\pi)$, and movement $m_{t_1,t_2}(\pi) \equiv \sum_{\tau=t_1}^{t_2-1} \pi_{\tau+1}(\gamma(\pi_\tau) - \gamma(\pi_{\tau+1})) + (1-\pi_{\tau+1})(\gamma(1-\pi_\tau) - \gamma(1-\pi_{\tau+1}))$. Following the same logic as above, $\mathbb{E}M_{t,t+1} - \mathbb{E}R_{t,t+1}$ can be rewritten as $\mathbb{E}[(\gamma(1-\pi_t) - \gamma(\pi_t)) \cdot (\pi_t - \pi_{t+1})]$, which implies that $\mathbb{E}M_{t,t+1} - \mathbb{E}R_{t,t+1} = 0$ for a Bayesian and therefore $\mathbb{E}M_{t_1,t_2} - \mathbb{E}R_{t_1,t_2}$. That is, each surprisal definition maps to a particular instrument $(\gamma(1-\pi_t) - \gamma(\pi_t))$ and $\mathbb{E}M_{t,t+1} - \mathbb{E}R_{t,t+1}$ continues to capture the tendency toward $\frac{1}{2}$, albeit using different weights which are a function of the surprisal function and the belief $\pi_t$. In subsequent work, Frankel and Kamenica (2019) show that some surprisal functions can lead to undesirable properties like negative movement and consequently refine the set of possible functions to remove this possibility. Our definitions of movement and uncertainty arise from the surprisal function $\gamma(\pi_t) = (1-\pi_t)^2$. An alternative natural surprisal function is $\gamma(\pi_t) = -\log_b(\pi_t)$ for some base $b$, such that the corresponding uncertainty measure is Shannon Entropy and the resulting movement measure is the sum of the Kullback–Leibler divergences of beliefs from one period to the next. We do not demonstrate nor suspect that our particular choice of surprisal functional form or resultant instrument is statistically "optimal" in any general way. As with similar tests (such as the variance-ratio test), we based our decision mainly due to its resultant transparent formulas and familiarity to economists.

## 2.3 Tests of Total Movement of Resolving Streams

In the next section, we will use the equivalence relationship above, in combination with the Central Limit Theorem (CLT), to create practical tests of too much or too little movement for streams of any length from multiple DGPs. Before turning to those tests, we explore some results that focus on resolving streams to illustrate the potential of the equivalence relationship and help lay out some of the principles and intuitions. Despite their conceptual usefulness, these tests do not employ the CLT and are consequently very conservative.

We start with another corollary to Proposition 1, which suggests the potential to construct a test of too much movement. Recalling that $u_0 = \pi_0(1 - \pi_0)$:

**Corollary 2** ***(Bounding Likelihood of Large Movements over One Stream)****: For all $\delta \in (0,1)$, $\Pr\left(M \geq \frac{u_0}{\delta}\right) \leq \delta$ for any DGP with prior $\pi_0$ .*

This is a simple application of Markov's inequality. As an example, consider a resolving DGP with an initial belief of $\frac{1}{2}$, which implies that $u_0 = \frac{1}{4}$. Plugging in $\delta = .05$, the corollary implies that movement cannot be greater than 5 more than 5% of the time. Intuitively, because movement is never negative, that movement distribution would imply that $\mathbb{E}M > .05 \cdot 5 = .25 = u_0$, which would violate Corollary 1. Therefore, if one observes a single empirical stream with movement greater than 5, one can reject the null hypothesis that the probability stream arose from Bayesian updating at the 95% confidence level.

We note that the possibility of rejecting a single stream of beliefs as being unlikely for a Bayesian is quite unsurprising. If a Bayesian in our model believes an event is only 5% likely, a shift to 100% should not happen more than 5% of the time. Consequently, if our only observation from a person is a movement from 5% to 100%, we can reject that the person is a Bayesian at the 95% confidence level. The crux of our tests is that the combination of more frequent yet less violent changes in beliefs is not likely either.

Here, it is useful to note that other obvious functional forms for movement will not produce similar results. For example, if movement is defined as the sum of Euclidean distances ($\sum_{\tau=1}^{T} |\pi_\tau - \pi_{\tau-1}|$), a statement similar to Corollary 2 is not possible. In fact, a known result is that there exists a DGP in which the sum of the Euclidean distances *in all streams* is (almost) always greater than any given critical value.[9]

[9] More formally, for all $\delta \in (0,1)$ and critical value $m^* > 0$, there is some DGP—one that induces beliefs to follow a simple random walk with a very small step size—with prior $\pi_0$ such that $\Pr\left(\sum_{t=1}^{T} |\pi_\tau - \pi_{\tau-1}| \geq m^*\right) \geq \delta$.

Next we improve on the bounds in Corollary 2, which only uses the fact that the movement distribution has a weakly positive support and has a specific expected value. By bounding the variance of this distribution in Proposition 2, we can state Corollary 3, which uses the one-sided Chebyshev's Inequality to further restrict the likelihood of observing a given level of movement:

**Proposition 2** ***(Bound on Variance of Movement Distribution)***: *For any resolving DGP with prior $\pi_0$, $var(M)$ must be less than $u_0(\frac{3}{2} - u_0)$.*

**Corollary 3** ***(Reduction of Bound in Corollary 2)***: *For all $\delta \in (0, 1)$,*
$$\Pr\left(M > u_0 + \sqrt{u_0(\tfrac{3}{2} - u_0)}\frac{\sqrt{(1-\delta)}}{\sqrt{\delta}}\right) \leq \delta \text{ for any DGP with prior } \pi_0.$$

This implies, for example, that we can bring down the 5% cutoff above for a resolving DGP with $\pi_0 = \frac{1}{2}$ from 5 to 2.69 and one with $\pi_0 = \frac{1}{10}$ from 1.8 to 1.64. We do not expect that any of these cutoffs are binding, and we can prove, via another proof technique, that the bound when $\pi_0 = \frac{1}{2}$ and $\delta = .05$ must be less than 2.065. We also believe, by numerically solving for all recursive possibilities, that the bound is closer to 1.31.[10,11]

Conversely, it is difficult to reject single streams for having too little movement (i.e. too much uncertainty reduction), because the movement can be very small with high probability for some data-generating processes. For example, consider the gaffe DGP discussed above with many periods and the likelihood of the gaffe chosen to be $\phi = 1 - \frac{1}{2}^{\frac{1}{T}}$, such that the politician makes no gaffe over all periods with probability $\pi_0 = \frac{1}{2}$. Whenever the politician wins, beliefs must have shifted upward by a tiny amount each period, leading the sum of squared movements to be tiny (and tending to zero as $T$ increases). Therefore, without an assumption ruling out this type of DGP, it is impossible to reject—with any reasonable level of confidence—that a person is Bayesian when observing even tiny total movement in one stream:

[10] The logic behind the proof of the 2.065 cutoff can be seen in the less-complex proof of a 4 cutoff. For the sake of contradiction, suppose there was a DGP where $M$ was greater than 4 more than 5% of the time. Consider some full history $H_T$ with $m_{0,T}(\boldsymbol{\pi}(H_T)) > 4$. Given that movement can rise by at most 1 after one period, there exists a subhistory $H_t \in H_T$ with $2 \leq m_{0,t}(\boldsymbol{\pi}(H_t)) \leq 3$. By Corollary 1, for all streams that share history $H_t$, the expected movement from time $t$ to $T$, $\mathbb{E}M_{t,T}|H_t$, must equal $\pi(H_t)(1 - \pi(H_t))$, which must be $\frac{1}{4}$ or lower. For the original history $H_T$ with movement more than 4, $m_{t,T}(H_T) \geq 1$ because $m_{0,T}(H_T) \geq 4$ and $m_{0,t}(H_T) \leq 3$. Therefore, for $\mathbb{E}M_{t,T}|H_t < \frac{1}{4}$, there must exist other streams – denoted $\boldsymbol{\pi}(H'_T)$ – with $m_{t,T}(\boldsymbol{\pi}(H'_T)) < 1$. Given these streams share history $H_t$, it must be that $2 \leq m_{0,t}(\boldsymbol{\pi}(H'_T)) \leq 3$, but given that $m_{t,T}(\boldsymbol{\pi}(H'_T)) < 1$, it must be that $2 \leq m_{0,T}(\boldsymbol{\pi}(H'_T)) < 4$. It is then possible to show that there must be so many of these streams that $\mathbb{E}M > \frac{1}{4}$, which contradicts Corollary 1.

[11] Upper bounding of movement in this way was possible given modification of previous results in the statistics literature, but with a much looser bound: Burkholder (1973) proved a general result for martingales which implies, in our environment, $\Pr(M > \lambda^2) \leq \frac{3}{\lambda}$ for any prior, leading a single-stream result that $\Pr(M > 3600) \leq .05$, which was improved by Cox (1982) to show that $\Pr(M > \lambda^2) \leq \frac{\sqrt{e}}{\lambda}$ , implying that $\Pr(M > 1087) \leq .05$. These results do not employ the restriction that beliefs only exist on the unit interval, which is why we are able to improve on the bound by nearly three orders of magnitude.

**Corollary 4** ***(Difficulty Bounding Small Movements over One Stream)****: For any* $\delta < \max(\pi_0, 1-\pi_0)$ *and critical value* $m^* > 0$*, there is some DGP with prior* $\pi_0$ *such that* $\Pr(M < m^*) \geq \delta$*.*

While we cannot make a meaningful statement about lower bounds with one stream, we can provide lower (and upper) bounds on the sum of the movement in multiple streams from potentially different DGPs:

**Corollary 5** ***(Bounding Likelihood of Movement over Multiple Streams)****: For all* $\delta \in (0,1)$, $\Pr\left(|\sum M - n \cdot u_0| > \frac{\sqrt{n \cdot u_0(\frac{3}{2} - u_0)}}{\sqrt{\delta}}\right) \leq \delta$*, for any set of* $n$ *independent streams from* $n$ *DGPs with prior* $\pi_0$*.*

Corollary 5 implies, for example, that the combined movement from 200 streams from any set of DGPs with $\pi_0 = \frac{1}{2}$ has a mean of 50 and lies between 14.6 and 85.4 with 95% probability (for 20 streams, there is no lower bound, but the upper bound is 16.2).

It is possible to use the results in this section as statistical tests of Bayesianess given resolving belief streams. However, we will largely not take this path in this paper, instead seeing these results as expositionally useful for demonstrating that it is possible to statistically reject certain belief streams based on movement even when the individual movements are not large. The two main issues with the empirical applicability of the tests are that (1) the bounds are constructed from layered worst-case scenarios and are therefore fairly conservative, and (2) creating similar tests that allow the aggregation of non-resolving belief streams from different DGPs with different priors is possible, but leads to even more conservative bounds.[12,13] We therefore focus on another test of the relationship in the next section.

## 2.4 Main Test with multiple streams and DGPs

We now move away from tests over entire streams and instead focus on more general tests of belief movements across different DGPs with potentially different priors. Recall that Proposition 1 establishes that each *one-period excess movement* random variable, $M_{t,t+1} - R_{t,t+1}$ for all $t$, has an expectation of zero for any DGP. Therefore, given observations from various independent

[12] As an example of the conservative conclusions, consider the symmetric noisy signal process in the first example above with a large number of periods and precision $\theta = \frac{3}{4}$. Numeric simulations suggest that the correct 5% cutoff is around 0.653, with only 0.2% of streams having movement of 1.31 or greater and less than 0.1% having movement greater than 2.07.

[13] For a similar test of a combination of one-period movements within and across DGPs, note that the one-period excess movement statistic $M_{t,t+1} - R_{t,t+1}$ given $\pi_t > \frac{1}{2}$ is bounded above by $(2\pi_t - 1)\pi_t$ (when $\pi_{t+1} = 0$) and below by $(2\pi_t - 1)(\pi_t - 1)$ (when $\pi_{t+1} = 1$), with these equations reversed when $\pi_t \leq \frac{1}{2}$, leading to an absolute difference of bounds of $|(2\pi_t - 1)|$. Therefore, the Hoeffding two-tailed concentration inequality implies that $P[|\overline{M_{t,t+1} - R_{t,t+1}}| \geq m^*] \leq 2\exp(\frac{-2n^2 m^{*2}}{\sum(2\pi_t - 1)^2})$ given $n$ single-period observations.

belief streams of potentially-different lengths from potentially-different DGPs, *all* of the one-period excess movement variables have an expectation of zero. Furthermore, it is easy to show that these variables must satisfy the conditions for a simple Martingale CLT, suggesting the possibility of using a standard means test to determine if the average excess movement in each period is equal to zero.[14] Specifically, given $n$ observations – indexed by $i$ – of one-period movement $m^i_{t,t+1}$ and one-period uncertainty reduction $r^i_{t,t+1}$ for any time period and for any DGP, we define the *one-period excess-movement statistic* as:

$$X = \overline{m}_{t,t+1} - \overline{r}_{t,t+1},$$

where $\overline{m}_{t,t+1} \equiv \frac{1}{n}\sum_i m^i_{t,t+1}$ and $\overline{r}_{t,t+1} \equiv \frac{1}{n}\sum_i r^i_{t,t+1}$. Under the null hypothesis that the streams were generated by Bayesian updating, the test statistic:

$$Z \equiv (\frac{\sqrt{n}}{s_{t,t+1}})(\overline{m}_{t,t+1} - \overline{r}_{t,t+1}) \rightarrow N(0,1) \text{ as } n \rightarrow \infty,$$

where $s_{t,t+1} \equiv \sqrt{\frac{1}{n}\sum(m^i_{t,t+1} - r^i_{t,t+1} - \overline{m}_{t,t+1} - \overline{r}_{t,t+1})^2}$.[15]

Using the test statistic in a finite sample requires the additional assumption that the CLT limiting distribution is a reasonable approximation. The CLT approximation is used in almost all empirical studies in economics, where authors either implicitly assume its appropriateness or justify its use with some simple rule-of-thumb based on the number of observations. However, in our setting, basing the rule solely on the number of observations is problematic, an issue highlighted by again considering the gaffe example discussed above with $\phi = 1 - \frac{1}{2}^{\frac{1}{T}}$. In this case, the Bayesian belief stream after the politician wins contains an arbitrarily large number $T-1$ of universally negative (albeit small) one-period excess-movement statistics, leading a means test to wrongly reject Bayesianess with arbitrarily high confidence. Therefore, we require a rule that additionally takes the *scale* of movements into account. That is, as we discuss in more detail below, the fact that belief movements are defined on the unit interval means that belief movements are not scale-invariant: 100 movements from 0.01 to 0.99 should lead to a different evaluation of a person's Bayesianess than 100 movements from 0.499 to 0.501. Following this observation, we searched for a rule-of-thumb combining the number of observations and scale, finding through numerical simulations that the CLT appears appropriate when the total movement across all streams is at

[14] For example, given that the one-period excess movement statistics are bounded and have finite variance, the sum of the statistics satisfies the Martingale CLT in Theorem 3.3 of Hall & Heyde (1980). The CLT requires one additional assumption that we have not explicitly stated: the observations must come from *changes* in beliefs. That is, the CLT obviously will not hold given an almost-surely constant belief over time.

[15] Under the null hypothesis, $\mathbb{E}[\overline{m}_{t,t+1} - \overline{r}_{t,t+1}] = 0$, such that $s_{t,t+1}$ could alternatively be defined as $\sqrt{\frac{1}{n}\sum(m^i_{t,t+1} - r^i_{t,t+1})^2}$. We use the more standard formula in the text as it is more easily implementable using statistical software.

least 3.[16,17]

While reliance on the CLT is a cost, it has great benefits. By no longer focusing on worst-case scenario distributions and DGPs, the test has much greater power as we demonstrate in Section 5 of the paper. Furthermore, we show in Section 4 that this power, at least for our example DGPs and using the rule-of-thumb above, does not come at the cost of high levels of false rejection of Bayesianess.

Finally, we note that interpreting any given magnitude of the excess-movement statistic might depend on the prior. Average excess movement of .01 may have a different meaning if the uncertainty reduction is .25 instead of .01, because the former provides more potential for movement. Without taking a firm position on this issue, we present the *normalized excess-movement* created by dividing the excess movement by the average uncertainty reduction:

$$X_{norm} \equiv \frac{\overline{m}_{t,t+1}}{\overline{r}_{t,t+1}} = \frac{X}{\overline{r}_{t,t+1}} + 1,$$

which has an asymptotic expectation of 1 under the null.[18] A normalized statistic of 1.1 suggests that there is 10% more movement than expected given uncertainty reduction. We propose that researchers report both the standard and normalized statistic along with the test statistic $Z$

[16] The rule-of-thumb is naturally not based on a formal result or proof. Instead, as with other rules of thumb – such as requiring $\frac{5}{\min(p,1-p)}$ observations to use the normal approximation of the binomial distribution given a success probability of $p$ – it is designed to work "good enough" for most situations. To determine a reasonable total-movement rule, we (1) choose over 300,000 mixtures of different one-period excess movement distributions possible under the null hypothesis, (2) for each mixture, we simulated 1000 samples that contained enough observations to satisfy different total-movement rules (3) for each rule-mixture-sample combination, we conducted a 95% confidence means test using the normal approximation. If the normal approximation is appropriate for a given rule-mixture, the rejection rates across samples should be "near" 5%. For total-movement rules of .5, 1, 1.5, 2, 2.5, and 3, the sample rejection rates were below 10% for 73%, 90%, 96%, 96%, 97%, and 97% of the mixtures and below 7.5% for 61%, 74%, 82%, 90%, 93%, 93% of the mixtures, respectively. That is, for that for the vast majority of mixtures, the normal approximation works reasonably well even given samples with total movement over 1 or 1.5. However, there are some "extreme" mixtures (particularly those with symmetric movements starting at .5 as well as large asymmetric movements starting at the poles) that have high rejection probabilities even as total movement rises past 5, suggesting that the normal approximation is inappropriate. We arbitrarily choose a rule of 3 to balance these situations given that the rejection rates appear to level off at this point.

[17] Given that expected total movement equals initial uncertainty for resolving streams, an alternative when using data of full resolving streams is to require $\sum_s \pi_0^s(1-\pi_0^s) \geq 3$ where streams are indexed by $s$.

[18] Given that $\overline{R}_{t,t+1} \to E[R_{t,t+1}] > 0$ and $\overline{M}_{t,t+1} \to E[M_{t,t+1}]$ in probability (and therefore in distribution) as $n \to \infty$, Slutsky's Theorem (Lemma 2.8 in Van der Vaart(2000)) implies that $\frac{\overline{M}_{t,t+1}}{\overline{R}_{t,t+1}} \to \frac{E[M_{t,t+1}]}{E[R_{t,t+1}]} = 1$ in probability as $n \to \infty$. It is useful to clarify that $X_{norm} \neq \frac{1}{n}\sum \frac{m^i_{t,t+1}}{r^i_{t,t+1}}$, which does *not* have a constant expectation across DGPs. For completeness, we note that given $n_s$ resolving streams indexed by $s$, it is possible to use a full-stream-ratio test statistic $\frac{1}{n_s}\sum_s \frac{m^s}{r^s}$. This statistic does have an expectation of 1 as $r^s$ is constant within a stream. We prefer $X_{norm}$ even given full streams because the variance of this alternative test statistic can be extremely high for some DGPs.

above.[19,20]

## 2.5 Effect of Measurement Error and Rounding

If beliefs are mismeasured either due to measurement error or rounding, the equivalence relationship will no longer hold. In this section, we demonstrate that while measurement error always leads to excess movement, the effect is relatively small for reasonable amounts of mismeasurement. Then, we numerically show that the effect of rounding is also reasonably small with erratic sign and size.

For a simple model of observational noise, we consider a person who holds a Bayesian belief, but reports a belief distorted by an additive error: $\widehat{\pi}_t = \pi_t + \varepsilon_t$, where the error term is mean-zero with variance $\sigma^2_{\varepsilon_t}$ and uncorrelated with recent belief and error realizations ($\mathbb{E}[\varepsilon_t \pi_t] = \mathbb{E}[\varepsilon_t \pi_{t-1}] = \mathbb{E}[\varepsilon_t \varepsilon_{t+1}] = 0$). Although we suppress the notation for readability, $\varepsilon_t$ can be function of history $H_t$: this is necessary to keep measured beliefs constrained to the unit interval. When beliefs become more extreme, the error must become small to maintain zero mean: that is, $\sigma^2_{\varepsilon_t} \to 0$ as $\pi_t \to \{0, 1\}$. Given this model of noise, the expectation of the excess movement statistic is not zero, but rather $2\sigma^2_{\varepsilon_t} > 0$.[21] For intuition on the magnitude of this effect, we simulate a simple DGP in which a person starts with a prior $\pi_0 = \frac{1}{2}$ and observes an independent, symmetric signal $s_t \in \{h, l\}$, with a conditional distribution $\theta \equiv \Pr(h|H) = \Pr(l|L) = .75$ for 20 periods until realization. We then assume that measured beliefs are drawn from the Beta Distribution centered at the Bayesian belief with parameters chosen such that at a Bayesian belief of $\frac{1}{2}$, 95% of measured beliefs fall within an interval of length $\Delta$. For distortion parameters $\Delta$ of 0, 0.01, 0.05, 0.10, and 0.20, the normalized excess movement is 1, 1.001, 1.006, 1.022, 1.084, respectively. That is, the statistic is relatively unaffected for reasonably small measurement distortions but does rise steeply if the person's beliefs move substantially due to noise. While this bias is concerning, it might be also appropriate to conclude that the person is not Bayesian if large fluctuations actually represent random changes in the person's genuinely-held beliefs rather than simple measurement

[19] We propose partnering the statistic $X_{norm}$ with the test statistic $Z$ while recognizing that $Z$ does not take into account the variance in $\overline{r}_{t,t+1}$ in the demoninator of $X_{norm}$. We do this instead of constructing a different test statistic $Z_{norm}$ for two reasons. First, the standard sample variance for $\overline{r}_{t,t+1}$ will be biased upwards in many cases. For example, suppose that $n$ one-period observations constitute different times in a set of resolving streams. Then, for each stream, $\sum r_{t,t+1}$ is always equal to the constant $\pi_0(1-\pi_0)$. and therefore $\overline{r}_{t,t+1}$ will be a constant with no variance, while the sample variance will not be zero. Second, even if there was a more appropriate statistic to measure variance, non-full-stream data will lead $|Z_{norm}| < |Z|$. However, we are not fundamentally interested in whether $X = 0$ or $X_{norm} = 1$, but rather whether we can reject the null hypothesis of Bayesian updating, which implies both of these equalities. Therefore, we choose to focus on the stronger test statistic $Z$.

