

A Model of Non-Belief in the Law of Large Numbers

Daniel J. Benjamin

Cornell University, University of Southern California, and NBER

Matthew Rabin

Collin Raymond*

Harvard University

University of Oxford

December 11, 2014

Abstract

People believe that, even in very large samples, proportions of binary signals might depart significantly from the population mean. We model this “non-belief in the Law of Large Numbers” by assuming that a person believes that proportions in any given sample might be determined by a rate different than the true rate. In prediction, a non-believer expects the distribution of signals will have fat tails. In inference, a non-believer remains uncertain and influenced by priors even after observing an arbitrarily large sample. We explore implications for beliefs and behavior in a variety of economic settings.

JEL Classification: B49, D03, D14, D83, G11

*We thank Nick Barberis, Herman Chernoff, Botond Köszegi, Don Moore, Ted O’Donoghue, Marco Ottaviani, Larry Phillips, Daniel Read, Larry Samuelson, Jeremy Stein, three anonymous referees, and seminar participants at UC Berkeley, Cornell University, Koc University, the LSE-WZB Conference on Behavioral IO, the Yale Behavioral Science Conference and the Behavioral Economics Annual Meeting for helpful comments. We are grateful to Samantha Cunningham, Ahmed Jaber, Greg Muenzen, Desmond Ong, Nathaniel Schorr, Dennis Shiraev, Josh Tasoff, Mike Urbancic, and Xiaoyu Xu for research assistance. For financial support, Benjamin thanks NIH/NIA grant T32-AG00186 to the NBER, and Raymond thanks the University of Michigan School of Information’s Socio-Technical Infrastructure for Electronic Transactions Multidisciplinary Doctoral Fellowship funded by NSF IGERT grant #0654014. E-mail: daniel.benjamin@gmail.com, rabin@econ.berkeley.edu, collin.raymond@economics.ox.ac.uk.

1 Introduction

Psychological research has identified systematic biases in people’s beliefs about the relationship between sample proportions and the population from which they are drawn. Following Tversky and Kahneman (1971), Rabin (2002) and Rabin and Vayanos (2010) model the notion that people believe in “the Law of Small Numbers (LSN),” exaggerating how likely it is that small samples will reflect the underlying population. LSN predicts that when inferring the population proportion that generated a given sample, people will be overconfident. Yet experimental evidence on such inference problems clearly indicates that when the sample contains more than a few observations, people’s inferences are typically *under*confident. Moreover, these inferences appear to be driven by people’s beliefs about the distribution of sample proportions that are too diffuse for medium-sized and large samples (see Appendix D for our review of the evidence on inference and sampling-distribution beliefs). In this paper, we develop a formal model of this bias in people’s sampling-distribution beliefs and show that it has a range of economic consequences, including causing under-inference from large samples, a lack of demand for large samples, and, as a function of the environment, either too little or too much risk-taking.

This bias co-exists with LSN; we discuss the relationship between the two in Section 6, and provide a model combining both biases in Appendix C. We call the bias “non-belief in the Law of Large Numbers,” abbreviated NBLLN¹, because we believe its source is the absence of sample size as a factor in people’s intuitions about sampling distributions. Our view is motivated by Kahneman and Tversky’s (1972) evidence and interpretation. They find that experimental subjects seem to think sample proportions reflect a “universal sampling distribution,” virtually neglecting sample size. For instance, independent of whether a fair coin is flipped 10, 100, or 1,000 times, the median subject thinks that there is about $\frac{1}{5}$ chance of getting between 45% and 55% heads, and about $\frac{1}{20}$ chance of between 75% and 85%. These beliefs are close to the right probabilities of $\frac{1}{4}$ and $\frac{1}{25}$ for the sample size of 10, but wildly miss the mark for the sample size of 1,000, where the sample is almost surely between 45% and 55% heads.

In Section 2, we develop our model of NBLLN in a simple setting, where a person is trying to predict the distribution of—or make an inference from—a sample of fixed size. Throughout, we refer to our modeled non-believer in the Law of Large Numbers as *Barney*, and compare his beliefs and behavior to a purely Bayesian information processor, *Tommy*.² Tommy knows that the likelihood of different sample distributions of an i.i.d. coin biased θ towards heads will be the

¹NBLLN is pronounced letter by letter, said with the same emphasis and rhythm as “Ahmadinejad.”

²“Tommy” is the conventional designation in the quasi-Bayesian literature to refer somebody who updates according to the dictums of the Reverend Thomas Bayes.

“ θ -binomial distribution.” But Barney, as we model him, believes that large-sample proportions will be distributed according to a “ β -binomial distribution,” for some $\beta \in [0, 1]$ that itself is drawn from a distribution with mean θ . This model directly implies NLLN: whereas Tommy knows that large samples will have proportions of heads very close to θ , Barney feels that the proportions in any given sample, no matter how large, might not be θ . Although the model largely reflects the “universal sampling distribution” intuition from Kahneman & Tversky (1972), it also embeds *some* sensitivity to sample sizes, consistent with other evidence, such as Study 1 of Griffin and Tversky (1992).³ Other models would share the basic features of NLLN that we exploit in this paper; we discuss in Section 6 the merits and drawbacks of our particular formulation.

After defining the model, Section 2 describes some of its basic features for Barney’s predictions about the likelihood of occurrence of different samples and his inferences from samples that have occurred. While Barney makes the same predictions as Tommy about sample sizes of 1, his beliefs about sample proportions are a mean-preserving spread of Tommy’s for samples of two or more signals. In situations of inference, we show that applying Bayesian updating based on his wrong beliefs about the likelihood of different sample realizations, NLLN implies under-inference from large samples: Barney’s posterior ratio on different hypotheses is less extreme than Tommy’s. Importantly, for any proportion of signals—including the proportion corresponding to the true state—Barney fails to become fully confident even after infinite data. Consequently, Barney’s priors remain influential even after he has observed a large sample of evidence.

Sections 3 and 4 illustrate some of the basic economic implications of NLLN. Section 3 examines willingness to pay for information. If Barney and Tommy can choose what sample size of signals to acquire, then Barney (because he expects to learn less from any fixed sample size) may choose a larger sample, and can therefore end up being *more* certain about the state of the world. But because Barney thinks that his inference would be limited even from an infinite sample, he unambiguously has a lower willingness to pay for a *large* sample of data than Tommy. This lack of demand for statistical data is a central implication of NLLN. We believe it contributes to explaining why people often rely instead on sources of information that provide only a small number of signals, such as anecdotes from strangers, stories from one’s immediate social network,

³Even though we are not aware of any evidence on people’s beliefs regarding sample sizes larger than 1,000, our model imposes—consistent with Kahneman and Tversky’s (1972) interpretation—that Barney puts positive probability on sample proportions other than θ even in an infinite sample. We conjecture that people’s beliefs regarding much larger samples do indeed resemble the same “universal sampling distribution” as for a sample size of 1,000. Nonetheless, we emphasize that even if the literal implications of our model for infinite sample sizes were not true, our large-sample limit results would still have substantial bite for the applications where we invoke them. This is because, as per the frequent reliance on large-sample limit results in econometrics, the Law of Large Numbers typically provides a good approximation for Tommy’s beliefs in the finite, moderately-sized samples that are realistic for those applications.

and limited personal experience. Indeed, direct real-world evidence of the propensity to over-infer from limited evidence might be more ubiquitous than evidence of under-inference precisely because people rarely choose to obtain a large sample.

Section 4 next explores how Barney’s mistaken beliefs about the likelihood of different samples matters for choice under risk. For example, Barney believes that the risk associated with a large number of independent gambles is greater than it actually is. This magnifies aversion to repeated risks, whether that risk aversion is due to diminishing marginal utility of wealth or (more relevantly) reference-dependent risk attitudes. Because he does not realize that the chance of aggregate losses becomes negligible, Barney may refuse to accept even infinite repetitions of a small, better-than-fair gamble. Even assuming a plausible model of risk preferences, such as loss aversion, that generates the intrinsic aversion to small risks, a person who is focusing on whether to accept a large number of independent risks would not exhibit the observed behavior if he believed in LLN. Benartzi and Thaler (1999), in fact, demonstrate clearly the role of both loss aversion and what we are calling NBLLN. However, in other contexts, where payoffs depend on extreme outcomes, Barney’s mistaken sampling beliefs could instead make him appear *less* risk averse than Tommy, such as playing a lottery in which whether he wins a prize depends on correctly guessing all of several numbers that will be randomly drawn.

Sections 3 and 4 assume that a person is analyzing all his information as a single sample of fixed size. In many settings, it seems more likely people will instead analyze the evidence dynamically as information arrives. In Section 5, we confront a conceptual challenge intrinsic to the very nature of NBLLN that arises in such settings: because Barney under-infers more for larger samples than smaller ones, he will infer differently if he lumps observations together versus separately. We discuss possible assumptions regarding how Barney “retrospectively groups” signals—how he interprets evidence once he sees it—and “prospectively groups” signals—how he predicts ahead of time he will interpret evidence he might observe in the future. We formalize these assumptions in Appendix A, and in Appendix B, we use the single-sample and multiple-sample models to draw out consequences of NBLLN in additional applications that cover some of the major areas in the economics of uncertainty, such as valuation of risky prospects, information acquisition, and optimal stopping.

