Title
Inference by Believers in the Law of Small Numbers

Author
Rabin, Matthew

Publication Date
2000-06-04

Inference by Believers in the Law of Small Numbers

Matthew Rabin
Department of Economics
University of California—Berkeley

January 27, 2000

Abstract

Many people believe in the "Law of Small Numbers," exaggerating the degree to which a small sample resembles the population from which it is drawn. To model this, I assume that a person exaggerates the likelihood that a short sequence of *i.i.d.* signals resembles the long-run rate at which those signals are generated. Such a person believes in the "gambler's fallacy", thinking early draws of one signal increase the odds of next drawing other signals. When uncertain about the rate, the person over-infers from short sequences of signals, and is prone to think the rate is more extreme than it is. When the person makes inferences about the frequency at which rates are generated by different sources — such as the distribution of talent among financial analysts — based on few observations from each source, he tends to exaggerate how much variance there is in the rates. Hence, the model predicts that people may pay for financial advice from "experts" whose expertise is entirely illusory. Other economic applications are discussed.

Keywords: Bayesian Inference, The Gambler's Fallacy, Law of Large Numbers, Law of Small Numbers, Over-Inference.

JEL Classifications: B49

Acknowledgments: I thank Kitt Carpenter, Erik Eyster, David Huffman, Chris Meissner, and Ellen Myerson for valuable research assistance, Jerry Hausman, Winston Lin, Dan McFadden, Jim Powell, and Paul Ruud for helpful comments on the first sentence of the paper, and Colin Camerer, Erik Eyster, and seminar participants at Berkeley, Yale, Johns Hopkins, and Harvard-M.I.T. for helpful comments on the rest of the paper. I thank the Russell Sage, Alfred P. Sloan, MacArthur, and National Science Foundations (Award 9709485) for financial support.

Contact: Mail: Department of Economics / 549 Evans Hall #3880 / University of California, Berkeley / Berkeley, CA 94720-3880. E-mail: rabin@econ.berkeley.edu. CB handle: "Game Boy". Web page: http://emlab.berkeley.edu/users/rabin/index.html.

1. Introduction

Loosely put, the law of large numbers tells us that the distribution of a large random sample from a population closely resembles the distribution of the overall population. But many people believe in the "law of small numbers": They exaggerate how likely it is that a *small* sample resembles the parent population from which it is drawn. This paper develops a simple model reflecting this error, and studies how people making this error differ from Bayesians in their inferences. The model mathematically and intuitively ties the law of small numbers together with other biases, such as the gambler's fallacy, the tendency to over-infer from short sequences, and the belief in non-existent expertise. I use the model to sketch out some possible economic implications of these biases.

In Section 2, I review in some detail the psychological evidence that people systematically depart from Bayesian reasoning in ways that resemble the law of small numbers, the gambler's fallacy, and over-inference of the sort modeled in this paper. The law of small numbers itself was first labeled and demonstrated by Tversky and Kahneman (1971). An immense literature identifies the existence and determinants of the related "gambler's fallacy" (the belief that recent draws of one signal increase the odds of next drawing a different signal) in both the laboratory and in the field. A smaller literature demonstrates over-inference from small samples.

In Section 3, I present the model. A person observes a sequence of binary signals of some underlying quality, such as a sequence of good or bad investments by a financial analyst that signal her underlying competence, a sequence of good or bad performances by a company that signals its long-run prospects, or a sequence of good or great movies starring Johnny Depp that signals his thespian virtues. I assume that each value of the signal is generated randomly from a stationary probability that I shall refer to as the "rate". The person is a Bayesian and has correct probabilistic priors about this rate. *But:* Whereas in reality these signals are generated by an *i.i.d.* process, the person believes they are generated by random draws *without replacement* from an "urn" of $N < \infty$ signals, where the urn contains the proportion of the two values of the signal corresponding to the rate. This captures belief in the law of small numbers, since it means that the person believes that the proportion of signals must balance out to the population rate before N signals are observed. When $N \to \infty$, the person becomes fully Bayesian; the smaller is N, the more he believes in the law of small numbers.[1]

The model leads directly to the "gambler's fallacy": People expect the second draw of a signal to be negatively correlated with the first draw. Because we exaggerate how likely it is that a small set of coin flips yields very close to half heads and half tails, if early flips are disproportionately heads, the "law of averages" tells us that the next flips are more likely to be tails. And if an observer is sure that a particular financial analyst invests successfully close to half the time even over short intervals, then he thinks that an analyst who is successful in her first year has a less than $\frac{1}{2}$ chance of being successful next year.

In Section 4, I turn to the crux of the model, which examines inference by a believer in the law of small numbers who is uncertain about the rate by which signals are generated. Because such a person exaggerates how likely it is that a short sequence of signals will closely resemble the rate, he is too confident that the underlying rate resembles a short sequence he observes. If he believes every pair of flips of a fair coin surely generates one head and one tail, then he believes that two heads in a row indicates a biased coin. If he believes that an average financial analyst is successful once every two years, then he believes that an analyst who is successful two years in a row must be unusually good. I formalize this over-inference result by showing that, after two signals, a believer in the law of small numbers always has stronger beliefs than he should, and — what is essentially the same result — that the probability distribution over his possible posterior beliefs after two signals has too high a variance.

To investigate what the person infers when he observes many signals, I assume he believes that the "urn" generating the signals is replaced every two periods. Such a deterministic and frequent replacement of the urn is of course highly artificial. But it serves to capture in a tractable way the fact that people expect small subsequences of a long sequence to yield signals in approximately the same proportions as the overall sequence. That is, just as people do not expect the composition of small samples to differ dramatically from overall population proportions, they also do not expect to see "streaks" of signals that are not representative of the overall frequency in a sequence.

The person's theory of streaks implies that the inferences he makes depend not just on the proportions of signals he observes, but also on the precise sequence of those signals. Unlike when

[1] While the psychological evidence shows the existence of a cognitive bias, the formal model also lends itself to a more literal Bayesian interpretation: People may completely understand the nature of *i.i.d* stochastic processes, but merely underestimate how common such processes are. Little depends on interpreting anything in the paper as a cognitive error rather than merely assuming that people have an empirical misconception about what random processes prevail in the world.

observing a small number of signals, after observing a long sequence of signals a believer in the law of small numbers is too likely to believe that the rate is *less* extreme than it is, and his beliefs may be too moderate even after an infinite sequence of signals. This is because he does not expect many long streaks, and is especially surprised to see streaks of rare signals. To explain why he is observing so many streaks of rare signals, he may come to believe that the true rate is close to 50/50 — even if such a moderate rate doesn't accord with the overall frequency of the signals. To illustrate, suppose an observer is initially uncertain whether a financial analyst is bad, average, or good, having successful investment years $\frac{1}{4}$, $\frac{1}{2}$, or $\frac{3}{4}$ of the time. Suppose the analyst is, in fact, good, investing successfully $\frac{3}{4}$ of the time. Eventually, the observer will see all possible pairs of performances, including an occasional two unsuccessful years in a row. If the person is an extremely strong believer in the law of small numbers, he'll believe that two unsuccessful years in a row by a good analyst is virtually impossible, and hence will eventually conclude that the analyst is average, since average analysts are the only ones who often have both two unsuccessful and two successful years in a row. He believes this despite his surprise that this supposedly average analyst is successful $\frac{3}{4}$ of the time. This example is extreme, but it does capture the intuition that the false inference people make from the frequency of streaks may dominate their inference from overall proportions of signals, to yield a false world view. If a person does not observe the precise sequence of signals, on the other hand, then he makes the proper inference after a large number of signals.

In Section 5, I suppose a person who believes in the law of small numbers observes a stream of signals from each of a series of different "sources", and from such observations makes inferences about the distribution of rates among a large population of sources. Consider again an observer of financial analysts, and suppose he observes two performances from a large number of analysts — as he might if he reads an article that lists the performances of a large number of mutual fund managers over the last couple of years, or if he observed a series of them he has hired for brief durations. The model predicts that if, in truth, all analysts are average — and a Bayesian with any initial beliefs would eventually figure this out — the believer in the law of small numbers will infer that some analysts are good and some are bad. Because he underestimates how often average analysts will have consecutive successful or unsuccessful years, he interprets what he sees as evidence of the existence of good and bad analysts. Such "fictitious variation" is a very direct corollary to the over-inference results, but I consider it to be one of the economically most important implications of the law of small numbers.

I extend and apply the fictitious-variation result by presenting a stylized model of investors who observe a short series of performances by firms, and predict near-term performance from these observations. While different patterns may also be consistent with the law of small numbers, the model can predict that—due to the gambler's fallacy—investors underpredict repetition of short strings of performances while—due to over-inference—they over-predict repetition of longer strings. This provides one plausible psychological account of a phenomenon in financial markets—short-term under-reaction to announcements by firms but medium-term over-reaction—that has recently been modeled using various rational and quasi-Bayesian models.

In Section 6, I examine inference by a person who decides what signals to observe based on his earlier observations, so that the sequence of signals a person observes is endogenous. Suppose a person employs financial analysts one at a time, and decides when to switch analysts based on both his beliefs about the talent of his current and other analysts. Assuming he observes only the performance of his current analyst, I show that such a person will eventually become convinced that average talent is less than it is. The investor switches quickly from an analyst who initially performs poorly—and when he does so he has over-inferred that the analyst is bad. But he sticks with an analyst who initially performs well—until he discovers (as he will) that she is average. Because he corrects his overly positive inference but not his overly negative inference, his beliefs are biased downward.

A second interesting possibility arises when a person's information is endogenously determined because of his decisions: There can be two different long-run steady-state beliefs that a person may converge on, depending on his initial beliefs and the early signals he observes. This is because different patterns of behavior in response to different initial beliefs that a believer in the law of small numbers thinks will both eventually reveal the true underlying distribution can in fact lead to two different long-run beliefs, even though the believer in the law of small numbers anticipates both behaviors will reveal the true distribution. If an investor initially believes (correctly) in relatively little quality dispersion among financial advisors, he will not switch advisors often, and hence will observe enough of each advisor to learn that she is average. But if he initially believes (falsely) in wide quality dispersion, he may frequently switch advisors after poor performance, and because he does not observe as much on average of each advisor, he may end up maintaining his belief in wide quality dispersion. More generally, different belief and behavior profiles may occur with high probability, even in single-person environments whose information structure is rich enough such that a Bayesian's steady-state beliefs and behavior would be deterministic and correct.

I conclude in Section 7 with a brief discussion of some of the limits of the model in this paper and ideas for modifications to rectify these limits and facilitate further economic applications. I also speculate on the relationship between the law of small numbers and another well-documented (and seemingly contradictory) bias, "the hot-hand fallacy".

2. Evidence for the Law of Small Numbers

The term "the law of small numbers" was coined by Tversky and Kahneman (1971) to describe how people exaggerate the degree to which the probability distribution in a small group will closely resemble the probability distribution in the overall population. Tversky and Kahneman relate the law of small numbers to another bias, *the representativeness heuristic*, and they and other researchers have emphasized the connection between the law of small numbers, the gambler's fallacy, regression errors, over-inference from short sequences, and other mistakes. In this section, I review some of broad array of evidence that supports the assumptions and results presented in the remainder of the paper about these various phenomena.

Tversky and Kahneman (1971) provide several examples. For instance, they asked a group of mathematical psychologists to forecast the likelihood of replication of results in a variety of scenarios presented to them. Participants were, for instance, told that a pattern of behavior matching a theory had been identified as statistically significant in an experiment with 20 subjects, and asked to predict the likelihood that the pattern of behavior would reappear as statistically significant in a subsequent experiment on 10 subjects. The respondents greatly exaggerated the likelihood of replication, apparently exaggerating the likelihood that true theories would show up as statistically significant in even small samples. Further illustrations provided by Tversky and Kahneman (1971) indicated that the source of the error was that people fundamentally expect that population proportions reliably show up even in small samples.

Both Tversky and Kahneman (1971) and Abraham and Schulz (1984) told a group of subjects that the "mean IQ of the population of eighth graders in a city is *known* to be 100. You have selected a random sample of 50 children for a study of educational achievements. The first child tested has an IQ of 150. What do you expect the mean IQ to be for the whole sample?" Tversky and Kahneman (1971) report that a "surprisingly large" number of subjects believe that the expected

IQ for the sample is still 100, and Abraham and Schulz (1984) found that 13 out of 22 subjects guessed 100, while only 3 said 101.

Kahneman and Tversky (1982a, p. 44) illustrate how people expect close to the same probability distribution of types in small groups as they do in large groups, asking a group of undergraduates the following question:

> A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?

Twenty-two percent of the subjects said that they thought that it was more likely that the larger hospital recorded more such days, and 56% said that they thought the number of days would be about the same. Only 22% of subjects answered correctly that the smaller hospital would report more such days.

As this example illustrates, the belief in the law of small numbers has often been demonstrated by showing that subjects are too inattentive to sample size. This inattentiveness, however, manifests itself in a second way. Though people believe in the law of small numbers, they *don't* necessarily believe in the law of large numbers: While overestimating the resemblance of small samples to the overall population, people *underestimate* the resemblance of large samples will have to the overall population. Kahneman and Tversky (1972), for instance, found that subjects on average thought that there was a greater than 1/10 chance that, of 1000 babies born on a given day, more than 750 would be male. The actual likelihood is (much) less than 1%. To overstate it a bit, people seem to have a universal probability distribution over sample means that is insensitive to the sample size. As Kahneman and Tversky (1972, p. 45) note, this has important implications for inference: people often infer a lot from statistics reported in percentage terms even from small sample sizes, but by the same token are not convinced when they should be by huge sample sizes. The results emphasized below pertain mostly to small-sample inference and prediction. In Section 7 I briefly discuss whether and how incorporating lack of belief in the law of large numbers might add to or change the analysis of this paper.

Some of the analysis below concerns inferences people make from observing just a few signals, but what they infer from long sequences of signals. Bar Hillel and Wagenaar (1991) review the extensive evidence for the existence of a "local representativeness" bias, whereby people expect too few streaks in random sequences. Most such evidence comes from two types of experiments, *production tasks* and *recognition tasks*. In production tasks, people are asked to produce "random sequences". Typically, they are instructed to produce a series of binary variables that look to them representative of a sequence that might be generated by a random process such as flipping a fair coin. Interpreting many of these experiments is potentially problematic, since participants are not given explicit instruction to produce *i.i.d.* samples. But participants are typically instructed in a way that should evoke the desire to produce sequences resembling *i.i.d.* sequences with which they are familiar, such as being told to "generate a random sequence, such as one might expect from a long sequence of flips of an unbiased coin."

In a series of papers that replicate and extend research on production tasks, Rapoport and Budescu (1992, 1997) and Budescu and Rapoport (1994) provide some additional confirmation for the general patterns reported in Bar Hillel and Wagenaar (1991). They also clarify some issues and provide a new perspective on such research. Rapoport and Budescu (1992, p. 355) asked subjects to "simulate the random outcome of tossing an unbiased coin 150 times in succession," while Rapoport and Budescu (1997, p. 612) asked subjects to "imagine a sequence of 150 draws with replacement from a well-shuffled deck, including five red and five black cards, and then call aloud the sequence of these binary draws."[2] The switching rate between successive elements of the sequence people produced around 58% in each of two experiments. Participants generated too few long streaks: Of all triplets of successive elements in the sequences subjects produced, around 15% were identical triplets (compared to the appropriate random average of 25%) and around 4% of four successive sequences were identical quadruplets (compared to the appropriate random average of 12.5%). Looked at another way, subjects were less and less likely to choose a signal when more and more of the preceding choices were those same signals. The probabilities that a subject would produce a signal given that the previous 0, 1, 2, or 3+ signals chosen were that same signal[3]:

[2] In the first experiment, subjects wrote down their sequences, and hence could examine what streaks they had developed, whereas calling out the sequence in the second experiment presumably made it harder for them to fix the best pattern for themselves.

[3] I derived these numbers from Table 7 of Rapoport and Budescu (1997) as follows. $\Pr(A \mid B)$ is simply percentage of two-tuples that were XY rather than XX; $\Pr(A \mid AB)$ is derived as the relative proportion of YXX sequences to XYX sequences, since this represents the percentage of time the subjects chose to repeat the second element rather than the first element in triplets where the first two elements differed. $\Pr(A \mid AAB)$ was derived as the relative frequency of YXXX sequences to XYYX sequences; and $\Pr(A \mid AAA...)$ was derived as the relative frequency of XXXX sequences to XXXY sequences. For ease of presentation, the numbers I report are just the simple average of these numbers as derived from the 'Observed' columns of Experiments 1 and 2.

$\Pr(A \mid B)$	58.5%
$\Pr(A \mid AB)$	46.0%
$\Pr(A \mid AAB)$	38.0%
$\Pr(A \mid AAA...)$	29.8%

Rapoport and Budescu (1997, p. 615) interpret their experiments as suggesting that about 70% of subjects exhibit the 'alteration bias' discussed here, whereas about 15% seem consistently to exhibit an 'inertia bias' opposite of what is discussed here; the remaining 15% seem inconsistent.

Rapoport and Budescu also develop interesting variations on the production tasks that more clearly than in previous studies provide subjects incentive to produce *i.i.d.* sequences. Experiments in Rapoport and Budescu (1992) and Budescu and Rapoport (1994) studied the sequence of moves chosen by subjects in variants of a 'matching pennies' game—two-player, zero-sum games where players have a clear incentive to choose unpredictably[4]. Participants made a series of binary choices against an anonymous other player, and payoffs were realized before they then made another move. They hypothesized that subjects would produce sequences that better resemble *i.i.d.* sequences than in their own and previous 'one-player' production tasks, because subjects had both a greater incentive to be successful at randomizing, and were motivated to be unpredictable rather than needing to interpret the meaning of any randomization instructions. Their hypothesis was confirmed: In comparing the typical 'produce-a-random-sequence' instructions to their binary-choice experiments, for instance, alternation probabilities were reduced from 59% to 53% in Rapoport and Budescu (1992) and from 58% to 52% in Budescu and Rapoport (1994). Hence, the direction of the bias uncovered in the literature on production tasks appears to be robust, with alteration rates reduced but remaining statistically significantly greater than 50%. Rapoport and Budescu (1992) ran a third condition: Where subjects once more were told they were playing a zero-sum matching pennies game, but each was asked to produce a sequence of 150 choices all at once—and then the outcome of the game would be determined by matching the sequences of two of the participants. In this condition, subjects had no incentive to appear random, and since they were also not instructed to produce the sequence randomly, it is not clear how to interpret their behavior. Subjects in this condition alternated choices 49.6% of the time, exhibiting a weak inertia effect. Indeed, many subjects produced a very long sequence of one choice followed by a long sequence of the other, and others alternated in a clear pattern. Such subjects were clearly not trying to be random. Though I am skeptical that it explains much if any of the pattern, one interpretation of the results might be

[4] See also O'Neill (1987) for similar results in a similar setting.

that subjects expected the others to play randomly. In that case, if subjects mistakenly believed in local representativeness by other players, then they might 'insure' themselves—lowering the variance in the number of rounds they win—by choosing longer streaks. The problem with this is if they are aware that others might do the same thing, choosing strongly non-random patterns could exacerbate variance in wins rather than reduce it.