[20] One nice feature is that the one-period normalized statistic is always equal to the natural full-stream normalized statistic: that is, if the dataset consists only of $n_s$ resolving streams indexed by $s$, then $\frac{\overline{m}_{t,t+1}}{\overline{r}_{t,t+1}} = \frac{\overline{m}}{\overline{r}}$ where $\overline{m} \equiv \frac{1}{n_s}\sum_s m^s$ and $\overline{m} \equiv \frac{1}{n_s}\sum_s m^s$. In contrast, $\overline{m} - \overline{r} \neq \overline{m}_{t,t+1} - \overline{r}_{t,t+1}$ for multi-period streams as the average excess movement across an entire stream is larger than the average one-period excess movement.

[21] Recall that the one-period excess movement statistic $M_{t,t+1} - R_{t,t+1}$ can be written as $(2\pi_t - 1)(\pi_t - \pi_{t+1})$. Simple algebra shows that $\mathbb{E}[(2\widehat{\pi}_t - 1)(\widehat{\pi}_t - \widehat{\pi}_{t+1})] = \mathbb{E}[(2\pi_t - 1)(\pi_t - \pi_{t+1})] + \mathbb{E}[4\pi_t\varepsilon_t + 2\varepsilon_t^2 - \varepsilon_t - 2\varepsilon_t\pi_t - 2\pi_t\varepsilon_{t+1} - 2\varepsilon_t\varepsilon_{t+1} + \varepsilon_{t+1}]$, which is equal to $2\sigma^2_{\varepsilon_t}$ given the zero-mean and zero-correlation assumptions on the error term.

error. Crucially, as we discuss in the empirical results in Section 5, to produce the normalized statistic of 1.2 from our individual-level dataset would require an error parameter $\Delta$ calibrated to be nearly .5, which seems far from reasonable measurement error.

Rounding is perhaps more concerning in individual-level data: it is likely that people report 65% rather than 66.2% and potentially 50% rather than 45%. Here, it more difficult to state formal results because the effect depends on the exact coarseness of rounding and the specific DGP. From numerical simulations, it appears the effect of rounding is commonly small, with both the size and sign changing erratically. For example, for the DGP above, the normalized excess movement for a person who rounds to the nearest 1%, 2%, 5%, and 10% is 0.998, 0.999, 1.001, and 1.094, respectively. But, this effect is unstable across DGPs: given precision parameter $\theta$ of 0.65, 0.70, 0.75, 0.80, and 0.85, the average total movement given 5% rounding is 1.050, 1.046, 0.999, 1.022, and 0.986, respectively. Therefore, while it is very appropriate to be concerned about rounding in individual-level data, the statistic does not appear to be dramatically affected and the direction of the bias is unclear.

## 2.6 Many States

It is relatively straightforward to generalize the model to allow for many states, such that $x \in \{0, 1, 2...\}$. That is, a person has a histogram of beliefs about different possible levels of the stock market in three months or the potential number of points that will be scored in a football game. In this case, the belief in state $x$ at time $t$ is denoted $\pi_t^x$ and a belief stream $\pi = (\pi_0^0, \pi_0^1, \pi_0^2...), (\pi_1^0, \pi_1^1, \pi_1^2, ...), (\pi_2^0, \pi_2^1, \pi_2^2, ...)....$ The adjusted movement and uncertainty statistics are constructed by simply adding the binary statistics over all states: that is, the *total movement* from period $t_1$ to period $t_2$ defined as $m_{t_1,t_2}^{\Sigma}(\pi) \equiv \sum_{x \in X} \sum_{\tau=t_1+1}^{t_2} (\pi_\tau^x - \pi_{\tau-1}^x)^2$, the *total uncertainty* at period $t$ as $u_t^{\Sigma} \equiv \sum_{x \in X} (1 - \pi_t^x)\pi_t^x$ and the total uncertainty reduction from period $t_1$ to period $t_2$ as $r_{t_1,t_2}^{\Sigma}(\pi) \equiv u_{t_1}^{\Sigma}(\pi) - u_{t_2}^{\Sigma}(\pi)$, with the related capital-letter random variables defined as above.[22] Then:

**Proposition 3** ***(Generalization of Proposition 1 with Many States)*** *For any DGP and for any periods $t_1$ and $t_2$, $\mathbb{E}M_{t_1,t_2}^{\Sigma} = \mathbb{E}R_{t_1,t_2}^{\Sigma}$.*

The intuition of this Proposition is simple: beliefs about each state $x$ can be interpreted as a binary belief, and so Proposition 1 implies that the expected movement must equal the expected

[22] Note one slight irritating mismatch with the binary-state definition: the binary statistics do not add movement and uncertainty reduction over both states, but rather only use one state (because the statistics are the same for both states). We did this so that the binary results were simpler and more transparent. A more consistent definition of the binary movement would have been $m_{t_1,t_2} = 2 \cdot \sum_{\tau=t_1+1}^{t_2} (\pi_\tau - \pi_{\tau-1})^2$ and binary uncertainty as $2 \cdot (1 - \pi_t)\pi_t$.

uncertainty reduction for those beliefs. But then the relationship must also hold when summing across all of the states. This implies, for example, that the combined movement of the full histogram of beliefs over many states must equal the combined initial uncertainties on average for resolving DGPs, as well as other results analogous to those discussed for the binary case.

## 2.7 Expectations and the Importance of Scale

The multi-state generalization requires the observation of the person's full distribution of beliefs in all periods. There are many empirical cases in which this information is not available, but it is possible to observe the person's expected value of some variable given these beliefs, such as the expected level of the stock market in three months or the expected number of points that will be scored in a football game.

To analyze this situation, we associate each state $x \in \{0, 1, 2, ...\}$ with a bounded outcome $v^x \in \mathbb{R}$ and study the person's expected value of the outcome at time $t$, $v_t = \sum_{x \in X} \pi_t^x v^x$, and the consequent *expectations stream* $v = [v_0, v_1, v_2, ...]$. We define the *movement in expectations* from period $t_1$ to period $t_2$ as $m^v_{t_1,t_2}(v) \equiv \sum_{\tau=t_1+1}^{t_2} (v_\tau - v_{\tau-1})^2$ and the related capital-letter random variables as above.

Given we can probabilistically bound the movement of beliefs, it is natural to suppose similar bounds on the movement of the expectation formed from these beliefs. Note that the binary results suggest that the supposition is true when the outcome value can be either 0 or 1, because the person's belief exactly corresponds with her expectation of the outcome. Our first result and corollary regarding the movement in expectations shows that the supposition, at least without additional information, is false:

**Proposition 4** ***(Any Expectations Stream Arbitrarily Likely Without Additional Assumptions)*** *For all* $\delta \in (0, 1)$ *and expectations stream* $v$*, there exists some DGP and set of state values such that* $\Pr(V = v) \geq \delta$*.*

**Corollary 6** ***(No Upper Bound on Movement of Expectations Without Additional Assumptions)*** *For all* $\delta \in (0, 1)$ *and critical value* $m^*$*, there is some DGP with initial expectation* $v_0$ *such that* $\Pr(M^v \geq m^*) \geq \delta$*.*

We see this negative result as critical to understanding our main result with binary beliefs. The proposition suggests that, not only is movement unbounded without further assumptions, but that any single expectations stream—no matter how seemingly volatile—is arbitrarily likely in some DGP. To understand the simple intuition, consider the expectations stream [0.01,0.99,0.01,0.99,...] in which the person's expectations bounce from 0.01 and 0.99 one hundred times. Given a binary state with values 0 and 1, these movements equal the person's binary beliefs and this stream would be extremely rare. However, if the state values are changed to $10^{10}$ and $-10^{10}$, there is a DGP—similar to the double-gaffe example above—in which a person always transitions from 0.99 to 0.01 with extremely high probability while moving to $10^{10}$ with extremely low probability (and vice versa from 0.01 to 0.99). Therefore, without understanding the *scale* of the movements relative to what is expected, arbitrarily large movements can be made arbitrarily common.[23] This insight implies that a sense of scale is crucial to any test focused on movement size. Implicitly, our previous results with beliefs were dependent on the scaling of beliefs, which always exist in the unit interval.

It is possible to make statements given information or assumptions about the scale of movements. Specifically, defining the *uncertainty* (or perceived variance) at period $t$ as $u_t^v \equiv \sum_{x \in X} \pi_t^x (v^x - v_t)^2$ and the *uncertainty reduction* from period $t_1$ to period $t_2$ as $r_{t_1,t_2}^v(v) \equiv u_{t_1}^v(v) - u_{t_2}^v(v)$, allows a result that mirrors our binary-beliefs result: as the person's expectation shifts, there must be a concomitant drop in uncertainty about the state value in expectation. Consequently, information about uncertainty can limit the expected amount of movement:

**Proposition 5** ***(Generalization of Proposition 1 with Expectations)*** *For any DGP and for any periods $t_1$ and $t_2$, $\mathbb{E}M_{t_1,t_2}^v = \mathbb{E}R_{t_1,t_2}^v$.*

Intuitively, the person's perceived variance reduction provides the necessary scaling information to limit movement. Consequently, while the full distribution of beliefs is not necessary to probabilistically bound the movement of expectations, some information or assumptions about the second-moment is required. Crucially, when working with binary beliefs, the variance is implied by the belief and therefore no addition information is needed.

Of course, it might also be difficult to gather second-moment information at every period. Following Corollary 7, it is possible to only gather this information at the initial period and compare it to movement in an entire resolving stream:

**Corollary 7** ***(Resolving Streams with Expectations):*** *For any resolving DGP, $\mathbb{E}M^v = u_0^v$.*

And, without this information, Corollary 8 shows it is also possible to make weaker statements by simply knowing the bounds on the possible outcomes, (defining $\underline{v} \equiv \min(v^1, v^2, ...)$ and $\overline{v} \equiv \max(v^1, v^2, ...)$), as this implicitly limits the variance[24]:

[23] Conversely, movements of beliefs that are very rare can be very common when "compressed": while it is extraordinarily unlikely to observe many belief vibrations between 0.01 and 0.99, a Bayesian observing a long-horizon double-gaffe model with a probability of a gaffe $\phi$ of $\frac{1}{1000}$ for one hundred periods will produce a belief stream that repeatedly bounces between $\frac{999}{1999}$ and $\frac{1000}{1999}$ more than 90% of the time.

[24] The variance can be further bounded by additional plausible distributional assumptions on beliefs, such as single-peakness or requiring the mean to equal to the mode.

**Corollary 8** ***(Expectations Given Bounds)****: For any DGP with bounded state values,* $\mathbb{E}M^v \leq (\overline{v} - v_0)(v_0 - \underline{v})$.

This Corollary again stresses the benefit of bounding outcomes to our statements. Bounds imply that the initial variance is bounded, which in turn bounds the movement of a resolving stream, which in turn bounds the movement of all streams.

## 2.8 Discussion and Properties of the Tests

In our simple framework, given a particular signal realization from a given DGP, the belief stream is unique. Knowing this *proper-belief stream*—as in a laboratory study with complete control of the DGP—would allow the rejection of the model whenever a person's beliefs deviate from this exact stream. We instead construct a test without knowledge of the proper stream (or even the signal realizations), basing rejection on the statistical violation of the equivalence relationship between movement and uncertainty reduction. As a result, our test does not reject all non-proper belief streams.

Most starkly, if a person reports appropriate Bayesian beliefs given signals from a completely different DGP than the one we believe she is watching, our test will not reject her Bayesianess because these "wrong" beliefs are still *consistent* with Bayesian updating. We can reject many of these streams if we observe the true final state of the world and append it to the end of the belief stream. However, even with this constraint, we highlight two specific types of non-Bayesian streams for which the relationship still holds. Interestingly, we argue that these non-rejections can be considered a benefit of our test rather than a cost in some situations, because they allow flexibility to not reject potentially reasonable belief streams.

First, in our abstract model, beliefs are observed at every time period after every signal. However, in most empirical settings, belief changes are only observed at certain periods, such as the points of belief elicitation or the times of voluntary disclosure. For example, by only observing beliefs in odd periods, the person's actual belief stream $[.4, .6, .7, .4, .2, .3...]$ is perceived as $[.4, .4, .7, .7, .2, .2, ...]$. Crucially, while the latter belief stream does not match the proper-belief stream, it does match the beliefs of a different DGP with the same state resolution in which the person receives no signal every even period, but receives the information from two signals in the odd periods. Consequently, even though the streams arise from a different DGP with different movement distributions, the expectation of the movement distribution is still equal to the uncertainty reduction.[25]

Second, if a person ignores a specific signal (or does not observe it), her belief streams will satisfy the tested relationship. Again, while this person's belief streams will not match the model

[25] Even if the person is *endogenously* disclosing beliefs only at certain periods (based on signals or beliefs), the whole-stream relationship $\mathbb{E}M = \mathbb{E}R$ continues to hold as long as the final resolution of the state is known and appended to the end of the stream.

belief stream, they do match those from an alternative DGP with the same state resolution in which the person doesn't receive some signals. Depending on the circumstance, this disregard may or may not be considered non-Bayesian: a Bayesian can miss a signal, but maybe not glaring evidence.[26]

Finally, we note that it is also possible to incorrectly not reject non-Bayesian behavior due to the need to *aggregate* data across situations to create a single test statistic when dealing with finite data. That is, if the statistic is positive given some conditions and negative in other conditions – suggesting non-Bayesian updating – the effects might balance in the aggregated statistic, such that Bayesian updating is not rejected. In fact, Section 5.3 presents a situation in which movement and uncertainty reduction are nearly balanced over the entire stream, but there is clear overreaction during earlier periods and clear underreaction in later periods.

# 3 Sources of Bias in Movement and Uncertainty Reduction

In this section, we sketch how some well-known psychological biases might lead to systematic distortions in our main statistic, $\mathbb{E}M - \mathbb{E}R$. In order to create sharp results, we focus on particular types of biases and DGPs, but we provide some intuition about why the connection may be more general. These results are suggestive that our statistics might be more likely to identify common psychological biases than other approaches, which is a theme we return to in Section 4. Additionally, we examine how these biases may affect the individual components, $\mathbb{E}M$ and $\mathbb{E}R$, and compare these with those of a Bayesian observing the same process, which we label $\mathbb{E}M^B$ and $\mathbb{E}R^B$. Because separately analyzing these components helps us distinguish more finely among these biases, we are able to provide hints about the potential for (and limits to) using our test to go beyond identifying *whether* people are departing from Bayesian updating to identifying *how* they might be doing so.

## 3.1 Setup and Simple Model of Four Biases

We examine two variants of the simple environment discussed in the first example of the previous section: given two states $H$ and $L$, a person observes an independent, symmetric signal $s_t \in \{h, l\}$, with a conditional distribution $\Pr(h|H) = \Pr(l|L) = \theta$, where we call $\theta$ the *precision*. In the first environment, the person starts with the correct prior and only observes one signal. In the second, the person starts with prior $\pi_0 = \frac{1}{2}$, and observes many signals over time. As we discuss below,

[26] Similarly, the relationship will continue to hold if the person cannot distinguish between certain sets of signal realizations such that she observes *coarser* information than in the model. For example, if a signal has three realizations, $l1$, $l2$, and $h$, but the person cannot distinguish between $l1$ and $l2$ (observing them as the same signal $l$), her belief streams will still satisfy our relationship (as long as she correctly interprets the coarse signal).

we restrict to this relatively simple dynamic model because once we allow for multiple signals, people's priors at any time are potentially distorted by the collection of their biased updates in previous periods.

We consider a very simple, portable framework for non-Bayesian updating that embeds variants of commonly proposed biases. We start by recasting the usual formulation of Bayesian updating in terms of likelihood ratios. To do so, we define a function $LR[\pi] \equiv \frac{\pi}{1-\pi}$ that maps a belief into its implied likelihood ratio. Likewise, we recast the signal into its likelihood ratio: In our environment, $LR[s_t] = \frac{\theta}{1-\theta}$ if $s_t = h$ and $LR[s_t] = \frac{1-\theta}{\theta}$ if $s_t = l$. Then given a prior $\pi_t$ and signal $s_{t+1}$, a Bayesian's posterior belief $\pi_{t+1}$ satisfies:

$$LR[\pi_{t+1}] = LR[\pi_t] \cdot LR[s_{t+1}].$$

For example, given a prior belief $\pi_t = \frac{1}{5}$ ($LR[\pi_t] = \frac{1}{4}$) and precision $\theta = \frac{4}{5}$, a signal $s_{t+1}$ of $l$ ($LR[s_{t+1}] = \frac{1}{4}$) leads to a posterior likelihood ratio of $\frac{1}{16}$, implying that $\pi_{t+1} = \frac{1}{17}$, while a signal of $h$ ($LR[s_{t+1}] = 4$) implies $\pi_{t+1} = \frac{1}{2}$.

In our formulation – which follows early work on representativeness by Grether (1980,1992) and the model of dynamic base-rate neglect of Benjamin, Bodoh-Creed, and Rabin (2017) – the person can under- or overweight the likelihood ratio of the prior or of the signal. The non-Bayesian updating can be formulated in terms of exponent parameters $\alpha$ and $\beta$:

$$LR[\pi_{t+1}] = LR[\pi_t]^{\alpha} \cdot LR[s_{t+1}]^{\beta}.$$

Always assuming that $\alpha > 0$ and $\beta > 0$, and examining the separate implications of $\alpha$ and $\beta$ by varying only one parameter while fixing the other parameter at 1, we focus on four types of biases: $\alpha < 1$ ("base-rate neglect"), $\alpha > 1$ ("confirmation bias",) $\beta > 1$ ("overreaction"), and $\beta < 1$ ("underreaction"). To demonstrate the effect of these four biases, consider the above example. When $\alpha = \frac{1}{2}$, the person effectively discounts prior evidence by contracting $LR[\pi_t]$ toward 1, leading to posteriors after signals $l$ and $h$ of $(\frac{1}{9}, \frac{2}{3})$, while a person with $\alpha = 2$ exaggerates the prior, leading to $(\frac{1}{65}, \frac{1}{5})$. In contrast, when $\beta = \frac{1}{2}$, the person discounts the signal, leading to $(\frac{1}{9}, \frac{1}{3})$, while a person with $\beta = 2$ exaggerates the signal and ends with $(\frac{1}{65}, \frac{4}{5})$.

It bears emphasizing that we do not interpret our analysis of these four cases as applying to different people with different psychological propensities, but rather in general to people who may suffer from all the psychological biases discussed—but where the environment determines what values of $(\alpha, \beta)$ apply.[27] One major point of our analysis is that, even though it might

[27] We very briefly outline of the psychological underpinnings and environments in which we might expect different values of $\alpha$ and $\beta$. As established by Kahneman and Tversky (1973) and others, and formally modeled in Rabin (2002) and Rabin and Vayanos (2010), people have a tendency to infer too much from small samples of independent variables, such that we expect $\beta > 1$ for people receiving information in small data sets. But as established by Tversky and Kahneman (1973) and others, and formally modeled in Benjamin, Rabin, and Raymond (2016), people have a tendency to infer (way) too little from large samples, such that we would expect $\beta < 1$ in these

intuitively appear that overreaction is the same as base-rate neglect (both involve the signal being overweighted relative to the prior) and underreaction the same as confirmation bias (both involve the prior being overweighted relative to the signal), these biases are different and create different signatures in terms of movement and uncertainty reduction.

## 3.2 Bias in $\mathbb{E}$M-$\mathbb{E}$R in Two Environments

We now state two propositions that suggest a close connection in our two environments between the four biases and our main statistic $\mathbb{E}M - \mathbb{E}R$:

**Proposition 6** *Consider the single-signal model with correct priors.*

*Given any signal precision $\theta$ and any initial prior $\pi_t \neq \frac{1}{2}$:*

$\beta > 1$ *(overreaction)* $\Rightarrow \mathbb{E}M_{t,t+1} > \mathbb{E}R_{t,t+1}$ *with* $\frac{\partial(\mathbb{E}M_{t,t+1} - \mathbb{E}R_{t,t+1})}{\partial \beta} > 0$

$\beta < 1$ *(underreaction)* $\Rightarrow \mathbb{E}M_{t,t+1} < \mathbb{E}R_{t,t+1}$

$\alpha > 1$ *(confirmation bias)* $\Rightarrow \mathbb{E}M_{t,t+1} < \mathbb{E}R_{t,t+1}$ *with* $\frac{\partial(\mathbb{E}M_{t,t+1} - \mathbb{E}R_{t,t+1})}{\partial \alpha} < 0$

$\alpha < 1$ *(base-rate neglect)* $\Rightarrow \mathbb{E}M_{t,t+1} > \mathbb{E}R_{t,t+1}$ *with* $\frac{\partial(\mathbb{E}M_{t,t+1} - \mathbb{E}R_{t,t+1})}{\partial \alpha} < 0.$

This proposition indicates that over-reactors and base-rate neglectors have positive expected excess movement in the single-signal setting, while the opposite is true for under-reactors and agents with confirmation bias. Furthermore, for all but under-reactors, there is a monotonic connection between the level of bias and the statistic. The monotonicity fails for under-reactors because as the bias becomes more extreme and the person perceives that the signal is nearly complete noise, leading to very little belief movements and a low-magnitude statistic.