In Section 6, we discuss why we think our model of NBLLN is more compelling than alternative modeling approaches and explanations for the phenomena we are trying to explain with NBLLN, including both fully rational and not fully rational alternatives. We also discuss the drawbacks of our particular formulation. Perhaps most importantly, while the model makes predictions about Barney’s sampling-distribution beliefs (e.g., 1 head out of 2 flips), it cannot be used to make

predictions regarding Barney’s beliefs about the likelihood of particular sequences (e.g., a head followed by a tail). In addition, our model ignores other important departures from Bayesian inference, such as belief in the Law of Small Numbers and base-rate neglect. To begin to understand how these various biases relate to each other, in Appendix C we present a (complicated) formal model embedding some of these other errors along with NLLN.

Section 7 concludes. We discuss how NLLN acts as an “enabling bias” for distinct psychological biases, such as “vividness bias” and optimism about one’s own abilities or preferences, that would otherwise be rendered irrelevant by the Law of Large Numbers. We also suggest directions for extending the model to non-i.i.d. and non-binomial signals.

In Appendix D, we review the extensive experimental literature on inference and the smaller body of evidence on sampling-distribution beliefs that motivate this paper. In light of this evidence—much of it from the 1960s—we do not fully understand why the bias we call NLLN has not been widely appreciated by judgment researchers or behavioral economists. We suspect it is largely because findings of under-inference have been associated with an interpretation called “conservatism” (e.g., Edwards, 1968)—namely, that people tend not to update their beliefs as strongly as Bayesian updating dictates—that does not mesh comfortably with other biases that often imply that people infer more strongly than Bayesian. Our interpretation of the under-inference evidence as NLLN differs from the conservatism interpretation; rather than being an intrinsic aversion to strong updating, NLLN is a bias in intuitive understanding of sampling distributions. In Appendix C, we show that NLLN can co-exist with base-rate neglect and other biases that can generate over-inference in some settings. Appendix E contains proofs.

2 The Single-Sample Model

Throughout the paper, we study a stylized setting where an agent observes a set of binary signals, each of which takes on a value of either a or b . Given a *rate* $\theta \in \Theta \equiv (0, 1)$, signals are generated by a binomial (i.i.d.) process where the probability of an a -signal is equal to θ . Signals arrive in *clumps* of size N . We denote the set of possible ordered sets of signals of size $N \in \{1, 2, \dots\}$ by $S_N \equiv \{a, b\}^N$, and we denote an arbitrary clump (of size N) by $s \in S_N$.⁴ Let A_s denote the total number of a ’s that occur in the clump $s \in S_N$, so that $\frac{A_s}{N}$ is the proportion of a ’s that occur in a clump of N signals. For a real number x , we will use the standard notations “ $\lceil x \rceil$ ” to signify the smallest integer that is weakly greater than or equal to x and “ $\lfloor x \rfloor$ ” to signify the largest integer

⁴Note that we forego the conventional strategy of providing notation for a generic signal, indexed by its number. It is less useful here because (within a clump) what matters to Barney is just the number of a signals, not their order. In Appendix A, when we formalize the multiple-sample model, we use t to index the clumps of signals.

that is weakly less than or equal to x . For any random variable y that takes as possible values the elements of set Y , let beliefs by Tommy (who believes in the Law of Large Numbers) be denoted by cumulative distribution function $F_Y(\cdot)$, implying probability density function $f_Y(\cdot)$, expectation $E_Y(\cdot)$, and variance $Var_Y(\cdot)$. Let corresponding beliefs by Barney (the non-believer in the LLN) be denoted by $F_Y^\psi(\cdot)$, $f_Y^\psi(\cdot)$, $E_Y^\psi(\cdot)$, and $Var_Y^\psi(\cdot)$, where ψ signifies Barney’s beliefs (and is a parameter for the degree of NBLLN in the parameterized special case of the model described in equation (3) below).

In this section, we develop our model of Barney for the case where he is considering a single clump of N signals. This case corresponds to most of the experimental evidence about NBLLN, which has been collected in settings where subjects were presented with a single, fixed sample of signals or outcomes, in which subjects presumably process all the information together. This special case also allows us to lay bare the essential features of how our model captures NBLLN. The domain of the model—and the focus of all of our applications—is Barney’s beliefs about the total number of a ’s that occur in a sample of N signals; as we discuss in Section 6, the model cannot be applied to beliefs about specific sequences of signals. The single-clump model of Barney is all we need for the applications we explore in the main text. In Section 5 and Appendix A, we discuss the conceptual challenges that arise when generalizing the model to multiple clumps, and in Appendix B, we use the multiple-clump model in the applications we explore.

According to the Law of Large Numbers, with probability 1 in the limit as the sample size gets large, the mean of a random sample equals the rate: For any interval $(\alpha_1, \alpha_2) \subseteq [0, 1]$,

$$\lim_{N \rightarrow \infty} \sum_{x=\lceil \alpha_1 N \rceil}^{\lceil \alpha_2 N \rceil} f_{S_N | \Theta}(A_s = x | \theta) = \begin{cases} 1 & \text{if } \theta \in (\alpha_1, \alpha_2) \\ 0 & \text{otherwise} \end{cases}.$$

How might we capture the possibility that Barney believes (say, as per an example in Kahneman and Tversky, 1972) that it is reasonably likely that at least 600 of 1000 births at a hospital in a given year are boys, even though he knows that boys are born at a rate of 50%? The essence of our model is to assume that Barney believes samples are generated as if a rate of θ , here 50%, means that the rate is θ on average, but might be higher or lower for any given sample; intuitively, there is common noise added to all signals in a given sample. For a given true rate θ , we model Barney as believing that for the sample he is considering: first, a “subjective rate” $\beta \in [0, 1]$ is drawn from a distribution centered at θ . Then the i.i.d. sample of 1000 babies is generated using rate β . The key implication is that if a given value of β were the actual rate, it would (by the Law of Large Numbers!) exactly determine the proportion of signals in the limit of a very large sample. Therefore, the probability density that Barney assigns to any proportion β of signals (say, 60% of babies are boys) in a large

sample is equal to the probability density that Barney assigns to the possibility that β equals that value. Although this modeling approach is “as if” Barney is unsure that the rate is θ —and indeed exactly our parametric version of the model described below is commonly used in statistics to capture unobserved heterogeneity—true parameter uncertainty is not at all our interpretation. Instead, consistent with the underlying NLLN psychology, we interpret it as Barney’s belief that even his certainty that the underlying rate is θ is not a guarantee that the proportion in very large samples will approximate θ .⁵

Formally, we assume that when Barney knows the rate is θ , he believes that signals are generated by a binomial (i.i.d.) process where the probability of an a -signal is equal to β . This $\beta \in [0, 1]$ is called the *subjective rate*, and it is drawn from a density $f_{\beta|\Theta}^\psi(\beta|\theta)$. We refer to $f_{\beta|\Theta}^\psi$ as Barney’s *subjective rate distribution* and assume that it has the following properties:

A1. For all β and θ , $F_{\beta|\Theta}^\psi(\beta|\theta)$ is absolutely continuous in β , and $f_{\beta|\Theta}^\psi(\beta|\theta)$ has full support on $(0, 1)$ and is point-wise continuous in θ . We will sometimes make the more restrictive assumption:

A1’. The same as A1, except $f_{\beta|\Theta}^\psi(\beta|\theta)$ has full support on $[0, 1]$.

A2. For all β and θ , $F_{\beta|\Theta}^\psi(\beta|1 - \theta) = 1 - F_{\beta|\Theta}^\psi(1 - \beta|\theta)$.

A3. For all θ and $\theta' > \theta$, $\frac{f_{\beta|\Theta}^\psi(\beta|\theta')}{f_{\beta|\Theta}^\psi(\beta|\theta)}$ is increasing in β .

A4. For all β and θ , $E_{\beta|\Theta}^\psi(\beta|\theta) = \theta$.

Assumption A1 is a mild technical assumption. A1 differs from A1’ in allowing the density functions for β to converge to zero at proportions 0 and 1. The available experimental evidence cannot distinguish between A1 and A1’ because people’s beliefs about the likelihood of the most extreme sample proportions have not been elicited. While A1’ generally allows us to draw sharper theoretical conclusions, A1 accommodates our parametric example of the beta distribution discussed below. Any results below that assume A1 also hold with A1’ a fortiori.