Research on the local-representativeness bias employing recognition tasks, where participants are asked to identify which from a menu of sequences appear "random," yields results very similar to those from production-task research. Bar Hillel and Wagenaar (1991) report that, in both production and recognition tasks, the average "switching rate" is about 60% in representative binary studies. But the evidence that is most directly relevant to my model is not production tasks or recognition tasks, but from *prediction tasks*. There is less research on this, but what research there is supports the view that people exhibit the gambler's fallacy and local-representativeness bias. Though he strongly emphasizes the view that production tasks exaggerate the local-representativeness bias, Kareev (1992), for instance, produces data supportive of the bias.[5] He gathered data from about 130 each of 2^{nd}-graders, 8^{th}-graders, and college students. He ran three conditions—the standard production task of the sort reported above, a "guessing" condition, and a "guessing-with-feedback" condition. His instructions were as follows:

> Introduction: "I will ask you to perform three short tasks. All involve the tossing of a coin. I have here a 1-shekel coin [both sides of the coin were shown]. It has a tree on one side and the number 1 on the other. In the following tasks you should say *tree* or 1, according to what you choose as the appropriate answer."
> Instructions for the standard task: "In this task I would like you to flip an imaginary coin. Let us say that I want to toss a coin 10 times, but I do not have a coin at my disposal. Tell me what you think might come out on the first toss? the 2^{nd}? ... The 10^{th}?"
> Instructions for the guessing task: "I will now toss the coin and cover it. You will have to guess the outcome and tell me. I will check whether or not you are correct. After 10 tosses I will tell you the number of correct guesses you made."
> Instructions for the guessing-with-feedback task: "I will now toss the coin and cover it. Before each toss you will have to guess the outcome and tell me. After each toss of the coin I shall tell you if you were correct or not. We shall repeat this 10 times."

Kareev (1992, p. 1193) reports average alternation rates for 2^{nd}-graders, 8^{th}-graders, and college students, as 65%, 62%, and 57% for the standard task, and 56%, 56%, and 53% for the guessing

[5] Kareev (1992) frames his results in terms of a theory, similar to Rapoport and Budescu's (1997) arguments, that the mispredictions come from memory limitations rather than misunderstanding of statistics.

task. He does not report the switching rate for the guessing-with-feedback task (the most relevant for many of the situations below); from graphs elsewhere in the article the results appear very close to the numbers for the standard task for 2^{nd}-graders and 8^{th}-graders, and very close to the guessing task for college students. Making estimates by eyeballing the graphs from Kareev (1992, p. 1192), support for the local-representativeness bias is even starker. The following table shows the percentage of participants in different conditions who chose a sequence of 10 coin flips that had exactly 5 heads and 5 tails.

Task	2^{nd}-graders	8^{th}-graders	college students
Actual	24%	24%	24%
Standard	47%	41%	48%
Guessing	33%	31%	33%
Guessing/Feedback	41%	42%	34%

Percentage of respondents choosing 5-head/5-tail combinations

While taken as a whole the experimental literature clearly supports the prevalence of the gambler's fallacy and local representativeness, the support is not universal. Experiments in Edwards (1961) and Lindman and Edwards (1961) provide harder-to-interpret evidence both for and against the gambler's fallacy. A group of 120 trainees in the air force were each shown a 1000-long sequence of signals L's and R's.[6] Participants were asked, after each of the first 999 signals, to predict what the next signal would be.[7] Some of the sequences of signals they were shown were generated by an *i.i.d.* random process, while others were carefully designed to mimic an *i.i.d.* sequence. Different trials were run where the signal proportions underlying probability of one of the signals was variously .5, .6, or .7. Participants were apparently told neither the probabilities of the two signals, nor were they told anything about the stochastic process by which the signals were generated. Hence, as with many other psychological experiments in this area, it is hard to interpret the results. Edwards reports that he found far less evidence of the gambler's fallacy than in previous research, saying that the gambler's fallacy appears in the first 200 trials, but that in the last 600 trials the

[6] Edwards (1961) notes that the population of basic airmen upon which the study was conducted contained few college attendees, making the set less educated (and presumably less smart) than in the other psychological experiments reported here. He did, however, screen out potential subjects who scored in the lowest category on the general intelligence test administered to all members of the air force.

[7] Edwards did not provide a full description of what the subjects were told. But he did report that each subject was told "Your purpose is to get as many predictions correct as possible. You will not be able to get all of them correct at any time during the test. There is no pattern or system you can use which would make it possible to get all of your answers correct. But you will find that you can improve your performance in the test if you pay attention and think about what you are doing."

opposite pattern appears.[8]

There are also problems with interpreting even the supportive data above as conclusive evidence for the gambler's fallacy. First, when the underlying probability of a binary signal is 50/50, all predictions have an equally good chance of being right. Hence, any observed patterns in prediction are not per se errors. In principle, there is a similar problem with interpreting behavior in zero-sum game; arguably such data indicate a statistical violation of Nash-equilibrium behavior, but they don't prove irrational (non-rationalizable) play, since any pattern of behavior is optimal against randomization by one's opponent. There is a more profound problem in interpreting the behavior in many of the prediction-task (and game-playing) experiments that involve predicting signals that participants realize are not equi-probable. A well-known experimental finding — discussed by Edwards (1961) — is a behavior called *probability matching*. When making a long series of predictions, or a long series of choices whose payoffs depend on the signals, participants tend to pick particular outcomes in proportion to the frequency with which they expect the corresponding signal rather that picking the highest payoff odds. So, for instance, in experiments like Edwards's sessions in which participants figure out that the proportions of signals are .7 and .3, they tend to pick the .7 signal 70% of the time rather than 100% of the time —which would maximize the number of correct guesses. Because probability matching is a tendency rather than a firm rule (some people clearly don't make the error)— it is therefore hard to determine participants' beliefs in this type of setting.

Despite such problems, I believe the evidence above strongly supports the prevalence of the gambler's fallacy. But there are some experiments that provide even stronger evidence of an error. In an unpublished manuscript, Gold and Hester (1987) also presented data on the gambler's fallacy in variants of a prediction task. Their paper both provides stronger evidence that predictions are really biased away from fifty-fifty, and provides fascinating evidence clarifying the mind-set of believers in the gambler's fallacy.

---

8 In fact, in the case where the true probability of each signal is 50%, my understanding of the data is that subjects are more likely to predict L than R following an L even in the first 200 trials, which suggests that his data contradicts the gambler's fallacy even more than claimed. By saying that the gambler's fallacy appeared in the first 200 trials, Edwards (1961) and Lindman and Edwards (1961) mean that participants were less likely to guess a signal for higher n' s when the previous $n > 0$ draws were that signal, whereas I have been interpreting the gambler's fallacy to meet this criterion for $n = 0$ as well. Of course, early in this prediction process, even a Bayesian who believed that the underlying process is *i.i.d.* might predict L more often following L than following R, since they will be updating their beliefs about the probabilities of L's and R's. Although Lindman and Edwards (1961) have some data that might support such as interpretation, it is not clear that this "inference stage" should or would last for 200 periods, and hence I would view these results as evidence against the gambler's fallacy as I have modeled it.

Consider the most interesting of Gold and Hester's (1987) experiments.[9] 120 University of Pittsburgh undergraduates were told that a coin would be flipped 25 times, and were awarded extra-credit points for their course for their performance in the experiments. While the coin was flipped by an experimenter in front of the participants, the coin landed too far for the subjects to see, and the experimenter actually reported a pre-determined sequence of outcomes. The sequence involved 9 red and 8 black realizations in the first 17 flips of a painted half dollar, a black realization in flip 18, followed by 4 red flips in flips 19-22. One set of participants spent 24 minutes before any of the flips working on some word-search puzzles, and then proceeded with the experiment. A second set of participants started right away, but the 23rd coin flip was delayed for 24 minutes (the pretext for doing so was not reported in the paper) to allow the participants to do the same word-search puzzles. To assure that this second set did not forget the previous flips, all participants were required to review the flips after flips 11 and 22, with the 'pause' group reviewing the realizations right after their pause and right before the flipping proceeded.

Participants could earn points after every flip. On each flip, they were given a choice between 70 points for sure, or a gamble in which they would get 100 points if the next flip was the winning color, and 0 points if it was the losing color.[10] But half of each group of subjects would get the 100 points if the next flip was a red—the color that had come up in the previous four flips—and half would get the 100 points if the next flip was black. Hence, the propensity to take the 70 points revealed beliefs about the odds that the next flip would be red or black, and hence differential rates for the groups as a function of whether red or black was the winning color would reveal the belief that a black flip was "due".

For the no-pause group, 24 of 29 subjects who would win with a red flip chose to take the sure thing, whereas only 8 of 30 who would win with the black flip chose the sure thing. This clearly indicated a belief that the next flip was more likely black, and hence that subjects exhibited the gambler's fallacy. In the pause group, however, 18 of 32 subjects who would win with a red flip chose to take the sure thing, whereas 13 of 29 who would win with the black flip chose the sure thing. This too indicated a belief in the gambler's fallacy—but this tendency was statistically significantly less than in the no-pause group. Subjects seemed to believe that letting a coin "rest"

9 All of their experiments used the basic structure of the experiment I describe, and all confirmed the presence of the gambler's fallacy among participants.

10 Pre-testing had indicated that subjects would be 'risk-taking', and would be roughly indifferent between 70 points for sure and such a gamble if the odds were 50/50.

makes it more plausible that a streak would continue.

Not all the evidence for local representativeness come from the laboratory. Walker and Wooders (1999) find some evidence consistent with the gambler's fallacy in the play of zero-sum games. They study final and semi-final matches at Wimbledon so as to get away from the inexperienced players and low stakes of experimental games. While their paper is not framed as a test of local representativeness it does provide some confirmation of the "over-switching" phenomenon found in the laboratory games. Although the authors argue that these top-ranked tennis players maximize their payoffs more effectively than experimental subjects, they conclude (p. 4) that

> Our tests indicate that the tennis players are not quite playing randomly: they switch their serves from left to right and vice versa somewhat too often to be consistent with random play. This is consistent with extensive experimental research in psychology and economics which indicates that people who are attempting to behave truly randomly tend to "switch too often."

There are several studies demonstrating the existence of the gambler's fallacy in lottery play.[11] Clotfelter and Cook (1993), for instance, examine the number of bets placed on numbers that won one of Maryland's lotteries in March and April of 1988. The particular lottery they examined was a pick-three lottery. Each day the state would randomly draw a triple-digit number. Lottery players pick a triple digit number, getting a payoff of $500 for every $1 bet if they guess all three digits correct, in the correct order. People are also allowed to bet on the unordered sequence, getting a payoff of $80 for every $1 bet if their triple-combination comes up in any order. In both cases, both the random draw by the state and the betting involved allowed repetition of numbers.

The gathered and reported total number of bets placed on each of the winning numbers during this period, on all days from two days before, one day before, and 1, 2, 3, 7, 28, 56, and 84 days after the win. The authors consider the ratio of the number of people betting on each of the winners for each of the chosen lags to the average of the two days before the number won. The authors

[11] We should be cautious in inferring too much from the behavior of lottery players—who are either stupider than average, or are not out to maximize pecuniary winnings. But I can think of no likely non-pecuniary motivation or non-expected utility preferences that could explain the particular pattern we see. This appears to be a manifestation of the gambler's fallacy. And while lottery players are apt to be stupider than average, playing the lottery is so common that the cognitive errors exhibited by lottery players are certainly economically relevant. Furthermore, Holtgraves and Keel (1992) found evidence of local representativeness (and other biases) in a lottery game played hypothetically by all students in two introductory psychology classes. Although about three fourths of the students reported having has played the lottery at some point, this group is clearly smarter than the average participant in the U.S. economy. (The form of local representativeness Holtgraves and Keel (1992) identify is somewhat different than the other lottery evidence I report here; they show that when asked to pick 3- or 4-digit numbers in a hypothetical lottery of randomly selected digits, subjects were more likely to pick sequences without repetition than with repetition.

report the 25th percentile, the median, and the 75th percentile of these ratios for the 52 winning numbers they gathered. The following is derived from Clotfelter and Cook's (1993, p. 1524) Table 2:

Days after Winning	25th Percentile	Median	75th Percentile
1 day	.77	.84	.97
2 days	.60	.66	.72
3 days	.58	.64	.76
1 week	.66	.75	.87
4 weeks	.69	.79	.92
8 weeks	.72	.90	1.09
12 weeks	.74	.93	1.11

Ratio of bets N days after winning to the two-day average before winning

That is, bettors stopped betting on the number immediately after it won, with winning numbers gradually recovering their popularity over the next 3 months.[12,13]

As with some of the laboratory data, the very fact that the lottery is random also implies that people cannot make any mistakes. Since players have no incentive to bet on one number over another, they might as well mimic a cognitive error. Perhaps, as well, they are actually making the cognitive errors but would get wise if stakes were involved. Stronger evidence comes from pari-mutuel betting. Terrell (1994), in fact, conducted a study examining Clotfelter and Cook's conclusions in a pari-mutual context. The State of New Jersey runs a pick-three lottery much like the Maryland lottery that Clotfelter and Cook (1993) examined, except that it is pari-mutual. In New Jersey, the state divides 52% of money bet on a given day's pick-three lottery evenly among those bettors who choose the winning number. Hence, the amount distributed to winners each day is an indicator of both the number of bettors on that number and the costliness of errors bettors are making. Tickets cost 50 cents each, so that if equal numbers of bettors choose all numbers, winnings on a given day should be $260. If the winnings are significantly higher than that, it indicates that

[12] I do not believe that the authors discuss the fact that the first-day drop was less dramatic than the second. Perhaps this occurred because of a lag in the news spreading about a particular number winning.

[13] These data provide some evidence distinguishing two hypotheses for behavior resembling what we observe here. A lot of evidence that people seem to bet on "losers" and avoid "winners" immediately after they lose and win is interpreted (often, clearly correctly) as reflecting loss aversion. As Thaler and Johnson (1990) show experimentally and Odean (1995) and Barberis, Huang, and Santos (1999) show in stock-market behavior, decision-makers in risky settings seek to get back to the status quo, which means they can ride losers. While in many cases one or the other interpretation is obvious, it is often hard to distinguish whether somebody holds onto a stock that has declined since they bought it because they hate to realize their losses, (or because they believe "the law of averages" means the stock will rebound. In this cae, it is implausible tht the numbers reflect loss aversion — at best it may be that those who play a particular number habitually refrain from betting for a while, more likely (I am surmising) the numbers must reflect a decreased tendency by losers to bet on assets that recently won.

too few bettors are betting that number. Terrel (1994) examines the data from 1785 daily drawings on New Jersey's pick-three numbers game from 1988 to 1992. Much in the spirit of Clotfelter and Cook, he examines betting on numbers that have recently won. The following is Table 1 from Terrel (1994, p. 311), with some rounding and omitting standard derivations, reporting the average winnings for numbers that repeated within various lags between 1 and 8 weeks, and comparing this to numbers that last won further ago:

	Number	Mean
Winners repeating within 1 week	8	$349
Winners repeating between 1 and 2 weeks	8	$349
Winners repeating between 2 and 3 weeks	14	$308
Winners repeating between 3 and 8 weeks	59	$301
Winners not repeating within 8 weeks	1622	$260
All Winners	1714	$262

Average payouts to winning numbers

The pattern clearly replicates that of Clotfelter and Cook (1993), though Terrel compares New Jersey behavior to Maryland behavior to conclude that the heightened incentives of New Jersey's pari-mutuel scheme led to a decrease in the gambler's fallacy.[14] For instance, he calculates that the one-week-lag winnings that Marylanders would have earned had they bet as they did but played by New Jersey rules would have been $396 rather than the $349 it was in New Jersey. Put differently, 13% more New Jerseyns bet on a number that won a week earlier than did the Marylanders. But he also emphasizes that the scope of the error in New Jersey is substantial. From Table 1 it can be inferred, for instance, that 25% fewer lottery players in New Jersey bet on a number that won a week ago than typically bet on that number.[15]

I am familiar with much less evidence on over-inference based on small samples than on the other biases related to my model. But there is some. A series of experiments by Grether (1980, 1992) and Camerer (1987) has subjects observe a series of draws coming from one of two underlying rates which they do not know, but which they make inferences from the draws. In these

[14] Terrell does not discuss the alternative hypothesis that lottery players in Maryland are inherently stupider than those in New Jersey.

[15] For another verification of the gambler's fallacy in pari-mutuel betting, consider Metzger's (1985) study of betting at the race track. Among other biases, she finds that the odds given for long-shot horses late in a day are higher when a long shot has won earlier in the day than when no long shots have won. Because the odds are continuously adjusted to reflect the amount of betting, this indicates the bettors anticipated that a long shot winning early means one won't win later. Presumably bettors take the view that lightening doesn't strike twice.

experiments, the subjects that indicate that making gambles or investments whose attractiveness depends on their beliefs about an underlying rate. These experiments are careful to explain all aspects of the environment to the subjects, involve incentives, markets, and allow for learning. They all, to varying degrees, provide support for the over-inference hypothesis.

Consider Camerer's (1987) experiment. A group of undergraduates at the Wharton School at the University of Pennsylvania, all of whom had both statistics and economics courses, participated in an asset-market experiment in which they had the incentive to correctly predict an underlying variable. Subjects observed three draws being drawn *with replacement* from one of two urns, urn X which contained exactly twice as many black balls as red balls, and urn Y which contained exactly twice as many red balls as black balls. The *ex ante* probability of urn X was .6. After observing the three draws, the subjects participated in an asset market for assets whose value depended on which urn was generating the draw of the balls. Hence, they were essentially betting on their posterior beliefs about the relative likelihood of X and Y.