To help interpret our analysis, recall from Section 2.2 that the excess-movement statistic, $\mathbb{E}M_{t,t+1} - \mathbb{E}R_{t,t+1}$, can be rewritten as $\mathbb{E}_{\pi_t}[(2\pi_t - 1)(\pi_t - \pi_{t+1})]$ and seen as summary statistic of the tendency to revert to the point of highest uncertainty. For base-rate neglect, the person "forgets" a portion of her prior, updating as if her prior were closer to $\frac{1}{2}$ than in reality. This error leads all of her posterior beliefs to be biased in the direction of $\frac{1}{2}$ from her prior, causing the expectation of her posterior to move in the direction of $\frac{1}{2}$, which leads to an excess-movement statistic. Conversely, a person with confirmation bias overweights her prior, causing her beliefs to be biased in the direction away from $\frac{1}{2}$, leading to a negative statistic. For overreaction, the case is more subtle as the biased person's posterior beliefs are higher than that of a Bayesian after a positive signal, but lower after a negative signal. However, the expectation of posteriors still in

cases. Identifying the situations leading to specific a $\alpha$ parameter is more complicated. Base-rate neglect ($\alpha < 1$)—articulated by Kahneman and Tversky (1972, 1973), Bar Hillel (1980) and others, and modeled by many researchers over the years including recently by Benjamin, Bodoh-Creed, and Rabin (2017)—is the tendency to under-use "base rates" (which we are assuming correspond to priors) when updating their beliefs. The least familiar possibility in this context is that $\alpha > 1$. We see it as a form of confirmation bias, as modeled for instance by Rabin and Schrag (1999), whereby people have a tendency to miscode evidence as favoring their current hypotheses, effectively exaggerating their prior.

the direction $\frac{1}{2}$ because the effect is greater for signals that contradict the prior (shift beliefs back toward $\frac{1}{2}$) than confirm the prior (shift beliefs toward certainty).[28] The opposite is true in the case of underreaction.

By assuming a correct prior, Proposition 6 isolates the effect of biased updating at period $t$ from the effect of biased updating from all previous periods, which is captured in an incorrect prior. The multi-period environment allows for endogenously-incorrect priors for every history, but comes at the cost of requiring a prior of $\frac{1}{2}$ and some parametric restrictions to obtain analytic proofs. While our main interest is about across-history expectations such as $\mathbb{E}M$ and $\mathbb{E}R$, these assumptions allow the results to hold at every possible history, which immediately implies the across-history results. Our suspicion—based on simulations—is that it is in fact possible to make the statements about $\mathbb{E}M$ and $\mathbb{E}R$ without the prior or parametric restrictions.[29]

**Proposition 7** *Consider the multi-period model with correct initial prior $\pi_0 = \frac{1}{2}$.*

*Consider any signal precision $\theta$, any $t_0 < t_1 < t_2$, and any history $H_{t_0}$:*

| | | |
|---|---|---|
| $\beta > 1$ | *(overreaction)* | $\Rightarrow \mathbb{E}M_{t_1,t_2} > \mathbb{E}R_{t_1,t_2}$ *given* $H_{t_0}$ |
| $\beta < 1$ | *(underreaction)* | $\Rightarrow \mathbb{E}M_{t_1,t_2} < \mathbb{E}R_{t_1,t_2}$ *given* $H_{t_0}$ |
| $\alpha > 2$ | *(strong confirmation bias)* | $\Rightarrow \mathbb{E}M_{t_1,t_2} < \mathbb{E}R_{t_1,t_2}$ *given* $H_{t_0}$ |
| $\alpha < \alpha^*(\theta)$ [30] | *(strong base-rate neglect)* | $\Rightarrow \mathbb{E}M_{t_1,t_2} > \mathbb{E}R_{t_1,t_2}$ *given* $H_{t_0}$ |

In the case of over- and underreaction, the wrong priors induced by the error actually amplify the effects in the previous proposition. For example, if a Bayesian observing a set of signals believes one state is more likely, an overreactor given the same signals will believe that state is even more likely. Given that the Bayesian has the appropriate assessment, the overreactor's expected posterior will move in the direction of the less-certain Bayesian prior (i.e. toward $\frac{1}{2}$), increasing her already-positive excess-movement statistic. As before, the opposite is true for the underreactor.[31] Confirmation bias and base-rate neglect are more complicated, because there is

[28] For instance, in the example above, a Bayesian adjusts from $\frac{1}{5}$ to $\frac{1}{17}$ when observing a confirming signal and from $\frac{1}{5}$ to $\frac{1}{2}$ when observing a disconfirming signal. An overreactor adjusts farther downward with a confirming signal to $\frac{1}{65}$, but the magnitude of the change induced by the bias is far greater for the disconfirming signal, which shifts beliefs to $\frac{4}{5}$.

[29] As an example of the issue, the proof strategy for Proposition 7 when $\alpha > 2$ is to demonstrate that $\mathbb{E}M_{t,t+1} < \mathbb{E}R_{t,t+1}$ for any history $H_t$ following $H_{t_0}$. An immediate consequence is that summing over time periods and taking expectations over histories implies $\mathbb{E}M_{t_1,t_2} < \mathbb{E}R_{t_1,t_2}$ given $H_{t_0}$. When $a \in (1, 2)$, this proof strategy fails. In particular, there is a rare signal history $H_t$ in which $\mathbb{E}M_{t,t+1} > \mathbb{E}R_{t,t+1}$. In this history, a particular set of signals cause the Bayesian to be nearly certain about the correct state of the world, while the the biased individual's still hovers around the incorrect state. Given the rareness of this history, we still believe that summing over periods and taking expectations over histories will yield $\mathbb{E}M_{t_1,t_2} > \mathbb{E}R_{t_1,t_2}$ given $H_{t_0}$. However, proving this statement requires a different strategy that quantifies the likelihood of different histories and the relative magnitudes of $\mathbb{E}M_{t,t+1} - \mathbb{E}R_{t,t+1}$ in those histories.

[30] $\alpha^*(\theta)$ is the solution to the (unfortunately ugly) implicit equation: $\theta \cdot p(\frac{1-\theta}{\theta}^{\alpha^* \frac{1-2\alpha^*}{1-\alpha^*}+1}) + (1-\theta) \cdot p(\frac{1-\theta}{\theta}^{\alpha^* \frac{1-2\alpha^*}{1-\alpha^*}-1}) - p(\frac{1-\theta}{\theta}^{\frac{1-2\alpha^*}{1-\alpha^*}}) = 0$ where $p(x) = \frac{x}{1+x}$. For reference, $\alpha^*(\theta) = 0.40, 0.32, 0.25$, and $0.18$ when $\theta = 0.6, 0.7, 0.8$, and $0.9$.

[31] More generally, consider a simple one-period environment with an agent that updates correctly ($\alpha = 1, \beta = 1$) but starts with a prior $\pi$ not equal to the Bayesian prior $\pi^*$. The expectation of the agent's posterior belief will

no clear connection between the person's beliefs and the Bayesian beliefs which holds at all signal histories. However, if the bias is strong enough, the same relationships hold because the bias overwhelms any difference in priors induced by the error. For example, for strong confirmation bias ($\alpha > 2$), the biased person will, in all periods, move in the direction implied by her initial signal $s_1$, regardless of the later signals.

Proposition 7 only provides directional results. To provide intuition about the mapping between the psychological bias parameters and the level of the resultant excess movement, we numerically simulate the DGP for $\theta = .75$ with resolution after 20 periods. Given $\alpha$ equal to .75, .85, .95, 1.05, 1.15, and 1.25, the normalized excess movement statistic is 2.44, 1.66, 1.14, 0.92, 0.87, and 0.90, respectively. That is, even small amounts of base-rate neglect can produce large amounts of movement and the mapping between the level of base-rate neglect and movement is highly convex, while the effect of confirmation bias is not very large or monotonic. Note that, for base-rate neglect, as the number of periods before resolution grows, the statistic will rise without bound as beliefs are ergodic. Alternatively, given $\beta$ equal to .75, .85, .95, 1.05, 1.15, and 1.25, the normalized excess movement statistic is 0.73, 0.84, 0.95, 1.05, 1.16, and 1.27, respectively. That is, for this simple DGP, the parameter $\beta$ and the normalized statistic happen to be nearly equal.[32]

## 3.3 Bias in $\mathbb{E}$M and $\mathbb{E}$R Separately

We now explore how the biases affect movement and uncertainty reduction *separately*, by comparing $\mathbb{E}M$ and $\mathbb{E}R$ to those of a Bayesian observing the same signals, which we label $\mathbb{E}M^B$ and $\mathbb{E}R^B$. This exercise is in sharp contrast to the rest of the paper where we assume that the Bayesian beliefs are unknown, which is made possible by exploiting the fact that $\mathbb{E}M^B = \mathbb{E}R^B$ universally. The four biases have four different implications for these comparisons in our simple (and restrictive) dynamic environment, as shown in our next Proposition and explained below:

**Proposition 8** *In the multi-period model, given any signal precision $\theta$:*

not equal her prior (as in a martingale), but will instead be biased in the direction of the Bayesian prior. Given the interpretation of the statistic $\mathbb{E}[M - R]$ as $\frac{1}{2}$ reversion, the sign of the agent's statistic then depends on whether the bias is in the direction of $\frac{1}{2}$.

[32] An alternative way to gauge the magnitude of these biases is welfare effects, which of course depend on the situation. For a simple calibration, we again consider a model using the DGP with $\theta = .75$ in which two risk-neutral non-discounting people—one biased with beliefs $\widehat{\pi}_t$ and one Bayesian with beliefs $\pi_t$— make simple prediction-market-like bets with each other at each of the 20 time periods. Specifically, if $\pi_t > \widehat{\pi}_t$, the Bayesian buys a security that pays 1 in state 1 for the midpoint price $\frac{\pi_t + \widehat{\pi}_t}{2}$ from the biased person, while if $\pi_t < \widehat{\pi}_t$, the biased person instead buys from the Bayesian. In the case that $\pi_t > \widehat{\pi}_t$, the biased person receives a payment of $\frac{\pi_t + \widehat{\pi}_t}{2}$, but must pay 1 with probability $\pi_t$, leading to an expected payoff of $-\frac{1}{2}|\pi_t - \widehat{\pi}_t|$. Using similar logic, it is simple to show that the expected total payoff of the biased person is $-\frac{1}{2}\mathbb{E}[\sum_t |\pi_t - \widehat{\pi}_t|]$. That is, the loss is proportional to the expected total absolute distance between biased and Bayesian beliefs. In this setup, $\alpha$ equal to .75, .85, .95, 1.05, 1.15, and 1.25 yields to an expected monetary loss of 1.04, 0.56, 0.14, 0.11, 0.33, and 0.55, respectively. Alternatively, given $\beta$ equal to .75, .85, .95, 1.05, 1.15, and 1.25, the loss is 0.21, 0.11, 0.03, 0.03, 0.08, and 0.13, respectively.

*For any $t$, a person with:*

$$\begin{array}{lll} \beta > 1 & \textit{(overreaction)} & \Rightarrow \mathbb{E}M_{0,t} > \mathbb{E}M^B_{0,t}, \quad \mathbb{E}R_{0,t} > \mathbb{E}R^B_{0,t} \\ \beta < 1 & \textit{(underreaction)} & \Rightarrow \mathbb{E}M_{0,t} < \mathbb{E}M^B_{0,t}, \quad \mathbb{E}R_{0,t} < \mathbb{E}R^B_{0,t} \end{array}$$

*There exists a time period $t^*$ such that for all $t > t^*$, a person with:*

$$\begin{array}{lll} \alpha > 2 & \textit{(strong confirmation bias)} & \Rightarrow \mathbb{E}M_{0,t} < \mathbb{E}M^B_{0,t}, \quad \mathbb{E}R_{0,t} > \mathbb{E}R^B_{0,t} \\ \alpha < 1 & \textit{(base-rate neglect)} & \Rightarrow \mathbb{E}M_{0,t} > \mathbb{E}M^B_{0,t}, \quad \mathbb{E}R_{0,t} < \mathbb{E}R^B_{0,t} \end{array}$$

The most intuitive of these results concerns base-rate neglect, where, as a result of "forgetting" information, the person shifts beliefs more than a Bayesian but ends up being less certain. We show, following Benjamin, Bodoh-Creed, and Rabin (2017), that base-rate neglect implies that in the long-run beliefs are ergodic (so that $\mathbb{E}M \to \infty$) but bounded away from certainty (so that $\mathbb{E}R < \pi_0(1-\pi_0)$), whereas $\mathbb{E}M^B$ and $\mathbb{E}R^B$ both converge to $\pi_0(1-\pi_0)$, so that in the limit necessarily $\mathbb{E}M > \mathbb{E}M^B$ and $\mathbb{E}R < \mathbb{E}R^B$. Formally, as stated in the proposition, there exists some period after which the two inequalities must hold.

Things are more complicated with confirmation bias. As stated in the proposition, if confirmation bias is strong enough, there exists some period where there is too little movement and too much uncertainty reduction: $\mathbb{E}M < \mathbb{E}M^B$ and $\mathbb{E}R > \mathbb{E}R^B$ (Of course, in the limit, Bayesian updating induces full resolution of uncertainty, so that both $\mathbb{E}R$ and $\mathbb{E}R^B$ approach $\pi_0(1-\pi_0)$). Confirmation bias can lead to $\mathbb{E}M > \mathbb{E}M^B$ in the short run if the signal is very weak relative to the degree of confirmation bias. This occurs because the person in our model always "drifts" toward the currently-percieved-as-likely outcome, leading her to move significantly even with a very weak signal. Interestingly, the issue of drift given little information does not occur using the model of confirmation bias in Rabin and Schrag (1999) in which a person's signals are replaced with a "confirming" signal with some probability. In both models, however, our result that $\mathbb{E}M < \mathbb{E}M^B$ in the long run depends on the assumption that the truth is never observed. If instead the person is assumed to learn the truth in a final revelation, strong confirmation bias leads to $\mathbb{E}M > \mathbb{E}M^B$, because this final revelation leads to a huge movement in the common case where the person becomes very certain in the wrong state.

The results for over-reaction and under-reaction are far simpler. Overreaction leads to $\mathbb{E}M > \mathbb{E}M^B$ and $\mathbb{E}R > \mathbb{E}R^B$ even in the short run, while underreaction leads to $\mathbb{E}M < \mathbb{E}M^B$ and $\mathbb{E}R < \mathbb{E}R^B$: over-interpreting information leads to both too much movement and causes the person to be too certain, while under-interpretation leads to too little movement and not enough certainty.

In the case of base-rate neglect, when $\mathbb{E}M > \mathbb{E}M^B$ and $\mathbb{E}R < \mathbb{E}R^B$, then our previous finding that $\mathbb{E}M > \mathbb{E}R$ must hold given our main result that $\mathbb{E}M^B = \mathbb{E}R^B$. Similarly, the confirmation-bias results $\mathbb{E}M < \mathbb{E}M^B$ and $\mathbb{E}R > \mathbb{E}R^B$ imply $\mathbb{E}M < \mathbb{E}R$. However, for over- and underreaction, the results on the separate components do not imply the previous findings that

$\mathbb{E}M > \mathbb{E}R$ and $\mathbb{E}M < \mathbb{E}R$. Instead, the combination of the individual component and combination results point to a sense in which the effect of overreaction on increasing belief movement is greater than on increasing uncertainty reduction, and the effect of underreaction in mediating uncertainty reduction is greater than in reducing belief movement.

# 4 Comparisons to Other Tests

Bayesian beliefs are bounded in the unit interval and satisfy the martingale property—the expectation of future beliefs equal the current belief regardless of the previous history of beliefs. There are many ways that the martingale property can be violated. For example, a non-Bayesian's beliefs might always trend in one direction, might exhibit positive or negative autocorrelation of different degrees, or might trend differentially based on some complicated function of past beliefs. For finite data sets, where it is unlikely that all fully defined situations appear often enough to test each deviation, it is necessary to aggregate behavior across different streams and situations to test for a *specific* type of deviation. To increase statistical power, a good test of non-Bayesian updating consequently focuses on *likely* types of deviations, while—of course—remaining a valid test by not falsely rejecting the null hypothesis more than specified. In Section 3, we argued that there is a relatively tight connection between our test and the types of biases people are prone to make, suggesting that our test might be efficient at detecting these biases. In this section, we support that conclusion by comparing its performance with that of other martingale tests using numerical simulations of Bayesian and biased beliefs in the multi-period environment with signals discussed in Section 3. The results suggest that some existing tests are not valid on belief data and, among the valid tests, our test has greater power to detect the four biases.[33] Additionally, we point to three attractive characteristics of our test when using belief data: *time-invariance, interpretability*, and *scale-variance*.

We start the comparison with *calibration*, perhaps the most intuitive and direct way of determining the accuracy of probabilistic beliefs. A person is "well-calibrated" if, when she assigns a given probability to an outcome, the proportion of outcomes that actually occur is equal to this probability. For example, a weather forecaster's statements that "there is an 80% chance of rain tomorrow" is well-calibrated if it turns out that rain occurred 80% of the days following these statements. To understand a fundamental difference between our test and calibration, it is useful to consider a person with severe base-rate neglect ($\alpha = 0$ in the model in Section 3) who perpetually overlooks previous signals and reports beliefs based only on most recent signal. While this person shifts beliefs continuously with no resolution of uncertainty and therefore strongly violates our test, she is—in fact—well-calibrated for all forecast horizons, because she appropriately updates given the most recent signal. This discrepancy occurs because calibration does not

[33] However, we once again stress that this does not imply that our test is universally better at detecting all deviations from the martingale property or all types of bias in updating.

test the relationship between internal beliefs in a belief stream, but rather between one internal belief and the final outcome. Given that calibration effectively ignores all-but-one of the internal beliefs, it requires many belief streams, such that calibration is simply infeasible in our numerical simulations.[34] Instead, we employ a simple and more feasible *drift* test – which we call a *simple martingale test* – of whether the average difference in belief changes each period $\pi_t - \pi_{t+1}$ is equal to zero.

The vast majority of martingale tests focus on the correlations between the current movement and one or many movements in previous periods.[35] To understand the difference between these *autocorrelation* tests and our test, recall from Section 2.2 that $\mathbb{E}[f(\pi_0, \pi_1, ..\pi_t) \cdot (\pi_t - \pi_{t+1})] = 0$ for any instrument $f$ and our test employs the instrument $(2\pi_t - 1)$, implying that the sign of the excess-movement statistic $\mathbb{E}M - \mathbb{E}R$ can be interpreted as the tendency to revert toward a belief of $\frac{1}{2}$. In contrast, a simple one-period autocorrelation test on beliefs employs the instrument $(\pi_{t-1} - \pi_t)$. Therefore, for example, while negative one-period autocorrelation implies the expected direction of current belief updating will be downward (rather than straight, as in a martingale) if the previous belief movement was upward, positive excess movement suggests that it will be downward if the current belief is above $\frac{1}{2}$.

The reversion to a fixed point rather than based on recent movements has two potential benefits. First, the biases studied in Section 3 revert to the fixed point $\frac{1}{2}$, so our test will be more powerful in rejecting these biases, which is useful to the extent that those biases accurately represent common patterns in non-Bayesian beliefs. Auto-correlation tests will pick up these errors with a large enough dataset because, on average, previous upward movements are correlated with prior beliefs being above $\frac{1}{2}$. Second, reversion to a fixed point allows our test to be *time-invariant* in that it does not depend on a specific dependence between reversion and time. In contrast, for example, one-period autocorrelation is designed to identify reversion that occurs exactly after exactly one "period" of time. But, it is hard to know a priori what a period is, and presumably both the amount of time required for belief reversion and the amount of time between belief observations are not consistent within or across datasets.

The simplest implementation of the one-period autocorrelation test is to calculate the sample Pearson Correlation $\widehat{\rho}_1$ between current movement and previous movement and test the hypothesis that it is zero. Multi-period autocorrelation tests, such as the *Ljung–Box* test and *Box–Pierce* test extend this logic to multiple lags of previous movement and test the joint hypothesis that all

[34] In practice, empiricists commonly aggregate statements with similar probability assignments, such as "there is an 80% chance of rain" and "there is a 76% chance of rain," into the same "bin" (statements with probabilities of 70-80%), comparing the empirical proportion that actually occurred to the average predicted probability in the bin. The size of the bins are generally chosen based on the size of the data set. The results across different bins are presented jointly using a "calibration curve," which compares the empirical proportion to the predicted probability for each bin (see, for example, Wolfers and Zitzewitz (2004) for a review of calibration in prediction markets).