Assumption A2 states that Barney’s beliefs when the bias is $1 - \theta$ are symmetric to his beliefs when the bias is θ . This assumption is consistent with the available experimental evidence on

⁵By comparison, Acemoglu, Chernozukov, and Yildiz (2009) analyze a model that is formally similar to our single-sample model but is a model of parameter uncertainty. Similar to our Proposition 2 below, they show that a Bayesian agent with parameter uncertainty fails to learn the state θ (which determines the distribution of β) with certainty even after observing an infinite number of signals. However, such an agent does learn with certainty the signal-generating rate β , which governs the rate of future signals. In contrast, under our interpretation of our model, any inferences about β would be useless because a new β is drawn with every new sample. In the single-sample applications we discuss in the main text, this distinction is purely interpretational. In the dynamic applications we develop in Appendix B, the distinction will be integral to the workings and predictions of the model. To see why, consider a situation where Barney must predict further signals after observing a sample of 100 signals. In our model, Barney’s expected proportion of a -signals remains still θ , rather than being influenced by the first 100 signals as it would be in a model of parameter uncertainty.

people’s beliefs.

Assumptions A3 and A4 are substantive assumptions that do not follow easily from the psychology. Assumption A3 is a monotone-likelihood-ratio property: fixing any two rates, Barney believes that the likelihood of drawing any particular subjective rate given the high rate relative to the low rate is increasing in the subjective rate. It is easy to imagine specifications of $f_{\mathcal{B}|\Theta}^\psi(\beta|\theta)$ —especially in the spirit of the type of diminishing-sensitivity evidence discussed in Appendix D—that would violate A3. But A3 is in accord with the most directly relevant evidence, namely Griffin and Tversky’s (1992) Study 3, which examines a range of parameters of the sort that seems most likely for violating it.⁶ It holds for our main example of the beta distribution and more generally is useful for establishing some of our results. Especially because the range of samples for which it is potentially false are inherently very unlikely, we think it is probably not an important caveat to our results.

Assumption A4 means that we restrict attention to cases where the mean of Barney’s subjective rate distribution is equal the known objective rate θ . Although we rely on A4 extensively in the analysis, it is in fact violated in existing data. Below, when discussing Figure 1, we discuss how A4 is violated—in particular the mean of the subjective sampling distribution is not equal to θ —and in Appendix C, we propose a more comprehensive model that captures the psychology that we believe underlies the violations.

When Barney knows the rate is θ , he believes the likelihood of observing a particular clump of N signals, $s \in S_N$, is

$$f_{S_N|\Theta}^\psi(s|\theta) = \int_{\beta \in [0,1]} f_{S_N|\mathcal{B}}(s|\beta) f_{\mathcal{B}|\Theta}^\psi(\beta|\theta) d\beta, \quad (1)$$

where $f_{S_N|\mathcal{B}}(s|\beta)$ is the (correct) probability of observing s if the rate were β , and this is averaged over the density of subjective rates, $f_{\mathcal{B}|\Theta}^\psi(\beta|\theta)$. Consequently, Barney’s belief that a large sample will have a proportion of a signals in some range $[\alpha_1, \alpha_2]$ is exactly equal to Barney’s belief that the subjective rate β is in that range.⁷

Lemma 1. *Assume A1-A4. Barney does not believe in LLN: for any $\theta \in \Theta$ and interval*

⁶In particular, Griffin and Tversky asked subjects to infer the likelihood that a coin is biased $\theta_A = .6$ in favor of heads rather than $\theta_B = .25$ in favor of heads, depending on different possible outcomes from flipping the coin 12 times. According to the (statistically erroneous) diminishing-sensitivity intuition, extreme samples, such as 10 heads out of 12, seem so unexpected in the case of either rate that they do not provide strong evidence about which rate is generating the flips. Yet consistent with A3, Griffin and Tversky find that subjects’ posterior beliefs in favor of the .6-biased coin are monotonically increasing in the number of heads.

⁷All proofs are relegated to Appendix E.

$$[\alpha_1, \alpha_2] \subseteq [0, 1],$$

$$\lim_{N \rightarrow \infty} \sum_{x=\lfloor \alpha_1 N \rfloor}^{\lceil \alpha_2 N \rceil} f_{S_N|\Theta}^\psi(A_s = x|\theta) = F_{\beta|\Theta}^\psi(\beta = \alpha_2|\theta) - F_{\beta|\Theta}^\psi(\beta = \alpha_1|\theta) > 0.$$

Because we assume that Barney’s beliefs about the distribution of β puts positive probability density on the entire interval $(0, 1)$, the subjective-rate model captures the essence of our interpretation of NBLLN: Barney believes that the proportion of heads from flipping a coin *known to be fair* may not be 50% in any given sample, no matter how large.

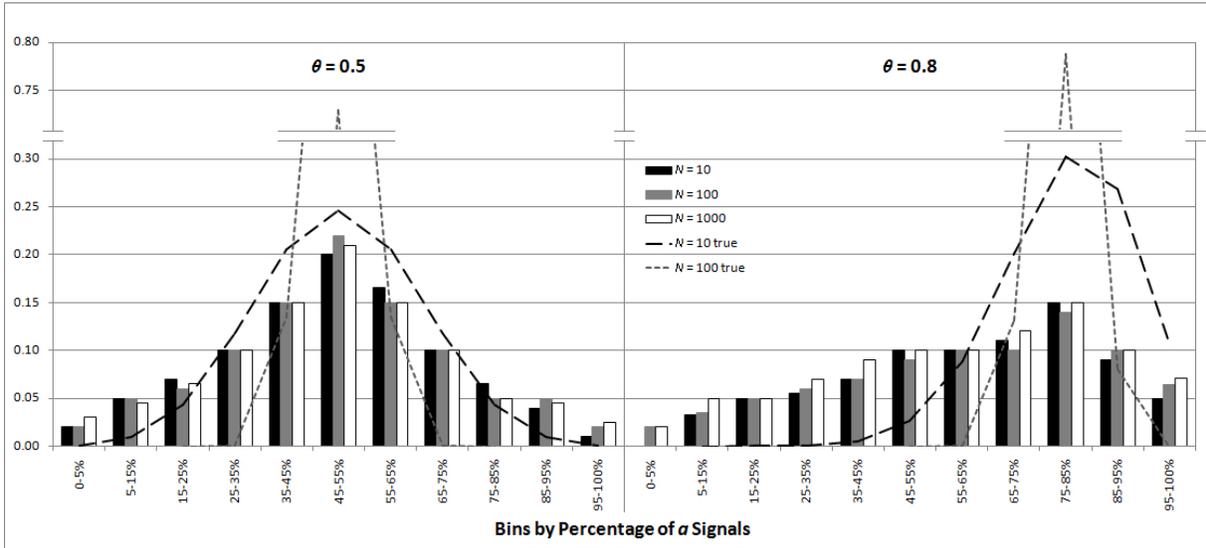


Figure 1: Evidence from Kahneman and Tversky (1972)

Since Barney’s belief about the distribution of signals in large samples coincides with his subjective-rate distribution, the most appropriate density of $\beta|\theta$ would correspond to the limiting distribution of people’s empirically observed beliefs about signals. In Figure 1, drawn from Kahneman and Tversky (1972), the black, gray, and white bars correspond to people’s reported beliefs regarding samples of size 10, 100, and 1000, respectively. Viewed through the lens of our model, the fact that the distributions are virtually identical implies that this common distribution corresponds to people’s large-sample beliefs about sample proportions. It could thus be directly assumed to be the density of $\beta|\theta$.

Although Assumptions A1 and A2 are consistent with the density of $\beta|\theta$ implied by Figure 1, A4 is not. Beliefs for $\theta = .5$, depicted in the left panel, naturally have a mean approximately equal to $.5$. However, beliefs for $\theta = .8$, depicted in the right panel, have a mean approximately equal to $.6$. The mean of the distribution of signals is displaced toward $.5$ apparently because the long

tail of the distribution is fat. As we discuss in Appendices C and D, we believe that the fatness of the tail is in turn due to even more *flatness* of the tail than implied by NBLLN alone. Because such extreme flatness is omitted from the model, it will not match some features of the empirical evidence, especially when the agent observes an extreme sample. We nonetheless assume A4 for two reasons: analytical convenience, and our contention that the violation of the assumption is due to psychological biases unrelated to NBLLN (see Appendix C for a model of flatness as resulting from a form of “diminishing sensitivity”), whose robustness and general properties are poorly understood.

A *subjective sampling distribution* specifies an agent’s belief about the likelihood of each possible combination of signals when the rate θ is known. Whereas Lemma 1 shows that Barney’s subjective sampling distribution (for the number of a -signals) in the large-sample limit equals his “subjective rate distribution,” Proposition 1 shows some implications of NBLLN for finite-sample subjective sampling distributions.