Results were complicated, but one pattern stood out: When either one out of three or two out of three balls were red—i.e., when the proportions of the three draws exactly reflected the proportions of either the X or Y urn—market behavior clearly indicated that subjects were exaggerating the likelihood that the urn was the one matching the proportion of red and black balls drawn. While significant over-inference was not found when either zero or three of the balls were red, the strong results for the more moderate cases indicate over-inference about the underlying rate based on small samples. Subjects persisted in making the mistake through fifty rounds of repetition.[16]

There is also field evidence for intertemporal behavior by investors that is interpretable as a combination of the gambler's fallacy and over-inference. In Section 5, I show how my model can help interpret and explain some phenomena in American financial markets that appear to be departures from Bayesian rational expectations. There is extensive research, well summarized by Barberis, Shleifer, and Vishny (1998), showing that investors in stock markets and other financial markets seem to 1) under-react in the short term to good and bad news about a firm's financial prospects, but 2) *over*-react to in the medium or longer term to such news. The underreaction is evidenced in the fact that, as found in Cutler, Poterba, and Summers (1991), the expected return

[16] The results also clearly indicate that the over-inference was due to something like the law of small numbers rather than the (related) phenomenon of base-rate neglect; Camerer notes that had subjects been to neglecting the base rates of the two urns, the over-inference following two-red/one-black draws should have been more severe than for one-red/two-black draws, since the 2:1 Y urn had lower likelihood. Grether (1980, 1992) found similar results, while emphasizing more base-rate neglect and the role of learning and incentives.

in many financial markets is positively autocorrelated from one period to the next, across periods of a month, a quarter, or a year. More studies show that the returns to stocks in the short period following better-than-average unexpected earnings by a firm are higher than the earnings following worse-than-average unexpected earnings, indicating that the prices of these stocks set be investors are not immediately taking into account good or bad news. But there is an opposite pattern when firms or markets perform consistently well or poorly over a longer horizon. Cutler, Poterba, and Summers (1991) show a slight negative correlation in the returns in markets over horizons of 3-5 years, and Campbell and Shiller (1988) also find that returns are negatively correlated over time. A series of articles beginning with De Bondt and Thaler (1985) have shown that the returns to specific portfolios of stocks that have very poor returns over a period tend to significantly outperform portfolios with good returns, indicating that investors are too pessimistic about the future prospects of portfolios that have performed poorly recently.[17]

As I will show below, such data are consistent and interpretable with my model. But there are several other models that also explain the data. Barberis, Shleifer, and Vishny (1998) construct a quasi-Bayesian model of the simultaneous underreaction and overreaction, where they assume that performance is really a random walk, but that investors believe in either of two false models of the world—that returns of negatively autocorrelated, or that they are positively autocorrelated. They show that, given this false model, investor behavior will track the observed pattern of underreaction and overreaction. While not claiming that their model from psychological evidence, Barberis, Shleifer, and Vishny (1998) note that it is consistent with the framework developed by Griffin and Tversky (1992) that combines conservatism—the tendency not to fully absorb new information ('underreaction') with representativeness ('overreaction'); the second component has an attention somewhat similar to the law of small numbers and over-inference results in my model. Daniel, Hirshleifer, and Subrahmanyam (1998), by contrast, build a model based on the well-established

[17] There is also some suggestive evidence gathered from financial markets constructed in the laboratory. Andreassen and Kraus (1990) test the investment patterns of undergraduates in an artificial laboratory experiment with stock series derived from real world stock performances. Investors in these markets tend to sell after stocks go up and buy when prices fall, consistent with the gambler's fallacy. This evidence is inconclusive for many reasons—such behavior is consistent, for instance, with rational expectations combined with loss aversion and a preference for grabbing gains. De Bondt (1993) finds evidence in a similarly constructed financial market that investors tend to extrapolate recent performance of the market, and also collects survey evidence indicating more directly that investors believe too strongly that trends will continue.

finding that people tend to be over-confident in the precision of the signals they get.[18] Hong and Stein (1999) are able to explain these patterns by combining that assumption of slow diffusion of firm-specific information with a relaxation of the rational-expectations assumption, and Hong, Lim, and Stein (1998) test and confirm a novel prediction of the Hong and Stein (1999) model that the other models don't predict. Berk, Green, and Naik (1999) and Barberis, Huang, and Santos (1999) each provide (very different) rational-choice explanations of the observed phenomena.

I do not know whether my model is psychologically more valid than these other models, nor do I have a strong intuition at this point that the law of small numbers does or should be expected to generate the precise patterns we see.[19] But just as it may help to unify seemingly disparate psychological data as deriving from the same underlying phenomenon, however, my model can help explain some of the anomalies in financial markets as deriving from the same underlying bias in reasoning. Moreover, the model connects this phenomenon with what may be a more important bias in financial markets—the exaggerated belief by investors in variance in financial insight, and the large amounts of money that people pay to obtain this insight.

In addition to laboratory and field statistical evidence, more anecdotal evidence of these biases abounds. In fact, it often provides grist for the pedagogical mill of statisticians and economists to cite all the ways that real people (rather than the inhabitants of our traditional economic models) make errors in their statistical reasoning. The gambler's fallacy and over-inference from small samples are very high on the list of most noticeable common errors, amply represented in daily conversation, as well as in newspaper articles.[20]

Rarely does a random fluctuation in the stock market get explained as random, nor does a good game or two by an athlete get explained as random. For instance, in a newspaper sports page, Ortiz (1999) reports on a basketball game as follows:

---

[18] This over-confidence could, of course, be interpreted in light of the over-inference generated by the law of small numbers. Daniel, Hirshleifer, and Subrahmanyam (1998) invoke the over-confidence in a different way than I do below, by assuming that expert investors over-infer from private signals they receive, rather than from the recent, public performance of the stock.

[19] My model makes at least one prediction about the nature of people's beliefs that destinguishes it from Barberis, Shleifer, and Vishny (1998): The short-term underreaction does not come from pessimism about a portfolio's average performance, but rather from the belief in the gambler's fallacy. Hence, evidence that people infer too little from the short-term performance about the long-term performance would reject my model as the only or primary explanation, whereas evidence that people are not exhibiting inertia in updating their beliefs, but rather implementing their belief in the gambler's fallacy in the short run, would indicate that "conservatism" in updating does not explain the phenomenon.

[20] The popular book *Innumeracy: Mathematical Illiteracy and its Consequences,* in which Paulos (1988) becries widespread misunderstanding of basic mathematics, is largely devoted to errors in statistical reasoning, including the types of errors discussed in this paper.

A Sudden Reversal for Coles

Warriors guard is now hitting jumpers with confidence

> The body filling out that No.12 in a Warriors uniform is the same one as in the last three years, the one prone to clanking jump shots and running the offense at a glacial pace. ... Ever since Bimbo Coles got a release from the deepest part of the bench a week ago ... the nine-year point guard has been a changed player. ... Considering Coles' [sic] shooting percentage has declined in each of the last five years, to a career-low of 37.9 in 1997-98, the turnaround is *stunning* [emphasis added]. In the last four games, Coles has shot 15-for-26 (57.7 percent) ... "The mind is a crazy thing," Coles said. "When you totally lose your confidence, you're not going to play well. I'm starting to regain my confidence and going out there and having fun." ... The confidence was especially evident on the late jumpers, the last of which tied the game 87-87 with 1:33 to play.

In the four games following the stunning four-game turnaround, Coles went 7 for 21 for 33%. After regained confidence led to his stunning turnaround, Coles apparently either re-lost his confidence — or became over-confident.[21] It is very plausible that a player can have significantly different success from season to season or team to team (Coles had switched teams the previous season), or even game to game because of changed team compositions or changed position, or different opponents (this headline was written after the Warriors defeated the L.A. Clippers, who were 0-17 at the time, and finished the season with the league's worst record). Indeed, for the remainder of the season Cole *did* improve over the previous year, and finished the season by shooting 137 for 303 (45.2%) following the turnaround, for a year end total of 156 for 348 (44.8%). But the article made an inference that was statistically unwarranted, yet typical of such articles. The chance of a 37.9% shooter making 15 out of a given 26 shots is about 2%. If (say) 200 NBA players a week have enough shots to warrant such a headline were they to perform comparably stunning four-game turnarounds, about one player per day would warrant this headline. It could also be noted that had Coles remained a 37.9%, *i.i.d.* shooter, then the chance of him having had a 15-for-26 streak at some point in his 348 shot season would be about 76 %. Roughly speaking, the standards by which a four-game performance gets labeled a "stunning turnaround" are such that if most players don't experience turnarounds in either direction throughout their career, they will be deemed to have had a stunning turnaround at least once a season.

While this example of over-inference is typical of sports and financial media analysis, a more

[21] Later in the article it was noted, "Coles was feeling so good, he even tried a play that could have ruined the Warriors' comeback. ... the Warriors made a defensive stop and Coles attempted a risky fast-break pass [that] was intercepted ..." Had the headline been written after such mishaps that actually *did* ruin the Warriors' comeback, surely it would have emphasized Coles's overconfidence.

gruesome and more worrying example of the gambler's fallacy comes from an article in the NY Times Magazine titled "How Not to Get Killed on Deadline." Fisher (1999) reports that "In 'hostile-environment school,' foreign correspondents learn how to improve their chances of surviving kidnappings, cross-fires and other perils of the workplace." The article is about advice given to journalists by a company called *Centurion Risk Assessment Services Ltd. Centurion* gives the following advice to journalists paying for war-survival training: "In mortar attacks, ... lie down. If you can, crawl into one of the holes made by a previous shell because — back to the thunderstorm analogy — lightning rarely strikes twice in the same place." I do not know if, within a journalist's crawling range, the pattern of mortar attacks exhibits positive or negative correlation.[22] But the metaphor chosen to be persuasive is a commonplace metaphor used to convey intuition for a commonplace misguided belief in a "law of averages" that says that once a rare event occurs, it becomes less likely it will reoccur, because such recurrence will throw averages out of whack. The lightening metaphor is striking: In actuality, heading for the same spot where lightning struck earlier is a bad idea in a thunderstorm, since lightening is more likely to hit a spot it has hit before than to hit a spot for the first time. Lightening rarely strikes twice, but only because it rarely strikes once.

3. The Model

Throughout the paper, I consider a situation where each of a finite number of possible *rates*, $\theta \in [0,1]$, at which an infinite sequence of *i.i.d.* signals, s_t, $t = 1, 2...$, is generated. Each signal s_t takes on a value of either a or b, where for each t, $prob(s_t = a) = \theta$. Let Θ denote the set of rates that occur with positive probability; the rate θ occurs with prior probability $\pi(\theta) > 0$, where $\sum_\Theta \pi(\theta) = 1$. Given θ, signals are generated by an *i.i.d.* process.

The model describes a person who begins with correct prior beliefs about the probability distribution π over possible rates Θ and is fully Bayesian. But to capture the belief in the law of small numbers, instead of understanding that the world is *i.i.d.*, there is a positive integer N such that for each rate θ, the person believes signals are drawn without replacement from an "urn" of size N consisting of exactly θN a signals and $(1-\theta)N$ b signals. To reconcile his belief in finite urns and to model his belief in "local representativeness" as discussed in Section 2, I assume the person

[22] I have sent research assistants to war zones to collect data on this question, but none have reported back.

believes this urn is renewed after every two draws. That is, in every odd period he believes a first signal is drawn from an N-signal urn, and in every even period he believes a second signal is drawn *without replacement* from the same urn he drew from in the previous period. Assuming instead that the person thinks there is a constant 50% chance each period that the urn is renewed would make many aspects of the model less artificial, but also less tractable. As I develop the model, the advantage of the deterministic renewal of the urn will be clear: While the person doesn't recognize the world is *i.i.d.*, he believes that *pairs* of signals are generated by an *i.i.d.* process, which vastly simplifies analysis of his belief formation. Throughout the paper I assume that the person believes that the first signal he observes is the first signal of a new urn, even if he is aware that previous, unobserved sequences have been generated.

When N is large, the person perceives the signals to be close to uncorrelated, and his inference and predictions become that of a Bayesian as $N \to \infty$. But when N is small, the person is very biased. Suppose, for instance, that an observer is positive that a particular financial analyst invests successfully with probability $\frac{1}{2}$. If $N = 4$, the observer thinks the analyst has an "urn" of 2 good years and 2 bad years. Then if the analyst is successful in her first year, the observer thinks that there is only a $\frac{1}{3}$ chance that she will have an above-average year the following year.[23]

To avoid tedious repetition of the phrase "a person who believes in the law of small numbers", for the remainder of the paper I shall refer to a believer in the law of small numbers as "Freddy," named for a guy I once knew who believed in the law of small numbers.[24]

To make the model fully coherent, it must be that Freddy's prior beliefs always put positive weight on some rate whose urn contains at least two of both signals. This is necessary and sufficient to ensure that Freddy believes all sequences of g's and b's are possible. Formally:

Definition 1 (π, N) is *compatible* if:

[23] My model closely resembles one previously developed by Rapoport and Busdescu (1997). Their purpose is to explain "production tasks" of the sort discussed in Section 2 in which people are asked to generate sequences of numbers that look random. They assume that people do so as if they were choosing signals without replacement from an urn, but have memories of what they have done shorter than the size of the urn. Hence, their model of production of signals is a (stochastic and stationary) variant of my model.

More broadly, this paper belongs to a small literature developing "quasi-Bayesian" models of biased information processing: A person is modeled as having a specific form of misreading of the world meant to correspond to a heuristic error, but then is assumed to operate as a Bayesian given this misreading. In this sense, it is related to the Barberis, Schleifer, and Vishny (1998) paper discussed in the previous Section, as well as papers like Rabin and Schrag (1999) and Mullainathan (1997) which assume that people have the correct model of the world, but misread (or misremember) the signals they observe.

[24] Actually, I have no recollection of Freddy believing in the law of small numbers. But since most people believe in it, probably Freddy did too.

1) For all $\theta \in \Theta$, θN is an integer; and
2) There exists $\theta \in \Theta$ such that $Min[\theta N, (1-\theta)N] \geq 2$.

Given compatibility, Freddy's "Bayesian" updating always uniquely determines his beliefs in all possible contingencies. Note that $N \geq 4$ is required for compatibility.

For given priors π, let $\pi_t^N(h_t)$ represent an N-Freddy's posterior beliefs after history of signals h_t. I will use the notation $\pi_t^N(\cdot)$ throughout to represent beliefs by an N-Freddy following the t-th signal, but depending on the context, I vary which variables are included as arguments in this function. Bayesian beliefs are $\pi_t^\infty(h_t) \equiv \lim_{N\to\infty} \pi_t^N(h_t)$.

The "gambler's fallacy" is a nearly tautological implication of the model: Because an N-Freddy believes there are only θN a signals and $(1-\theta)N$ b signals when the rate is θ, if in an odd period an a signal occurs, then if he is positive he is facing rate θ, he thinks the probability of an a signal next time is less than θ.[25]

Lemma 1 Consider N and θ such that θN is an integer and $\pi(\theta) = 1$. For all even $t \geq 2$ and histories h_{t-2}, $\pi_t^N(s_t = a | s_{t-1} = b, h_{t-2}) = \frac{\theta N}{N-1} > \theta$ and $\pi_t^N(s_t = a | s_{t-1} = a, h_{t-2}) = \frac{\theta N - 1}{N-1} < \theta$. For all odd t, and histories h_{t-2}, $\pi_t^N(s_t = a | s_{t-1} = b, h_{t-2}) = \pi_t^N(s_t = a | s_{t-1} = a, h_{t-2}) = \theta$.

Lemma 1 shows how the Freddier the person is, the more severe is the gambler's fallacy. It also makes manifest the stark contrast between odd- and even-numbered signals. While this contrast will sometimes assist interpretation of the model by permitting a crisp separation of different effects, the distinction between even and odd periods is of course completely artificial, and will not be a focus in the analysis below.[26]

4. Inference About the Rate

The most interesting implications of the law of small numbers come when Freddy is uncertain about the true rate and makes inferences from the signals he observes. Suppose, for instance, that an observer believes there is equal chance a financial analyst can be any of three types, bad, average, or good, having successful investment years $\frac{1}{4}$, $\frac{1}{2}$, or $\frac{3}{4}$ of the time, respectively. What

[25] All proofs are in the Appendix.
[26] To verify that Freddy expects an average proportion of θ a signals if the true state is θ, note that he thinks the probability of getting two a's out of two signals in $\theta \cdot \frac{\theta N - 1}{N-1}$, and the probability of getting one a is $2\theta \cdot \frac{(1-\theta)N}{N-1}$, yielding an average of 2θ a's after two signals.

does he infer from two successful years of investment in a row by a particular analyst? A Bayesian thinks such a sequence occurs with probability $\frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$ for bad analysts, $\frac{2}{4} \cdot \frac{2}{4} = \frac{4}{16}$ for average analysts, and $\frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}$ for good analysts. But an $N = 4$-Freddy believes the probabilities are $\frac{1}{4} \cdot \frac{0}{3} = \frac{0}{12}$ if the analyst is bad, $\frac{2}{4} \cdot \frac{1}{3} = \frac{2}{12}$ if the analyst is average, and $\frac{3}{4} \cdot \frac{2}{3} = \frac{6}{12}$ if the analyst is good. Notice that for each rate, Freddy assigns a lower probability to a streak of two a's than a Bayesian assigns because he believes that no matter the rate drawing the first a means there are fewer a's left for the second draw. Of greater interest, however, is that Freddy's beliefs are too skewed towards believing the analyst is good, since making one less a available for the second draw has a proportionately greater impact when there are fewer a's to begin with. From his priors, Freddy forms probabilistic beliefs about the rate given an observed sequence of signals using a sort of warped Bayes Law — applying Bayes Law with his mistaken beliefs about how likely each sequence is given an underlying rate. While a Bayesian believes the probability that the analyst is good is $\frac{18}{28}$, Freddy believes the probability is $\frac{21}{28} > \frac{18}{28}$.