[35] A more comprehensive review of martingale tests can be found in Escanciano and Lobato (2009). Lichtenstein et al (1982) reviews the early literature on calibration tests.

correlations $\rho_1, \rho_2, \ldots$ are zero.[36] Unfortunately, these correlation tests are not designed to account for the existence of heteroskedastic errors, which is common in beliefs data, and therefore the tests commonly incorrectly reject Bayesian updating in our simulations.[37] This problem is alleviated by regressing current change on a number of past changes and relying on the heteroskedastic-robust F-statistic, which we use in the simulations below.

An issue with multi-period autocorrelation tests is that the intuitive interpretation of a rejection given a joint null hypothesis is cloudy as the exact source of the violation of the joint of hypothesis is unclear. It is possible to construct a lower-powered test that potentially identifies particular lags that exhibit conspicuous non-zero autocorrelations or to assess the individual autocorrelations informally, but the interpretation of magnitudes is still challenging and depends on the period lag length in the particular environment. Therefore, an additional benefit of the (normalized) excess movement test statistic is that it has a *consistent* and *portable* interpretation that arises from the core theory: a statistic of 1.2 implies that the person's movement was 20% more than expected, regardless of the timing or dataset.

It is possible to create a more interpretable test statistic using a function of the correlations, although the interpretation is still dependent on the frequency of data. Specifically, the *Variance-ratio* (VR) test considered by Lo and MacKinlay (1988) (and reviewed in Charles and Darne (2009)) is based on the observation that, when changes are uncorrelated over time, the variance of the one-period changes is proportional to the variance of the change over an interval of $k$ periods: $k \cdot \mathbb{E}[(\pi_{t-1} - \pi_t)^2] = \mathbb{E}[(\pi_{t-k} - \pi_t)^2]$.[38,39]

[36] The test statistics of the Ljung–Box and Box-Pierce tests for $k$ lags are $n(n+2)\sum_{i=1}^{k} \frac{\widehat{\rho}_i^2}{n-i}$ and $n\sum_{i=1}^{k} \widehat{\rho}_i^2$ , respectively, both of which are rejected at significance level $\alpha$ if the statistic is greater than $\chi^2_{1-\alpha,k}$.

[37] Heterskedasticity of belief movement naturally depends on the DGP, but appears common because beliefs are bounded. When beliefs are close to certainty, present and future belief movements are likely to be small. Consequently, variation in belief movement is likely small when previous movements were small. The effect of heteroskedasticity on the validity of the non-adjusted tests is large: given simulations of 100 belief streams arising from Bayesian updating in the multi-period environment from Section 3 with $\theta = .75$, the simple correlation test, Ljung-Box and Box-Pierce with 95% confidence incorrectly reject the null of Bayesian updating 67%, 100%, 100%, of the time, respectively.

[38] There is an intuitive connection between our test and the simple VR test, which is apparent when focusing on resolving DGP with an even number of periods. Our test exploits that $\mathbb{E}[\sum_{\tau=0}^{T} m_{\pi_\tau,\pi_{\tau+1}}] = \pi_0(1-\pi_0)$. However, as we discuss in Section 2.8, the same equality holds when focusing on a subset of time periods, such as even-period-only beliefs: $\mathbb{E}[\sum_{\tau=0}^{\frac{T}{2}-1} m_{\pi_{2\tau},\pi_{2\tau+2}}] = \pi_0(1-\pi_0)$. These equalities imply that $\mathbb{E}[\sum_{\tau=0}^{T} m_{\pi_\tau,\pi_{\tau+1}}] = \mathbb{E}[\sum_{\tau=0}^{\frac{T}{2}-1} m_{\pi_{2\tau},\pi_{2\tau+2}}]$, which is the relationship exploited by the VR test. Our test is intuitively more powerful in beliefs because we exploit the fact that, when dealing with beliefs, the variance of outcomes $\pi_0(1-\pi_0)$ is a known quantity implied by beliefs rather than something that needs to be estimated. The cost is that our test is not applicable for variables in which the outcome variance is unknown. Finally, we note an issue with the more-powerful version of the VR test that uses overlapping estimates of the $k-$interval variance estimate by, for example, also employing movement in the odd periods $\mathbb{E}[\sum_{\tau=1}^{\frac{T}{2}-1} m_{\pi_{2\tau},\pi_{2\tau+2}}]$. In belief data, this test is inappropriate as the expectation underlying the "overlapping" version of the VR test does not hold. Intuitively, the movement in odd periods does not equal $\pi_0(1-\pi_0)$ but rather $\mathbb{E}[\pi_1(1-\pi_1)]$.

[39] Both our test and the VR test use squared deviations to construct the test statistic. As we discuss in Section 2, squared deviations arise from the use of a particular surprisal function and we could just as well used a different definition to create a slightly different form of our test. The same logic holds for the VR test: for every surprisal function, there is an associated valid test that is similar to the standard VR test.

Consequently, a test statistic constructed by the ratio of the sample analogs of these variance measures has an expectation equal to one for any $k$ under the null hypothesis of zero autocorrelation. When $k = 2$, the test statistic (minus one) can be shown to be approximately equal to $\widehat{\rho}_1$ and, more generally, approximately equal to a linear combination of the first $k-1$ correlation coefficients $\widehat{\rho}_1, \widehat{\rho}_2, ...\widehat{\rho}_{k-1}$ with linearly-declining weights. Whereas our (normalized) test statistic being 1.2 means that the person's movement was 20% more than expected, a VR test statistic of 1.2 with a lag of 2 suggests that the movement calculated using every other observation is 20% more than the movement calculated using every observation. Although the VR test statistic has a more intuitive interpretation than an F-statistic, the interpretation and portability still depends on the timing in the dataset.

A final important difference between our test and other martingale test relates to a core theme of the paper: appreciation of the inherent *scale of beliefs*. Standard tests of the martingale property are invariant to scale changes in these differences, and therefore invariant to affine transformations of the data. This scale-invariance is a desirable property in many cases because a test's conclusion should not be influenced by the unit of measurement: a change in the price of a stock of 900¢ holds the same meaning as a change of $9. But, it is *undesirable* when studying beliefs: a movement of 0.99 holds inherently different information than a movement of 0.0099. The importance of this observation to testing can be made more formal with a simple Proposition similar to Proposition 4 in Section 2.7:

**Proposition 9** ***(Some Affine Transformation of any Belief Stream is Arbitrarily Likely)*** *For all $\delta \in (0, 1)$ and belief stream $\pi$, there exists some $b_0$ and $b_1$ such that the transformed belief stream $\pi' = b_0 + b_1\pi$ occurs with probability greater than $\delta$ for some DGP.*

That is, any belief stream can be subjected to an affine transformation that leads to another belief stream that is arbitrarily likely given some DGP. Consequently, if a scale-invariant test rejects *any* belief stream as non-Bayesian, it *necessarily* also rejects belief streams that are arbitrarily likely for a Bayesian in some DGP. Given the same logic, if a scale-invariant test rejects any dataset of belief movements, the test must also incorrectly reject a dataset that is arbitrarily likely for a Bayesian given some DGPs. In contrast, our tests are not scale invariant. For example, the simple implication of Corollary 2 that total movement in one stream cannot be greater than 5 obviously inherently depends on the scale of the movements. Alternatively, when we use the seemingly scale-invariant mean test, the additional rule-of-thumb imposing the restriction that total movement be greater than 3 builds in scale-dependence by preventing application of our test to data with many small movements. While it might be possible to create a similar rule-of-thumb that requires enough observations to employ another martingale test, any rule we can imagine implicitly channels the core concept of movement from our approach.

We now shift to a numerical evaluation of the power of our test in comparison to the above autocorrelation tests. We simulate 100,000 draws of various sample sizes (10, 30, and 100) of belief-streams following the multi-period symmetric noisy-signal DGP with a starting prior of $\frac{1}{2}$ discussed in the previous section.[40] In the top panel of Table 2, we report the estimated type I error—the likelihood that the test incorrectly rejects that the stream is a martingale—for streams generated by a Bayesian for five different signal strengths ($\theta \in \{.55, .65, .75, .85, .95\}$).[41] We calculate the critical value for each test using a 5% level of significance, implying that the expected level of rejection should be close to 5%. For visual ease, we put in bold those cases where the estimated chance that a data set will wrongly reject Bayesian updating is higher than 10%. We do not include the non-heteroskedastic-robust autocorrelation tests, as the rejection rates are far higher than 10% and therefore the tests all appear misspecified in this environment. All of the robust autocorrelation tests (we report those with 1, 2 and 5 lagged periods) perform well given low signal strengths, but allowing for more lags appears to invalidate the test when the signal strength is high.

In the lower panel of Table 2, we report the estimated power—the likelihood that the test correctly rejects non-Bayesian updating—for streams generated by various parameterizations of non-Bayesian behavior discussed in the previous section, holding the signal strength constant at .75. Here, because some parameters (such as $\alpha = .8$ or $\beta = 1.5$) lead to rejection by nearly all of the tests all of the time, we choose parameters for which there is variation across tests and sample sizes.[42] The simple martingale test appears to have very little power, typically rejecting non-Bayesian beliefs less than 10% of the time. The excess-movement test has universally higher power than the autocorrelation tests, rejecting at around twice the best autocorrelation test for rejection rates that are less than 1. Finally, given that the person updates appropriately ($\alpha = 1, \beta = 1$), but starts with the wrong prior of $\pi_0 = .6$, our test and the autocorrelation tests do not appear well-suited to detect this error. This failure might be a positive feature insofar as researchers are more interested in testing flawed reasoning rather than bad priors.

There are multiple imperfections in the above analysis. We have only focused on specific biases in a specific environment and other tests might better detect other biases in other environments. For example, the simple martingale test will presumably be more powerful than our test to detect a bias of always moving in the direction of one specific state. However, better detection of this bias would presumably require some additional knowledge about the normative meaning of the state space that would drive a person to believe in one particular state. Similarly, if a person only temporarily overreacts to signals and shortly recognizes the appropriate likelihood-ratio of

[40] In a noisy signal environment, beliefs never reach certainty. Therefore, we end each belief stream after 100 draws, by which a Bayesian is almost always nearly certain.

[41] Although Table 2 only shows the results for five symmetric signal strengths, the conclusions below are robust to using different signal strengths and allowing for asymmetric signal strengths.

[42] The rejection rates change smoothly across these parameters and the signal strength, such that we do not expect the conclusions to change given different parameter choices.

Table 2: Type-I Error and Power of Various Martingale Tests

**Type 1 Error**

| | $\theta = .55$ Bayesian $(\alpha = 1, \beta = 1)$ | | | $\theta = .65$ Bayesian $(\alpha = 1, \beta = 1)$ | | | $\theta = .75$ Bayesian $(\alpha = 1, \beta = 1)$ | | | $\theta = .85$ Bayesian $(\alpha = 1, \beta = 1)$ | | | $\theta = .95$ Bayesian $(\alpha = 1, \beta = 1)$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of Streams | 10 | 30 | 100 | 10 | 30 | 100 | 10 | 30 | 100 | 10 | 30 | 100 | 10 | 30 | 100 |
| Excess Movement | .05 | .05 | .05 | .05 | .05 | .05 | .06 | .05 | .05 | .08 | .06 | .05 | **.35** | **.15** | .05 |
| Autocorrelation-1 (R) | .05 | .05 | .05 | .05 | .05 | .05 | .07 | .05 | .05 | **.12** | .07 | .06 | **.37** | **.13** | .07 |
| Autocorrelation-2 (R) | .05 | .05 | .05 | .05 | .05 | .05 | .05 | .05 | .05 | .05 | .05 | .05 | **.23** | .04 | .04 |
| Autocorrelation-5 (R) | .05 | .05 | .05 | .05 | .05 | .05 | .06 | .05 | .05 | .06 | .05 | .05 | **.14** | **.13** | .05 |
| Simple Martingale | .05 | .05 | .05 | .06 | .05 | .05 | .06 | .05 | .05 | .05 | .05 | .05 | .05 | .05 | .05 |

**Power**

| | $\theta = .75$ Underreaction $(\alpha = 1, \beta = .85)$ | | | $\theta = .75$ Overreaction $(\alpha = 1, \beta = 1.2)$ | | | $\theta = .75$ B-R Neglect $(\alpha = .95, \beta = 1)$ | | | $\theta = .75$ Conf. Bias $(\alpha = 1.2, \beta = 1)$ | | | $\theta = .75$ Wrong Prior $(\alpha, \beta = 1, \pi_0 = .6)$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of Streams | 10 | 30 | 100 | 10 | 30 | 100 | 10 | 30 | 100 | 10 | 30 | 100 | 10 | 30 | 100 |
| Excess Movement | **.13** | **.25** | **.63** | **.10** | **.23** | **.62** | **.09** | **.25** | **.70** | **.38** | **.76** | **1** | .05 | .05 | .06 |
| Autocorrelation-1 (R) | .08 | .08 | **.14** | .07 | .08 | **.14** | .07 | **.10** | **.24** | **.23** | **.43** | **.87** | .07 | .06 | .06 |
| Autocorrelation-2 (R) | .06 | .06 | .08 | .06 | .07 | .08 | .07 | .08 | **.14** | **.11** | **.19** | **.51** | .06 | .05 | .05 |
| Autocorrelation-5 (R) | .06 | .06 | .06 | .06 | .06 | .06 | .06 | .06 | .07 | .05 | .04 | .04 | .06 | .05 | .05 |
| Simple Martingale | .09 | .08 | .08 | .04 | .03 | .03 | .03 | .03 | .03 | **.10** | **.10** | **.10** | **.11** | **.20** | **.52** |

Note: This table reports the results of numerical simulations of the performance of various martingale tests given sample sizes of 10, 30, and 100 belief streams created using a simple multi-period symmetric-signal DGP. The upper panel reports the type 1 error: the likelihood of (incorrect) rejection of the null of Bayesian updating given belief streams created by Bayesian updating. The cutoff of each test is chosen to reject with 95% confidence, such that this error should be 5%. The lower panel displays the power - the likelihood of (correct) rejection of the null - given different types of non-Bayesian updating. In both panels, percentages over 10% are emphasized with bold text. For the autocorrelation tests, the number after the dash is the number of lags, and "(R)" emphasizes that the test is heteroskedastic-robust as the non-robust test appears misspecified.

the signal, an auto-correlation test might be more powerful.

# 5 Empirical Applications

As an exploration of the potential usefulness of the excess-movement statistics, we apply our main test to data. We first apply our tests to a data set of beliefs by a panel of individual volunteers about global events, and then to an algorithm used to map events in baseball games to the likelihood of each team winning the game. There are very few domains where individuals' probabilistic beliefs are elicited or can be directly inferred—and even fewer where the dynamics of beliefs are observed. Consequently, we additionally extend the analysis to (binary) prediction markets, with the caveat that there is a less clear mapping from our theory of individual beliefs to inferred "market beliefs." This issue is discussed in greater detail in Augenblick & Lazarus (2018), who adapt the individual theory to market outcomes.

Table 3: Excess movement in individual, algorithmic, and prediction markets

| | (1) Individual | (2) FanGraphs | (3) Betfair |
|---|---|---|---|
| Excess Movement $X$ | 0.0137*** | -0.00022*** | 0.00036*** |
| $Z$ | (4.22) | (9.40) | (19.55) |
| Norm. Excess M. $X_{norm}$ | 1.20 | 0.931 | 1.046 |
| Belief Observations | 593,218 | 926,083 | 7,422,530 |
| Event Observations | 291 | 11,753 | 286,257 |
| Individuals | 3,408 | - | - |
| Total Movement | 33,178 | 2,724 | 61,382 |

Note: This table reports statistics for the three belief datasets. Excess movement is the average difference between movement and uncertainty reduction for each belief change, which is equal in expectation to zero under the null of Bayesian updating. Normalized excess movement is average movement divided by average uncertainty reduction, which is equal to one under the null of Bayesian updating. The standard errors of the individual data are clustered at both the individual and event-level. The other standard errors are heteroskedasticity robust.

## 5.1 Individual Belief Streams

We begin by applying the approach to a large dataset, provided by and explored previously in Mellers et al. (2014) and Moore et al. (2017), that tracks individual probabilistic beliefs over an extended period of time. For the study, the authors recruited thousands of forecasters from email lists of scientific and professional societies, research centers, and alumni associations to continuously report beliefs about geopolitical events throughout 2011-2014.[43] For each question, forecasters reported probability estimates for events of their choice and were actively encouraged to update them over time. Forecasters were paid for participation, but not for accuracy. Because most forecasters did not record their (presumably) certain belief after the resolved, we add a final belief to each stream matching this ex-post realization.[44] This leaves a final dataset with 593,218 probability assessments by 3,408 forecasters about 291 events, with each forecaster making an average of 119 assessments on an average of 55 events.

We start with the analysis of excess movement in the data, shown in column (1) of Table 3. Because common information will potentially make belief movements correlated across individuals and potentially-related events, we cluster the standard errors at both the individual and event-level . The mean aggregate excess-movement statistic is 0.0137, which is statistically different from zero (Z=4.22, p<0.001), with 64% of forecasters having positive individual statistics. Normalizing the aggregate statistic by the average aggregate level of uncertainty reduction in each movement

[43] Following Moore et al. (2017), we do not include an additional year (2015) of data in which the general public was invited to participate. As might be expected, the additional year includes much more sporatic involvement by many more individuals. This data contains higher levels of excess movement.

[44] Without resolution included, the excess movement statistic drops (0.0098 compared to 0.0137). However, as the average reduction of uncertainty falls from 0.0683 to 0.0311, the normalized excess movement rises from 1.20 to 1.32.

(0.0683) leads to a normalized aggregate excess-movement statistic of 1.20, implying there is 20% too much belief movement given the uncertainty reduction in the dataset. This provides evidence that, as a group, these forecasters are not fully rational Bayesian updaters. To understand if measurement error could drive our results, we calibrate a Beta Distribution by choosing the first shape parameter to provide enough variance to match the excess-movement statistic in the data while choosing a period-specific second shape parameter such that the expectation of the distribution matches $\pi_t$ from the data at each period. The required parameters imply a standard deviation of measurement error of .13 when $\pi_t = .5$, such the error required to rationalize the data appears implausibly high.[45]

Further analysis provides evidence of systematic non-Bayesian behavior at the individual level. For a Bayesian, excess movement should be uncorrelated with excess movement in previous periods. However, there appears to be a strong persistence in the tendency toward excess movement within individuals across the dataset: an individual's excess-movement statistic in the second half of the year is positively ($r = .230$) and strongly significantly ($p < 0.001$) correlated with their excess-movement statistic in the first half of the year. This correlation appears strongly deflated by a large subset of individuals with very few assessments: the correlation rises to $r = 0.549$ ($p < 0.001$) when weighing individuals by their number of assessments, and similarly rises to $r = 0.709$ ($p < 0.001$) when focusing on individuals with an above-average number of assessments.

## 5.2 Algorithmic Evidence: Fangraphs

In some domains, such as sports and politics, algorithms have been developed to make dynamic probabilistic predictions about certain events. In this section, we examine the predictions of a popular baseball-statistics website called *Fangraphs*, which uses a variety of state variables—such as inning number, number of outs, and pitch count—to estimate a play-by-play probability that a team will win the game.[46] For example, in the May 25, 2008 game of the Cincinnati Reds versus the San Diego Padres, Kevin Kouzmanoff of the Padres tied the score at 6-6 by hitting a solo home run in the eighth inning. According to Fangraph's algorithm, this play changed the chance of the Padres winning from 26.8% to 59.6%. The algorithm apparently works by simply predicting the observed likelihood of winning in previous games with the same (or very similar) state variables.[47]

We collected 937,836 probability assessments from 11,753 games in the 2006-2010 baseball seasons directly from the Fangraphs website. Column (3) of Table 3 displays the excess-movement

[45] Intuitively, measurement error at each period leads to excess movement of $2\sigma_\varepsilon^2$, so to match the empirical excess movement statistic of 0.0137, the required average error variance is 0.0069, which corresponds to a standard deviation of around 0.08. The average error variance is a mix between necessarily-small variances near beliefs of certainty and larger variances at points such as $\pi_t = .5$, which is why we calibrate a larger standard deviation at that point.

[46] We thank Devin Pope for suggesting that we look at Fangraphs.

[47] Fangraph's algorithm is described as "identify[ing] all similar situations in the last ten years or so...and then find[ing] the winning percentage of teams who found themselves in those situations."

statistic ($-0.000219$), which is highly statistically different from zero ($Z = 9.4, p < 0.001$) and the normalized excess-movement statistic (.931). Therefore, without any information about the algorithm or the data-generating process of each game, we can conclude that the algorithm is "too tame" with high probability.

This result naturally raises the issue of how a seemingly-well designed algorithm could be (even slightly) "non-Bayesian." We point to three potential explanations. The first concerns the stability of the DGP: as the algorithm works by averaging outcomes from similar games in the past ten years, it will obviously produce poor predictions if the connection between situations and outcomes has changed over that period. For example, if teams up 3-1 in the 9th inning won 90% of games on average in the past, but win 100% of those games in today's environment, the algorithm will exhibit excess uncertainty reduction. However, even with a stable DGP, the necessity of aggregating finite past data leads to two additional issues. If the state variable is defined very precisely, there will be few "similar" previous games, leading to excessively-volatile predictions due to simple estimation error. Conversely, however, a too-coarse state definition can also be problematic. While the algorithm will pass our test if it *consistently* "ignores" information (such as the identity of the home team), it will potentially fail if it takes advantage of information that it later ignores. For example, if team A wins 100% of games when up 2-0 but only wins 75% of games when up by 2-1, the algorithm will shift from 100% to 75% if it partitions games based solely on current score because it "forgets" the full history of scores. Interestingly, while the algorithm is transparently non-Bayesian, it is—as with the example of the extreme base-rate neglector in Section 4—still well-calibrated.