Proposition 1. *Assume A1-A4. For any $\theta \in \Theta$ and $N \in \{1, 2, \dots\}$:*

1. $E_{S_N|\Theta}^\psi \left(\frac{A_s}{N} | \theta \right) = E_{S_N|\Theta} \left(\frac{A_s}{N} | \theta \right) = \theta$. Therefore, for $N = 1$, $F_{S_1|\Theta} (A_s | \theta) = F_{S_1|\Theta}^\psi (A_s | \theta)$.
2. $F_{S_N|\Theta} (A_s | \theta)$ second-order stochastically dominates (SOSD) $F_{S_N|\Theta}^\psi (A_s | \theta)$, and $\text{Var}_{S_N|\Theta}^\psi \left(\frac{A_s}{N} | \theta \right) \geq \text{Var}_{S_N|\Theta} \left(\frac{A_s}{N} | \theta \right)$ with strict inequality for $N > 1$.
3. $\text{Var}_{S_N|\Theta}^\psi \left(\frac{A_s}{N} | \theta \right)$ is strictly decreasing in N .
4. $F_{S_N|\Theta}^\psi (A_s | \theta')$ first-order stochastically dominates (FOSD) $F_{S_N|\Theta}^\psi (A_s | \theta)$ whenever $\theta' > \theta$.

Part 1 states that Barney’s subjective sampling distribution gets the mean right—and therefore, Barney’s beliefs coincide with Tommy’s when $N = 1$. The latter follows because for a sample size of 1, the mean of an agent’s subjective sampling distribution is simply equal to his perceived probability of an a -signal; thus Barney, like Tommy, thinks the probability of a single a -signal is θ .

Part 2 states that Barney has a riskier subjective sampling distribution than Tommy. Combined with the fact that the mean of Barney’s subjective sampling distribution is the same as Tommy’s, this implies that Barney’s subjective sampling distribution is a mean-preserving spread of Tommy’s. This naturally implies that the variance of Barney’s subjective sampling distribution is larger than Tommy’s.

Part 3 states that the variance of Barney’s subjective sampling distribution (for the sample proportion) is strictly decreasing in N . Part 4 states that a higher true rate generates a rightward shift in Barney’s entire subjective sampling distribution in the sense of first-order stochastic dominance. Thus Parts 3 and 4 show that important qualitative comparative statics regarding Tommy’s beliefs also hold true regarding Barney’s beliefs.

We now turn to inference problems, where an agent with prior beliefs must infer from observed signals what the underlying rate is; e.g., determining the likelihood that a coin is head-biased rather than tail-biased, after observing a sample of coin flips. Let $\Theta \subseteq (0, 1)$ denote the set of rates that have positive prior probability. For simplicity, we assume Θ is a finite set. Without loss of generality, we consider the agent’s beliefs about the relative likelihood of two of the rates $\theta_A > \theta_B$, given priors $f_\Theta(\theta_A), f_\Theta(\theta_B) > 0$ and $f_\Theta(\theta_A) + f_\Theta(\theta_B) \leq 1$.

We maintain the conventional assumption that an agent draws inferences by applying Bayes’ Rule to his subjective sampling distributions. We do so both to highlight the role played per se by NBLLN, and because (as we discuss in Appendix D) our reading of the experimental evidence is that except for the well-established phenomenon of “base-rate neglect” (i.e., underweighting of priors), people’s inferences are actually well-approximated by Bayes’ Rule applied to their subjective sampling distributions.⁸ Consequently, Barney’s beliefs after observing a particular clump $s \in S_N$

$$\text{are } f_{\Theta|S_N}^\psi(\theta_A|s) = \frac{f_{S_N|\Theta}^\psi(s|\theta_A)f_\Theta(\theta_A)}{\sum_{\theta \in \Theta} f_{S_N|\Theta}^\psi(s|\theta)f_\Theta(\theta)} \text{ and } f_{\Theta|S_N}^\psi(\theta_B|s) = \frac{f_{S_N|\Theta}^\psi(s|\theta_B)f_\Theta(\theta_B)}{\sum_{\theta \in \Theta} f_{S_N|\Theta}^\psi(s|\theta)f_\Theta(\theta)}.$$

Due to the LLN, after observing a sufficiently large number of signals, Tommy will be arbitrarily close to certainty on the true rate. In contrast, the central implication for inference of Barney’s NBLLN—which plays a large role in many of the applications later in this paper—is that Barney remains uncertain even after observing an *infinite* number of signals. To boot:

Proposition 2. *Assume A1-A4. Let $\theta \in \Theta$ be the true rate. Then for any $\theta_A, \theta_B \in \Theta$ and prior $f_\Theta(\theta_A), f_\Theta(\theta_B) \in (0, 1)$, Barney draws limited inference even from an infinite sample: as $N \rightarrow \infty$, Barney’s posterior ratio converges almost surely (with respect to the true probability distribution over events) to a positive, finite number:*

$$\frac{f_{\Theta|S_N}^\psi(\theta_A|s)}{f_{\Theta|S_N}^\psi(\theta_B|s)} \xrightarrow{a.s.} \frac{f_{\mathbb{B}|\Theta}^\psi(\beta = \theta|\theta_A) f(\theta_A)}{f_{\mathbb{B}|\Theta}^\psi(\beta = \theta|\theta_B) f(\theta_B)}. \quad (2)$$

Because Barney’s asymptotic sampling distribution coincides with the subjective-rate distribution, his limit inference depends on the relative weights that the pdfs of the subjective-rate distributions for θ_A and θ_B assign to the proportion θ of a ’s. Since the subjective-rate distributions put positive density on every proportion in $(0, 1)$, Barney’s likelihood ratio will be finite. An immediate and important implication is that Barney’s priors—i.e., his ex ante theories about the world—influence his beliefs even in the limit of an infinite sample.

⁸Therefore, in applications where priors are equal—and hence base-rate neglect is neutralized as a factor—our assumption of Bayesian inference is fully appropriate. In applications where we do not assume equal priors, however, base-rate neglect could modify some of our results. Appendix C discusses and formalizes how to combine NBLLN with base-rate neglect.

Tommy not only will learn the true rate for sure after observing a sufficiently large number of signals, he also correctly *anticipates* that a sufficiently large number of signals will make him certain of the true rate. In contrast, Barney mistakenly thinks that his posterior probability of rate θ_A after observing an infinite number of signals is a stochastic function of the true rate. The reason is that, even though Barney knows that his inferences in a large sample will be pinned down by the proportion of a 's, he incorrectly thinks the proportion of a 's is determined by the subjective rate, which could be any number between 0 and 1.

Proposition 3. *Assume A1-A4. Fix rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A > \theta_B$ and any prior $f_\Theta(\theta_A) = 1 - f_\Theta(\theta_B) \in (0, 1)$. Before having observed any data, Tommy believes: if the rate is θ_A , then his limit posterior probability that the rate is θ_A is 1. In contrast, before having observed any data, Barney believes: if the rate is θ_A , then his limit posterior probability that the rate is θ_A is a random variable that has positive density on a nondegenerate interval in $[0, 1]$. If we strengthen assumption A1 to A1', then, in addition, the interval is closed and is a strict subset of $[0, 1]$.*

Moreover, since Barney updates correctly given his mistaken belief that common noise is added to any given sample, he mistakenly believes that his subjective beliefs satisfy the “Law of Iterated Expectations”: Barney expects that for any sample size, the mean of his posterior beliefs will equal the mean of his prior beliefs. Formally, for any $N \geq 1$, $E_{S_N}^\psi \left[E_{\Theta|S_N}^\psi(\theta|s) \right] = E_\Theta^\psi(\theta)$. In fact, however, Barney’s actual beliefs do not have this martingale property. As per Proposition 2, Barney’s posterior beliefs converge to a limit distribution, and the mean of this limit distribution is in general not equal to the mean of Barney’s prior distribution.

We next turn to inference in finite samples. Because (as per Part 1 of Proposition 1) Barney, like Tommy, correctly believes that the probability of a single a -signal is θ , he will draw correct inferences about θ_A vs. θ_B in that case.

Proposition 4. *Assume A1-A4. Fix rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A > \theta_B$ and prior $f_\Theta(\theta_A) = 1 - f_\Theta(\theta_B) \in (0, 1)$. For $N = 1$, Barney and Tommy infer the same. If $\theta_A = 1 - \theta_B$, then for any set of $N \in \{1, 2, \dots\}$ signal realizations $s \in S_N$, neither Tommy’s beliefs nor Barney’s beliefs change from the priors when $\frac{A_s}{N} = \frac{1}{2}$.*

In many of the inference experiments reviewed in Appendix D and in many of our applications involving inference, the two rates are “symmetric” in the sense that $\theta_A = 1 - \theta_B$, e.g., an urn might have either 60% red balls or 40% red balls. In that case, when exactly half the signals are a -signals, the sample is uninformative for both Barney and Tommy, and neither updates his beliefs about the rate.

For further analysis of the finite-sample case—as well as for some theoretical applications and

empirical analysis—it is useful to have a parametric model of Barney’s subjective rate distribution. For some of our results (Proposition 5 and multiple-sample applications in Appendix B), we impose the functional form of the beta distribution:

$$f_{\beta|\theta}^{\psi}(\beta|\theta) = \beta^{\theta\psi-1} (1-\beta)^{(1-\theta)\psi-1} \frac{\Gamma(\psi)}{\Gamma(\theta\psi)\Gamma((1-\theta)\psi)}, \quad (3)$$

where $0 < \psi < \infty$ is the exogenous parameter of the model, and $\Gamma(x) \equiv \int_{[0,\infty)} y^{x-1} e^{-y} dy$, defined on $x > 0$.⁹ The properties A1-A4 are satisfied (see Appendix E, Lemma 6), and this family of beta densities shares many qualitative features of people’s empirically-observed large-sample beliefs about the distribution of signals. A major advantage of this formulation is tractability: since the beta distribution is the conjugate prior for the binomial distribution, standard results from probability theory can be used to characterize Barney’s beliefs.^{10,11} “Parameterized-Barney” is more biased for smaller ψ —with more dispersed subjective sampling distributions in the sense of SOSD—and Barney coincides with Tommy in the parameter limit $\psi \rightarrow \infty$.