Freddy's beliefs following longer sequences can also be calculated. While Freddy wrongly believes that there is negative correlation within odd-even pairs of signals, he believes consecutive odd-even pairs of signals *are* distributed *i.i.d.* Hence, there is a reasonably simple formula determining Freddy's beliefs as a function of the number of aa, ab, and bb pairs of signals he observes, where throughout the paper, the odd-even pair "ab" is meant to be unordered, representing both ab and ba. Suppose that, after either $2(q + r + s)$ or $2(q + r + s) + 1$ signals, Freddy observes q aa pairs, r ab pairs, and s bb pairs, in some fixed order, possibly followed by one unpaired signal, consisting of $y \in \{0, 1\}$ a signals and $1 - y$ b signals. Freddy's beliefs about the likelihood of that particular sequence if the rate is θ are given by:

Lemma 2 Freddy believes that state θ generates the ordered sequence of q aa pairs, r ab $(= ba)$ pairs, s bb pairs, followed by $y \in \{0, \}$ isolated a signals or $z \in \{0, 1\}, z \leq 1 - y$, isolated b signals is $\pi_x^N(q, r, s, y, z|\theta) = \left(\theta \cdot \frac{\theta N-1}{N-1}\right)^q \left(2\theta \cdot \frac{(1-\theta)N}{N-1}\right)^r \left((1-\theta) \cdot \frac{(1-\theta)N-1}{N-1}\right)^s \theta^y (1-\theta)^z$, where $x = 2(q + r + s) + y + z$.

The formula in Lemma 2 is directly derived from the formula in Lemma 1, and is a generalization of the above example. Implicit in Lemma 2 is the key intuition for how Freddiness affects belief formation and hence the core intuition in the paper: In each even period, Freddy thinks that one of whatever signal he observed in the previous period has been removed from the urn, making that signal less likely. If that signal *does* occur again, then Freddy exaggerates how strongly it indicates that the true rate is one that generates many such signals.

Given his theory of how signal sequences are generated, Freddy's beliefs are formed by plugging the formula in Lemma 2 into Bayes' Law:

Lemma 3 Freddy's beliefs about the likelihood of rate θ^* following a (q, r, s, y, z) sequence of signals are $\pi_x^N(\theta^*|q, r, s, y, z) = \frac{\pi_x^N(q,r,s,y,z|\theta^*)\pi(\theta^*)}{\sum_{\theta\in\Theta}\pi_x^N(q,r,s,y,z|\theta)\pi(\theta)}$, where $x = 2(q + r + s) + y + z$ and $\pi_x^N(q, r, s, y, z|\theta)$ is derived in Lemma 2.

A simple and uninteresting result is that Freddiness does not affect a person's inferences from the first signal:

Lemma 4 For all N, and π, for all θ^*, $\pi_1^N(\theta^*|s_1 = a) = \frac{\theta^*\pi(\theta^*)}{\sum_{\theta\in\Theta}\theta\pi(\theta)}$ and $\pi_1^N(\theta^*|s_1 = b) = \frac{(1-\theta^*)\pi(\theta^*)}{\sum_{\theta\in\Theta}(1-\theta)\pi(\theta)}$.

That is, Freddy's beliefs after one signal are the same as a Bayesian's with the same priors. Combining this result with Lemma 1, Lemma 5 says the 'gambler's fallacy' still holds for the second signal despite uncertainty about the rate:

Lemma 5 For all N and π, for all θ^*, $\pi_1^N(s_2 = a|s_1 = a)$ and $\pi_1^N(s_2 = b|s_1 = b)$ are increasing in N.

Three comments help to interpret Lemma 5 and the other comparative statics on N presented in the paper. First, because as $N \to \infty$ Freddy becomes a Bayesian, a result on how the degree of Freddiness affects beliefs is also a comparison of Freddies to Bayesians. When the number is increasing in N, it means that Freddy has a lower value for that number than does a Bayesian. Second, the wording in all these results is loose, since changing N typically affects the compatibility of the model. Noting that if (π, N_0) is compatible, so is (π, kN_0) for all positive integers k, all comparative statics on N are proven using a more precise interpretation: Increasing N means increasing k for a fixed N_0, and the limit as $N \to \infty$ corresponds to the limit as $k \to \infty$ for a fixed N_0. Third, most results in this paper can be stated in terms of precise formulas that have qualitative features of interest. While I state some of the results in terms of these precise formulas (as in Lemmas 2, 3, and 4), I present others, such as Lemma 5, in terms of their qualitative features. The proofs contain the precise formulas.

The first "over-inference" result is that Freddy infers too much about the likely extremity of the rate from an extreme sequence of signals. Proposition 1 formalizes the result by showing that when all of the signals are of one type, Freddy exaggerates the relative likelihood of rates that are most likely to generate that signal. Let h_t^a be the sequence of t a signals, and h_t^b be the sequence of t b signals. Then:

Proposition 1 For all $t > 1$, and $\theta, \widehat{\theta} \in \Theta$ such that $\theta > \widehat{\theta}$, $\frac{\pi_t^N(\theta|h_t^a)}{\pi_t^N(\widehat{\theta}|h_t^a)}$ and $\frac{\pi_t^N(\widehat{\theta}|h_t^b)}{\pi_t^N(\theta|h_t^b)}$ are both strictly decreasing in N.

That is, following an extreme sequence of signals, the Freddier is Freddy the more skewed are his beliefs towards those rates where the signals are more likely. There is a simple corollary concerning his predictions of signals in odd periods, when the gambler's fallacy does not kick in:

Corollary 1 For all odd t, $\pi_t^N(s_{t+1} = a \mid h_t^a)$ and $\pi_t^N(s_{t+1} = b \mid h_t^b)$ are both decreasing in N.

Proposition 1 says that following an extreme sequence of signals Freddy infers too strongly that he is facing an extreme rate. Proposition 2 shows that a similar bias holds after any sequence of signals where exactly half are a's and half b's: In such cases, Freddy exaggerates the likelihood that the true rate is close to $\frac{1}{2}$. Let $H_t^{\frac{1}{2}}$ be the set of all t-sequences with exactly the same number of a's and b's. Then:

Proposition 2 For all even t and all $h_t \in H_t^{\frac{1}{2}}$, and for all $\theta, \widehat{\theta} \in \Theta$ such that either $\theta > \widehat{\theta} \geq \frac{1}{2}$ or $\theta \leq \widehat{\theta} < \frac{1}{2}$, $\frac{\pi_t^N(\theta|h_t)}{\pi_t^N(\widehat{\theta}|h_t)}$ is weakly increasing in N.

It turns out that $H_t^{\frac{1}{2}}$, h_t^a, and h_t^b are special types of sequences: It is not true generally that Freddy necessarily exaggerates the likelihood that the true rate resembles the proportion of signals he has received when those signals are mixed. For many sequences, the opposite is true. I return to that issue below.

Propositions 1 and 2 characterize Freddy's beliefs following given sequences of signals. But what sequence Freddy observes is stochastic, with likelihoods of observing sequences determined by the true rate. I now turn to characterizing Freddy's possible beliefs as a function of the true rate. This will allow us to compare the mean and variance of Freddy's beliefs to a Bayesian's. Let $E_t^N(h_t) \equiv \sum_{\theta\in\Theta} \pi_t^N(\theta|h_t)\cdot\theta$ be the mean value of Freddy's probabilistic beliefs about the rates following sequence of signals h_t. That is, $E_t^N(h_t) \in [0,1]$ is Freddy's perception of the expected value of the rate given he has observed h_t.[27] Let $f_{t,\pi,\theta}^N$ be the probability distribution over the values of $E_t^N(h_t)$ if the true rate is θ and prior beliefs are π. Let $E_t^N(\pi) \equiv \sum_{\theta\in\Theta}\sum_{h_t\in H_t} \pi(\theta)\cdot\pi_t^N(\theta|h_t)\cdot\theta$ be the expected mean value of Freddy's probabilistic beliefs about the rate given the actual probabilistic distribution of rates, and let $f_{t,\pi}^N$ be the probability distribution over the values of $E_t^N(h_t)$ given the probability distribution π over the rates. Note that the mean of $f_{t,\pi}^N$ is $E_t^N(\pi)$. While somewhat cumbersome, notationally and conceptually, the distributions $f_{t,\pi}^N$ play an important role in intuiting the implications of Freddiness. They represent the distribution of Freddy's expected beliefs about the underlying rate and the actual distribution of rates. If $f_{t,\pi}^N$ has a higher variance than the corresponding Bayesian distribution, $f_{t,\pi}^\infty$, then Freddy's beliefs are too dispersed due to overinference. If $E_t^N(\pi) \neq E_t^\infty(\pi)$, then belief in the law of small numbers causes biased belief formation.

It follows from Proposition 1 that for all t, $E_t^N(h_t^a)$ is decreasing in N — Freddy has too high an estimate of θ following a string of a signals. Similarly, $E_t^N(h_t^b)$ is increasing in N. Hence, it is trivial that the range of possible beliefs that Freddy might have is decreasing in N: For all t, π, and θ, the size of the supports of $f_{t,\pi,\theta}^N$ and $f_{t,\pi}^N$ are decreasing in N. In this sense the variation in Freddy's beliefs is too great.

Beyond this, little general can be said about the distribution of beliefs. One of the problems with doing so reflects a feature of the model that is important in its own right: For many prior probabilities of rates, there may be predictable drift in Freddy's average beliefs as he gets information.[28] Suppose, for instance, that the prior probability of rates is π^*, where $\pi^*(\theta = \frac{1}{2}) = \pi^*(\theta = \frac{3}{4}) = \frac{1}{2}$. By the law of iterated expectations, a Bayesian's expectation of his future beliefs are his current beliefs: $E_t^\infty(\pi^*) = \frac{5}{8} = E_0^\infty(\pi^*)$ for all t. Yet it can be shown that $E_2^4(\pi^*) = \frac{319}{512} < \frac{5}{8}$. That is, *on*

[27] And hence $E_t^N(h_t)$ is Freddy's estimated probability that the signal in the next odd period will be an a.

[28] Freddy, of course, does not understand this drift: Because Freddy is a Bayesian with the wrong model of the world, he obeys the law of iterated expectations in the sense that his own expectation of future beliefs are his current beliefs.

average Freddy will after two signals believe that θ is lower than he believed when he started.[29]

To disentangle overinference from such drift in beliefs, many results will be easier to formulate and interpret when limiting analysis to symmetric prior beliefs, defined as follows:

Definition 2 A distribution π is *symmetric* if for all $d \in [0, \frac{1}{2}]$, $\pi(\frac{1}{2} + d) = \pi(\frac{1}{2} - d)$.

It is easy to verify that $E_t^N(\pi) = \frac{1}{2}$ for all t, N, and symmetric prior distributions π, since Freddy forms opposite beliefs for h_t and h_t' that switch a's and b's, and h_t and h_t' are equally likely given symmetric priors. That is, we know in such cases that $f_{t,\pi}^N$ will be symmetric with mean $\frac{1}{2}$. Indeed, focusing on symmetric distributions, a strong result can be stated for two or three signals. Namely, if one person (Freddy) is Freddier than a second person (Freddy), then the distribution the former's (that is to say, Freddy's) expected beliefs after either two or three signals is unambiguously more dispersed:

Proposition 3 For all symmetric distributions π and all $\widehat{N} > N$, $f_{2,\pi}^N$ is a mean-preserving spread of $f_{2,\pi}^{\widehat{N}}$ and $f_{3,\pi}^N$ is a mean-preserving spread of $f_{3,\pi}^{\widehat{N}}$.

Propositions 1, 2, and 3 together constitute the main "over-inference" results of the paper, which say that Freddy infers too much from a *short* sequence of signals. Propositions 1 and 2 indicate that for all possible combinations of two signals, Freddy believes too strongly that the underlying rate is that which most resembles the observed signals. Proposition 3 then says that the distribution of the mean of Freddy's beliefs has too high a variance after two or three observations of the signals.

The analysis of Freddy's beliefs after longer sequences of signals is considerably more complicated. Propositions 1 and 2 show that Freddy over-infers that the rate is extreme from an extreme sequence of signals, and over-infers that the rate is close to $\frac{1}{2}$ from a 50/50 sequence of signals. When Freddy has observed just two signals, these are the only types of sequences he can observe. But longer sequences of signals typically don't fall into either of these categories. For many other

[29] To verify both of these assertions, it can be shown that a Bayesian with priors $\pi\left(\theta = \frac{1}{2}\right) = \pi\left(\theta = \frac{3}{4}\right) = \frac{1}{2}$ will form posterior beliefs $\pi_2^\infty\left(\theta = \frac{1}{2}|aa\right) = \frac{4}{13}$, $\pi_2^\infty\left(\theta = \frac{1}{2}|ab\right) = \frac{4}{7}$, and $\pi_2^\infty\left(\theta = \frac{1}{2}|bb\right) = \frac{4}{5}$, where a 4-Freddy will form beliefs $\pi_2^4\left(\theta = \frac{1}{2}|aa\right) = \frac{1}{4}$, $\pi_2^4\left(\theta = \frac{1}{2}|ab\right) = \frac{4}{7}$, and $\pi_2^4\left(\theta = \frac{1}{2}|bb\right) = 1$. Given π, the sequence aa will actually be generated $\frac{1}{2}\left(\frac{1}{2}\right)^2 + \frac{1}{2}\left(\frac{3}{4}\right)^2 = \frac{13}{32}$ of the time, ab $(= ba)$ will be generated $\frac{1}{2}2\left(\frac{1}{2}\right)^2 + \frac{1}{2}2\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{14}{32}$, and bb will be generated $\frac{1}{2}\left(\frac{1}{2}\right)^2 + \frac{1}{2}\left(\frac{1}{4}\right)^2 = \frac{5}{32}$ of the time. Hence, the Bayesian's expected posterior beliefs about how likely $\theta = \frac{1}{2}$ is $\pi_2^\infty\left(\theta = \frac{1}{2}|\pi^*\right) = \frac{13}{32}\left(\frac{4}{13}\right) + \frac{14}{32}\left(\frac{4}{7}\right) + \frac{5}{32}\left(\frac{4}{5}\right) = \frac{1}{2}$, whereas the 4-Freddy's expected beliefs are $\pi_2^4\left(\theta = \frac{1}{2}|\pi^*\right) = \frac{13}{32}\left(\frac{1}{4}\right) + \frac{14}{32}\left(\frac{4}{7}\right) + \frac{5}{32}(1) = \frac{65}{128}$, so that $E_2^4(\pi^*) = \frac{65}{128} \cdot \frac{1}{2} + \frac{63}{128} \cdot \frac{3}{4} = \frac{319}{512}$.

types of sequences, in fact, Freddy may underestimate the likelihood that the rate closely corresponds to the proportion of signals in the sequence.

To illustrate this, consider again an observer who thinks an analyst might be any of three types, bad, average, or good, having successful investment years $\frac{1}{4}$, $\frac{1}{2}$, or $\frac{3}{4}$ of the time. Suppose that the analyst is in reality, good — she invests successfully $\frac{3}{4}$ of the time. If Freddy knew just this aggregate statistic, then he would reach the obvious and correct conclusion — that this analyst is good. Suppose that Freddy observes the sequence of 6 successful followed by 2 unsuccessful performances, $aaaaaaabb$. Despite the fact that $\frac{6}{8} = \frac{3}{4}$ of the signals are a's, a 4-Freddy perceives this sequence as surely coming from $\theta = \frac{1}{2}$ rather than $\theta = \frac{3}{4}$. This is because he thinks an odd-even streak of 2 straight b's is *impossible* when $\theta = \frac{3}{4}$, when these two signals are drawn from an urn containing 3 a's and 1 b. In a 4-Freddy's mind, good analysts simply aren't unsuccessful two years in a row.

The fact that Freddy infers too much from unexpected "streaks" of two signals in a row means that after observing a good analyst for a long time, Freddy will almost surely observe all possible pairs of signals — two successful years of investing, one successful and one bad, and (less often) two unsuccessful years in a row. Hence, he will almost surely infer that the analyst is average — since the only type of analyst who can have both two unsuccessful years and two successful years in a row are average ones. He believes this despite his surprise that this supposedly average analyst is successful $\frac{3}{4}$ of the time.

In fact, quite generally Freddy is prone to believe the rate is less extreme than it is after observing a very large number of signals. To see this, note that the proportion of aa, ab, and bb pairs given the true rate θ^* and an infinite number of observations equals almost exactly $(\theta^*)^2$ aa's, $2\theta^*(1-\theta^*)$ ab's, and $(1-\theta^*)^2$ bb's. Having received these proportions, therefore, Freddy thinks that this distribution was generated by the rate θ that is most likely to generate such proportions. Lemma 6 derives Freddy's limit beliefs from his maximum-likelihood estimate of the rate:

Lemma 6 Suppose the true rate is θ^*. Then $\lim_{t\to\infty} \pi_t^N(\widehat{\theta} \mid \pi) = 1$ for all π such that $\pi(\widehat{\theta}) > 0$, where $\widehat{\theta} = \arg\max_{\theta\in\Theta} \left[\left(\theta \cdot \frac{\theta N-1}{N-1}\right)^{(\theta^*)^2} \left(2\theta \cdot \frac{(1-\theta)N}{N-1}\right)^{2(1-\theta^*)\theta^*} \left((1-\theta) \cdot \frac{(1-\theta)N-1}{N-1}\right)^{(1-\theta^*)^2} \right]$.

Lemma 6 says that as the number of signals observed becomes arbitrarily large, Freddy's beliefs converge to certainty about the rate.[30] More interestingly it can be shown that $\widehat{\theta}$ is never further

[30] The Lemma is of course well-posed only if there is a unique maximand, which I shall assume there is.

away from $\frac{1}{2}$ than is θ^*. Combined with the example above, this shows that Freddy never thinks that the rate may be more extreme than it is, but sometimes thinks that it may be strictly less extreme. Because it is an important result, and surprisingly hard to prove, this corollary to Lemma 6 is worth stating as Proposition 4:

Proposition 4 Suppose N is even. If the true rate is $\theta^* > \frac{1}{2}$, $\lim_{t\to\infty} \pi_t^N(\widehat{\theta}) = 1$ for some $\widehat{\theta} \in [\frac{1}{2}, \theta^*]$. If the true rate is $\theta^* < \frac{1}{2}$, $\lim_{t\to\infty} \pi_t^N(\widehat{\theta}) = 1$ for some $\widehat{\theta} \in [\frac{1}{2}, \theta^*]$. If the true rate is $\theta^* = \frac{1}{2}$, $\lim_{t\to\infty} \pi_t^N(\widehat{\theta} = \frac{1}{2}) = 1$.