## 5.3 Market Evidence: Betfair

Our final data comes from Betfair, a large British prediction market that matches individuals who wish to make opposing financial bets about a binary event. For example, if the current market "fractional odds" that the Denver Broncos win the Super Bowl is 2:1 (50%), individuals can either buy or sell a contract for $1 that pays $2 if Denver wins.[48] Just as with a stock market, the odds adjust depending on supply and demand. The data then consist of the time and odds of a *transaction* (matched bet on January 12th 2014 at 1:53:19 pm MST with odds of 1.21:1 (82.6%)) from a binary *contract* (Denver wins the game) with a given *selection* (Denver Broncos) focused on a particular *event* (Denver Broncos vs. San Diego Chargers in the 2014 Playoffs).

Of course, our theory is about an individual's beliefs rather than a price constructed from a complicated market interaction. We somewhat ignore this issue and follow the literature on prediction markets by constructing the "market beliefs" implied by the market odds, noting that our tests below can then be seen as the *joint* test that the market price can be interpreted as

[48] Sell!

beliefs and that these beliefs are Bayesian.[49]

We restrict the data in a variety of ways. First, we focus on major sporting events (soccer, American football, baseball, basketball, hockey) and only include transactions that occur when the game is being played. Second, we focus on contracts about the final winner of the game, dropping secondary (likely correlated) contracts such as the leader of the game at halftime.[50] For simplicity, if there are multiple selections in a contract (i.e. team A wins; team B wins; or, the teams draw), we focus on the binary beliefs of the event with the first transaction odds closest to .5.[51] Finally, as is common in trading data (see the discussion in Ait-Sahalia and Yu (2009)) we drop all but the first transaction for each one-minute increment to remove high-frequency trading noise, recalling that our theoretical results hold regardless of sampling time. These restrictions leave a raw dataset of over 7 million transactions from nearly three-hundred thousand events from 2006-2014.

The normalized excess movement, shown in column (3) of Table 3 is 1.046, which is highly statistically significantly positive, although reasonably close to 1. It appears that extremely-small trades are adding volatility to the odds: when we remove trades with volume less than 1% of the average trading volume, the statistic drops to 1.012. In this case, the normalized statistics for the five individual sports are 1.01, 1.16, 1.02, 0.99, and 1.04 for soccer, American football, baseball, basketball, hockey, respectively.[52] We conclude that markets prices imply beliefs that are reasonably close to satisfying the relationship between movement and uncertainty reduction, at least on average.

# 6 Conclusion

Our paper introduces a simple relationship between movement of beliefs and the reduction of uncertainty for a Bayesian updater, examines its implications, uses it to create statistical tests of rationality, studies the connection between the tests and four common psychological biases,

[49] See Manski (2005) and Wolfers and Zitzewitz (2006) for a debate about the validity of this assumption. Augenblick & Lazarus (2018) discuss conditions under which this type of market price can be perceived as a probability. Very loosely, the transformation requires standard assumptions in the finance literature, such as no arbitrage, and an additional assumption that, in equilibrium, participants' marginal value of an additional dollar in one state (Denver wins) is the same as the marginal value in another state (Denver doesn't win). We find these assumptions reasonably plausible in our domain.

[50] We ignore the possibility that belief movement for contracts *across* events are correlated (one unexpected storm might lead to multiple particularly volatile sporting events in close locations), but note that our results are robust to focusing on one contract per day.

[51] The criteria must be observable ex-ante: for example, it is not appropriate to focus on the selection with the most ex-post volume as volume might be correlated with movement. We have tried a variety of other criteria, such as choosing the selection with the highest volume of very early transactions or the selection with the earliest transactions (excluding these early transactions from the analysis). The results are robust to these choices.

[52] When we further remove trades with volume less than 5% and 10% of the average trading volume, the normalized statistic drops further to 1.007 and 1.004, respectively, with the individual sports statistics changing to (1.01,1.10,1.00,0.98,1.04) and (1.01,1.07,1.00,0.98,1.05), respectively.

compares the principles and performance of the main test with other tests of the martingale property, and then employs this test on datasets of beliefs.

We believe that our method can be expanded and improved in a variety of ways. First, while we have listed many theoretical results in Section 2, we guess that there are methods to tighten some of the bounds, either by improving the general results we present or by imposing some minimal restrictions on DGPs. Second, our analysis in Section 3 suggests that the theoretical concepts of movement and uncertainty reduction might help differentiate and categorize different biases. We believe that this intuitive method of categorization is promising, but there is much work to be done to expand the results to different biases and to different environments. Third, our empirical results from betting data hint that the market-aggregated "beliefs" might dampen the excess movement of individuals. Fourth, our emphasis on belief data makes salient the general lack of data about people's beliefs over time. As others have recognized, this is perhaps surprising given the core role of belief change from information in both economic and psychological theory, and efforts are underway to remedy this deficit. But along the lines we briefly developed in Section 2, we also see possibilities to apply our results to non-belief data, such as changes in asset prices or willingness-to-pay over time, which can be used to infer bounds on changes in beliefs.

Finally, and perhaps most importantly, the simple version of the test determines *whether* there is a violation of Bayesian updating and provides an intuitive magnitude of the deviation, but it does not determine *how* the violation occurred. While our results in Section 3 of differentiating among psychological biases by their effects separately on $\mathbb{E}M$ and $\mathbb{E}R$ departed from the theme of the rest of the paper by requiring us to know the correct Bayesian updating process, we suspect that it is possible to differentiate the biases (and even identify bias-model parameters) without much knowledge of the true information they are receiving by exploiting the fact that different biases may lead to different average belief-movement patterns across different time periods or situations.

# 7 References

Ait-Sahalia, Yacine and Jialin Yu, "High Frequency Market Microstructure Noise Estimates and Liquidity Measures," The Annals of Applied Statistics, Vol. 3, No. 1 (2009): 422–457

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold. "Parametric and nonparametric volatility measurement." Handbook of Financial Econometrics: Tools and Techniques. North-Holland, 2010. 67-137.

Augenblick, Ned and Eben Lazarus. "Restrictions on Asset-Price Movements Under Rational Expectations: Theory and Evidence." (2018), Working Paper.

Bar-Hillel, Maya. "The base-rate fallacy in probability judgments." Acta Psychologica 44, no. 3 (1980): 211-233.

Barndorff-Nielsen, Ole E., and Neil Shephard. "Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63.2 (2001): 167-241.

Benjamin, Daniel, Aaron Bodoh-Creed, and Matthew Rabin. “The Dynamics of Base-Rate Neglect.” Manuscript (2017)

Benjamin, Daniel J., Matthew Rabin, and Collin Raymond. “A model of nonbelief in the law of large numbers.” Journal of the European Economic Association 14, no. 2 (2016): 515-544.

Berg, J. E., Forsythe, R., Nelson, F. D., & Rietz, T. A. (2008). Results from a dozen years of election futures markets research. in C.R. Plott and V.L. Smith (Eds.), Handbook of Experimental Economic Results. Amsterdam: North Holland.

Bondt, Werner FM, and Richard Thaler. “Does the stock market overreact?.” The Journal of finance 40, no. 3 (1985): 793-805.

Burkholder, Donald L. “Distribution function inequalities for martingales.” the Annals of Probability (1973): 19-42.

Campbell, John Y., and Robert J. Shiller. “Cointegration and tests of present value models.” Journal of political economy 95, no. 5 (1987): 1062-1088.

Charles, Amelie and Olivier Darne. “Variance ratio tests of random walk: An overview.” Journal of Economic Surveys, Wiley, 23 3 2009: 503-527.

Chernoff, Herman. “A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations.” The Annals of Mathematical Statistics (1952): 493-507.

Cox, David C. “The Best Constant in Burkholder’s Weak-L^1 Inequality for the Martingale Square Function.” Proceedings of the American Mathematical Society (1982): 427-433.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica. “Suspense and surprise.” Journal of Political Economy 123, no. 1 (2015): 215-260.

Escanciano, J. Carlos, and Ignacio N. Lobato. “Testing the martingale hypothesis.” In Palgrave handbook of econometrics, pp. 972-1003. Palgrave Macmillan UK, 2009.

Frankel, Alex and Emir Kamenica.“Quantifying Information and Uncertainty.” Working paper (2018).

Grether, David M., “Bayes Rule as a Descriptive Model: The Representativeness Heuristic,” Quarterly Journal of Economics (1980), 95 (3), 537–557

Grether, David M, “Testing Bayes rule and the representativeness heuristic: Some experimental evidence,” Journal of Economic Behavior and Organization, (1992), 17 (1), 31–57.

Hall, Peter, and Christopher C. Heyde. “Martingale limit theory and its application.” Academic press, 1980.

Hoeffding, Wassily. “Probability inequalities for sums of bounded random variables.” Journal of the American statistical association 58, no. 301 (1963): 13-30.

Kahneman, Daniel, and Amos Tversky. “Subjective probability: A judgment of representativeness.” Cognitive psychology 3, no. 3 (1972): 430-454.

Kahneman, Daniel, and Amos Tversky. “On the psychology of prediction.” Psychological review 80, no. 4 (1973): 237.

LeRoy, S.F. and Porter, R.D., 1981. The present-value relation: Tests based on implied variance bounds. Econometrica: Journal of the Econometric Society, pp.555-574.

Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence D. Phillips. Calibration of probabilities: The state of the art. Decisions and Designs, 1976.

Lo, Andrew W., and A. Craig MacKinlay. “Stock market prices do not follow random walks: Evidence from a simple specification test.” The review of financial studies 1, no. 1 (1988): 41-66.

Manski, Charles F. “Interpreting the predictions of prediction markets.” Economics Letters 91, no. 3 (2006): 425-429.

Mellers, Barbara, Lyle Ungar, Jonathan Baron, Jaime Ramos, Burcu Gurcay, Katrina Fincher, Sydney E. Scott et al. “Psychological strategies for winning a geopolitical forecasting tournament.”

Psychological science 25, no. 5 (2014): 1106-1115.

Mertens, Jean-François, and Shmuel Zamir. "The maximal variation of a bounded martingale." Israel Journal of Mathematics 27, no. 3 (1977): 252-276.

Moore, Don A., Samuel A. Swift, Angela Minster, Barbara Mellers, Lyle Ungar, Philip Tetlock, Heather HJ Yang, and Elizabeth R. Tenney. "Confidence calibration in a multiyear geopolitical forecasting competition." Management Science 63, no. 11 (2017): 3552-3565.

Rabin, Matthew. "Inference by believers in the law of small numbers." The Quarterly Journal of Economics 117, no. 3 (2002): 775-816.

Rabin, Matthew, and Joel L. Schrag. "First impressions matter: A model of confirmatory bias." The Quarterly Journal of Economics 114, no. 1 (1999): 37-82.

Rabin, Matthew, and Dimitri Vayanos. "The gambler's and hot-hand fallacies: Theory and applications." The Review of Economic Studies 77, no. 2 (2010): 730-778.

Shannon, Claude E. "A mathematical theory of communication." ACM SIGMOBILE Mobile Computing and Communications Review 5, no. 1 (2001): 3-55.

Shiller, Robert J. "The volatility of long-term interest rates and expectations models of the term structure." Journal of political Economy 87, no. 6 (1979): 1190-1219.

Shiller, Robert J. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" American Economic Review, 71 (1981): 421–436

Stein, Jeremy. "Overreactions in the options market." The Journal of Finance 44, no. 4 (1989): 1011-1023.

Tversky, Amos, and Daniel Kahneman. "Availability: A heuristic for judging frequency and probability." Cognitive psychology 5, no. 2 (1973): 207-232.

Van der Vaart, Aad W. "Asymptotic statistics. Vol. 3." Cambridge university press, 2000, p.11

White, Halbert. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." Econometrica: Journal of the Econometric Society (1980): 817-838.

White, Halbert, and Ian Domowitz. "Nonlinear regression with dependent observations." Econometrica: Journal of the Econometric Society (1984): 143-161.

Wolfers, Justin, and Eric Zitzewitz. "Prediction markets." The Journal of Economic Perspectives 18, no. 2 (2004): 107-126.

Wolfers, Justin, and Eric Zitzewitz. Interpreting prediction market prices as probabilities. No. w12200. National Bureau of Economic Research, 2006.

# 8 Appendix

## 8.1 Proofs

**Proof of Proposition 1:**

There are many ways to prove the statement. One could also use the proof strategy in the Proof of Proposition 5 (Proposition 1 is, in fact, a Corollary of that Proposition). We therefore mainly present this proof for intuition.

We will first prove the statement that $\mathbb{E}[M_{t,t+1}] = \mathbb{E}[R_{t,t+1}]$ for any $t$:

$$
\begin{aligned}
\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}] &= \mathbb{E}[(\pi_{t+1} - \pi_t)^2] - \mathbb{E}[((1-\pi_t)\pi_t - (1-\pi_{t+1})\pi_{t+1})] \\
&= \mathbb{E}[(\pi_{t+1} - \pi_t)^2 - (1-\pi_t)\pi_t + (1-\pi_{t+1})\pi_{t+1}] \\
&= \mathbb{E}[(2\pi_t - 1)(\pi_t - \pi_{t+1})] \\
&= \mathbb{E}[(2\pi_t - 1)(\pi_t - \mathbb{E}[\pi_{t+1}|\pi_t])] \\
&= 0,
\end{aligned}
$$

where the first line is by the definition of $\mathbb{E}[M_{t,t+1}]$ and $\mathbb{E}[R_{t,t+1}]$, the second line uses the fact that expectation is a linear operator, the third line is simple algebraic simplification, the fourth line uses the law of iterated expectations, and the fifth line exploits that beliefs being a martingale: $\mathbb{E}[\pi_{t+1}|\pi_t] = \pi_t$.

Now, note that then:

$$
\begin{aligned}
\mathbb{E}[M_{t_1,t_2}] - \mathbb{E}[R_{t_1,t_2}] &= \sum_{\tau=t_1}^{t_2-1} \mathbb{E}[M_{\tau,\tau+1}] - \sum_{\tau=t_1}^{t_2-1} \mathbb{E}[R_{\tau,\tau+1}] \\
&= \sum_{\tau=t_1}^{t_2-1} (\mathbb{E}[M_{\tau,\tau+1}] - \mathbb{E}[R_{\tau,\tau+1}]) \\
&= 0,
\end{aligned}
$$

where the first line is by definition, the second line uses the the fact that expectation is a linear operator, the third line is from the previous proof.

**Proof of Corollary 1:**

Note that:

$$
\begin{aligned}
\mathbb{E}[M] &\equiv \mathbb{E}[M_{0,T}] \\
&= \mathbb{E}[R_{0,T}] \\
&= \sum_{\tau=0}^{T-1} \mathbb{E}[(1-\pi_t)\pi_t - (1-\pi_{t+1})\pi_{t+1}] \\
&= \sum_{\tau=0}^{T-1} \mathbb{E}[(1-\pi_t)\pi_t] - \mathbb{E}[(1-\pi_{t+1})\pi_{t+1}] \\
&= \mathbb{E}[(1-\pi_0)\pi_0] - \mathbb{E}[(1-\pi_T)\pi_T]
\end{aligned}
$$

where the first line is by definition, the second line is by the results above, the third line is by definition, the fourth line uses the the fact that expectation is a linear operator, and the fifth line cancels similar terms. But, for resolving DGPs, $\pi_T(H_T) \in \{0,1\}$ and therefore $\mathbb{E}[(1-\pi_T)\pi_T] = 0$ so $\mathbb{E}[M] = \mathbb{E}[(1-\pi_0)\pi_0] = (1-\pi_0)\pi_0 \equiv u_0$.

**Proof of Corollary 2:**

For the sake of contradiction, assume that $\Pr(M > \frac{u_0}{\delta}) > \delta$. Then:

$$
\begin{aligned}
\mathbb{E}[M] &= \Pr(M > \frac{u_0}{\delta}) \cdot \mathbb{E}[M|M > \frac{u_0}{\delta}] + \Pr(M \le \frac{u_0}{\delta}) \cdot \mathbb{E}[M|M \le \frac{u_0}{\delta}] \\
&> \delta \cdot \mathbb{E}[M|M > \frac{u_0}{\delta}] + (1-\delta) \cdot \mathbb{E}[M|M \le \frac{u_0}{\delta}] \\
&> \delta \cdot \frac{u_0}{\delta} + (1-\delta) \cdot \mathbb{E}[M|M \le \frac{u_0}{\delta}] \\
&> \delta \cdot \frac{u_0}{\delta} \\
&> u_0,
\end{aligned}
$$

where the first line is a simple conditional expectations, the second line comes from our maintained assumption that $\Pr(M > \frac{u_0}{\delta}) > \delta$, the third line comes from the fact that $\mathbb{E}[M|M > \frac{u_0}{\delta}] > \frac{u_0}{\delta}$, the fourth line from the fact that $\mathbb{E}[M|M \leq \frac{u_0}{\delta}] > 0$ as movement is always positive, and the final line is a simple simplification.

**Proof of Statement about Euclidean Distance:**

First, we will construct a simple DGP that produces a beliefs that follow a simple symmetric binary random walk with step size $\varepsilon$. Consider the following DGP with two signals $h$ and $l$ periods and for notational ease, define the state as $x = \{L, H\}$. Let $\Pr(s_{t+1} = h|x = H, \pi_t) = \frac{1}{2} + \frac{\varepsilon}{2\pi_t}$ and $\Pr(s_{t+1} = l|x = L, \pi_t) = \frac{1}{2} + \frac{\varepsilon}{2(1-\pi_t)}$. Given this DGP, a Bayesian with beliefs $\pi_t$ will have beliefs of $\pi_{t+1}$ equal to $\pi_t - \varepsilon$ or $\pi_t + \varepsilon$ with equal probability.

But, then standard results of symmetric random walks imply that, for any prior $\pi_0$ and for any length of path $m^*$, there is some $T$ and $\varepsilon$ such that paths will exceed that length with arbitrarily high probability, which proves the statement.

**Proof of Proposition 2:**

Recall that the random variable $M$ is defined as $\sum_{\tau=0}^{T-1}(\pi_{\tau+1} - \pi_\tau)^2$. For notational ease, define $(\pi_{\tau+1} - \pi_\tau)^2 = \Delta_\tau^2$ so $M = \sum_{\tau=0}^{T-1} \Delta_\tau^2$. We are trying to solve for the variance of $M$ :

$$\sigma^2[M] = \mathbb{E}[M^2] - \mathbb{E}[M]^2 \tag{1}$$

The previous result implies that $\mathbb{E}[M] = \pi_0(1 - \pi_0) = u_0$, so $\mathbb{E}[M]^2 = {u_0}^2$. Now, consider the term $\mathbb{E}[M^2] = \mathbb{E}[\left(\sum_{\tau=0}^{T-1} \Delta_\tau^2\right)^2]$ and break it into two parts:

$$\mathbb{E}[M^2] = \mathbb{E}[\textstyle\sum_{\tau=0}^{T-1}(\Delta_\tau^2)^2] + 2 \cdot \mathbb{E}[\sum_{\tau=0}^{T-1} \left(\Delta_\tau^2 \cdot \sum_{\upsilon=\tau+1}^{T-1} \Delta_\upsilon^2\right)] \tag{2}$$

Consider the first part of (2). As $\Delta_\tau^2 < 1$ for any history, it must be that $\sum_{\tau=0}^{T-1}(\Delta_\tau^2)^2 < \sum_{\tau=0}^{T-1} \Delta_\tau^2 = u_0$.

Consider the second part of (2). The law of iterated expectations implies that:

$$\begin{aligned}
2 \cdot \mathbb{E}[\textstyle\sum_{\tau=0}^{T-1} \left(\Delta_\tau^2 \cdot \sum_{\upsilon=\tau+1}^{T-1} \Delta_\upsilon^2\right)] &= 2 \cdot \mathbb{E}[\mathbb{E}[\textstyle\sum_{\tau=0}^{T-1} \left(\Delta_\tau^2 \cdot \sum_{\upsilon=\tau+1}^{T-1} \Delta_\upsilon^2\right)|H_{t+1}]] \\
&= 2 \cdot \mathbb{E}[\textstyle\sum_{\tau=0}^{T-1} \Delta_\tau^2 \cdot \mathbb{E}[\sum_{\upsilon=\tau+1}^{T-1} \Delta_\upsilon^2|H_{t+1}]] \\
&\leq 2 \cdot \frac{1}{4} \cdot \mathbb{E}[\textstyle\sum_{\tau=0}^{T-1} \Delta_\tau^2] \\
&= \frac{1}{2} \cdot u_0
\end{aligned}$$

Where the inequality arises from the fact that $\mathbb{E}[\sum_{\upsilon=\tau+1}^{T-1} \Delta_\upsilon^2|H_{t+1}] \leq \frac{1}{4}$ because uncertainty is at most $\frac{1}{4}$. Therefore,

$$\mathbb{E}[M^2] < \frac{3}{2}u_0$$

And therefore:

$$\sigma^2[M] < \frac{3}{2}u_0 - {u_0}^2 = u_0(\frac{3}{2} - u_0)$$

**Proof of Corollary 3:**

Consider Chebyshev's One-Sided Inequality (also called Cantelli's inequality): Let $M$ be a random variable with finite expected value $\mu_0$ and finite non-zero variance $\sigma^2$. Then for any real number $a > 0$,

$P(M \geq \mu_0 + a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$. Rewritten to match the style of the theorem, this implies that, for any $\delta \in (0,1)$, $P(M \geq \mu_0 + \sigma\frac{\sqrt{(1-\delta)}}{\sqrt{\delta}}) \leq \delta$ where $\sigma \equiv +\sqrt{\sigma^2}$.