Although we do not conduct a careful structural estimation, we estimate from studies that elicit subjects’ subjective sampling distribution as well as inference studies that ψ falls within a range of 7-15. This parameterized model of Barney gives a sense of magnitudes for how Barney’s under-inference depends on the rates θ_A and θ_B . Suppose $\psi = 10$, Barney begins with equal priors on the two states, and the true rate is θ_A . If the difference between the rates is relatively large—with $\theta_A = 1 - \theta_B = .8$ —then the role of NLLN is relatively small. In an infinite sample, Barney’s subjective posterior probability of rate θ_A will converge to .9998. However, if the two rates are closer together—with $\theta_A = 1 - \theta_B = .6$ —then in an infinite sample, Barney’s subjective posterior probability of rate θ_A will converge to only .69 (which is what Tommy’s posterior would be after only 6 heads and 4 tails!). As a reminder about the role of priors, this means if Barney initially had beliefs more extreme than 2.25:1 in favor of rate θ_B , he will, in an infinite sample, surely end

⁹The more common way of writing this beta density is $f_{\beta|\theta}^{\psi}(\beta|\theta) = \beta^{\theta\psi-1} (1-\beta)^{(1-\theta)\psi-1} \frac{(\psi-1)!}{(\theta\psi-1)!((1-\theta)\psi-1)!}$. Our formulation is equivalent, except it allows for non-integer values of ψ . Recall that the Gamma Function, $\Gamma(x)$, is the standard generalization of the factorial function: it has the properties that $\Gamma(x+1) = x\Gamma(x)$ and $\Gamma(1) = 1$, so that for any positive integer x , $\Gamma(x) = (x-1)!$.

¹⁰The functional form (3) has a few implications about asymmetric inference (i.e., inference problems where $\theta_A \neq 1 - \theta_B$) that do not have general intuitions related to NLLN. These are presented in Lemma 5 in Appendix E. In the text, we avoid stating implications of these properties of the functional form that would not generalize to other models that are equally consistent with existing evidence.

¹¹To interpret the “Barneyness parameter,” ψ , consider the problem of inferring the rate from a sample of signals. The beta density (3) corresponds to what a Bayesian’s posterior about the rate would be if the Bayesian had begun with an improper beta prior on $[0, 1]$ corresponding to the $\psi \rightarrow 0$ limit of (3), and had observed a total of ψ signals with $\theta\psi$ a ’s and $(1-\theta)\psi$ b ’s. For $\psi = 0$, the posterior would remain the improper prior on $[0, 1]$, while for $\psi \rightarrow \infty$, the posterior would converge to a point mass at the true rate θ .

up believing rate θ_B is more likely, even when θ_A is the true rate.

While most dramatic in large samples, NBLLN has implications for all sample sizes larger than 1. Since Barney’s subjective sampling distribution is too dispersed when $N > 1$, Barney will generally under-infer when the sample size is larger than 1. In order to make that claim precise, we measure Barney’s (and Tommy’s) “change in beliefs” by the absolute difference between his posterior probability that θ_A is the true rate and his prior probability: $\left| f_{\Theta|S_N}^{\psi}(\theta_A|s) - f_{\Theta}(\theta_A) \right|$. Unlike in large samples, in small samples it is no longer universally true that Barney under-infers relative to Tommy. For particular realizations Barney can over-infer or under-infer relative to Tommy—or even infer in the opposite direction, so that a sample that causes Tommy to think rate θ_A is more likely, causes Barney to think rate θ_B is more likely.¹² Nonetheless, we believe that Barney under-infers *in expectation*, taken with respect to the true sampling distribution. Proposition 5 proves this statement for the case of ψ sufficiently small, but we conjecture that it holds for any $0 < \psi < \infty$.¹³

Proposition 5. *Assume Barney has the beta-distribution functional form given by equation (3). Fix rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A > \theta_B$, prior $f_{\Theta}(\theta_A) = 1 - f_{\Theta}(\theta_B) \in (0, 1)$ and a set of $N \in \{1, 2, \dots\}$ signal realizations $s \in S_N$. Regardless of whether the true rate is θ_A or θ_B , for ψ sufficiently small, the expected change in Barney’s beliefs is smaller than the expected change in Tommy’s beliefs. Furthermore, suppose $\theta_A = 1 - \theta_B$. Then for any sample of $N > 1$ signals such that $\frac{A_s}{N} \neq \frac{1}{2}$ and any ψ , Barney under-infers relative to Tommy. In addition, while Tommy’s inference depends solely on the difference in the number of a and b signals, Barney’s change in beliefs is smaller from larger samples with the same difference.*

Intuitively, on average Barney under-infers because he partially attributes the information in the realized sample to the subjective rate, rather than extracting all of the information about the true rate.

In symmetric inference problems (i.e., $\theta_A = 1 - \theta_B$), Proposition 5 shows that stronger comparisons can be made between Barney and Tommy: as long as the realized sample is informative, parameterized-Barney will under-infer, not just in expectation. Proposition 5 also notes a key feature of Barney’s updating that shows how it leads to a bias toward “proportional thinking” in inference along the lines suggested by researchers such as Griffin and Tversky (1992). Consider

¹²For example, using the parameterized model, set $\psi = 10$, $\theta_A = .7$, and $\theta_B = .6$, and assume equal priors on the two states. Then if the realizations of 80 signals are 53 *a*-signals and 27 *b*-signals, then Tommy believes that state A is more likely, while Barney believes state B is more likely.

¹³We have simulated Barney’s and Tommy’s expected change in beliefs for a range of parameter values: for each of $\psi \in \{1, 2, \dots, 30\}$ and $N \in \{5, 10, 15, 20\}$, we examined each of $\theta_A, \theta_B \in \{.5, .6, .7, .8, .9\}$. We also ran a number of simulations for $\theta_A = .99$ and $.999$ and for $\psi = 100$. In every case we examined, Barney’s expected change in beliefs was smaller than Tommy’s.

samples where the difference between the number of a -signals and the number of b -signals is the same, e.g., (2 a , 0 b) and (5 a , 3 b). Tommy will draw the same inference from the two samples. But because his asymptotic sampling distributions depend on the proportion of a and b signals rather than their number, Barney infers less from the larger sample.

3 Lack of Demand for Large Samples

This section and the next explore the implications of the simple model from the last section. Here we explore what is perhaps the most direct and important economic implication of NBLLN: because people do not expect to learn much from large samples, they are more likely to rely on small-sample sources of information than to incur the cost of obtaining a large-sample data source.

Suppose Barney is trying to decide what make of car to buy, a Volvo or a Lada.¹⁴ The state is $\omega = A$ if the Volvo is superior and $\omega = B$ if the Lada is superior. Barney is choosing whether to acquire information by asking a friend, which will provide him with a single signal at cost $c_f > 0$, or by purchasing *Consumer Reports*, which will provide him with the aggregate information from a large number N of signals, at cost $c_r > c_f$. After observing the information, Barney must take an action μ , either buying the Volvo ($\mu = \mu_A$) or the Lada ($\mu = \mu_B$). Barney's payoff is $u(\mu, \omega)$, which equals 1 if the action matches the state and 0 otherwise.

The comparison between Barney's and Tommy's valuations of an intermediate-sized sample is ambiguous: even though Barney might expect to infer less on average than Tommy would, Barney also overestimates the probability of an extreme outcome that would allow for stronger inferences. If such stronger inferences are valuable (because they induce the agent to switch to the superior action), Barney may be willing to pay more than Tommy for a sample of a given size. Furthermore, perhaps counterintuitively, if the agent can choose the total number of signals to purchase, Barney may choose to purchase a larger sample size than Tommy if doing so makes it possible to observe an extreme outcome that is valuable—an extreme outcome that Barney thinks is more likely than it is.

For a sufficiently large sample, however, Barney unambiguously has a lower willingness to pay than Tommy does because he anticipates making a weaker inference for any realized outcome. Tommy correctly understands that—due to the Law of Large Numbers—reading *Consumer Reports* will make him virtually certain about the true state, but as per Proposition 3, Barney expects to remain uncertain. Because Barney expects to learn less from *Consumer Reports* than Tommy does, Barney is more likely to ask the friend.

¹⁴A Lada is a type of car. So is a Volvo.