The logic behind these results is that Freddy observes more streaks than he expects, and that the frequency of streaks of rare signals is especially surprising to him. While Freddy thinks such streaks are unlikely no matter the underlying rate, he believes that a moderate rate better explains the streaks on both sides than does an extreme rate. Thus, the unexpected streakiness of the signals can outweigh the mean frequency of signals in determining Freddy's beliefs.[31]

Despite its generality and (once we've retrained our intuitions) intuitive basis, my guess is that the "conservatism" identified by Proposition 4 is not, in pragmatic terms, that important. More generally, there are reasons to be cautious about interpreting the relevance of limit results, and in fact this may be a good juncture to point out an important feature of the model that applies to all the limit results in this paper. When Freddy's limit beliefs are different than a Bayesian's the difference depends on the fact that Freddy places *exactly* zero probability on the stochastic structure of the world being as it actually is. If Freddy placed any positive probability on the world being *i.i.d.*, then he would eventually come to believe it is *i.i.d.* This is because the pattern Freddy observes surprises him immensely in almost every case, and all limit results involve Freddy choosing the least implausible of two unlikely explanations, rather than providing an explanation he finds plausible.

This observation in turn brings up an important methodological issue that comes up in many quasi-Bayesian models. Taking these models at face value, there are typically many ways for a sufficiently sophisticated person to see that he must be making an error. In the model in this paper, for instance, Freddy should figure out that his theory of negative autocorrelation is wrong. There are three related reasons why, in my view, this does not render the model irrelevant. First,

[31] In the model of this paper, with the urn renewal after every two periods, I have not found examples where Freddy's beliefs do not converge to the true rate θ^* when $\min[\theta^* N, (1-\theta^*)N] \geq 1$. I do not know if there are such examples. But in the considerably more complicated model that has 3-period renewal and $N = 100$, $\theta^* = .97$ will eventually be rejected in favor of beliefs that $\theta = .96$, even though Freddy will never observe a sequence that he considers impossible if $\theta = .97$.

though obscured by the modeling techniques, the quasi-Bayesian approach is meant as a model of a boundedly-rational person. The reasoning needed to correct the error is often as difficult as, or more difficult than, that needed to avoid the error in the first place. Second, the model itself is simplified to keep it tractable for the analysts; in a realistically complicated model identifying the mistake is likely to be much more complicated. Third, empirically people don't correct this error. The evidence in Section 2 is from the behavior of either smarter-than-average, more-educated-than-average 20 year olds in reasonably naturalistic experimental settings, or from "the real world." Existing evidence shows not that these people make these errors before they know better, but rather that they make these mistakes given their past experience. The hypothesis that reasoning will be ubiquitously Bayesian given realistic experience levels *is* what is tested and rejected by these data.

So far I have shown that Freddy's beliefs are on average too extreme after a small number of signals, and not extreme enough after a large number of signals, as formalized by the notion that Freddy is biased towards believing in an overly moderate rate after an infinite number of signals. But there is one form of over-inference that holds even after observing a long sequence of signals: While Freddy may be underconfident that the rate is as extreme as the proportion of signals he gets would seem to indicate, he is prone to be overconfident that the true rate is consistent with the overall direction of the signals rather than the opposite. Formally, Proposition 5 states that given any symmetric prior distribution π, whenever the majority of signals in the sequence h_t are a's, Freddy will, for any $\theta > \frac{1}{2}$, exaggerate the relative likelihood that the rate is θ rather than $1 - \theta$:

Proposition 5 For all symmetric π, rates $\theta > \frac{1}{2}$ such that $\pi(\theta) > 0$, and histories h_t yielding more a signals than b signals, $\frac{\pi_t^N(\theta|h_t)}{\pi_t^N(1-\theta|h_t)}$ is decreasing in N.

The intuition for Proposition 5 is that when comparing θ to $1 - \theta$, Freddy is comparing two equally extreme rates, and hence does not favor either as an explanation for the surprising pattern of streaks being observed. With no countervailing effect, the over-inference from the relative frequency determines the nature of the error.

Of course, the pattern of signals can only affect Freddy's beliefs if he observes this pattern. I now turn to a different case — where Freddy does not observe, or does not attend to, the precise sequence in which his signals arrive. In this case, Freddy can only infer from the frequency of the two signals, and hence he will eventually discover the true rate. To formalize this, let $\pi_{\{t\}}^N(\cdot|x_a, x_b)$ be an N-Freddy's beliefs from a set of x_a a signals and x_b b signals, $x_a + x_b = t$, when he does not observe the sequences of those signals. Let $f^N_{\{t\},\pi,\theta}$ be an N-Freddy's distribution of mean beliefs after t signals given priors π and rate realization θ. Proposition 6 says that if Freddy does not observe the sequence of signals, he will come to be certain and correct about what the true rate is.

Proposition 6 For all π, θ^* and N, $\lim_{t\to\infty} f^N_{\{t\},\pi,\theta^*} = \pi^*$ where $\pi^*(\theta^*) = 1$.

Beyond this, I have little to say about what Freddy believes when he doesn't observe the sequence of signals.[32] For ease of analysis, and because the more interesting results occur when Freddy can observe the sequence of signals, and because I am often more interested in Freddy's inferences based on a small number of signals, I shall in the following two sections consider the case where Freddy does observe the signals.

5. Inference about the Distribution of Rates

In this section I explore what Freddy comes to believe about the distribution of rates when he observes signals from a large number of *different* rates. For instance, Freddy may over the course of his lifetime form beliefs about the distribution of talent among financial analysts based on observing a small number of performances from each of many analysts. Let p be Freddy's prior beliefs over the set of probability distributions π, that might prevail. Let h be a sequence of signals Freddy observes, M be the number of signals Freddy observes for each draw of a rate, and assume that Freddy observes infinitely many different draws. I shall refer to each draw of a rate as a *source*. Let $\pi^N_{p,M}(h)$ be Freddy's beliefs about the possible distribution of rates after observing an infinite sequence h of M sources. Let $g^N_M(p,\pi)$ be the probability distribution over Freddy's beliefs $\pi^N_{p,m}(h)$ when Freddy observes M signals per source if the true probability distribution of rates is π.

[32] I do not know, for instance, if Freddy necessarily over-infers the extremity of the rate if he does not observe the sequence of signals. I conjecture but have not proven that the following is true. Consider $\theta, \widehat{\theta} \in \Theta$ such that $\theta > \widehat{\theta}$. For all $\{x_a, x_b\}$ such that $\frac{x_a}{x_a+x_b} > \theta$, $\frac{\pi^N_{x_a+x_b}(\theta|\{x_a,x_b\})}{\pi^N_{x_a+x_b}(\widehat{\theta}|\{x_a,x_b\})}$ is strictly decreasing in N. For all $\{x_a, x_b\}$ such that $\frac{x_a}{x_a+x_b} < \widehat{\theta}$, $\frac{\pi^N_{x_a+x_b}(\widehat{\theta}|\{x_a,x_b\})}{\pi^N_{x_a+x_b}(\theta|\{x_a,x_b\})}$ is strictly decreasing in N. In two cases Freddy's inferences clearly do not depend on whether or not he observes the precise sequence: when the signals are all the same, or when all but one are the same. In either case, Freddy can figure out exactly which combinations of pairs of signals he has received even if he does not directly observe them. Likewise, when Freddy observes exactly two signals, it clearly does not matter if he observes the order.

If M is very large, then after any given history of signals Freddy may form accurate or conservative beliefs about the distribution of rates. To return once more to our earlier example, if the true distribution of talent has positive proportions of good, average, and bad analysts, who perform well $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ of the time, then for reasons outlined in Section 4, a 4-Freddy who observes an infinite number of performances by each analyst will eventually come to believe that they all are average. An 8-Freddy, on the other hand, will come to believe the correct distribution, since if he observes an infinite number of predictions from an analyst, he will figure out that analyst's true type. If Freddy observes an infinite number of signals, but not their sequence, from an infinite number of people, then he will come to discover the true distribution for sure, precisely paralleling Proposition 4.

Of greater interest is what happens when Freddy observes a large number of analysts, but only observes a small number of predictions from each analyst. Suppose that Freddy only observes 2 predictions from each of a large number of analysts. The result about over-inference following two signals translates readily into a belief that there is more variation in rates than there really is:

Proposition 7 For all $N < \infty$ and symmetric π such that $\pi(\theta = 1) < \frac{1}{2}$, there exists a strict mean-preserving spread of $\pi, \widehat{\pi}$, such that if $p(\widehat{\pi}) > 0$ and $p(\pi) = 1 - p(\widehat{\pi})$, then $g_2^N(p, \pi) = p^*$ where $p^*(\widehat{\pi}) = 1,$ and there does not exist a $\widetilde{\pi}$ such that π is a strict mean-preserving spread of $\widetilde{\pi}$ such that $p(\pi) > 0$ but $g_2^N(p, \pi) = p^{**}$ where $p^{**}(\widetilde{\pi}) > 0$.

Proposition 7 says that when observing two signals per source, Freddy may come to believe there is more dispersion in rates than there really is, but will never come to believe there is less dispersion in rates. Hence, when Freddy does not observe many predictions by each financial analyst, he will believe in more variation in expertise than really exists. By contrasting Proposition 7 to Proposition 4, we can see beliefs depend crucially on the number of signals observed for each rate. Proposition 7 shows that Freddy always exaggerates variance if he observes only two signals per source, whereas Proposition 4 indicates he underestimates variance if he observes a large number of signals per rate.[33]

For the remainder of the paper I turn to specific examples and applications of the model. To illustrate more concretely the intuition of some of the above results, and to facilitate analysis of some economic applications, consider the simple class of symmetric distributions of the sort used

[33] I do not know what happens if there is no variance in the world and Freddy observes a large but finite number of signals from each source. I conjecture that the following result might hold: For all $N < \infty$, for all $M < \infty$, then if the true distribution is π^* such that $\pi^*\left(\frac{1}{2}\right) = 1$, there exists a $\widehat{\pi}$ such that $\widehat{\pi}\left(\frac{1}{2}\right) < 1$ and a p such that $g_M^N(p, \pi) = p^*$ where $p^*\left(\widehat{\pi}\right) = 1$.

in all of the above illustrations: Freddy has symmetric beliefs over probability distribution of $\Theta = \{\frac{1}{2} - d, \frac{1}{2}, \frac{1}{2} + d\}$, where $d \in (0, \frac{1}{2})$, and beliefs $\pi(\theta = \frac{1}{2} - d) = \pi(\theta = \frac{1}{2} + d) = q$ and $\pi(\theta = \frac{1}{2}) = 1 - 2q$. Let q^* be the true distribution, and let Freddy's prior beliefs over the possible probability distributions over types be $p(q)$. Assume Freddy observes two signals per source.

Even more specifically, consider $\Theta = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$. If Freddy observes proportion r aa pairs, s bb pairs, and $1-r-s$ ab pairs, what does he infer? The following chart indicates the true probability of each of these pairs of signals for each rate, and an N-Freddy's perception of those probabilities. In the last line is the true proportion of the pairs of signals as a function of q, and Freddy's perceptions of how many pairs there should be as a function of q.

	True Frequency of Pairs			N-Freddy's Predicted Frequency		
	aa	ab	bb	aa	ab	bb
If $\theta = \frac{3}{4}$	$\frac{9}{16}$	$\frac{6}{16}$	$\frac{1}{16}$	$\frac{9N-12}{16N-16}$	$\frac{6N}{16N-16}$	$\frac{N-4}{16N-16}$
If $\theta = \frac{1}{2}$	$\frac{4}{16}$	$\frac{8}{16}$	$\frac{4}{16}$	$\frac{4N-8}{16N-16}$	$\frac{8N}{16N-16}$	$\frac{4N-8}{16N-16}$
If $\theta = \frac{1}{4}$	$\frac{1}{16}$	$\frac{6}{16}$	$\frac{9}{16}$	$\frac{N-4}{16N-16}$	$\frac{6N}{16N-16}$	$\frac{9N-12}{16N-16}$
If $\pi\left(\frac{1}{4}\right) = \pi\left(\frac{3}{4}\right) = q$	$\frac{1}{4} + \frac{q}{8}$	$\frac{1}{2} - \frac{q}{4}$	$\frac{1}{4} + \frac{q}{8}$	$\frac{Nq+2N-4}{8N-8}$	$\frac{2N-Nq}{4N-4}$	$\frac{Nq+2N-4}{8N-8}$

This chart can be used to derive Freddy's perceived distribution of rates, $\widetilde{q}$, as a function q, the actual distribution of rates. If Freddy believes that $\pi\left(\theta = \frac{1}{4}\right) = \pi\left(\theta = \frac{3}{4}\right) = \widetilde{q}$, then he expects to observe $\frac{2N-N\widetilde{q}}{4N-4}$ ab signals. Given he observes $\frac{1}{2} - \frac{q}{4}$ ab pairs where q is the probability of $\theta = \frac{1}{4}$ and $\theta = \frac{3}{4}$, Freddy's beliefs satisfy $\frac{2N-N\widetilde{q}}{4N-4} = \frac{1}{2} - \frac{q}{4}$, yielding the result that $\widetilde{q} = \frac{N-1}{N}q + \frac{2}{N}$.

Notice that $\widetilde{q}$ converges to q as N converges to infinity: As Freddy becomes closer and closer to Bayesian, his beliefs are closer and closer to accurate. But for all $N < \infty$, $\widetilde{q} > q$: Freddy always thinks there is more variation than there is. Note in particular that even if $q = 0$, $\widetilde{q} > 0$ for all $N < \infty$. In the extreme, when $N = 4$, Freddy believes that no analyst is average even when they all are.[34]

I now turn to a starker and even more tractable example that will serve as a template for many further examples. Suppose that there is a distribution over three possible rates, $\Theta = \{0, \frac{1}{2}, 1\}$, and $\pi(\theta = 0) = \pi(\theta = 1) = q$, $\pi(\theta = \frac{1}{2}) = 1 - 2q$ is the true distribution, for $q \in \left[0, \frac{1}{2}\right]$. Consider a 4-Freddy. Freddy understands that rate $\theta = 1$ always generates a's and rate $\theta = 0$

[34] Again, these results and intuitions are only for the case where M is small. If Freddy observes more than two signals from each source, on the other hand, he may eventually come to believe there is no variance even when there is. Indeed, given $\Theta = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$, if Freddy's beliefs after observing M signals each from an infinite number of sources is represented by $\widetilde{q}(M, q)$, then for all q, $\lim_{M\to\infty} \widetilde{q}(M, q) = 0$.

always generates b's. They differ only in their beliefs about the probability which $\theta = \frac{1}{2}$ generates sequences. Whereas $\theta = \frac{1}{2}$ actually generates pairs $aa, ab,$ and bb in proportions $\frac{1}{4}, \frac{1}{2},$ and $\frac{1}{4}$, 4-Freddy thinks it generates them in proportions $\frac{1}{6}, \frac{2}{3},$ and $\frac{1}{6}$. Hence, Freddy thinks that when the distribution is $\pi(\theta = 0) = \pi(\theta = 1) = \widetilde{q}$, he'll see aa or bb pairs $\widetilde{q}(1) + (1 - 2\widetilde{q})\frac{1}{6} + \widetilde{q}(0) = \frac{1}{6} + \frac{2}{3}\widetilde{q}$ of the time, and ab pairs $\frac{2}{3} - \frac{4}{3}\widetilde{q}$ of the time. If the distribution is q, then he will actually observe aa pairs $q(1) + (1 - 2q)\frac{1}{4} + q(0) = \frac{1}{4} + \frac{1}{2}q$ of the time. Hence, setting $\frac{1}{6} + \frac{2}{3}\widetilde{q} = \frac{1}{4} + \frac{1}{2}q$, we see that in the limit as Freddy observes two signals each from an infinite number of rates, he will come to believe $\widetilde{q} = \frac{1}{8} + \frac{3}{4}q$, as we can see setting once more, for all $q \in [0, \frac{1}{2})$, Freddy exaggerates how common the extreme rates are.

But now suppose that, in the same population of sources, Freddy observes four signals each from an infinite number of sources.[35] Freddy's beliefs about the likelihood of all combinations of four signals are as follows:

			Probability of the sequence if $\theta = \frac{1}{2}$	
		# Permutations	Actual Likelihood	4-Freddy's Perceived Likelihood
aa	aa	1	$\frac{1}{16}$	$1 \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$
bb	bb	1	$\frac{1}{16}$	$1 \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$
aa	ab	4	$\frac{4}{16}$	$4 \times \frac{1}{6} \times \frac{1}{3} = \frac{8}{36}$
bb	ab	4	$\frac{4}{16}$	$4 \times \frac{1}{6} \times \frac{1}{3} = \frac{8}{36}$
ab	ab	4	$\frac{4}{16}$	$4 \times \frac{1}{3} \times \frac{1}{3} = \frac{16}{36}$
aa	bb	2	$\frac{2}{16}$	$2 \times \frac{1}{6} \times \frac{1}{6} = \frac{2}{36}$

In the next Table, Freddy expects the frequencies in the right-hand column when $\pi(\theta = 0) = \pi(\theta = 1) = \widetilde{q}$, whereas he actually observes the frequencies in the left-hand column when the true distribution is q:

[35] Studying the case where $\theta = 1$ or $\theta = 0$ obscures the more general possibility that Freddy underestimates the frequency of extreme rates, since in this case he will never observe any surprising counter-signals if the true rate is extreme, and hence here he will never come to underestimate the variance in distribution.