For our situation, the above statements imply that $\mu_0[M] = u_0 = \pi_0(1-\pi_0)$ and $\sigma^2[M] < u_0(\frac{3}{2} - u_0)$. Therefore: $P(M > u_0 + \sqrt{u_0(\frac{3}{2} - u_0)}\frac{\sqrt{(1-\delta)}}{\sqrt{\delta}}) \leq \delta$

**Proof of Corollary 4:**

First, consider any prior $\pi_0 \geq \frac{1}{2}$. Consider the gaffe DGP discussed in the paper with $\phi = 1 - \pi_0^{\frac{1}{T}}$, which has this prior. State 1 (the politician wins the election) occurs with probability $\pi_0$, and in this case, the beliefs of the person at time $t$ are: $(\pi_0^{\frac{1}{T}})^{T-t}$, and therefore the movement in the DGP with $T$ periods is $M(T) = \sum_{\tau=0}^{T-1}((\pi_0^{\frac{1}{T}})^{T-\tau+1} - (\pi_0^{\frac{1}{T}})^{T-\tau})^2 = \sum_{\tau=0}^{T-1}((\pi_0^{\frac{1}{T}})^{T-\tau}(\pi_0^{\frac{1}{T}} - 1))^2$. As $T \to \infty$, $M(T) \to 0$. Therefore, for any $M^*$, there exists some $T$ such that $M(T) < M^*$. Therefore, given this $T$, $\Pr(M < M^*) = \pi_0$. This proves the statement for $\pi_0 \geq \frac{1}{2}$. For priors $\pi_0 < \frac{1}{2}$, consider a gaffe DGP, but in which the person has beliefs over the likelihood of state 0 rather than 1. In this case, the same proof shows that, for and $M^* > 0$, there is some $T$ such that $\Pr(M < M^*) = 1 - \pi_0$. This proves the statement for $\pi_0 < \frac{1}{2}$.

**Proof of Corollary 5:**

Consider Chebyshev's Inequality: Let $M$ be a random variable with finite expected value $\mu_0$ and finite non-zero variance $\sigma^2$. Then for any real number $a > 0$, $P(M \geq \mu_0 + a \cdot \sigma) \leq \frac{1}{a^2}$. Rewritten to match the style of the theorem, this implies that, for any $\delta \in (0,1)$, $P(M \geq \mu_0 + \frac{\sigma}{\sqrt{\delta}}) \leq \delta$ where $\sigma \equiv +\sqrt{\sigma^2}$.

For our situation with $n$ streams, denote the random variable representing the movement statistic from independent stream $i$ as $M(i)$ and define $\mathbf{M} = \sum_{i=n}^{N} M(i)$. As the mean of $M(i) = u_0$ for all $i$, the variances are bounded by $u_0(\frac{3}{2} - u_0)$, and the variables are assumed independent: $\mathbb{E}[\mathbf{M}] = n \cdot u_0$ and $\sigma^2[\mathbf{M}] < n \cdot u_0(\frac{3}{2} - u_0)$.Therefore: $\Pr(|\mathbf{M} - n \cdot u_0| > \frac{\sqrt{n \cdot u_0(\frac{3}{2} - u_0)}}{\sqrt{\delta}}) \leq \delta$

**Proof of Proposition 3:**

The statement is nearly a corollary of Proposition 1. Consider the beliefs $\pi_t^x$. This is the binary belief that state $x$ occurs or does not occur. Define $m_{t_1,t_2}^x(\pi) \equiv \sum_{\tau=t_1}^{t_2-1}(\pi_\tau^x - \pi_{\tau-1}^x)^2$, $u_t^x = (1-\pi_t^x)\pi_t^x$, and $r_{t_1,t_2}^x(\pi) \equiv u_{t_1}^x(\pi) - u_{t_2}^x(\pi)$, with the related capital-letter random variables defined as normal. Proposition 1 implies: $\mathbb{E}M_{t_1,t_2}^x = \mathbb{E}R_{t_1,t_2}^x$. But then:

$$\begin{aligned}
\sum_{x \in X} \mathbb{E}M_{t_1,t_2}^x &= \sum_{x \in X} \mathbb{E}R_{t_1,t_2}^x \\
\sum_{x \in X} \sum_{\tau=t_1}^{t_2-1} \mathbb{E}[(\pi_\tau^x - \pi_{\tau-1}^x)^2] &= \sum_{x \in X} (u_{t_1}^x - u_{t_2}^x)] \\
\mathbb{E}[\sum_{x \in X} \sum_{\tau=t_1}^{t_2-1} (\pi_\tau^x - \pi_{\tau-1}^x)^2] &= \mathbb{E}[\sum_{x \in X} (u_{t_1}^x - u_{t_2}^x)] \\
\mathbb{E}M_{t_1,t_2}^\Sigma &= \mathbb{E}R_{t_1,t_2}^\Sigma,
\end{aligned}$$

where the first and third lines follows from the fact the expectation is a linear operator, ad the second and fourth lines are by definition.

**Proof of Proposition 4:**

Consider some expectations stream $v = [v_0, v_1, v_2, ...v_T]$ and $\delta \in (0,1)$.

Intuitive Note: The basic idea will be to construct a simple binary DGP that moves from each $v_t$ to $v_{t+1}$ in the stream $v$ with arbitrarily high probability. The simple idea is to keep the expectations stream a martingale by making the movement in the opposite direction very large, so that even though this movement occurs with small probability, $E[v_{t+1}] = v_t$.

Formally, consider a situation with two states (which for notation ease) we call $x \in \{L, H\}$ with some state values $v^H > \max(v)$ and $v^L < \min(v)$ and two signals $h$ and $l$. Consider the following DGP at time $t$ that is determined by the assumed stream $v$.

If $v_{t+1} > v_t$, define: $\Pr(s_{t+1} = h|x = H, v_t) = 1, \Pr(s_{t+1} = l|x = L, v_t) = \frac{(v_{t+1}-v_t)(v^H-v^L)}{(v_{t+1}-v^L)(v^H-v_t)}$. Given this signal distribution, a Bayesian will update expectations from $v_t$ to $v_{t+1}$ in the case of a signal $h$ and will update from $v_t$ to $v^L$ in the case of signal $l$, where the likelihood of observing a signal $h$ is:

$$\begin{aligned}
\Pr(s_{t+1} &= h|v_t) \\
&= \Pr(x = H|v_t) \cdot \Pr(s_{t+1} = h|x = H, v_t) + \Pr(x = L|v_t) \cdot \Pr(s_{t+1} = h|x = L, v_t) \\
&= \Pr(x = H|v_t) \cdot 1 + \Pr(x = L|v_t) \cdot (1 - \frac{(v_{t+1} - v_t)(v^H - v^L)}{(v_{t+1} - v^L)(v^H - v_t)}) \\
&= \frac{v_t - v^L}{v^H - v^L} \cdot 1 + (1 - \frac{v_t - v^L}{v^H - v^L}) \cdot (1 - \frac{(v_{t+1} - v_t)(v^H - v^L)}{(v_{t+1} - v^L)(v^H - v_t)}) \\
&= \frac{v_t - v^L}{v_{t+1} - v^L}.
\end{aligned}$$

Importantly, this probability of transition from $v_t$ to $v_{t+1}$ can be made arbitrarily high by reducing $v^L$.

Similarly, if $v_{t+1} < v_t$, one can define a signal distribution such that a Bayesian will update expectations to $v_{t+1}$ in the case of a signal $l$ and will update to $v^H$ in the case of signal $h$, with the signal $l$ occuring with probability $\frac{v^H-v_t}{v^H-v_{t+1}}$, which can be made arbitrarily high by increasing $v^H$.

Finally, if $v_{t+1} = v_t$, then set $\Pr(s_{t+1} = h|x = H, v_t) = \frac{1}{2}, \Pr(s_{t+1} = l|x = L, v_t) = \frac{1}{2}$, such that the person always updates to $v_{t+1}$.

Now, as each of the probabilities of transitioning from $v_t$ to $v_{t+1}$ can be made arbitrarily high in this DGP, we can make the likelihood of the entire stream $v$ arbitrarily high. That is, for any $\delta \in (0, 1)$, there exists $v^H$ and $v^L$ such that the above DGP produces expectations stream $v$ with probability above $\delta$.

**Proof of Corollary 6:**

This is a straightforward corollary. For a given value $m^*$ and initial expectation $v_0^*$ there is obviously some stream $v$=$[v_0^*, v_1, v_2, ...v_T]$ with $m^v(v) > m^*$. Occuring to the previous Proposition, there exists a DGP such that this stream is arbitrarily likely, and therefore $\Pr(M^v \geq m^*)$ can be made arbitrarily high.

**Proof of Proposition 5 and Corollary 7:**

(The proof of the Proposition includes a proof of the Corollary.)

Bayesian updating implies that expected value streams are a martingale and therefore the martingale difference sequence $\Delta_t = V_{t+1} - V_t$ is a serially uncorrelated random variable with a mean of zero. Given that the expectation is bounded between $\min(v^1, v^2, ...,)$ and $\max(v^1, v^2, ...,)$, $Var(\Delta_\tau)$ is bounded.

We first consider resolving DGPs (those in which $u_T^v = 0$ for all histories). We proceed in three steps:

1) We start by proving that $\mathbb{E}[M_{0,T}^v] = \mathbb{E}[R_{0,T}^v] = u_0^v$ (Corollary 7**).** Consider the following transfor-

mations:

$$\begin{aligned}
\sum_{\tau=0}^{T-1} Var(\Delta_\tau) &= Var(\sum_{\tau=0}^{T-1} \Delta_\tau) \\
\sum_{\tau=0}^{T-1} Var(V_{t+1}-V_t) &= Var(\sum_{\tau=0}^{T-1} (V_{t+1}-V_t)) \\
\sum_{\tau=0}^{T-1} Var(V_{t+1}-V_t) &= Var(V_T-V_0) \\
\sum_{\tau=0}^{T-1} \mathbb{E}[V_{t+1}-V_t]^2 &= \mathbb{E}[V_T-V_0]^2 \\
\mathbb{E}[\sum_{\tau=0}^{T-1} (V_{t+1}-V_t)^2] &= \mathbb{E}[V_T-V_0]^2 \\
\mathbb{E}[\sum_{\tau=0}^{T-1} (V_{t+1}-V_t)^2] &= \sum_{x\in X} \pi_0^x (v^x - v_0)^2 \\
\mathbb{E}M_{0,T}^v &= u_0^v,
\end{aligned}$$

where the first equality arises because the difference variables are uncorrelated, implying that the variance of the sum of the variables equals the sum of the variances of the variables, the second line is by definition, the third line uses the simple fact that $\sum_{\tau=0}^{T-1}(V_{t+1}-V_t) = V_T - V_0$, the four line uses (i) the definition of variance $Var(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$ and (ii) that fact that (because $V_t$ is a martingale), $\mathbb{E}[V_{t+1} - V_t] = 0$ for any $t$ and $\mathbb{E}[V_T - V_0] = 0$, the fifth line uses the fact that expectation is a linear operator, the sixth line uses the fact that (i) the stream is resolving and (ii) the person has correct priors, such that with probability $\pi_0^x$, $V_T = v^x$, and the seventh line is by definition.

2) Next, we show that $\mathbb{E}[M_{t_1,T}^v] = \mathbb{E}[R_{t_1,T}^v] = \mathbb{E}[u_{t_1}^v]$. To see this, note that—making the dependency on history explicit—it must be that $\mathbb{E}[M_{t_1,T}^v|H_{t_1}] = u_{t_1}^v|H_{t_1}$ for any history $H_{t_1}$ because the signals after $H_{t_1}$ can be seen as a new DGP with $T+1-t_1$ periods starting with prior $v(H_{t_1})$ such that the expected movement following $H_{t_1}$ must equal the initial uncertainty following $H_{t_1}$. Therefore $\mathbb{E}_{H_{t_1}}[\mathbb{E}[M_{t_1,T}^v|H_{t_1}]] = \mathbb{E}_{H_{t_1}}[u_{t_1}^v|H_{t_1}]$, which, (suppressing the dependencies again) is $\mathbb{E}[M_{t_1,T}^v] = \mathbb{E}[u_{t_1}^v] = \mathbb{E}[R_{t_1,T}^v]$.

3) Finally, showing $\mathbb{E}[M_{t_1,t_2}^v] = \mathbb{E}[R_{t_1,t_2}^v]$ from the previous results is transparent:

$$\begin{aligned}
\mathbb{E}[M_{t_1,t_2}^v] &= \mathbb{E}[M_{t_1,T}^v] - \mathbb{E}[M_{t_2,T}^v] \\
&= \mathbb{E}[R_{t_1,T}^v] - \mathbb{E}[R_{t_2,T}^v] \\
&= \mathbb{E}[R_{t_1,t_2}^v]
\end{aligned}$$

where the first line is a simple transformation of $\mathbb{E}[M_{t_1,T}^v] = \mathbb{E}[M_{t_1,t_2}^v] + \mathbb{E}[M_{t_2,T}^v]$, which uses the fact that expectation is a linear operator, the second line uses the previous result, and the third is uses $\mathbb{E}[R_{t_1,T}^v] = \mathbb{E}[R_{t_1,t_2}^v] + \mathbb{E}[R_{t_2,T}^v]$.

Finally, showing the same result for non-resolving DGPs is fairly straightforward. Consider some DGP with $T$ periods that is non-resolving. There exists a DGP' with $T-1$ periods with the same prior and signal distributions as the original DGP for the first $T$ periods with an added resolving signal (a signal that reveals the state) in period $T+1$. For DGP', the previous results suggest that $\mathbb{E}[M_{t_1,t_2}^v] = \mathbb{E}[R_{t_1,t_2}^v]$, but then this also must be true for the original DGP, which must share the same Bayesian expected values up to period $T$.

**Proof of Corollary 8:**

The proof is a result of the previous corollary and the fact that, for a random variable with expectation $v_0$, a low bound $\underline{v}$ and an upper bound $\overline{v}$, the largest variance is $(\overline{v} - v_0)(v_0 - \underline{v})$. Therefore, $u_0^v < (\overline{v} - v_0)(v_0 - \underline{v})$.

For a resolving stream, the previous corollary suggests that then: $\mathbb{E}M^v = u_0^v < (\overline{v} - v_0)(v_0 - \underline{v})$.

For a non-resolving stream, the logic follows that at the end of the previous proof: given a non-resolving DGP with $T$ periods, there exists a similar resolving DGP' with $T-1$ periods. For this resolving DGP', $\mathbb{E}M^v_{0,T} < \mathbb{E}M^v_{0,T+1} = u^v_0 < (\overline{v} - v_0)(v_0 - \underline{v})$, but the expected movement for the original DGP is equal to $\mathbb{E}M^v_{0,T}$, so the result holds.

−−−

For the next set of results, we will use the following two Lemmas:

**Lemma A:**

Given that $\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}] = \mathbb{E}[(1-2\pi_t)(\mathbb{E}[\pi_{t+1}|\pi_t] - \pi_t)]$ :
If $\pi_t > \frac{1}{2}$ and $\mathbb{E}[\pi_{t+1}] - \pi_t \gtreqless 0$, then $\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}] \lesseqgtr 0$.
If $\pi_t < \frac{1}{2}$ and $\mathbb{E}[\pi_{t+1}] - \pi_t \gtreqless 0$, then $\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}] \gtreqless 0$.
If $\pi_t = \frac{1}{2}$ and $\mathbb{E}[\pi_{t+1}] - \pi_t \lesseqgtr 0$, then $\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}] = 0$.

**Proof of Lemma A:**

Following the proof of Proposition 1, $\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}] = \mathbb{E}[(2\pi_t - 1)(\pi_t - \mathbb{E}[\pi_{t+1}|\pi_t])] = \mathbb{E}[(1 - 2\pi_t)(\mathbb{E}[\pi_{t+1}|\pi_t] - \pi_t)]$. Note that, if $\pi_t > \frac{1}{2}$, then $1 - 2\pi_t < 0$, and so $\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}]$ has the opposite sign as $\mathbb{E}[\pi_{t+1}] - \pi_t$. If $\pi_t < \frac{1}{2}$, then $1 - 2\pi_t > 0$, and so $\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}]$ has the same sign of $\mathbb{E}[\pi_{t+1}] - \pi_t$. If $\pi_t = \frac{1}{2}$, the excess-movement statistic is zero regardless of $\mathbb{E}[\pi_{t+1}] - \pi_t$.

**Setup for Propositions 6-8:**

We now discuss the four biases. We start with some setup used in the next few proofs. For notational simplicity, define $p(l) \equiv \frac{l}{1+l}$ (as a function that maps likelihood ratios to probabilities) and define $l(p) \equiv \frac{p}{1-p}$ (a function that maps a probability into a likelihood ratio). The likelihood ratio associated with a signal $h$ is $l(\theta)$ and a signal $l$ is $\frac{1}{l(\theta)}$. We sometimes abuse notation by using $l(s_t)$ to denote $l(\theta)$ or $\frac{1}{l(\theta)}$ depending on the signal realization.

Consider a Bayesian observer with beliefs $\pi^*_t$. She will transition into either $\pi^{*h}_{t+1} = p(l(\pi^*_t) \cdot l(\theta))$ if $s_{t+1} = h$ or $\pi^{*l}_{t+1} = p(l(\pi^*_t) \cdot \frac{1}{l(\theta)})$ if $s_{t+1} = l$. As the person is Bayesian, the martingale property must hold, implying that (given algebraic simplification):

$$\Pr(s_{t+1} = h|H_t) = \frac{\pi^*_t - \pi^{*l}_{t+1}}{\pi^{*h}_{t+1} - \pi^{*l}_{t+1}} = 1 - \theta + \pi^*_t(2\theta - 1). \tag{3}$$

Next, consider a biased person with beliefs $\pi_t$ who transitions into either $\pi^h_{t+1} = p(l(\pi_t)^\alpha \cdot l(\theta)^\beta)$ or $\pi^l_{t+1} = p(l(\pi_t)^\alpha \cdot (\frac{1}{l(\theta)})^\beta)$. We will focus on the true expectation of the biased person's period $t+1$ belief minus the belief at period $t$:

$$\mathbb{E}[\pi_{t+1}|H_t] - \pi_t = \Pr(s_{t+1} = h|H_t)\pi^h_{t+1} + (1 - \Pr(s_{t+1} = h|H_t))\pi^l_{t+1} - \pi_t \tag{4}$$

Note that (4) and $(2\pi_t - 1)$, maps to $\mathbb{E}[M_{t,t+1} - R_{t,t+1}]$ by Lemma A.

**Proof of Proposition 6:**

We now consider the one-period adjustment to $\pi_{t+1}$ when the biased prior is the same as the Bayesian prior ($\pi_t = \pi^*_t$).

Intuitive discussion of next section: We now consider the case in which the person has the correct prior and either has base-rate neglect or confirmation bias. To understand the basic idea of the proof, consider a prior greater $\frac{1}{2}$. In this case, given the ideas above about .5 reversion, if the expectation of the posterior beliefs following the signal is greater than the prior, we have reversion away from .5 and too little movement, while if the expectation is lower, we have .5 reversion and too much movement. First, consider confirmation bias. Because the person overweights her higher-than-.5-prior, she will have posterior beliefs that are higher than a Bayesian's beliefs for every signal. As the expectation of the Bayesian's posterior beliefs must be equal to her prior, the expectation of the biased beliefs must be higher than the prior (which is also the biased person's prior), leading to too little movement. Furthermore, as confirmation bias increases, this effect increases, increasing the deviation of posterior expectation from prior, thereby increasing the absolute magnitude of the excess-movement statistic. Conversely, for base-rate neglect, the person underweights her prior, leading to biased beliefs that are always lower than the Bayesian beliefs for every signal, leading to too much movement.

Formally, we first consider the case in which $\alpha \neq 1$ and $\beta = 1$ and $\pi_t = \pi_t^*$. Given that, by definition, $\Pr(s_{t+1} = h|H_t)\pi_{t+1}^{*l} + (1 - \Pr(s_{t+1} = h|H_t))\pi_{t+1}^{*l} = \pi_t^*$ and that $\pi_t = \pi_t^*$, (4) can be rewritten as:

$$\mathbb{E}[\pi_{t+1}|H_t] - \pi_t = p(s_{t+1} = h|H_t)(\pi_{t+1}^h - \pi_{t+1}^{*h}) + (1 - p(s_{t+1} = h|H_t))(\pi_{t+1}^l - \pi_{t+1}^{*l}) \tag{5}$$

Focus on the case of $\pi_t^* > .5$ and $\alpha < 1$. In this case, $\pi_{t+1}^h = p(l(\pi_t)^\alpha \cdot l(\theta)) < p(l(\pi_t) \cdot l(\theta)) = \pi_{t+1}^{h*}$ as $l(\pi_t)^\alpha < l(\pi_t)$ and $p(l)$ is monotonically increasing in $l$. Similarly, $\pi_{t+1}^l = p(l(\pi_t)^\alpha \cdot \frac{1}{l(\theta)}) < p(l(\pi_t) \cdot \frac{1}{l(\theta)}) = \pi_{t+1}^{l*}$. Therefore, $\mathbb{E}[\pi_{t+1}|H_t] - \pi_t < 0$. Using very similar logic, one can show if $\alpha < 1$ and $\pi_t = \pi_t^* < .5$, then $\mathbb{E}[\pi_{t+1}|H_{t+1}] - \pi_t > 0$. Given Lemma A, note that, when $\alpha < 1$, then either ($\mathbb{E}[\pi_{t+1}] < \pi_t$ and $\pi_t > .5$) or ($\mathbb{E}[\pi_{t+1}] > \pi_t$ and $\pi_t < .5$). In both of these cases, the expected excess-movement statistic must be positive.