Proposition 6. *Assume A1-A4. Fix payoffs $u(\mu, \omega)$, rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A > \theta_B$, prior $f_\Theta(\theta_A) = 1 - f_\Theta(\theta_B) \in (0, 1)$, and the cost of asking a friend $c_f > 0$. Suppose that c_f is small enough that if Consumer Reports were not available, Tommy would ask the friend. If the number of signals N in Consumer Reports is sufficiently large, then there exist thresholds c'_r and c''_r with $c_f < c'_r < c''_r$ such that: if $c_r < c'_r$, then both Tommy and Barney buy Consumer Reports; if $c_r > c''_r$, then both Tommy and Barney ask the friend; and if $c_r \in (c'_r, c''_r)$, then Tommy buys Consumer Reports while Barney asks the friend.*

Barney’s lack of demand for statistical data is a central implication of NBLLN. We believe it is consistent with obvious facts: we live in a world in which people are not persuaded by statistical evidence that should be convincing, people do not demand such information, and such information is therefore rarely supplied by the market. The car purchaser who actually consults Consumer Reports is the exception rather than the rule. Moreover, the flip side of people’s failure to demand large numbers of signals is their willingness to rely instead on sources of information that provide only a small number of signals. Indeed, given the amount of other information people may be able to obtain at relatively low cost, NBLLN helps explain why they nonetheless often instead rely on limited personal experience, stories from one’s immediate social network, or anecdotes from strangers.

Of course, in many situations there are other contributing factors. In some cases, such as restaurant satisfaction ratings, people may correctly expect large samples not to be very useful because preferences are heterogeneous. And due to the Law of Small Numbers, people may expect a small sample to be sufficiently persuasive. But in cases where preference heterogeneity is less of an issue and where getting statistical data would be barely more costly than collecting anecdotes—for example, finding out the frequency of car battery failure—the lack of demand for statistical data (and the consequent lack of supply) is a major “dog that didn’t bark” clue that implicates NBLLN.¹⁵

The lack of demand for large samples generated by NBLLN is especially severe when Barney is initially confident about the state or when each individual signal is relatively uninformative, i.e., θ_A is close to θ_B . Tommy understands that a sufficiently large number of such signals will nonetheless reveal the state. In contrast, when Barney has a confident prior or when signals are relatively

¹⁵In principle, another alternative explanation for lack of demand for large samples is that people are concerned about sample selection bias. For example, the reason that statistics showing that vaccines do not increase autism are unpersuasive to people may be because people worry that those getting vaccines are less prone to autism than those not getting vaccines. In general, we do not believe that this alternative explanation plays much of a role because it seems much more common that people attend too little to selection issues; indeed, we would live in a very different world if concerns with selection bias pervaded public reaction to statistical evidence.

uninformative, Barney may be unwilling to pay *any* positive cost for even an infinite number of signals.

Proposition 7. *Assume A1' and A2-A4. Suppose that the agent is deciding whether to buy Consumer Reports at cost c_r or not obtain any signals. For Tommy: for all rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A > \theta_B$ and priors $f_\Theta(\theta_A) = 1 - f_\Theta(\theta_B) \in (0, 1)$, there exists a threshold $c_r^* > 0$ such that if $c_r < c_r^*$, then as long as the number of signals N in Consumer Reports is sufficiently large, he buys Consumer Reports. In contrast, for Barney: (i) for all rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A > \theta_B$, there exist priors $f_\Theta(\theta_A)$ such that for any $c_r > 0$ and any N , he does not buy Consumer Reports; and (ii) for all priors $f_\Theta(\theta_A)$ at which he is not indifferent between μ_A and μ_B , there exist rates $\theta_A, \theta_B \in \Theta$, where $\theta_A > \theta_B$, such that for any $c_r > 0$ and any N , he does not buy Consumer Reports.*

Unlike Proposition 6, Proposition 7 relies on A1' (not just A1): Barney thinks that he will draw a limited inference no matter how extreme the sample proportions turn out to be. If his priors are extreme enough or the rates are close enough together, then he thinks that the information provided by an infinite number of signals will not affect whether he buys the Volvo or the Lada. Consequently, his willingness to pay for an infinite number of signals is zero.

On the flipside, even though Barney often under-infers when presented with information, he may nonetheless purchase information even when it will not have any objective value for him. This is because Barney believes that his posterior after observing a large sample is a random variable, and his willingness to pay is positive whenever he thinks he might make an extreme enough inference to switch his action from what he would do given only his prior. Yet for a large sample, Barney's posterior is in fact deterministic. Hence if Barney's priors are extreme enough, and the cost of information small enough, he may incur the cost of purchasing *Consumer Reports* even though the information will almost surely not affect his action. In Appendix B, we illustrate how such a “learning trap” phenomenon can also occur in a dynamic setting through a different mechanism—namely, Barney's incorrect beliefs about how he will process future signals.

4 Risky Gambles and Investments

In this section, we examine how NBLLN's basic implications for an agent's subjective sampling distribution play out in various gambling and investment environments.

4.1 Perceived Aggregate Risk

The comparison between Barney's and Tommy's valuation of a gamble is especially simple in settings where an agent experiences a binary win or loss depending on whether a particular outcome occurs. In such cases, Barney values the gamble more than Tommy does if and only if he thinks that the winning outcome is more likely.

A leading special case is where the winning outcome is an extreme sample proportion, which in general Barney thinks is more likely to occur than Tommy does. For example, suppose Barney plays a lottery where he must guess N numbers. He has probability θ of guessing any single number correctly. He receives a prize if he guesses all N numbers correctly, but he gets nothing otherwise. Notice that guessing all the numbers correctly is equivalent to the extreme sample proportion of 100% a -signals being realized. Thus, if $N > 1$, then Barney believes his chance of guessing all the numbers correctly is higher than it is. Hence Barney's willingness to pay to play this game is higher than Tommy's. The logic of NBLLN implies that the difference from Tommy is larger for larger N .

In contrast, in situations where Barney wins as long as an extreme outcome does *not* occur, Barney's behavior can appear to be especially risk averse. For example, suppose Barney attends a job fair at which he applies to $N > 1$ equally-valuable jobs, his chance of getting any particular job is θ , and he cannot accept more than one job. In this case, not receiving any job offers is equivalent to the extreme sample proportion of 0% a -signals being realized. Since Barney exaggerates the probability of this extreme outcome, his willingness to pay to attend the job fair is lower than Tommy's.

For the remainder of this subsection and the next, we turn to settings in which the agent's outcome is not binary but instead varies with the number of a -signals. Here we consider classical risk preferences as represented by a utility function $u(w)$ that is smooth and increasing in final wealth w (in the next subsection, we consider instead loss-averse preferences). Final wealth, $w(A_S)$, is an increasing function of the realized number of a -signals (good outcomes) out of N draws, where as usual the rate of a -signals is θ .

Proposition 8 states the key implication of NBLLN for valuation of such a gamble, which depends on whether the agent's utility over the number of good draws, $u(w(A_S))$, is concave or convex in A_S .¹⁶ Note that, because we will consider cases where $w(A_S)$ is not linear, the shape of the agent's utility depends on both the agent's risk preferences $u(w)$ and the manner in which the random outcomes translate into monetary outcomes $w(A_S)$.

¹⁶Because the outcome space is discrete, in both this subsection and the following one we use the standard definition that $u(w(A_S))$ is a convex function of A_S if $u(w(A_S - 1)) + u(w(A_S + 1)) - 2u(w(A_S)) \geq 0$, and concave if $u(w(A_S - 1)) + u(w(A_S + 1)) - 2u(w(A_S)) \leq 0$.

Proposition 8. *Assume A1-A4. Fix a risky gamble (θ, N) . If $u(w(A_s))$ is a concave (resp., convex) function of A_s , then Barney's willingness to pay for the risky investment is less than (resp., greater than) Tommy's.*

Given the fact that, as per Proposition 1, Barney's subjective sampling distribution is a mean-preserving spread of Tommy's, Proposition 8 follows directly from standard results in the theory of choice under risk. Its implications in different risky-choice contexts are determined jointly by $u(w)$ and $w(A_s)$.

Consider a repeated gamble over coin flips, in which the agent earns a fixed dollar amount h for each a -signal and a different dollar amount $t < h$ for each b -signal. In that case, $w(A_s)$ is linear, so $u(w(A_s))$ is concave in A_s if and only if $u(w)$ is concave in w . Thus, NBLLN reduces a risk-averse agent's willingness to pay for a repeated gamble.

Now suppose an agent is considering whether or not to invest in a diversified portfolio of N identical stocks for a single year. Any given stock does well with probability θ , in which case it pays off h , or badly with probability $1 - \theta$, in which case it pays off $t < h$. Since this example is mathematically equivalent to a repeated gamble, NBLLN can help explain why people fail to fully recognize the benefits of diversification: if investors face a fixed cost of diversification, Barney may find diversification not worth the cost, even when Tommy does.