	Actual distribution of sequences given q	Freddy's perceived distribution given $\widetilde{q}$
$aa \quad aa$	$q(1) + (1-2q)\frac{1}{16} = \frac{1}{16} + \frac{7}{8}q$	$\widetilde{q}(1) + (1-2\widetilde{q})\frac{1}{36} = \frac{1}{36} + \frac{17}{18}\widetilde{q}$
$bb \quad bb$	$q(1) + (1-2q)\frac{1}{16} = \frac{1}{16} + \frac{7}{8}q$	$\widetilde{q}(1) + (1-2\widetilde{q})\frac{1}{36} = \frac{1}{36} + \frac{17}{18}\widetilde{q}$
$aa \quad ab$	$\frac{4}{16}(1-2q) = \frac{1}{4} - \frac{1}{2}q$	$\frac{8}{36}(1-2\widetilde{q}) = \frac{2}{9} - \frac{4}{9}\widetilde{q}$
$bb \quad ab$	$\frac{4}{16}(1-2q) = \frac{1}{4} - \frac{1}{2}q$	$\frac{8}{36}(1-2\widetilde{q}) = \frac{2}{9} - \frac{4}{9}\widetilde{q}$
$ab \quad ab$	$\frac{4}{16}(1-2q) = \frac{1}{4} - \frac{1}{2}q$	$\frac{16}{36}(1-2\widetilde{q}) = \frac{4}{9} - \frac{8}{9}\widetilde{q}$
$aa \quad bb$	$\frac{2}{16}(1-2q) = \frac{1}{8} - \frac{1}{4}q$	$\frac{2}{36}(1-2\widetilde{q}) = \frac{1}{18} - \frac{1}{9}\widetilde{q}$

To derive what Freddy's eventual beliefs will be from the above table, we must confront a possibility not seen in the earlier examples: That Freddy cannot form *any* beliefs to explain the distribution he sees. While Freddy expects, for instance, to see twice as many foursomes of $\{ab, ab\}$ than $\{aa, ab\}$, irrespective of q, he is perplexed to see roughly equal numbers. Hence, Freddy must form beliefs by choosing among an array of very implausible explanations for the pattern he observes. To see what this inference process involves, imagine that Freddy receives 16 quadruplets of signals in exactly the expected proportions according to q, from which he tries to estimate q. It can be shown that Freddy would update his beliefs towards the $\widetilde{q}$ that maximizes the likelihood function

$$L(\widetilde{q}) \equiv \left[\left(\frac{1}{36} + \frac{17}{18}\widetilde{q} \right)^{2(\frac{1}{16}+\frac{7}{8}q)} \cdot Z \cdot (1-2\widetilde{q})^{\frac{7}{8}-\frac{7}{4}q} \right]^X ,$$

where Z is a term that does not depend on $\widetilde{q}$. When he observes 1600 or 16,000 quadruplets in these proportions — which he will — then he comes to believe firmly in $\widetilde{q}$. This yields the solution that after a very large number of observations, Freddy believes:

$$\widetilde{q} = \frac{5}{136} + \frac{63}{68}q.$$

Notice that if $q = 0$, then $\widetilde{q} = \frac{5}{136}$: If there is no variation, Freddy comes to believe that there is. Indeed, for all $q < \frac{1}{2}$ Freddy exaggerates variance. To use an example I will return to, if $q = \frac{1}{7}$, Freddy will believe $\widetilde{q} = \frac{23}{136} > \frac{1}{7}$. Freddy gets things exactly right when $q = \frac{1}{2}$. While again noting that the result might be reversed if Freddy both makes lots of observations from each source, and his priors assign positive probability to some $\theta \in \left(\frac{1}{2}, 1\right)$, the result does capture the intuition that Freddy is likely to exaggerate the prevalence of extreme rates.

The fact that Freddy tends to exaggerate the variance in rates has implications for the assumption, emphasized throughout Sections 3 and 4, that Freddy starts with correct priors and has mistaken beliefs solely due to erroneously updating his beliefs after observing signals. But results here suggest that we should not expect Freddy's "priors" to be correct after all: When Freddy has previously inferred the distribution of rates from small numbers of observations of each source, he will have overly dispersed prior beliefs about the rates for each new source he is facing. Hence, Freddy not only over-infers because of a bad updating rule, but because of bad priors. He'll infer too much about the extreme talent of a newly-hired financial analyst not only because he'll infer too much from small samples, but because when he hires her he exaggerates how likely it is that she is talented. Hence, Freddy's propensity to over-infer is even more severe than suggested in Section 4. This is most interesting after one signal: While Lemma 4 of Section 4 indicates that Freddy's inferences after one signal are the same as a Bayesian's, here we see that Freddy is likely to have over-dispersed beliefs even after one signal.

To illustrate this point (and to build the foundation for a later point), consider the situation just discussed, where 4-Freddy has observed four signals each from a large number of sources, where $\pi(\theta = \frac{1}{2}) = \frac{5}{7}, \pi(\theta = 0) = \pi(\theta = 1) = \frac{1}{7}$, and had original priors with support $\Theta \in \left\{0, \frac{1}{2}, 1\right\}$. Freddy will believe after observing one a from a new source that the probability he is facing rate $\theta = 1$ is $\frac{\frac{23}{136}\cdot 1}{\frac{23}{136}\cdot 1+\frac{90}{136}\cdot\frac{1}{2}} = \frac{23}{68} \approx 33.8\%$, whereas a Bayesian with correct priors of $q = \frac{1}{7}$, believes she is facing rate $\theta = 1$ with probability $\frac{\frac{1}{7}\cdot 1}{\frac{1}{7}\cdot 1+\frac{5}{7}\cdot\frac{1}{2}} = \frac{2}{7} \approx 28.6\%$.

I use this example to turn to an issue I have de-emphasized since Section 3: what Freddy predicts about future signals. Freddy's mispredictions about coming signals implicates not only over-inference but also the gambler's fallacy. The numbers 33.8% and 28.6% represent Freddy's beliefs about the rate, not his prediction of the next signal. Freddy predicts that the 2nd signal following one a will be an a with probability $\frac{23}{68} \cdot 1 + \frac{45}{68} \cdot 3 \approx 55.9\%$. A Bayesian, by contrast, predicts that the next signal will be a with probability $\frac{2}{7} \cdot 1 + \frac{5}{7} \cdot \frac{1}{2} \approx 64.3\%$. The reason that Freddy underestimates the probability of an ensuing a is because of his belief in the gambler's fallacy; he figures that *if* the analyst is average, then her next performance will be b with probability $\frac{2}{3}$.

Following 2 a's, Freddy will believe he is facing rate $\theta = 1$ with probability $\frac{\frac{23}{136}\cdot 1}{\frac{23}{136}\cdot 1+\frac{90}{136}\cdot\frac{1}{6}} = \frac{23}{38} \approx 60.5\%$, and hence predict a third a with probability $\frac{23}{38}(1) + \frac{15}{38}(\frac{1}{2}) \approx 80.3\%$. A Bayesian thinks he is facing $\theta = 1$ with probability $\frac{\frac{1}{7}\cdot 1}{\frac{1}{7}\cdot 1+\frac{5}{7}\cdot\frac{1}{4}} = \frac{4}{9} \approx 44.4\%$, and hence predicts a third a with probability $\frac{4}{9}(1) + \frac{5}{9}(\frac{1}{2}) \approx 72.2\%$. That is, while Freddy underestimates the probability of a second a in a row, he *exaggerates* the probability of a third a in a row. This is partly because in predicting the signal following renewal of the urn, the gambler's fallacy does not kick in. But the second reason Freddy now exaggerates the likelihood of a repeat signal is that the bias towards over-inference is even more severe with two signals than one. Indeed, consider Freddy's prediction after three initial a's. Now he thinks he is facing $\theta = 1$ with probability $\frac{\frac{23}{136}\cdot 1}{\frac{23}{136}\cdot 1+\frac{90}{136}\cdot\frac{1}{12}} = \frac{46}{61} \approx 75.4\%$, and hence predicts the next signal to be an a with probability $\frac{46}{61}(1)+\frac{15}{61}(\frac{1}{3}) \approx 83.6\%$. The Bayesian thinks 3 a's implies he is facing $\theta = 1$ with probability $\frac{\frac{1}{7}\cdot 1}{\frac{1}{7}\cdot 1+\frac{5}{7}\cdot\frac{1}{8}} = \frac{8}{13} \approx 61.5\%$, and predicts a fourth a with probability $\frac{8}{13}(1) + \frac{5}{13}(\frac{1}{2}) \approx 80.8\%$. After this longer sequence, even when the gambler's fallacy is operative, over-inference about the rate overwhelms it and leads to exaggerated prediction of repeated recent signals.[36]

If we disentangle odd-even effects from the trend, it seems that from a short extreme sequence Freddy underestimates repetition of a signal, but from a longer extreme sequence he *exaggerates* the likelihood of repetition. This indicates that the law of small numbers might provide an intuitive explanation for a financial anomaly that has recently received attention: DeBondt and Thaler (1990) and others have argued that investors seem irrationally to under-react in the short run to news about a firm's financial health, and over-react in the medium run to news. While there are several alternative rational-choice and quasi-Bayesian models, noted in Section 2, that might explain this phenomenon, the law of small numbers can also provide an account of it. To see how my model may lead to similar predictions about behavior, consider the following environment. I assume all investors live infinitely, and they invest at random in one stock for 4 months, and then move on to another stock, etc., never reinvesting in earlier stocks. Their eventual beliefs about the distribution of underlying quality of stocks is determined as derived above, where an a signal is a postitive shock to a firm's value and b is a negative shocks, where in actuality these shocks to not predict more positive or negative shocks. A given company lives infinitely, with the same number of potential investors each month observing them, with a turnover rate of $\frac{1}{4}$ of investors each period. The performance of all stocks is i.i.d., with $\frac{5}{7}$ of stocks having underlying quality $\theta = \frac{1}{2}$, and $\frac{1}{7}$ each are $\theta = 1$ and $\theta = 0$.

I now examine average beliefs by investors observing a company, as a function of the company's recent history. Consider all the possible histories of the company that next year's investors can ob-

[36] Once again, investigating the case of $\Theta \in \left\{0, \frac{1}{2}, 1\right\}$ obscures a key countervailing intuition. If the sequence observed represents the true rate $\theta \in (\frac{1}{2}, 1)$, eventually a Bayesian will come to believe in that rate, and Freddy will either believe it or underestimate its extremity. In either case, combined with the gambler's fallacy, Freddy will under-predict repetition of the dominant signal relative to a Bayesian. With that substantial caveat in mind, the lesson here is perhaps that Freddy's propensity to over-predict repetition of recent signals rather than under-predict it increases in the medium-run observation of streaks.

serve in which the most recent performance has been an a : $\{\emptyset a, \emptyset aa, \emptyset ba, \emptyset aaa, \emptyset baa, \emptyset aba, \emptyset bba\}$, where $\emptyset$ represents the fact that the investor did not observe the previous sequence. The following table summarizes the relevant numbers derived from the results calculated above:

	$\pi_t^\infty(\theta = 1 \mid \cdot)$	$\pi_t^4(\theta = 1 \mid \cdot)$	$\pi_t^\infty(s_t = a \mid \cdot)$	$\pi_t^4(s_t = a \mid \cdot)$
$\emptyset a$	$\frac{2}{7} \approx 28.6\%$	$\frac{23}{68} \approx 33.8\%$	$\frac{9}{14} \approx 64.3\%$	$\frac{19}{34} \approx 55.9\%$
$\emptyset aa$	$\frac{4}{9} \approx 44.4\%$	$\frac{23}{38} \approx 60.5\%$	$\frac{13}{18} \approx 72.2\%$	$\frac{61}{76} \approx 80.3\%$
$\emptyset ba$	0%	0%	$\frac{1}{2} = 50\%$	$\frac{1}{2} = 50\%$
$\emptyset aaa$	$\frac{8}{13} \approx 61.5\%$	$\frac{46}{61} \approx 75.4\%$	$\frac{21}{26} \approx 80.8\%$	$\frac{51}{61} \approx 83.6\%$
$\emptyset baa$	0%	0%	$\frac{1}{2} = 50\%$	$\frac{1}{3} \approx 33.3\%$
$\emptyset aba$	0%	0%	$\frac{1}{2} = 50\%$	$\frac{1}{3} \approx 33.3\%$
$\emptyset bba$	0%	0%	$\frac{1}{2} = 50\%$	$\frac{1}{3} \approx 33.3\%$

From this table, and from the assumption that equal numbers of new investors regenerate every 4th period, we can compare the beliefs of Bayesians versus 4-Freddies averaged among the four relevant cohorts of observers, about the next signal The following table derives the average likelihood that the next signal is an a among investors who have observed 0, 1, 2, or 3 of the most recent performances, as a function of what those performances have been.

					Bayesian					4-Freddy
	0	1	2	3	*Avg*	0	1	2	3	*Avg*
bba	$\frac{1}{2}$	$\frac{9}{14}$	$\frac{1}{2}$	$\frac{1}{2}$	.5357	$\frac{1}{2}$	$\frac{19}{34}$	$\frac{1}{2}$	$\frac{1}{3}$	.4730
aba	$\frac{1}{2}$	$\frac{9}{14}$	$\frac{1}{2}$	$\frac{1}{2}$	.5357	$\frac{1}{2}$	$\frac{19}{34}$	$\frac{1}{2}$	$\frac{1}{3}$	.4730
baa	$\frac{1}{2}$	$\frac{9}{14}$	$\frac{13}{18}$	$\frac{1}{2}$	.5913	$\frac{1}{2}$	$\frac{19}{34}$	$\frac{61}{76}$	$\frac{1}{3}$	.5487
aaa	$\frac{1}{2}$	$\frac{9}{14}$	$\frac{13}{18}$	$\frac{21}{26}$	.6682	$\frac{1}{2}$	$\frac{19}{34}$	$\frac{61}{76}$	$\frac{51}{61}$	.6744

While not numerically dramatic, the table reveals the pattern discussed: For "short" sequences of recent performance — a streak of one or two a's — an investment pool of Freddies will under-react to the string. But for a "long" sequence — 3 or more a's in a row — an investment pool of Freddys will over-react — exaggerating the likelihood that the observed firm is good.

I do not wish to claim that this pattern inheres in the logic of the law of small numbers; it depends on the many parameters of the model.[37] But the logic of the model seems to lead naturally to this

[37] On the other hand, neither do we know whether empirically observed investor behavior inheres in the logic of the market — it too may exist only because of the prevailing constellation of market parameters.

pattern if the true variance in performance is small and investors do not make too many observations of each firm.

6. Inference Based on Endogenous Observations

Thus far in the paper I have assumed that what Freddy observes is independent of the realization of the signals he has observed earlier. But people often choose what to observe, and do so in part based on their beliefs they have formed from earlier observations. Because the inferences they have made influence their information gathering, therefore, the law of small numbers can influence not only how people interpret signals, but also which signals they observe. In the examples I work out in this section, Freddy will no matter his behavior observe an infinite number of signals, and these signals will be consistent with only one underlying distribution of rates. Hence, a Bayesian would, independent of his behavior, always converge on full understanding about the world, and a unique long-run behavioral pattern.

To illustrate how Freddy's choice of behavior may influence his belief formation, suppose he believes that there is some variance in the usefulness of interacting with certain people; he thinks, for instance, that some financial analysts provide profitable advice while others don't. Now suppose Freddy quits employing analysts when he thinks the expected benefit of searching for a better one exceeds the benefit (net of transactions costs) of sticking with the current one.

In such contexts, it is likely that Freddy will come to believe that analysts are worse than they truly are. The intuition is straightforward: Freddy is likely to switch analysts after a short sequence of negative signals and stay with his current analyst after a short sequence of good signals. Because Freddy switches after over-inferring an analyst is bad, but sticks around to correct his beliefs when he over-infers that an analyst is good, he will end up exaggerating the prevalence of bad analysts.

Consider again the example of a world where all financial analysts are successful $\frac{1}{2}$ of the time, but Freddy believes it is possible that there are also some analysts who are always or never successful. Suppose that Freddy employs an infinite sequence of analysts to help him invest, where he has the opportunity to switch analysts after two signals, and *must* switch after four signals. Assume—crucially—that Freddy observes the performance of only those analysts that he hires.

Let the signal $s_t = a$ correspond to successful investment by the analyst, and $s_t = b$ correspond to unsuccessful investment. Assume that Freddy wishes to maximize $\sum_{t=1}^{\infty} \delta^t i(s_t)$, where $i(a) = 1$

and $i(b) = 0$. That is, Freddy wishes to maximize the present discounted sum of his money, where he earns more if the analyst he has hired is successful than if she is unsuccessful. Assume that δ is very close to 1, so that Freddy wants to maximize average per-period payoff.[38]

Before investigating which switching behavior Freddy chooses, I first address the question of what he will come to believe as a function of his switching behavior. Freddy forms some beliefs $(\widetilde{q}_a, \widetilde{q}_b)$, where he believes that proportion $\widetilde{q}_a$ of the analysts are good $(\theta = 1)$, $\widetilde{q}_b$ are bad $(\theta = 0)$, and $1 - \widetilde{q}_a - \widetilde{q}_b$ are average $(\theta = \frac{1}{2})$. These beliefs will be determined by what Freddy observes.

If Freddy never switches after two signals, he will always observe four signals, and Section 5 shows that his eventual beliefs will be $\widetilde{q}_a = \widetilde{q}_b = \frac{5}{136}$. To see what happens when Freddy switches after a bb pair, but not otherwise, notice that out of every 16 analysts Freddy employs, he observes on average one (aa, aa), one (aa, bb), two (ab, bb), four (aa, ab), four (ab, ab), and four (bb) combinations. The fact that Freddy abandons an analyst after bb means that one fourth of the time he will observe just bb from an analyst, rather than four signals. Eleven of these sixteen sequences involve mixes of a's and b's and hence can only be generated by $\theta = \frac{1}{2}$. The sequence (aa, aa) can be generated by either $\theta = \frac{1}{2}$ or $\theta = 1$, and bb can be generated by $\theta = \frac{1}{2}$ or $\theta = 0$. If 4-Freddy believes that the distribution is (q_a, q_b), he believes the frequency of the 11 mixed sequences is $\frac{29}{36}(1 - q_a - q_b)$, the frequency of (aa, aa) combinations is $\frac{1}{36}(1 - q_a - q_b) + 1q_a = \frac{1}{36}(1 + 35q_a - q_b)$, and the frequency of bb pairs is $\frac{1}{6}(1 - q_a - q_b) + q_b = \frac{1}{6}(1 - q_a + 5q_b)$. From this it can be shown that Freddy will come to believe $\widetilde{q}_a = \frac{9}{232}$ and $\widetilde{q}_b = \frac{25}{232}$.[39] Freddy's beliefs will be the same if he instead switched after both bb and ab, because Freddy will see the same proportions of (aa, aa) and (bb) combinations as when he switches only on bb; he'll see different mixed combinations of signals, but all mixed combinations mean the same thing to Freddy—that he is observing rate $\theta = \frac{1}{2}$ for sure. Because these are the only non-mixed sequences he observes, they are the only thing that determines his beliefs.