Similarly logic can show that (i) if $\alpha > 1$ and $\pi_t = \pi_t^* > .5$, then $\mathbb{E}[\pi_{t+1}|H_{t+1}] - \pi_t > 0$, and (ii) if $\alpha > 1$ and $\pi_t = \pi_t^* < .5$, then $\mathbb{E}[\pi_{t+1}|H_{t+1}] - \pi_t < 0$. Therefore, the similar logic to above, Lemma A shows that the excess-movement statistic must be negative.

We now show that the effect on the excess-movement statistic is monotonic in $\alpha$. Again, start by focusing on the case in which $\pi_t^* > .5$. Recalling that $\pi_{t+1}^h = p(l(\pi_t)^\alpha \cdot l(\theta))$, then this case, $\frac{\partial \pi_{t+1}^h}{\partial \alpha} > 0$ as $\frac{\partial (l(\pi_t)^\alpha)}{\partial \alpha} > 0$ and $p(l)$ is monotonically increasing in $l$. Similarly, $\frac{\partial \pi_{t+1}^l}{\partial \alpha} > 0$. But, then $\frac{\partial(\mathbb{E}[\pi_{t+1}|H_t] - \pi_t)}{\partial \alpha} = p(s_{t+1} = h|H_t)\frac{\partial \pi_{t+1}^h}{\partial \alpha} + (1 - p(s_{t+1} = h|H_t))\frac{\partial \pi_{t+1}^l}{\partial \alpha} > 0$ as $p(s_{t+1} = h|H_t) > 0$ and $(1 - p(s_{t+1} = h|H_t)) > 0$. Given that $(1 - 2\pi_t) < 0$, this implies that $\frac{\partial(\mathbb{E}[\pi_{t+1}|H_t] - \pi_t)(1-2\pi_t)}{\partial \alpha} < 0$ which is equivalent to $\frac{\partial(\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}])}{\partial \alpha} < 0$. Similar logic shows that, when $\pi_t^* < .5$, $\frac{\partial(\mathbb{E}[\pi_{t+1}|H_t] - \pi_t)}{\partial \alpha} < 0$ such that $\frac{\partial(\mathbb{E}[\pi_{t+1}|H_t] - \pi_t)(1-2\pi_t)}{\partial \alpha} < 0$, and therefore $\frac{\partial(\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}])}{\partial \alpha} < 0$. Consequently, when $\pi_t^* \neq .5$, $\frac{\partial(\mathbb{E}[M_{t,t+1}] - \mathbb{E}[R_{t,t+1}])}{\partial \alpha} < 0$.

Intuitive discussion of next section: We now consider the case in which the person has the correct prior and either has overreaction or underreaction. To understand the basic idea of the proof, consider a prior greater $\frac{1}{2}$. In this case, given the ideas above about .5 reversion, if the expectation of the posterior beliefs following the signal is greater than the prior, we have reversion away from .5 and too little movement, while if the expectation is lower, we have .5 reversion and too much movement. First, consider overreaction. We cannot use the same proof strategy as with base-rate neglect and confirmation bias, because the biased posteriors will be higher than the Bayesian posteriors when the signal is $h$, but lower when the signal is $l$. Therefore, the essence of the proof is that, because the prior is above $\frac{1}{2}$, the biased belief after the $h$ signal does not deviate as much from the Bayesian's belief as the biased belief from a $l$ signal deviates from the Bayesian's belief. As the expectation of the Bayesian's posterior beliefs must be equal to her prior, the expectation of the biased beliefs must be lower than the Bayesian prior (which is also the biased person's prior), leading to too much movement. Furthermore, as overreaction increases, this effect increases, increasing the deviation of posterior expectation from prior, thereby increasing the absolute magnitude of the excess-movement statistic. Conversely, for underreaction, the person underweights her signal, leading to the opposite effect. However, the absolute magnitude of the effect is not monotonic in $\beta$ when $\beta < 1$ because as $\beta \to 0$, the biased posteriors become very close to the prior and therefore the absolute deviation of the posterior expectation from the prior becomes very small.

Formally, we now consider the case in which $\alpha = 1$ and $\beta \neq 1$ and $\pi_t = \pi_t^*$.

We define:

$$p = \frac{\pi_t - \pi^l_{t+1}}{\pi^h_{t+1} - \pi^l_{t+1}} = 1 - \widehat{\theta} + \pi_t(2\widehat{\theta} - 1), \text{ where } \widehat{\theta} \equiv \frac{1}{1 + (\frac{1}{\theta} - 1)^\beta}, \tag{6}$$

which is the probability of a high signal required for the biased person beliefs to be a martingale (that is, $p\hat{\pi}^h_{t+1} + (1-p)\pi^l_{t+1} = \pi_t$). Intuitively, $\widehat{\theta}$ represents perceived precision of the biased person. Subbing $\pi_t = p\hat{\pi}^h_{t+1} + (1-p)\pi^l_{t+1}$ into (4) yields:

$$\mathbb{E}[\pi_{t+1}|H_t] - \pi_t = (p(s_{t+1} = h|H_t) - p)(\pi^h_{t+1} - \pi^l_{t+1}). \tag{7}$$

Given our assumption of the correct prior, $\pi_t = \pi^*_t$, so $p = 1 - \widehat{\theta} + \pi^*_t(2\widehat{\theta} - 1)$. Subbing this and $p(s_{t+1} = h|H_t)$ from (3) into (7) and simplifying yields:

$$\mathbb{E}[\pi_{t+1}|H_t] - \pi_t = (2\pi^*_t - 1)(\theta - \widehat{\theta})(\pi^h_{t+1} - \pi^l_{t+1}). \tag{8}$$

Focus on the case of $\pi^*_t > .5$ and $\beta > 1$. The other cases have the symmetric proofs. We show that $\mathbb{E}[\pi_{t+1}|H_t] - \pi^*_t < 0$. First, $\hat{\theta} = \frac{1}{1+(\frac{1}{\theta}-1)^\beta} > \theta$ because $\theta > \frac{1}{2}$ by assumption. Then $\mathbb{E}[\pi_{t+1}|H_t] - \pi_t > 0$ as $\pi^h_{t+1} > \pi^l_{t+1}$, $\hat{\theta} > \theta$, and $2\pi^*_t - 1 > 0$ because $\pi^*_t > .5$ by assumption. Using very similar logic, one can show that, if $\pi^*_t < .5$, then $\mathbb{E}[\pi_{t+1}|H_{t+1}] - \pi_t > 0$. Given Lemma A, note that, when $\beta > 1$, then either ($\mathbb{E}[\pi_{t+1}] < \pi_t$ and $\pi_t > .5$) or ($\mathbb{E}[\pi_{t+1}] > \pi_t$ and $\pi_t < .5$). In both of these cases, the expected excess-movement statistic must be positive.

Similar proofs show that when $\beta < 1$ (i) if $\pi_t = \pi^*_t > .5$, then $\mathbb{E}[\pi_{t+1}|H_{t+1}] - \pi_t > 0$ and (ii) if $\pi_t = \pi^*_t < .5$, then $\mathbb{E}[\pi_{t+1}|H_{t+1}] - \pi_t < 0$. Then, similar application of Lemma A shows that the excess-movement statistic must be negative.

We now show monotonicity of the statistic when $\beta > 1$. Again, we first focus on the situation in which $\pi^*_t > \frac{1}{2}$. Consider $\frac{\partial(\mathbb{E}[\pi_{t+1}|H_t]-\pi_t)}{\partial\beta} = \frac{\partial(\theta-\widehat{\theta})(\pi^h_{t+1}-\pi^l_{t+1})(2\pi^*_t-1)}{\partial\beta} = (2\pi^*_t - 1)\frac{\partial(\theta-\widehat{\theta})}{\partial\beta}(\pi^h_{t+1} - \pi^l_{t+1}) + (2\pi^*_t - 1)\frac{\partial(\pi^h_{t+1}-\pi^l_{t+1})}{\partial\beta}(\theta - \hat{\theta})$. It is easy to show that $\frac{\partial\pi^h_{t+1}}{\partial\beta} > 0$ and $\frac{\partial\pi^l_{t+1}}{\partial\beta} < 0$ and therefore $\frac{\partial(\pi^h_{t+1}-\pi^l_{t+1})}{\partial\beta} > 0$. Similarly, $\frac{\partial\widehat{\theta}}{\partial\beta} > 0$ and therefore $\frac{\partial(\theta-\widehat{\theta})}{\partial\beta} < 0$. Given that $\theta - \hat{\theta} < 0$ and $\pi^h_{t+1} - \pi^l_{t+1} > 0$ and $(2\pi^*_t - 1) > 0$, this implies that $\frac{\partial(\mathbb{E}[M_{t,t+1}]-\mathbb{E}[R_{t,t+1}])}{\partial\beta} < 0$. This implies that $\frac{\partial(\mathbb{E}[\pi_{t+1}|H_t]-\pi_t)(1-2\pi_t)}{\partial\beta} > 0$ which is $\frac{\partial(\mathbb{E}[M_{t,t+1}]-\mathbb{E}[R_{t,t+1}])}{\partial\beta} > 0$. Using very similar logic, one can show that if instead $\pi_t = \pi^*_t < .5$, $\frac{\partial(\mathbb{E}[M_{t,t+1}]-\mathbb{E}[R_{t,t+1}])}{\partial\beta} > 0$, which implies

**Proof of Proposition 7:**
We now consider the case in which $\pi_0 = \frac{1}{2}$ and updating occurs over multiple periods.

Intuitive discussion of next section: If confirmation bias is strong enough—the person will update in the direction suggested by their prior (i.e. towards 1 if the $\pi_t > \frac{1}{2}$) regardless of the signal once they see the first signal. This implies that the expected posterior moves away from .5 relative to the prior, which implies negative excess movement.

Formally, consider the case of "strong" confirmation bias ($\alpha \geq 2$). (here, we prove the result for the weak inequality. The statement in the paper uses $\alpha > 2$ for consistency with later statements that require the strict inequality). The setup follows that of the previous proof.

Claim: given an initial signal of $s_1 = h$, at every period $t \geq 1$, (i) $\pi_t \geq \theta$. In the process we prove that (ii) $\pi^h_{t+1} > \pi_t$ and (iii) $\pi^l_{t+1} \geq \pi_t$.

The proof is by induction.

First, consider $t = 1$. As $s_1 = h$, $\pi_1 = \theta$ so (i) is satisfied. Next, $\pi^h_2 = p(l(\pi_t)^\alpha \cdot l(\theta)) = p(l(\theta)^\alpha \cdot l(\theta)) = p(l(\theta)^{\alpha+1}) = 1 - \frac{1}{1+(\frac{\theta}{1-\theta})^{\alpha+1}} > \theta = \pi_1$, so (ii) is satisfied. Finally, $\pi^l_2 = p(l(\pi_t)^\alpha \cdot \frac{1}{l(\theta)}) = p(l(\theta)^\alpha \cdot \frac{1}{l(\theta)}) = p(l(\theta)^{\alpha-1}) = 1 - \frac{1}{1+(\frac{\theta}{1-\theta})^{\alpha+1}} \geq \theta = \pi_1$ as $\alpha \geq 2$, so (iii) is satisfied. (ii) and (iii) imply that $\pi_2 \geq \theta$.

Now, consider some $t$. By induction $\pi_t \geq \theta$. Then, similar logic shows: $\pi^h_{t+1} = p(l(\pi_t)^\alpha \cdot l(\theta)) \geq p(l(\theta)^\alpha \cdot l(\theta)) = p(l(\theta)^{\alpha+1}) = 1 - \frac{1}{1+(\frac{\theta}{1-\theta})^{\alpha+1}} > \theta = \pi_t$ where the weak inequality follows from the fact that $p(\cdot)$ and $l(\cdot)$ are monotonically increasing in its argument and $\alpha \geq 2$, therefore $l(\pi_t) \geq l(\theta)$,so $l(\pi_t)^\alpha \geq l(\pi_t)^\alpha$, and so $p(l(\pi_t)^\alpha \cdot l(\theta)) \geq p(l(\theta)^\alpha \cdot l(\theta))$ because $l(\theta) \geq 0$. Similarly, $\pi^l_{t+1} = p(l(\pi_t)^\alpha \cdot \frac{1}{l(\theta)}) \geq p(l(\theta)^\alpha \cdot \frac{1}{l(\theta)}) = p(l(\theta)^{\alpha-1}) = 1 - \frac{1}{1+(\frac{\theta}{1-\theta})^{\alpha+1}} \geq \theta = \pi_t$.

Therefore, the claim is true. Furthermore, as $\pi^h_{t+1} > \pi_t$ and $\pi^l_{t+1} \geq \pi_t$, then the expectation of the posterior must be greater than the prior ((4) is positive) as (3) implies that $\Pr(s_{t+1} = h|H_t) = 1 - \theta + \pi^*_t(2\theta - 1) \in (0, 1)$ regardless of $\pi^*_t$.

Similar logic shows that given an initial signal of $s_1 = l$, (i) at every period $t \geq 1$, $\pi_t \leq 1 - \theta$, (ii) $\pi^l_{t+1} < \pi_t$ and (iii) $\pi^h_{t+1} \leq \pi_t$, and therefore the expectation of the posterior must be less than the prior.

Using the same logic as the proofs above, this shows—for any history in which $\pi_t \neq \frac{1}{2}$—reversion away from .5, which combined with Lemma A shows that the excess-movement statistic is negative.

Intuitive discussion of next section: Base-rate neglect causes the person to downweight her prior. Given this, even with an infinite amount of signals in one direction, her beliefs are bounded away from certainty. Given this and strong enough base-rate neglect ($\alpha < \frac{1}{2}$), the signal that is opposite the direction of the person's current prior (i.e. $s_{t+1} = l$ given $\pi_t > \frac{1}{2}$) will cause the person to update to the other side of .5 ($\pi_{t+1} < \frac{1}{2}$). This can obviously leads to strong .5 reversion. However, it does not lead to .5 reversion if, for example, $\pi_t > \frac{1}{2}$ but is close to $\frac{1}{2}$, while $\pi^*_t$ (the Bayesian's belief given the signals) is close to 1. In this case, there is a relatively high likelihood ($\simeq \theta$) of observing signal $h$, which causes the person to update away from .5. The key to removing this situation is to bound beliefs away from .5, just as they are bounded away from 1.

Formally, next, consider the case of "strong" base-rate neglect. We maintain the assumption that $\alpha < \frac{1}{2}$ throughout. The setup follows that of the previous proof.

Consider $\pi_t > \frac{1}{2}$.

Claim 1: $\pi_t \leq p(l(\theta)^{\frac{1}{1-\alpha}})$.

Proof of Claim 1: Note that the person's belief at period 1 is $p(l(\pi_0)^\alpha \cdot l(s_1))$, at period 2 is $p((l(\pi_0)^\alpha \cdot l(s_1))^\alpha \cdot l(s_2)) = p(l(\pi_0)^{\alpha^2} \cdot l(s_1)^\alpha \cdot l(s_2))$ and therefore at period $t$ is

$\pi_t(H_t) = p(l(\pi_0)^{\alpha^t} \cdot l(s_1)^{\alpha^{t-1}} \cdot l(s_2)^{\alpha^{t-2}} \cdot ... \cdot l(s_t))$. Given that $\pi_0 = \frac{1}{2}$, it must be that $l(\pi_0) = 1$ and $l(\pi_0)^\alpha = 1$, so:

$$\pi_t(H_t) = p(l(s_1)^{\alpha^{t-1}} \cdot l(s_2)^{\alpha^{t-2}} \cdot ... \cdot l(s_t)).$$

Now, consider grouping likelihood ratios of similar signals:

$$\pi_t(H_t) = p(l(\theta)^{\sum_{\tau|s_\tau=h} \alpha^{t-\tau}} \cdot \frac{1}{l(\theta)}^{\sum_{\tau|s_\tau=l} \alpha^{t-\tau}})$$

Or:

$$\pi_t(H_t) = p(l(\theta)^{(\sum_{\tau|s_\tau=h} \alpha^{t-\tau} - \sum_{\tau|s_\tau=l} \alpha^{t-\tau})})$$

$p$ is monotonic in its arguments and $l(\theta) > 1$, so this is maximized when $s_t = h$ for all $t$. So, the highest belief is a sequence of only high signals,

$$\begin{aligned} \pi_t(H_t) &= p(l(\theta)^{1+...+\alpha^{t-2}+\alpha^{t-1}}) \\ &\leq p(l(\theta)^{\frac{1}{1-\alpha}}). \end{aligned}$$

This proves Claim 1.

Claim 2: $\pi_t \geq p(l(\theta)^{\frac{1-2\alpha}{1-\alpha}})$.

Proof of Claim 2; First, note that given that $\alpha < \frac{1}{2}$, $s_t = h$ implies that $\sum_{\tau|s_\tau=h} \alpha^{t-\tau} - \sum_{\tau|s_\tau=l} \alpha^{t-\tau} > 0$, such that $l(\theta)^{(\sum_{\tau|s_\tau=h} \alpha^{t-\tau} - \sum_{\tau|s_\tau=l} \alpha^{t-\tau})} > 1$, such that $\pi_t(H_t) = p(l(\theta)^{(\sum_{\tau|s_\tau=h} \alpha^{t-\tau} - \sum_{\tau|s_\tau=l} \alpha^{t-\tau})}) > \frac{1}{2}$. Similarly, $s_t = l$ implies that $\sum_{\tau|s_\tau=h} \alpha^{t-\tau} - \sum_{\tau|s_\tau=l} \alpha^{t-\tau} < 0$ and so $\pi_t(H_t) < \frac{1}{2}$. Given that we assumed $\pi_t > \frac{1}{2}$, it must be that $s_t = h$. To minimize the belief given this constraint, we set all other signals to $l$, such that:

$$
\begin{aligned}
\pi_t(H_t) &= p(\theta^{1-\sum_{\tau=1}^{t-1} \alpha^{t-\tau}}) \\
&\geq p(l(\theta)^{(1-\frac{\alpha}{1-\alpha})}).
\end{aligned}
$$

This proves Claim 2.

Now, we will find a condition under which (4) is negative regardless of the signal history. To do so, we will bound (4) above and show the bound is negative. First, note that as $\pi_{t+1}^h > \pi_{t+1}^l$, (4) is maximized when $\Pr(s_{t+1} = h|H_t)$ is maximized. (3) notes that $\Pr(s_{t+1} = h|H_t) = 1 - \theta + \pi_t^*(2\theta - 1)$, which given $\pi_t^* \in [0, 1]$, is maximized at $\theta$. Therefore:

$$
\begin{aligned}
\mathbb{E}[\pi_{t+1}|H_t] - \pi_t &\leq \theta\pi_{t+1}^h + (1-\theta)\pi_{t+1}^l - \pi_t \\
&= \theta(\pi_{t+1}^h - \pi_t) + (1-\theta)(\pi_{t+1}^l - \pi_t).
\end{aligned}
$$

Given $\pi_t > \frac{1}{2}$, $\pi_{t+1}^h - \pi_t$ is maximized when $\pi_t$ is minimized and $\pi_{t+1}^l - \pi_t$ is minimized when $\pi_t$ is minimized. Therefore, $\theta\pi_{t+1}^h + (1-\theta)\pi_{t+1}^l - \pi_t$ is maximized when $\pi_t$ is minimized, but then given that $\pi_t \geq p(l(\theta)^{\frac{1-2\alpha}{1-\alpha}})$,

$$
\theta\pi_{t+1}^h + (1-\theta)\pi_{t+1}^l - \pi_t \leq \theta p(l(\theta)^{\alpha\frac{1-2\alpha}{1-\alpha}+1}) + (1-\theta)p(l(\theta)^{\alpha\frac{1-2\alpha}{1-\alpha}-1}) - p(l(\theta)^{\frac{1-2\alpha}{1-\alpha}}).
$$

Given $p(\cdot)$ is monotonically increasing in its arguments, $l(\theta) > 1$ and $\alpha < \frac{1}{2}$, then the right side is monotonically decreasing in $\alpha$. Therefore, given the solution $\alpha^*(\theta)$ of the implicit equation:

$$
\theta p(l(\theta)^{\alpha^*\frac{1-2\alpha^*}{1-\alpha^*}+1}) + (1-\theta)p(l(\theta)^{\alpha^*\frac{1-2\alpha^*}{1-\alpha^*}-1}) - p(l(\theta)^{\frac{1-2\alpha^*}{1-\alpha^*}}) = 0,
$$

$\alpha < \alpha^*(\theta)$ must lead to a negative ride side, and therefore $\mathbb{E}[\pi_{t+1}|H_t] - \pi_t$ must be negative. For reference, $\alpha^*(\theta) = [0.45, 0.40, 0.36, 0.32, 0.28, 0.25, 0.22, 0.18.0.15]$ given $\theta = [0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95]$. Similar arguments show that if $\pi_t < \frac{1}{2}$, $\alpha < \alpha^*(\theta)$ implies that $\mathbb{E}[\pi_{t+1}|H_t] - \pi_t$ must be positive. Then, using the same logic as the proofs above, this shows reversion toward .5, which combined with Lemma A shows that—for any history in which $\pi_t \neq \frac{1}{2}$—the excess-movement statistic is positive.