As a final example, suppose an agent is considering whether or not to invest in a stock for N years. Any given year the stock does well with probability θ in which case it earns a gross *rate of return* $1 + r_h$, or badly with probability $1 - \theta$ in which case it earns a gross rate of return $1 + r_t$ with $r_t < r_h$. (Equivalently, we could assume that the stock pays off a fixed dollar amount, but all earnings are re-invested.) Because of compounding, $w(A_s)$ is now a convex function. Whether Barney or Tommy has a greater willingness to pay for this investment opportunity depends on the shape of $u(w)$. If the agent is risk-neutral, or more generally not sufficiently risk-averse, then Barney will find the long-term investment more attractive than Tommy does. In reality, however, we believe that people generally perceive long-term investments as *less* attractive than they should because of the combined effects of NBLLN and loss aversion, as we discuss next.

4.2 Samuelson's colleague

Proposition 8 from the previous subsection highlights the effect on risk attitudes of the fact that Barney's beliefs are a mean-preserving spread of Tommy's. NBLLN's most direct prediction, that Barney will put positive probability on extreme outcomes even in very large samples, also has important implications for risky choice.

Consider first a famous example: Paul Samuelson (1963) reports the story of an economics professor colleague at MIT telling Samuelson that, whereas he would reject a bet for even odds to gain \$200 or lose \$100, he would accept 100 repetitions of that bet. Even though such behavior sounds reasonable to most of us, Samuelson proves that it violates classical expected-utility theory. That is, a Tommy with expected-utility preferences defined over final wealth who does not exhibit unrealistically large wealth effects should be willing to take a single bet if and only if he is willing to take $N \geq 1$ independent plays of that bet. To see this, note that preferring $K + 1$ bets to K bets is the same thing as preferring 1 bet on top of any realization of the K bets. By induction, preferring to take any positive number of the bets is the same as preferring to take one bet.

Yet, it is not just the “switching” that violates classical expected-utility preferences, but the aversion to the single bet to begin with. Rabin (2000a, 2000b) and Rabin and Thaler (2001) have followed others in noting that the degree of concavity required for expected-utility preferences defined over wealth to explain risk-averse behavior over small stakes is calibrationally implausible. Loss aversion—the tendency to feel a loss more intensely than an equal-sized gain—probably explains why the majority of people who turn down the one-shot gamble do so.¹⁷ Somebody with a simple, piecewise-linear loss-averse utility function

$$u(w_0, z) = \begin{cases} w_0 + z & \text{if } z \geq 0 \\ w_0 + \lambda z & \text{if } z < 0 \end{cases}, \quad (4)$$

where w_0 is initial wealth and z is a monetary gain or loss, will refuse the one-shot bet as long as the coefficient of loss aversion, λ —often set equal to 2.25 (e.g., Tversky and Kahneman, 1991)—is greater than 2.¹⁸ However, a Tommy with typical loss-averse preferences would be extremely happy to accept 100 repetitions of the same gamble: while the expected gain is \$5,000, the chance of a net loss is only 1/700, and the chance of losing more than \$1,000 is only 1/26,000.

Despite being loss averse enough to turn down the one bet, nobody with fully rational beliefs would turn down 100 repetitions of this bet. Yet, unlike Samuelson’s colleague, many people would do so! In hypothetical questions from one study in Benartzi and Thaler (1999), for instance, 36%

¹⁷Samuelson himself had speculated that it was the willingness to accept repeated plays of the bet that was the mistake, rather than the refusal to accept a single gamble. Samuelson’s conjecture that his colleague’s willingness to accept the repeated gamble was the result of a “fallacy of large numbers”—a mistaken belief that the riskiness of the gamble evaporates with a sufficiently large number of repetitions—is the *opposite* of NBLLN, and is contradicted by Benartzi and Thaler’s (1999) evidence, reported below, that people exaggerate the probability of a loss in the repeated bet.

¹⁸Although essentially correct for small gambles, assuming linear consumption utility can become problematic if bets are repeated so many times as to involve large amounts of wealth. However, if in our limit results below, we halve the stakes every time we double the number of repetitions, the linearity assumption is unobjectionable. Similarly, linearity is an appropriate assumption for studying diversification of a fixed amount of wealth among many assets, with a small amount invested in each asset.

of participants said they would turn down a single scaled-down Samuelson type bet (win \$100 or lose \$50), but fully 25% also reject the 100-times repeated gamble.¹⁹

NBLLN helps explain why many people turn down these gambles.²⁰ Barney exaggerates the probability that the repeated bet will turn out badly. Indeed, Benartzi and Thaler report evidence consistent with this explanation: when asked the probability of losing money after 150 repetitions of a 90%/10% bet to gain \$0.10/lose \$0.50, 81% of subjects overestimated the probability—and by an enormous margin. While the correct answer is .003, the average estimate was .24. To show that subjects’ mistaken beliefs about the probabilities of the different possible outcomes were driving their choices, Benartzi and Thaler compared subjects’ willingness to accept the repeated bet with their willingness to accept a single-play bet that had the histogram of money outcomes implied by the repeated bet. While only 49% of the college-student subjects accepted 150 repetitions of the bet, 90% accepted the equivalent single-play bet, suggesting that the repeated bet would have been very attractive if subjects had correctly understood the distribution of outcomes.^{21,22}

A leading alternative explanation for turning down the repeated bet is “probability weighting,” a psychological bias in which people assign an unlikely event (such as losing money in the repeated bet) a higher decision weight than warranted by its objective probability (Kahneman and Tversky, 1979). However, probability weighting cannot explain the evidence that subjects misperceive the

¹⁹In two other subject pools, they find 34% and 23% turn down a simple \$20/\$10 gamble, and *more* people—57% and 50%—turn down the repeated gamble. Keren (1991) finds similar results in incentivized single bets vs five-times-repeated bets; for related hypothetical evidence, see Keren and Wagenaar (1987) and Redelmeier and Tversky (1992). Klos, Weber and Weber (2005) replicate and extend Benartzi and Thaler’s findings. They present subjects with four lotteries, each of which may be played singly, repeated 5 times, or repeated 50 times. Subjects generally prefer the repeated gambles but vastly overestimate the probability of loss as well as the expected loss conditional on losing money. Klos, Weber and Weber also find that subjects incorrectly believe the probability of the monetary outcome ending up within a given interval around the expected value increases with the number of repetitions. This last finding is inconsistent with our model of NBLLN and may reflect a bias from focusing subjects’ attention on the expected value, or it may be consistent with “exact representativeness,” a bias we discuss in Appendix D.

²⁰Our emphasis on how NBLLN helps explain why loss-averse people turn down the repeated bet is because of its calibrational relevance, but it is worth noting that NBLLN also has implications for how expected-utility-over-wealth agents respond to repetitions of bets. We can extend the “if” part of Samuelson’s theorem: if Barney rejects a bet at all initial wealth levels w_0 , then he would also reject any $N \geq 1$ independent plays of that bet. The “only if” direction does not extend, and a Barney who is just indifferent between accepting and rejecting a simple bet would, because he exaggerates the risk, strictly prefer to reject repeated versions of the gamble.

²¹Note also that something more than the type of “narrow bracketing” stressed by authors such as Tversky and Kahneman (1986), Kahneman and Lovallo (1993), Benartzi and Thaler (1995), Read, Loewenstein and Rabin (1999), Barberis, Huang and Thaler (2006), and Rabin and Weizsäcker (2009) seems to be playing a role. Those papers emphasize that people often react to a combination of risky bets as if they were deciding about each risky bet in isolation from all the others. While such neglect of the effects of aggregating risks may help explain why people reject the repeated gamble, it seems clear that even people who attend to the aggregate effects *misunderstand* these aggregate effects. Benartzi and Thaler (1999) make this especially clear by demonstrating directly that people asked the probability of aggregate loss of independent bets exaggerate along the lines predicted by NBLLN.

²²Benartzi and Thaler also elicited the effects of showing the histogram in the above hypothetical examples and showed that it reduces rejections from 25%, 57%, and 50% to, respectively, 14%, 10%, and 17%.

objective probability of losing money, nor the evidence that subjects would accept the repeated bet if they had the correct beliefs.

Proposition 9 formalizes the claim that NLLN and loss aversion taken together predict turning down the repeated bet. While a loss-averse Tommy will always accept a better-than-fair bet if it is repeated enough times, a loss-averse Barney may—depending on how favorable the bet is and how loss-averse he is—turn down an infinitely-repeated bet.