Just as when Freddy never voluntarily switches, when Freddy fires analysts after bad perfor-

[38] The assumption that Freddy can or must switch after even periods is important, and perhaps leads to misleading conclusions. Suppose Freddy could switch after one signal, and had to switch after 2 signals. Then the Gambler's Fallacy is likely to dominate his behavior. As such, depending on his beliefs, he is likely to voluntarily switch after observing an a signal, but not after observing a b — since he thinks his current analyst is likely to revert to mean. If Freddy were observing many more signals per analyst, and (more importantly) were making decisions about whether to stay with an analyst for a while, then even if the exact timing of Freddy's switch is determined by his belief in the Gambler's Fallacy, his bigger-scale decision about how long to remain with an analyst will likely be dominated by his belief about the analyst's general merits. Fleshing out this logic in a more complicated model would be difficult, and hence focusing on the two-period/four-period model serves as a useful way to capture these issues.

[39] These are the beliefs that maximize $(1 - q_a - q_b)^{11} (1 - q_a + 4q_b)^4 (1 + 35q_a - q_b)$.

mance, he believes there is variance in expertise where there is none. But here Freddy forms biased beliefs, because he is switching from "bad" analysts before he has found out that they aren't really that bad, but sticking with "good" analysts — long enough to discover they aren't that good. Notice also that now Freddy exaggerates the variance even more: Both $\widetilde{q}_a$ and $\widetilde{q}_b$ have gone up. Intuitively, Freddy is now observing fewer signals from each analyst, which generally raises his perception of variance.

To figure out Freddy's expected average payoffs from different behaviors is somewhat complicated, and has a somewhat complicated connection to his actual payoffs.[40] Letting C be the cost of premature switches, and assuming forced switches are free, then it can be shown that Freddy's perceptions of payoffs are as follows:

If never switch:	$1+\widetilde{q}_a-\widetilde{q}_b-\frac{C}{2}$.
If switch after bb:	$\frac{11+13\widetilde{q}_a-11\widetilde{q}_b-6C}{11+\widetilde{q}_a-5\widetilde{q}_b}$
If switch after bb or ab:	$\frac{7+17\widetilde{q}_a-7\widetilde{q}_b-6C}{7+5\widetilde{q}_a-\widetilde{q}_b}$

Finally, these payoffs can be calculated when $\widetilde{q}_a=\widetilde{q}_b=\frac{5}{136}$, and for $\widetilde{q}_a=\frac{9}{232}$, $\widetilde{q}_b=\frac{25}{232}$:

	$\widetilde{q}_a=\widetilde{q}_b=\frac{5}{136}$	$\widetilde{q}_a=\frac{9}{232}$, $\widetilde{q}_b=\frac{25}{232}$
Never switch	$\frac{2-C}{2}$	$\frac{99-58C}{116}$
Switch after bb	$\frac{1506-816C}{1476}$	$\frac{399-232C}{406}$
Switch after ab or bb	$\frac{1002-816C}{972}$	$\frac{267-232C}{274}$

From this last table, Freddy's potential switching strategies can be determined. The table can be used to make three relevant statements. When Freddy has beliefs $\widetilde{q}_a=\widetilde{q}_b=\frac{5}{136}$, he refrains from voluntary switching if and only if $C>\frac{25}{78}\approx .32$. When Freddy has beliefs $\widetilde{q}_a=\frac{9}{232}$ and $\widetilde{q}_b=\frac{25}{232}$, he refrains from switching if and only if $C>\frac{1923}{5510}\approx .35$. Finally, if Freddy prefers to

[40] First, if Freddy switches after the first pair proportion x of the time, then the proportion of pairs, y, that are first pairs will be given by $y=yx+(1-y)1$, since there is a y chance his pair was new last time, yet he switched, and $1-y$ chance he was old last time, and forced to switch. Hence, $y=\frac{1}{2-x}$. If Freddy plans to switch on bb only, then he expects to switch proportion $\frac{1}{6}(1-\widetilde{q}_a-\widetilde{q}_b)+\widetilde{q}_b$ of the time, so that his perceived proportion of new pairs will be $\frac{6}{11+\widetilde{q}_a-5q_b}$. Freddy perceives the average payoff from a new pair to be $\widetilde{q}_a\cdot 2+(1-\widetilde{q}_a-\widetilde{q}_b)\cdot 1=1+\widetilde{q}_a-\widetilde{q}_b$. He perceives the payoff from an old pair that followed an ab initial pair to be 1. He perceives the expected payoff of an old pair following an initial aa pair to be $\frac{1+11\widetilde{q}_a-\widetilde{q}_b}{1+5\widetilde{q}_a-\widetilde{q}_b}$. Freddy thinks the proportions of overall pairs that will be 2^{nd} pairs following initial aa and ab pairs will be $\frac{1+5\widetilde{q}_a-\widetilde{q}_b}{11+\widetilde{q}_a-5\widetilde{q}_b}$ and $\frac{4-4\widetilde{q}_a-4\widetilde{q}_b}{11+\widetilde{q}_a-5\widetilde{q}_b}$ respectively. So Freddy's expected average payoff from switching following bb pairs can be calculated to be $\frac{11+13\widetilde{q}_a-11\widetilde{q}_b-6C}{11+\widetilde{q}_a-5\widetilde{q}_b}$. Now suppose Freddy's strategy is to switch except after aa. Then Freddy believes that proportion $\frac{6}{7+5\widetilde{q}_a-\widetilde{q}_b}$ of pairs will be new pairs, with expected payoff $1+\widetilde{q}_a-\widetilde{q}_b$, and proportion $\frac{1+5\widetilde{q}_a-\widetilde{q}_b}{7+5\widetilde{q}_a-\widetilde{q}_b}$ of signals will be 2^{nd} pairs following aa initial pairs, with expected payoffs $\frac{1+11\widetilde{q}_a-\widetilde{q}_b}{1+5\widetilde{q}_a-\widetilde{q}_b}$. Hence, Freddy's perception of payoffs from this switching strategy will be $\frac{7+17\widetilde{q}_a-7\widetilde{q}_b-6C}{7+5\widetilde{q}_a-\widetilde{q}_b}$. Freddy's expected average payoff from never switching voluntarily will be $1+\widetilde{q}_a-\widetilde{q}_b-\frac{C}{2}$.

switch, then he does so if and only if he sees bb—he never switches following ab or aa. Taking these statements together, when $C < .32$, Freddy switches when the analyst he hires performs at bb, and when $C > .35$, Freddy never voluntarily switches. But: When $C \in (\frac{25}{78}, \frac{1923}{5510}) \approx (.32, .35)$, Freddy strictly prefers never switching *if* he has not been switching, but prefers to switch at bb if he has been switching. This shows that there are two steady-state belief-behavior combinations for the same parameters—one where Freddy switches a lot because he thinks there is variance in expertise that merits shopping around, and one where he doesn't. This is driven by the endogeneity of beliefs, which would not arise for a Bayesian. Because in one of these steady states Freddy is incurring more-than-necessary search costs, in addition to showing how errors in belief-formation can lead to lead to multiple steady-state belief-behavior combinations, this example illustrates how belief in the law of small numbers may lead to inefficient expenditures by people in pursuit of entirely illusory expert opinions.[41]

7. Discussion and Conclusion

The model in this paper helps see how several different phenomena logically derive from the same underlying judgmental bias. In doing so, it also ties together the *scale* of these phenomena; the strength of the gambler's fallacy determines the degree of over-inference, the scope of the false-variation bias, *etc*. This tight structure makes the model precise and refutable, and it would be of some interest to see how well the model does in simultaneously explaining the scope of these phenomena in relevant economic circumstances. But I suspect that the simple model of this paper will not calibrate well. As it stands, there is only one parameter of the model—how big an "urn" the person believes in—that provides a degree of freedom in the specifying the nature of a person's belief in the law of small numbers.[42] Allowing a more general (and more complicated) model that allows more parameters while preserving the qualitative features of the law of small numbers will

[41] This, in turn, raises the possibility that seemingly harmful interventions that interfere with choices people make can in fact be beneficial; in the examples here, for instance, raising the cost of search can make Freddy better off.

[42] One degree of freedom that could be added concerns the renewal periodicity: The choice of assuming it is every other period was made entirely for simplicity. Freddy could believe that the urn is renewed every K periods, where K is any positive integer that is small enough relative to N to keep the model coherent. For a fixed N, modifying K changes the intensity of the gambler's fallacy and over-inferences; changing K and N jointly may allow flexibility in fixing the average degree of gambler's fallacy while changing how sensitive it is to long versus short streaks. For instance, increasing both N and K in certain ways can decrease Freddy's expected beliefs that about the probability of a b signal following one a signal in a row, but greatly increase beliefs about the probability of a b signal following several a's in a row, since with a larger K Freddy believes that an urn is nearly run out.

likely be needed to allow greater ability to de-link the precise scale of the phenomena associated with the law of small numbers.

Other modifications to the model are needed to make the model more realistic. The most obvious is modifying the artificial distinction between even and odd periods. The best way to fix this artificial feature may sometimes be the one I used in the applications above—simply taking an average over the odd and even periods. But for other applications this may not be adequate, and a more stationary model would need to be developed.

Applying the law of small numbers to many economic situations of interest may also require an understanding of how it is offset or reinforced by other well-established cognitive biases. Several of these biases, in fact, suggest caution in concluding too much in practical terms from the analysis of Freddy's inference from a very large sequence of observations.[43] First, as reviewed in Section 2, while people believe in the law of small numbers, they tend *not* to believe in the law of *large* numbers, unconvinced that very large samples will very surely closely resemble the overall population. This will affect limit results. Roughly speaking, while the model predicts that Freddy, like a Bayesian, will always converge to complete (but often mislplaced) confidence in his beliefs after receiving lots of information, the lack of belief in the law of large numbers means people will be uncertain even after getting a huge sample.[44]

Other biases also enter the picture in understanding inference from long sequences. While confirmatory bias, as modeled in Rabin and Schrag (1999), reinforces over-inference after a short sequence of signals, its over-inference after a long sequence can counteract the conservatism generated by the law of small numbers. Also, because the limit results in this paper clearly require Freddy to correctly understand the pattern he is seeing, they are suspect in light of evidence that people are poor at judging such patterns; the model's prediction about inferences people make based on the patterns they observe in long sequences may be misleading because the patterns they "observe" aren't the patterns that are there.

[43] One reason to be cautious about the limit results doesn't concern any other biases. Rather, as demonstrated in Section 4, the inferences Freddy reaches from a large number of observations of a single source are very sensitive to whether he observes the precise sequence. If he doesn't observe the whole sequence, then his inference from a large number of observations will likely be over-inference rather than under-inference—and in the limit he will reach the correct conclusions. The incorrect, "under-inference" results arise only if Freddy can observe—and pays attention to—the pattern of signals. (And note that, while Freddy always believes there is some information in paying attention to the precise pattern of signals, he doesn't think that there is any information in the pattern in the limit—and hence may optimally decide to keep track only of the frequencies rather than patterns.)

[44] As such, the lack of belief in the law of large numbers may reinforce Freddy's conservatism in moving away from strong preconceptions that the rate is not extreme.

One such pattern-recognition bias stands out as having special significance in the context of this paper: The hot-hand fallacy. This is the tendency for people to perceive a "hot hand" (positive autocorrelation) in what are actually *i.i.d* sequences of signals. Variants of this misperception sometimes show up in experiments. In "prediction-task" experiments—in which participants predict coming signals as a function of recent signals—there are some cases where the predominant pattern is for people to over-predict continuation of recent signals rather than to commit the gambler's fallacy. The clear majority of experiments, however, are more consistent with the gambler's fallacy.[45]

But belief in the hot hand has been documented much more widely in the field. Gilovich, Vallone, and Tversky (1985) and Tversky and Gilovich (1989a, 1989b) have demonstrated that, while basketball fans believe that basketball players are streak shooters whose "on" and "off" nights cannot be explained by randomness, such a hot hand does not in fact exist (or at least not nearly to the degree that people believe in it). Camerer (1989) shows that organized gambling on basketball games exhibits a small hot-hand bias, insofar as betting indicates a belief that winning streaks and losing streaks are more likely to continue than they actually are.

At first blush, the hot-hand fallacy may seem in contradiction to the gambler's fallacy, since it suggests that people expect to see too many long strings of the same signal rather than too much alternation. Some cases where the hot-hand fallacy prevails do indeed represent an important caveat to the findings of this paper.[46] As many researchers have intuited, however, the hot-hand fallacy may in fact *derive from* the law of small numbers rather than contradict it. The hot hand fallacy, in most of the accounts I have seen, is interpreted as coming from people's perception that observed streaks are too long to be due to chance. That is, it is precisely because people expect to see more switching among signals than they actually will, they mistake true *i.i.d.* randomness for streakiness.

A model of how people develop a belief in non-existent hot hands could be developed building from the model in this paper. In this paper I have assumed Freddy believes firmly that the string of urns that lead to local representativeness all have the same rate of signal generation. Suppose,

[45] Moreover, the experiments I am familiar with in which subjects seem to be predicting over-long streaks suffer the problem discussed in Section 2. Because they do not control subjects' prior beliefs about the underlying signal probability, predictions by subjects of repetition of signals may simply be inference about the generic likelihood of those signals.

[46] But for many questions of economic significance, including those emphasized in this paper, the "long-wave" positive autocorrelation of the hot-hand bias does not undermine the predictions of the gambler's fallacy. For instance, it may not matter much whether somebody who observes an analyst who has recently done well over-infers her intrinsic talent, or merely infers that she is on a hot or cold streak; so long as she think the streak is likely to continue fora while, she may treat hot or cold average analysts as if they were good or bad analysts.

however, that he thinks it is possible that the underlying rate of the urn might stochastically change, but do so more rarely than the urn is renewed. This belief in "long-wave" positive autocorrelation may be quite reasonable. Explanations abound, for instance, for why a basketball player might get hot or cold. Maybe when a player is shooting well, he becomes confident, rather than tentative, in taking shots, and this improves his game. When he is doing poorly, he is nervous, and forces bad shots, *etc*.[47] There surely *is* a hot hand in some sports phenomena. But the law of small numbers provides a natural intuition for why somebody who begins with the belief that a stochastic process might or might not involve long-wave positive autocorrelation will over time come to believe in such autocorrelation even when none exists. Faced with actual independence of signals, people develop a bogus belief in a form of positive autocorrelation in signal generation that to them explains the missing negative autocorrelation they expected due to the gambler's fallacy.[48] Such a model would predict the gradual development of belief in the hot hand in those settings—such as basketball shooting—where people find it *a priori* plausible, but continue believing solely in the gambler's fallacy in contexts where they do not find streakiness plausible.[49]

[47] This suggests an intriguing possibility if confidence and underconfidence do influence performance: There may be a hot hand if and only if people *believe* in the hot hand. If a person accepts his performance is *i.i.d.*, he'll never gain nor lose confidence even in the face of streaks, and so his performance will be *i.i.d.* But if he believes in the hot hand, his fluctuations in confidence will extend what would otherwise be random streaks.

[48] For this type of model to work it would be crucial that Freddy not believe it possible that the urns' rate changes as often as the urns are renewed, since then the countervailing positive and negative autocorrelation he comes to believe in would generate a *de facto* *i.i.d.* signal process.

[49] The findings in Edwards (1961), discussed in Section 2, lend support to this interpretation of the hot-hand fallacy. He finds that in the first two hundred trials of a flip of a coin people's prediction correspond to the gambler's fallacy (as he defines it), but that in the last 800 trials their error switches to the hot-hand fallacy. This could be because participants observe less negative autocorrelation than (given their belief in the law of small numbers) they had predicted, and hence over time came to believe they were observing more long streaks of signals than they really were.

Appendix: Proofs

Proof of Lemma 1: Algebra.

Proof of Lemma 2: Algebra.

Proof of Lemma 3: Bayes Rule.

Proof of Lemma 4: Bayes Rule.

Proof of Lemma 5: $\pi_1^N(s_2 = a|s_1 = a) = \sum_{\theta\in\Theta} \pi_1^N(\theta|s_1 = a) \cdot \pi_1^N(s_2 = a|\theta, s_1 = a) = \sum_{\theta\in\Theta} \pi_1^N(\theta|s_1 = a) \cdot \frac{\theta N-1}{N-1}$. Since Lemma 4 says that $\pi_1^N(\theta|s_1 = a)$ is independent of N and for all $\theta \in \Theta$, $\frac{\theta N}{N-1}$ is increasing in N, the Lemma is established. Mutatis mutandi for $\pi_1^N(s_2 = b|s_1 = a)$.

Proof of Proposition 1: For even t, $\frac{\pi_t^N(\theta|h_t^a)}{\pi_t^N(\widehat{\theta}|h_t^a)} = \frac{\left(\theta\frac{\theta N-1}{N-1}\right)^{\frac{t}{2}}}{\left(\widehat{\theta}\frac{\widehat{\theta}N-1}{N-1}\right)^{\frac{t}{2}}} = \left(\frac{\theta(\theta N-1)}{\widehat{\theta}(\widehat{\theta}N-1)}\right)^{\frac{t}{2}}$, which is decreasing in N iff $\frac{\theta(\theta N-1)}{\widehat{\theta}(\widehat{\theta}N-1)}$ is decreasing in N. This is true iff $\theta > \widehat{\theta}$. The argument easily extends for odd t, and for $\frac{\pi_t^N(\widehat{\theta}|h_t^b)}{\pi_t^N(\theta|h_t^b)}$.

Proof of Corollary 1: Trivial.