Intuitive discussion of next section: We showed above that overreaction with the correct prior leads to reversion to .5 and excess movement. With multiple periods, the same proof cannot be used because the biased person's prior does not match the Bayesian prior at every signal history. However, it is straightforward to show that—given any signal history—the biased person's prior is always on the same side of .5, but closer to certainty, than the Bayesian's agents prior. For example, if $\pi_t^* > \frac{1}{2}$, then $\pi_t > \pi_t^* > \frac{1}{2}$. Given that the correct prior is in the direction of .5, she is likely to revert back to .5, providing an additional force of .5 reversion. Given that both forces (overreaction and the wrong prior) lead her to .5 reversion, the combination leads to .5 reversion. For underreaction, the opposite is true.

Consider the previous proof for the one-period excess-movement statistic given the correct prior. In this proof, we used the fact that $\pi_t = \pi_t^*$ to create equation (8). In this proof, we do not assume the correct prior at every history. Therefore, the equivalent equation adds an additional term:

$$\mathbb{E}[\pi_{t+1}|H_t] - \pi_t = (2\pi_t^* - 1)(\theta - \widehat{\theta})(\pi_{t+1}^h - \pi_{t+1}^l) + (\pi_t^* - \pi_t)(2\widehat{\theta} - 1)(\pi_{t+1}^h - \pi_{t+1}^l). \tag{9}$$

Focus on the case of $\pi_t > .5$ and $\beta > 1$. The other cases have the symmetric proofs.

Claim: $\pi_t > .5 \Rightarrow \pi_t^* > .5$ and $\pi_t > \pi_t^*$.

Proof of claim: Note that the person's belief at period 1 is $p(l(\pi_0) \cdot l(s_1)^\beta)$, at period 2 is $p((l(\pi_0) \cdot l(s_1))^\beta \cdot l(s_2)^\beta)$ and therefore at period $t$ is $\pi_t(H_t) = p(l(\pi_0) \cdot l(s_1)^\beta \cdot l(s_2)^\beta \cdot ... \cdot l(s_t)^\beta)$, which, given that $\pi_0 = \frac{1}{2}$ and therefore $l(\pi_0) = 1$, is equal to: $\pi_t(H_t) = p((l(s_1) \cdot l(s_2) \cdot ... \cdot l(s_t))^\beta)$. This is in contrast to the Bayesian belief: $\pi_t^*(H_t) = p((l(s_1) \cdot l(s_2) \cdot ... \cdot l(s_t)))$. We assumed that $\pi_t > \frac{1}{2}$, so $p((l(s_1) \cdot l(s_2) \cdot ... \cdot l(s_t))^\beta) = \pi_t > \frac{1}{2}$, which requires $(l(s_1) \cdot l(s_2) \cdot ... \cdot l(s_t))^\beta > 1$. As $\beta > 1$, this implies that $(l(s_1) \cdot l(s_2) \cdot ... \cdot l(s_t)) > 1$ and therefore that $\pi_t^* > \frac{1}{2}$. Finally, as $\beta > 1$ and $(l(s_1) \cdot l(s_2) \cdot ... \cdot l(s_t)) > 1$, it must be that $(l(s_1) \cdot l(s_2) \cdot ... \cdot l(s_t))^\beta > (l(s_1) \cdot l(s_2) \cdot ... \cdot l(s_t))$, which implies that $p((l(s_1) \cdot l(s_2) \cdot ... \cdot l(s_t))^\beta) > p(l(s_1) \cdot l(s_2) \cdot ... \cdot l(s_t))$ because $p(\cdot)$ is monotonically increasing. This implies that $\pi_t > \pi_t^*$. The claim is proved.

In the previous proof, we showed that $(2\pi_t^* - 1)(\theta - \widehat{\theta})(\pi_{t+1}^h - \pi_{t+1}^l) < 0$ given that $\pi_t^* > \frac{1}{2}$, which must be true given the assumption of $\pi_t > \frac{1}{2}$ and the claim that was just proved. Now, we also show that $(\pi_t^* - \pi_t)(2\widehat{\theta} - 1)(\pi_{t+1}^h - \pi_{t+1}^l) < 0$: (i) $(2\widehat{\theta} - 1) > 0$ because $\hat{\theta} = \frac{1}{1+(\frac{1}{\theta}-1)^\beta} > \theta > \frac{1}{2}$, (ii) $\pi_{t+1}^h - \pi_{t+1}^l > 0$, and (iii) the claim just proved $(\pi_t^* - \pi_t) < 0$ given the assumption of $\pi_t > \frac{1}{2}$.

Following the same logic, we can similarly show that (i) $\mathbb{E}[\pi_{t+1}|H_t] - \pi_t > 0$ when $\pi_t < \frac{1}{2}$ and $\beta > 1$, (ii) $\mathbb{E}[\pi_{t+1}|H_t] - \pi_t > 0$ when $\pi_t > \frac{1}{2}$ and $\beta < 1$, and (iii) $\mathbb{E}[\pi_{t+1}|H_t] - \pi_t < 0$ when $\pi_t < \frac{1}{2}$ and $\beta < 1$.

Then, using the same logic as the proofs above, this shows—for any history in which $\pi_t \neq \frac{1}{2}$ – there is reversion toward .5 when $\beta > 1$, which combined with Lemma A shows that the excess-movement statistic is positive, with the opposite statements when $\beta < 1$.

**Proof of Proposition 8:**

We now consider $\mathbb{E}[R] - \mathbb{E}[R^B]$ and $\mathbb{E}[M] - \mathbb{E}[M^B]$ in the case above in which $\pi_0 = \frac{1}{2}$ and updating occurs over multiple periods.

Intuitive discussion of next section: We showed above that overreaction always leads to positive excess movement at any history. We also showed that it leads the overreactor to always be more certain that a Bayesian. As a result, the expected uncertainty reduction from period 0 (where the Bayesian and biased person agree on the uncertainty) to period t (where the biased person is more certain) must be greater than that of the Bayesian. But, if the biased person has more uncertainty reduction than a Bayesian and the Bayesian's uncertainty reduction is equal to her movement, then the biased person must have more uncertainty reduction than the Bayesian's movement. But, as the biased person also has more movement than uncertainty reduction, her movement must be greater than the Bayesians movement.

We discuss the case in which $\beta > 1$ and $\alpha = 1$.

In the previous proof, we showed that $\mathbb{E}[M_{t,t+1} - R_{t,t+1}|H_t] > 0$ for any history $H_t$ in which $\pi_t(H_t) \neq \frac{1}{2}$. Therefore, by Lemma B, $\mathbb{E}[M_{0,t} - R_{0,t}] > 0$ for any $t > 1$ as $\pi_1(H_t) \neq \frac{1}{2}$. Now, consider $\mathbb{E}[R_{0,t} - R_{0,t}^B]$. In the previous proof, we showed that $\pi_t(H_t) > .5 \Rightarrow \pi_t^*(H_t) > .5$ and $\pi_t(H_t) > \pi_t^*(H_t)$ and a similar proof shows $\pi_t(H_t) < .5 \Rightarrow \pi_t^*(H_t) < .5$ and $\pi_t(H_t) < \pi_t^*(H_t)$. But, then $\pi_t(1 - \pi_t) < \pi_t^*(1 - \pi_t^*)$ for any history if $\pi_t(H_t) \neq \frac{1}{2}$. But as $\pi_0(1 - \pi_0) = \pi_0^*(1 - \pi_0^*)$ as by assumption $\pi_0 = \pi_0^* = \frac{1}{2}$, then it must be that $\pi_0(1 - \pi_0) - \pi_t(1 - \pi_t) > \pi_0^*(1 - \pi_0^*) - \pi_t^*(1 - \pi_t^*)$ for any history $\pi_t(H_t) \neq \frac{1}{2}$. But then, $\mathbb{E}[\pi_0(1 - \pi_0) - \pi_t(1 - \pi_t)] = \mathbb{E}[R_{0,t}] > \mathbb{E}[R_{0,t}^B] = \mathbb{E}[\pi_0^*(1 - \pi_0^*) - \pi_t^*(1 - \pi_t^*)]$ for $t > 0$ because for every $t > 0$, there is at least one history with $\pi_t(H_t) > 0$ that occurs with positive probability (for example, the history with all high signals).

But, (i) by our main proposition $\mathbb{E}[R_{0,t}^B] = \mathbb{E}[M_{0,t}^B]$, and (ii) we previously found that $\mathbb{E}[M_{0,t}] - \mathbb{E}[R_{0,t}] > 0$ for $t > 1$, and (iii) we just determined $\mathbb{E}[R_{0,t}] > \mathbb{E}[R_{0,t}^B]$. Combining these: $\mathbb{E}[M_{0,t}] - \mathbb{E}[R_{0,t}] > 0 \Rightarrow \mathbb{E}[M_{0,t}] - \mathbb{E}[R_{0,t}^B] > 0 \Rightarrow \mathbb{E}[M_{0,t}] - \mathbb{E}[M_{0,t}^B] > 0$ for any $t > 1$.

Collecting our conclusions: for any $t > 1$, (i) $\mathbb{E}[M_{0,t}] > \mathbb{E}[R_{0,t}]$ , (ii) $\mathbb{E}[R_{0,t}] > \mathbb{E}[R^B_{0,t}]$, and (iii) $\mathbb{E}[M_{0,t}] > \mathbb{E}[M^B_{0,t}]$

Similar logic shows that for $\beta < 1$ and $\alpha = 1$: for any $t > 1$, (i) $\mathbb{E}[M_{0,t}] < \mathbb{E}[R_{0,t}]$ , (ii) $\mathbb{E}[R_{0,t}] < \mathbb{E}[R^B_{0,t}]$, and (iii) $\mathbb{E}[M_{0,t}] < \mathbb{E}[M^B_{0,t}]$

Intuitive discussion of next section: We showed that a person with base-rate must have beliefs that are bounded away from certainty. Therefore, as the Bayesian tends to certainty over time, there must be some time $T$ in which the Bayesian has reduced more uncertainty from the (shared) prior than the biased person. For movement, Benjamin, Bodoh-Creed, and Rabin (2017) show that the movement must be ergodic and therefore is unbounded in expectation, while a Bayesian's expected movement must be bounded (at the initial uncertainty).

We discuss the case in which $\alpha < 1$ and $\beta = 1$.

Recall that above, we showed that if $\alpha < 1$, $\pi_t \leq p(l(\theta)^{\frac{1}{1-\alpha}}) < 1$. Therefore, $\mathbb{E}[R_{0,t}] < \pi_0(1-\pi_0) - p(l(\theta)^{\frac{1}{1-\alpha}})(1 - p(l(\theta)^{\frac{1}{1-\alpha}})) < \pi_0(1-\pi_0)$. But, as a Bayesian must approach certainty as $t \to \infty$, $\mathbb{E}[R^B_{0,t}] \to \pi^*_0(1-\pi^*_0) = \pi_0(1-\pi_0)$ as $t \to \infty$. But, then, there must be some point $t^*$ at which $\mathbb{E}[R_{0,t^*}] < \mathbb{E}[R^*_{0,t^*}]$. To address movement, we rely on a result in Benjamin, Bodoh-Creed, and Rabin, who show that the biased person's movement is ergodic and therefore unbounded in expectation as $t \to \infty$. Given that $\mathbb{E}[M^B_{0,t^*}] \to \pi^*_0(1-\pi^*_0) = \pi_0(1-\pi_0)$ as $t \to \infty$, it must be that there must be some $t^*$ such that $\mathbb{E}[M_{0,t^*}] > \mathbb{E}[M^B_{0,t^*}]$.

Intuitive discussion of next section: A person with strong enough confirmation bias must always move in the direction of the state suggested by her initial signal. For strong confirmation bias, the effect of the initial signal will become more powerful and all possible signals observed by a Bayesian. Therefore, her uncertainty will always be lower than a Bayesian's uncertainty at some point, leading her to have more uncertainty reduction from that point onward. For movement, as a biased person always moves in the same direction, the sum of her squared deviations (which are all less than 1) will always be less then the square of the total deviation. But, the Bayesian's expected movement approaches the square of the total deviation (as it is equal to the initial uncertainty). Therefore, the biased person's movement is lower than the Bayesian's movement at some point.

We discuss the case of "strong" confirmation bias ($\alpha > 2$) and $\beta = 1$.

Note that in the proof, we relabel $t^*$ in the statement to $\mathbf{t}$ as $^*$ is associated with the Bayesian belief in the proof.

First, we focus on uncertainty reduction. We show that there exists some time period $\mathbf{t}$ such that $u_t(H_t) = \pi_t(H_t)(1-\pi_t(H_t)) < u^*_t(H_t) = \pi^*_t(H_t)(1-\pi^*_t(H_t))$ for all $H_t$ with $t \geq \mathbf{t}$, which implies that $\mathbb{E}[R_{0,t}] = u_0 - \mathbb{E}[u_t] > u^*_0 - \mathbb{E}[u^*_t] = \mathbb{E}[R^B_{0,t}]$ for all $t \geq \mathbf{t}$ as $u_0 = u^*_0$.

To show this, first consider the biased beliefs following history $H_t$ in which $s_1 = h$. As shown in the proof to Proposition 7: given an initial signal of $s_1 = h$, at every period $t \geq 1$, (i) $\pi_t \geq \theta > \frac{1}{2}$. Therefore, the history(s) at time $t$ with the highest uncertainty (with a belief closest to $\frac{1}{2}$) has the lowest belief possible. This is clearly the history with all $l$ signals (except the assumed first signal) which (following the logic in the proof of base-rate neglect in Proposition 7) is:

$$\pi_t(\{s_1 = h, s_2 = l, s_3 = l...\}) = p(l(\theta)^{\alpha^{t-1} - \sum_{\tau=2}^{t} \alpha^{t-\tau}}).$$

Alternatively, the history(s) at time $t$ with the highest uncertainty for the Bayesian is one with all confirming signals, leading to a belief of:

$$\pi^*_t(\{s_1 = h, s_2 = h, s_3 = h...\}) = p(l(\theta)^t)$$

But, as $\alpha > 2$, the sequence $\alpha^{t-1} - \sum_{\tau=2}^{t} \alpha^{t-\tau} \to \infty$ as $t \to \infty$ at an exponential rate, and therefore there must be some $\mathbf{t}$ such that $\alpha^{t-1} - \sum_{\tau=2}^{t} \alpha^{t-\tau} > t$ for all $t \geq \mathbf{t}$. But, then, for $t \geq \mathbf{t}$,

$l(\theta)^{\alpha^{t-1}-\sum_{\tau=2}^{t}\alpha^{t-\tau}} > l(\theta)^t$ as $l(\theta) > 1$ and therefore $p(l(\theta)^{\alpha^{t-1}-\sum_{\tau=2}^{t}\alpha^{t-\tau}}) > p(l(\theta)^t)$ for $t \geq \mathbf{t}$ as $p(\cdot)$ is monotonically increasing. But, then $u_t(\{s_1 = h, s_2 = l, s_3 = l...\}) < u_t^*(\{s_1 = h, s_2 = h, s_3 = h...\})$ for $t \geq \mathbf{t}$ . But, as $u_t(H_t) \leq u_t(\{s_1 = h, s_2 = l, s_3 = l...\})$ for all $H_t$ and $u_t^*(\{s_1 = h, s_2 = h, s_3 = h...\}) \leq u_t^*(H_t)$ for all $H_t$, so $u_t(H_t) < u_t^*(H_t)$ for all $H_t$ when $t \geq \mathbf{t}$, and the statement is proved when $s_1 = h$. A symmetric proof shows the same for $s_1 = l$, and therefore as outlined above $\mathbb{E}[R_{0,t}] > \mathbb{E}[R_{0,t}^B]$.

For movement, note that, for a Bayesian, a standard result is that $\mathbb{E}[u_t^*] \to 0$ as $t \to \infty$. Therefore, as $\mathbb{E}[R_{0,t}^B] \to u_0^*$ as $t \to \infty$, and therefore, as $\mathbb{E}[M_{0,t}^B] = \mathbb{E}[R_{0,t}^B]$, $\mathbb{E}[M_{0,t}^B] \to u_0^*$ as $t \to \infty$. The proof to Proposition 7 shows that the biased-person's beliefs are monotonically increasing if $s_1 = h$ (and conversely monotonically decreasing if $s_1 = l$). Therefore, the sum of squared deviations $m_{0,t}(H_t)$ must be strictly bounded away from $u_0 = u_0^*$ for every history $H_t$ at every period $t$. But, then $\mathbb{E}[M_{0,t}]$ is bounded away from $u_0^*$ and therefore, there is some $\mathbf{t}_M$ such that $\mathbb{E}[M_{0,t}] < \mathbb{E}[M_{0,t}^B]$ for $t > \mathbf{t}_M$.

Combining the statements, taking there is some combined critical period greater than both of the critical periods above (labeled $\mathbf{t}$ and $\mathbf{t}_M$) such that for all periods after the combined critical period, $\mathbb{E}[M_{0,t}] < \mathbb{E}[M_{0,t}^B]$ and $\mathbb{E}[R_{0,t}] > \mathbb{E}[R_{0,t}^B]$.

**Proof of Proposition 9:**

Consider some belief stream $\pi = [\pi_0, \pi_1, \pi_2, ...\pi_T]$ and $\delta \in (0, 1)$.

Intuitive Note: The basic idea will be to shrink the stream while starting at $\pi_0$ to create $\pi' = [\pi'_0, \pi'_1, \pi'_2, ...\pi'_T]$. Then, following the idea in the proof of Proposition 4, we will construct a DGP that can make that stream very likely: the DGP either moves the belief from $\pi_t$ to $\pi_{t+1}$ or moves in the opposite direction to 0 or 1. Given that beliefs are a martingale, the probability of moving from $\pi_t$ to $\pi_{t+1}$ can be made arbitrarily high as the distance between $\pi_t$ to $\pi_{t+1}$ shrinks. Consequently, the likelihood of observing all movements in the stream can be made arbitrarily likely.

Formally, consider the affine transformation of stream $\pi$ to $\pi' = b_0 + b_1\pi = [\pi_0, \pi_0 + b_1(\pi_1 - \pi_0), \pi_0 + b_1(\pi_2 - \pi_0), ..., \pi_0 + b_1(\pi_T - \pi_0)]$ where $b_0 = \pi_0 - \pi_0 b_1$ and $b_1 < 1$. Note that because $\pi_t \in [0, 1]$, it must be that $(\pi_t - \pi_0) \in [-\pi_0, 1 - \pi_0]$ and therefore that $\pi'_t = \pi_0 + b_1(\pi_t - \pi_0) \in [0, 1]$, so $\pi'$ is an appropriate belief stream. Now consider the following DGP at time $t$ that is determined by this stream $\pi'$. This DGP follows that in the Proof 4. Label the states 0 and 1, $H$ and $L$ respectively for notational ease, and consider a DGP with two signals $h$ and $l$.

If $\pi'_{t+1} > \pi'_t$, define: $\Pr(s_{t+1} = h|x = H, \pi'_t) = 1, \Pr(s_{t+1} = l|x = L, \pi'_t) = \frac{(\pi'_{t+1} - \pi'_t)}{(\pi'_{t+1})(1 - \pi'_t)}$. Given this signal distribution, a Bayesian will update expectations from $\pi'_t$ to $\pi'_{t+1}$ in the case of a signal $h$ and will update from $\pi'_t$ to 0 in the case of signal $l$, where the likelihood of observing a signal $h$ is (following the Proof 4) $\frac{\pi'_t}{\pi'_{t+1}}$. Importantly, this probability of transition from $\pi'_t$ to $\pi'_{t+1}$ can be made arbitrarily high as $\pi'_{t+1} - \pi'_t$ tends to zero, or as, $\pi'_{t+1} - \pi'_t = (\pi_0 + b_1(\pi_{t+1} - \pi_0)) - (\pi_0 + b_1(\pi_t - \pi_0)) = b_1(\pi_{t+1} - \pi_t)$ tends to zero, which occurs as $b_1$ tends to zero. Similarly, if $\pi'_{t+1} < \pi'_t$, one can define a signal distribution such that a Bayesian will update expectations to $\pi'_{t+1}$ in the case of a signal $l$ and will update to 1 in the case of signal $h$, with the signal $l$ occuring with probability $\frac{1 - \pi'_t}{1 - \pi'_{t+1}}$, which can again be made arbitrarily high by reducing $b_1$ to zero. Finally, if $\pi'_{t+1} = \pi'_t$, then set $\Pr(s_{t+1} = h|x = H, \pi'_t) = \frac{1}{2}, \Pr(s_{t+1} = l|x = L, \pi'_t) = \frac{1}{2}$, such that the person always updates to $\pi'_{t+1}$.

Now, as each of the probabilities of transitioning from $\pi'_t$ to $\pi'_{t+1}$ can be made arbitrarily high in this DGP by reducing $b_1$, we can make the likelihood of the entire stream $\pi'$ arbitrarily high by reducing $b_1$. That is, for any $\delta \in (0, 1)$, there exists $b_1$ such that the above DGP produces belief stream $\pi'$ with probability above $\delta$.