Proof of Proposition 2: All $h_t \in H_t^{\frac{1}{2}}$ correspond to some sequence of ab $(= ba)$ pairs and an equal number of aa and bb pairs. Given this and Lemma 2, it is sufficient to show that both $\frac{\pi^N(ab|\theta)}{\pi^N(ab|\widehat{\theta})}$ and $\frac{\pi^N(aa,bb|\theta)}{\pi^N(aa,bb|\widehat{\theta})}$ are nondecreasing in N. The first expression equals $\frac{\frac{\theta(\theta(1-\theta)N)}{N-1}}{\frac{\widehat{\theta}(1-\widehat{\theta})N}{N-1}} = \frac{\theta(1-\theta)}{\widehat{\theta}(1-\widehat{\theta})}$ which doesn't depend on N. The second expression equals $\frac{\frac{\theta(\theta N-1)}{N-1}\frac{(1-\theta)((1-\theta)N-1)}{N-1}}{\frac{\widehat{\theta}(\widehat{\theta}N-1)}{N-1}\frac{(1-\widehat{\theta})((1-\widehat{\theta})N-1)}{N-1}}$. This is equal to $\frac{\theta(1-\theta)}{\widehat{\theta}(1-\widehat{\theta})}\frac{\theta(1-\theta)N^2-N+1}{\widehat{\theta}(1-\widehat{\theta})N^2-N+1}$, which can be shown to be increasing in N iff $\left(\widehat{\theta}(1-\widehat{\theta}) - \theta(1-\theta)\right)(N-2) > 0$. If $N > 2$ — which it must be to make the model coherent — then this inequality holds iff $\widehat{\theta}(1-\widehat{\theta}) > \theta(1-\theta)$. This is always true when either $\theta > \widehat{\theta} > \frac{1}{2}$ or $\theta < \widehat{\theta} < \frac{1}{2}$.

Proof of Proposition 3: It is trivial to show that for all π: $E_2^N(aa) > E_2^N(ab) > E_2^N(bb)$ for all N, $E_2^N(ab) = \frac{1}{2}$ for all N, $E_2^N(aa)$ is decreasing in N, and $E_2^N(bb)$ is increasing in N. If π is symmetric, then for all N $E_2^N(\pi) = \frac{1}{2}$. Since $\pi(aa) = \pi(bb)$, and these values are independent of N, this establishes the proposition. Irrespective of N, the third signal will be used appropriately for updating beliefs, so it will not affect the result. (Once four or more signals occur, it is possible to get a mixture of both aa and bb odd-even pairs.)

Proof of Lemma 6: By the law of large numbers, after an infinite sequence of signals, Freddy will get proportions very close to $(\theta^*)^2$ aa odd-even pairs, $2(1-\theta^*)\theta^*$ ab pairs, and $(1-\theta^*)^2$ bb pairs. Hence, Freddy's beliefs will converge to putting full weight on the beliefs $\widehat{\theta}$ that maximize the likelihood of observing such proportions. Applying Lemma 2, therefore, we get the maximization stated in this Lemma.

Proof of Proposition 4: We must show that the $\widehat{\theta}$ that maximizes the likelihood function in Lemma 6 has the specified properties for all θ^* and N. Taking the derivative of the logarithm of the likelihood function in Lemma 6 with respect to θ we get:

$$\frac{\partial L}{\partial \theta} = \frac{(\theta^*)^2}{\theta} + \frac{(\theta^*)^2 N}{\theta N - 1} + \frac{2(1-\theta^*)\theta^*}{\theta} - \frac{2(1-\theta^*)\theta^*}{1-\theta} - \frac{(1-\theta^*)^2}{1-\theta} - \frac{(1-\theta^*)^2 N}{(1-\theta)N - 1}$$

It can be shown that $\frac{\partial^2 L}{\partial \theta^2} < 0$ for all θ^*, N, and θ, by observing that all terms are strictly decreasing in θ. By observing further that the likelihood function in Lemma 6 is zero in the limit as either $\theta \searrow 0$ or $\theta \nearrow 1$, we know that the unique $\widehat{\theta}$ is an interior solution satisfying the first-order condition $\frac{\partial L}{\partial \theta} = 0$. Moreover, because $\frac{\partial L}{\partial \theta}$ is strictly decreasing, if we can show that, for $\theta^* > \frac{1}{2}$, $\frac{\partial L}{\partial \theta}\big|_{\theta=\theta^*} < 0$, and $\frac{\partial L}{\partial \theta}\Big|_{\theta=\frac{1}{2}} > 0$, we will have established that $\widehat{\theta} \in [\frac{1}{2}, \theta^*]$ for even N and $\theta^* > \frac{1}{2}$.

Algebra shows that $\frac{\partial L}{\partial \theta}\big|_{\theta=\theta^*} = 1 - 2\theta^* + \frac{(\theta^*)^2 N}{\theta^* N - 1} - \frac{(1-\theta^*)^2 N}{(1-\theta^*)N - 1}$, which can be shown to be negative when $\theta^* > \frac{1}{2}$. Algebra shows that $\frac{\partial L}{\partial \theta}\Big|_{\theta=\frac{1}{2}} = 2(\theta^*)^2 - 2(1-\theta^*)^2 + (\theta^{*2} - (1-\theta^*)^2)\frac{N}{\frac{1}{2}N - 1}$, which is positive for all N, when $\theta^* > \frac{1}{2}$.

To show that $\widehat{\theta} \in [\theta^*, \frac{1}{2}]$ when $\theta^* < \frac{1}{2}$ is the same. And it is easy to establish that when $\theta^* = \frac{1}{2}$, $\frac{\partial L}{\partial \theta} = 0$ is solved by $\widehat{\theta} = \frac{1}{2}$.

Proof of Proposition 5: Since π is symmetric, $\frac{\pi_t^N(\theta|h_t)}{\pi_t^N(1-\theta|h_t)} = \frac{\pi^N(h_t|\theta)}{\pi^N(h_t|1-\theta)}$. If h_t involves a total of r aa pairs, s bb pairs, and t ab pairs, then $\frac{\pi^N(h_t|\theta)}{\pi^N(h_t|1-\theta)} = \frac{(\theta \cdot \frac{\theta N-1}{N-1})^r ((1-\theta)\frac{(1-\theta)N-1}{N-1})^s (\theta \frac{(1-\theta)N}{N-1})^t}{((1-\theta)\frac{(1-\theta)N-1}{N-1})^r (\theta \cdot \frac{\theta N-1}{N-1})^s ((1-\theta)\frac{\theta N}{N-1})^t} = \left[\frac{\theta(\theta N-1)}{(1-\theta)((1-\theta)N-1)}\right]^r \left[\frac{(1-\theta)((1-\theta)N-1)}{\theta(\theta N-1)}\right]^s = \left[\frac{\theta(\theta N-1)}{(1-\theta)((1-\theta)N-1)}\right]^{r-s} = \left(\frac{\theta}{1-\theta}\right)^{r-s} \left(\frac{\theta N-1}{(1-\theta)N-1}\right)^{r-s}$. We know that if $x_a > x_b$ then $r > s$. Hence, the result is established by the fact that $\frac{\theta N-1}{(1-\theta)N-1}$ is decreasing in N when $\theta > \frac{1}{2}$.

Proof of Proposition 6: A law of large numbers.

Proof of Proposition 7: Let $d(\pi, N)$ be an N-Freddy's beliefs about the proportion each of aa and bb pairs that the symmetric distribution π will generate. First notice that if π' is a mean-preserving spread of π, then $d(\pi', N) > d(\pi, N)$ for all N (including $N = \infty$). Second, notice that for all

π, $d(\pi, N)$ is increasing in N. Notice further yet that if $\widehat{\pi} = (1-k)\pi + k\pi'$ for $k \in (0,1)$, then $d(\widehat{\pi}, N) = (1-k)d(\pi, N) + kd(\pi', N)$.

Now choose π generating real distribution $(s, 1-2s, s)$ of aa, ab, and bb pairs. Since $S = d(\pi, \infty)$, $d(\pi, N) < S$ for all $N < \infty$. Then by choosing any mean-preserving spread, π', of π such that $d(\pi', N) > S$ (π' such that $\pi'(\theta = 1) = \pi'(\theta = 0) = \frac{1}{2}$ always works), we can choose the $k \in (0,1)$ such that $\widehat{\pi} = k\pi' + (1-k)\pi$ generates $S = kd(\pi', N) + (1-k)d(\pi, N)$ aa pairs. Since $\widehat{\pi}$ is a mean-preserving spread of π whenever π' is, this proves the first part of the Proposition. Since any $\widetilde{\pi}$ where π is a mean-preserving spread of $\widetilde{\pi}$ yields $d(\widetilde{\pi}, N) < d(\pi, N)$, which can be used to show the second part of the Proposition.

References

Abraham, I. L. and S. Schultz (1984). "The 'law of small numbers': An unexpected and incidental replication." *Journal of Psychology* 117(2): 183-188.

Adams, R. M. (1992). "The 'hot hand' revisited: Successful basketball shooting as a function of intershot interval." *Perceptual & Motor Skills* 74(3, Pt 1): 934.

Barberis, N., M. Huang, and T. Santos (1999), "Prospect Theory and Asset Prices," mimeo, University of Chicago, November.

Barberis, N., A. Shleifer and R. W. Vishny (1998). "A Model of Investor Sentiment." *Journal of Financial Economics* September 49(3).

Bar-Hillel, M. and W. A. Wagenaar (1991). "Perceptions of randomness." *Advances of Applied Mathematics* 12: 428-454.

Beach, L. R. and R. G. Swensson (1967). "Instructions about Randomness and Run Dependency in Two-Choice Learning." *Journal of Experimental Psychology* 75(2): 279-282.

Berk, J., R. Green, and V. Naik (1999), "Optimal Investment, Growth Options, and Security Returns," *Journal of Finance* 54(5), October: 1553-1607.

Braun, P. A. and I. Yaniv (1992). "A case study of expert judgment: Economists' probabilities versus base-rate model forecasts." *Journal of Behavioral Decision Making* 5(3): 217-231.

Brown, W. and R. D. Sauer (1993). "Does the Basketball Market Believe in the Hot Hand? Comment." *American Economic Review* 83(5): 1377-1386.

Budescu, D. V. (1987). "A Markov model for generation of random binary sequences." *Journal of Experimental Psychology: Human Perception & Performance* 13(1): 25-39.

Budescu, D. V. and A. Rapoport (1994). "Subjective randomization in one- and two-person games."

Journal of Behavioral Decision Making 7(4): 261-278.

Camerer, C. F. (1987). "Do Biases in Probability Judgements Matter in Markets? Experimental Evidence." *American Economic Review* 77(5): 981-997.

Camerer, C. F. (1989). "Does the basketball Market Believe in the 'Hot Hand'?" *American Economic Review* 79(5): 1257-1261.

Clotfelter, C. and P. Cook. (1993). "The 'Gambler's Fallacy' in Lottery Play." *Management Science* 39(12): 1521-1525.

Colle, H. A., R. M. Rose and H. A. Taylor (1974). "Absence of the gambler's fallacy in psychophysical settings." *Perception & Psychophysics* 15(1): 31-36.

Daniel, K., D. Hirshleifer and A. Subrahmanyam (1998), "Investor Psychology and Security market Under- and Over-reactions," *Journal of Finance* 53: 1839-85.

De Bondt, W. F. M. and R. H. Thaler (1987). "Further Evidence on Investor Overreaction and Stock Market Seasonality." *Journal of Finance* 42(3): 557-581.

De Bondt, W. F. M. and R. H. Thaler (1990). "Do Security Analysts Overreact?" *American Economic Review* 80(2): 52-57.

Derks, P. L. and M. I. Paclisanu (1967). "Simple Strategies in Binary Prediction by Children and Adults." Journal of Experimental Psychology 73(2): 278-285.

Edwards, W. (1961). "Probability learning in 1000 trials." *Journal of Experimental Psychology* 62(4): 385-394.

Erev, I., T. S. Wallsten and D. V. Budescu (1994). "Simultaneous over- and underconfidence: The role of error in judgment processes." *Psychological Review* 101(3): 519-527.

Fisher, Ian. "How Not to Get Killed on Deadline." *NY Times Magazine (* Jan. 31, 1999): pp.36-37.

Gilovich, T., R. Vallone and A. Tversky (1985). "The Hot Hand in Basketball: On the Misperception of Random Sequences." *Cognitive Psychology* 17: 295-314.

Grether, D. M. (1980). "Bayes' rule as a descriptive model: The representativeness heuristic." *Quarterly Journal of Economics* 95: 537-57.

Grether, D. M. (1992). "Testing Bayes rule and the representativeness heuristic." *Journal of Economic Behavior and Organization* 17: 31-57.

Holtgraves, T. and J. Skeel (1992). "Cognitive biases in playing the lottery: Estimating the odds and choosing the numbers." *Journal of Applied Social Psychology* 22(12): 934-952.

Hong, H., T. Lim, and J. Stein (1998), "Bad News Travels Slowly: Size, Analyst Coverage and the Profitability of Momentum Strategies," mimeo, MIT Sloan School, April.

Hong, H. and J. Stein (1999), "A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets," *Journal of Finance* , forthcoming.

Jarvik, M. E. (1951). "Probability learning and a negative recency effect in the serial anticipation of alternative symbols." *Journal of Experimental Psychology* 41: 291-297.

Kahneman, D. and A. Tversky (1973), "On The Psychology of Prediction." *Psychological Review* 80:237-251.

Kareev, Y. (1992). "Not that bad after all: Generation of random sequences." *Journal of Experimental Psychology: Human Perception & Performance* 18(4): 1189-1194.

Kareev, Y. (1995). "Positive bias in the perception of covariation." *Psychological Review* 102(3): 490-502.

Keren, G. B. and W. A. Wagenaar (1985). "On the psychology of playing blackjack: Normative and descriptive considerations with implications for decision theory." *Journal of Experimental Psychology: General* 114(2): 133-158.

Keren, G. and C. Lewis (1994). "The two fallacies of gamblers: Type I and Type II." *Organizational Behavior & Human Decision Processes* 60(1): 75-89.

Kunda, Z. and R. E. Nisbett (1986). "Prediction and the partial understanding of the law of large numbers." *Journal of Experimental Social Psychology* 22(4): 339-354.

Ladouceur, R., M. Mayrand, A. Gaboury and M. St-Onge (1987). "Comportements des acheteurs de billets de loteries passives et pseudo-actives: Etude comparative. (Behavior of purchasers of passive and pseudo-active lottery tickets: A comparative study.)." *Canadian Journal of Behavioural Science* 19(3): 266-274.

Ladouceur, R., C. Paquet, N. Lachance and D. Dube (1996). "Examen d'une erreur fondamentale dans la perception du hasard. (Study of a basic error in the perception of chance.)." *International Journal of Psychology* 31(2): 93-99.

Ladouceur, R., C. Paquet and D. Dube (1996). "Erroneous perceptions in generating sequences of random events." *Journal of Applied Social Psychology* 26(24): 2157-2166.

Lepley, W. M. (1963). "'The maturity of the chances': A gambler's fallacy." *Journal of Psychology* 56(1): 69-72.

Lindman, H. and W. Edwards (1961). "Supplementary report: Unlearning the gambler's fallacy." J*ournal of Experimental Psychology* 62(6): 630.

Lopes, L. L. (1982). "Doing the impossible: A note on induction and the experience of randomness." *Journal of Experimental Psychology: Learning, Memory, & Cognition* 8(6): 626-636.

Lopes, L. L. and G. C. Oden (1987). "Distinguishing between random and nonrandom events." *Journal of Experimental Psychology: Learning, Memory, & Cognition* 13(3): 392-400.

Metzger, M. A. (1985). "Biases in betting: An application of laboratory findings." *Psychological*

Reports 56(3): 883-888.

Mullainathan, Sendhil (1997), "A Memory Based Model of Bounded Rationality," mimeao, Harvard, November.

O'Neill, B. (1987), "Nonmetric Test of the Minimax Theory of Two-Person Zero-sum games," *Proceedings of the National Academy of Sciences* 84: 2106-2109.

Oritz, Joge L. (1999), "A Sudden Reversal for Coles." *The San Francisco Examiner*, February 24: D1-D5.

Paulos, J. A. (1988), *Innumeracy: Mathematical Illiteracy and its Concequences*, Hill and Wang: New York.

Rabin, M. (1998), "Psychology and Economics," *Journal of Economic Literature* XXXVI, March, 11-46.

Rabin, M. and J. Schrag (1999), "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics* February 114(1): 37-82.

Rapoport, A. and D. V. Budescu (1992). "Generation of random series in two-person strictly competitive games." *Journal of Experimental Psychology: General* 121(3): 352-363.

Rapoport, A. and D. V. Budescu (1997). "Randomization in Individual Choice Behavior," *Psychology: Review* 104(3): 603-617.

Ross, B. M. and N. Levy (1958). "Patterned predictions of chance events by children and adults." *Psychological Reports* 4: 87-124.

Ross, B. M. and J. F. DeGroot (1982). "How adolescents combine probabilities." *Journal of Psychology* 110(1): 75-90.

Savoie, D. and R. Ladouceur (1995). "Evaluation et modification de conceptions erronees au sujet des loteries. (Evaluation and modification of misconceptions concerning lotteries.)." *Canadian Journal of Behavioural Science* 27(2): 199-213.

Shafir, E. B., E. E. Smith and D. N. Osherson (1990). "Typicality and Reasoning Fallacies." *Memory & Cognition* 18(3): 229-239.

Terrell, D. (1994), "A Test of the Gambler's Fallacy – Evidence from Para-Mutuel Games," *Journal of Risk and Uncertainty* 8(3): 309-317.

Tversky, A. and D. Kahneman (1971). "Belief in the law of small numbers." *Psychological Bulletin* 76(2): 105-110.

Tversky, A. and D. Kahneman (1973). "Judgment under uncertainty: Heuristics and biases", Eugene, Ore.: Oregon Research Inst., Vol. 13(1), 1973.

Tversky, A. and D. Kahneman (1982). "Judgments of and by Representativeness", in *Judgment under Uncertainty: Heuristics and Biases*. D. Kahneman, P. Slovic and A. Tversky, Cambridge

University Press: 84-98, Chap. 6.

Tversky, A. and D. Kahneman (1983). "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment." *Psychological Review* 90(4): 293-315.

Tversky, A. and T. Gilovich (1989). "The Cold Facts about the 'Hot Hand' in Basketball." *Chance* 2(1): 16-21.

Tversky, A. and T. Gilovich (1989). "The Hot Hand: Statistical Reality or Cognitive Illusion?" *Chance* 2(4): 31-34.

Wagenaar, W. A. (1972). "Generation of random sequences by human subjects: A critical survey of literature." *Psychological Bulletin* 77(1): 65-72.

Walker, M. and Wooders, J. (1999), "Minimax Play at Wimbledon," mimeo, University of Arizona.

Wallsten, T. S. (1983). "The Theoretical Status of Judgment Heuristics", in *Decision Making Under Uncertainty*, R. W. Scholz ed., Elsevier Science Publishers: 21-37.

Winefield, A. H. (1966). "Negative recency and event-dependence." *Quarterly Journal of Experimental Psychology* 18(1): 47-54.