Proposition 9. *Assume A1-A4. Suppose Barney and Tommy have simple, piecewise-linear loss-averse preferences as specified in (4). Fix any gamble (θ, h, t) , paying off $h > 0$ with probability θ and $-t$ with probability $1 - \theta$, that is better than fair: $\theta h > (1 - \theta)t$. For any $\lambda \geq 1$, there is some $N' \geq 1$ such that if $N > N'$, then Tommy will accept N repetitions of the gamble. In contrast, for Barney there is some threshold level of loss aversion $\hat{\lambda} > 1$ such that: if $\lambda < \hat{\lambda}$, then there is some N' sufficiently large such that Barney will accept N repetitions of the gamble for all $N > N'$; and if $\lambda \geq \hat{\lambda}$, then there is some N'' sufficiently large such that Barney will reject N repetitions of the gamble for all $N > N''$.*

Moreover, it is possible for Barney to exhibit the *opposite* pattern from Samuelson’s colleague, accepting the single bet but rejecting the 100-times repeated bet. Consistent with this possibility, the evidence cited above from Benartzi and Thaler (1999) found behavior in the opposite direction of Samuelson’s colleague in two out of their three studies.²³

Using the one-parameter functional form for Barney, calibrations suggest that NLLN goes much of the way, but *not* all of the way, in explaining why 25-57% of participants turned down the repeated bets in Benartzi and Thaler’s studies. For Tommy, the coefficient of loss aversion required to explain this data is absurdly high, in excess of 32,000. For Barney with $\psi = 10$, the required loss aversion is approximately 15—many orders of magnitude closer to reality but still much larger than reasonable estimates of λ .²⁴

Similarly as with Proposition 8, Proposition 9’s results on repeated bets carry over directly to diversification: it would go through essentially unchanged if, rather than repeating a gamble N

²³Preferences that generate an aversion for multi-stage resolution of risk—such as the preferences proposed by Koszegi and Rabin (2009) or Dillenberger (2010)—could also predict rejections of repeated gambles. The psychology underlying this prediction, however, only seems plausible when the outcomes of each individual gamble are observed separately. In contrast, NLLN predicts rejection precisely due to mistaken beliefs about the *combined* outcomes of the gambles. Furthermore, NLLN predicts risk-seeking behavior in settings where these other models would not, e.g., in the lottery example in the previous subsection.

²⁴Incorporating other known biases into the model, such as probability weighting (and sampling-distribution-tails diminishing sensitivity bias (SDTDS) described in Appendix C), would help reduce the required level of loss aversion even more—although still not to a reasonable level of λ such as 2.25. To see this, consider the most extreme form of overweighting losses: putting equal decision weight on every possible outcome of the repeated gamble. In this case, because only 1/3 of the outcomes are losses, the required level of loss aversion would be around 4.

times, the agent were mixing N independent gambles. Tommy would always accept a portfolio of positive-expected-value gambles if N is sufficiently large. In contrast, if Barney is sufficiently loss averse, then regardless of how large N is, Barney may prefer not to hold this portfolio.

Also similarly to above arguments, the application to long-term investing is complicated by compounding, which if strong enough could reverse the comparison between Barney and Tommy. Benartzi and Thaler (1999) reported evidence consistent with NLLN in the context of long-term investing: university employees vastly overestimated the probability that equities would lose money over a thirty-year horizon. Moreover, the employees stated a far greater willingness to invest in equities when they were explicitly shown the thirty-year returns, suggesting that the net effect of NLLN is to reduce the attractiveness of long-term investing. While there are many reasons why individuals may invest less in equities for retirement than recommended by standard finance models, we suspect that NLLN is an important contributing factor. As such, just as in other settings researchers may underestimate risk aversion by ignoring overconfidence when inferring risk preferences from investment behavior, researchers might therefore exaggerate the risk aversion of investors by ignoring NLLN.

5 NLLN in Multiple-Sample Settings

For Tommy, it does not matter whether he treats 20 independent signals as one sample of 20, two samples of 10, or 20 samples of 1. In contrast, and intrinsic to the very meaning of NLLN, Barney's beliefs about the distribution of and inference from signals depend on how he divides them up into samples. In this section, we provide a framework for formulating plausible assumptions regarding how Barney groups signals into samples. Doing so serves three purposes. First, adopting particular assumptions makes it possible to study the implications of NLLN in dynamic decision-making environments, as we illustrate in Appendix B. Second, testing between alternative assumptions constitutes a valuable research agenda for future experimental work; our review in Appendix D makes it clear that existing evidence on these assumptions is scant and somewhat contradictory. Third, the framework we develop may prove useful more broadly for studying other non-Bayesian models of judgment biases, such as base-rate neglect, that similarly exhibit sensitivity to how data are framed. We keep our discussion in this section informal, relegating formal definitions and further discussion to Appendix A.

Several distinctions will be useful. *Clumping* refers to how signals are objectively delivered to Barney by his environment. For example, when Barney asks a sequence of friends about their experience driving a Volvo, each friend's report arrives as a separate clump, but when Barney

reads a summary of 10,000 individuals’ experiences in *Consumer Reports*, the 10,000 signals arrive as a single clump. We refer to how Barney subjectively processes these clumps for the purposes of making forecasts and inferences as how he *groups* the signals. Barney forms beliefs regarding each group of signals as if a subjective rate β were drawn that applies only to that group—and hence the single-clump model from Section 2 can be applied to each group. Although economic models of decisionmaking do not traditionally specify how signals are clumped, our aspiration is that assumptions about clumping be pinned down by observable characteristics of a situation. How clumps are grouped, on the other hand, must be a feature of any complete model of a departure from Bayesian information processing, and the aspiration there would be to develop a general principle that maps objective clumping into subjective grouping.

We distinguish two facets of how Barney groups data. The first is how he processes clumps into groups *retrospectively*—how he processes clumps he has already received. The second is how he processes clumps into groups *prospectively*. Prospective grouping determines his forecast about what data he will observe given his current beliefs *and* his forecast of what he will infer from that data after he observes it.

In each of the retrospective and prospective directions, we focus attention on two ways that Barney might process clumps of signals. If Barney groups the signals the same way he receives them from the environment, we call him *acceptive*. Acceptive Barney would process *Consumer Reports* as a single sample of 10,000, and then each of his friends’ reports as a separate sample. If Barney processes all of the clumps of signals he observes as a single, large group, we call him *pooling*. For example, pooling Barney would treat *Consumer Reports* data and his friends’ stories as a single, larger sample.²⁵

In principle, one could imagine Barney’s beliefs in a dynamic environment as being either retrospective-acceptive or retrospective-pooling, combined with being either prospective-acceptive or prospective-pooling. We argue that Barney cannot be prospective-pooling, however, in any environment where he expects to make a decision before receiving the entire sample. More generally, we impose the following constraint on any model of NLLN: at any date where the agent makes a decision, he processes signals before and after that date as being in separate groups—and before that date, he knows he will do so. We consider this to be a modeling-coherence constraint, because it ensures that Barney’s NLLN from the single-clump model in Section 2 generalizes to every

²⁵While we conjecture that these two grouping processes cover a wide range of typical situations, we acknowledge that in certain situations, other grouping processes are psychologically plausible. For example, as we mention in Appendix B, in the context of social learning it may be natural to group one’s own signal separately from everyone else’s, even if all the signals occur simultaneously. As another example, Barney may group signals according to the perceived similarity of the information source. After observing 10,000 datapoints from *Consumer Reports* followed by 10 friends’ reports obtained sequentially, for instance, Barney may retrospectively process the information as a group of 10,000 followed by a group of 10.

decision node in a multiple-clump setting.

To see this, suppose Barney knows that $\theta = .5$, reads a summary of 10,000 individuals' experiences in *Consumer Reports*, and then must make a prediction about the next 1,000 signals he will observe. If, in violation of our modeling constraint, he were planning after observing the first 10,000 signals to process all 11,000 signals together as a single group, then he would believe that the same subjective rate β applies to all 11,000 signals. Using the first 10,000 signals, he would update his belief about β from $f_{\beta|\Theta}^{\psi}(\beta|\theta = .5)$ to a density that puts almost all the probability mass on the observed proportion of a -signals, say 50%. Since the next clump is grouped with the earlier clump, his subjective sampling distribution for the next clump will put negligible weight on a proportion of a -signals outside a neighborhood of 50%. In his predictions about future signals, Barney would no longer exhibit NBLLN. In contrast, our modeling constraint requires that Barney forms beliefs as if a new β is drawn from $f_{\beta|\Theta}^{\psi}(\beta|\theta)$ before the next 1,000 signals, so his subjective sampling distribution is exactly as in the single-clump model for $N = 1,000$. Precisely because it rules out learning about the subjective rate, this constraint distinguishes our model of NBLLN from the generalization of the model from Section 2 as one of rational uncertainty about the parameter β .

6 Alternative Models

In this section, we explain why we believe NBLLN is a better interpretation of the evidence reviewed in Appendix D than other possible interpretations.

As noted in Section 1, the bulk of evidence from inference experiments has been associated with an interpretation called “conservatism bias,” according to which people have a generic tendency to update less than would be prescribed by Bayes' Rule. This interpretation, however, contradicts evidence on base-rate neglect and other cases in which over-inference is often observed. In addition to explaining the experimental evidence on inference, moreover, the NBLLN interpretation also matches the experimental evidence on subjective sampling distributions and connects it as a source for under-inference.

When it induces people to believe low-probability events are more likely than they are, NBLLN bears a superficial resemblance to “probability weighting” (Kahneman and Tversky, 1972), a phenomenon in which small probabilities are overweighted in risky decisionmaking. Our view is that the two phenomena are psychologically distinct and unrelated. Most fundamentally, probability weighting is generally interpreted (see, e.g., Barberis, 2013) as a phenomenon about “decision weights” in which people care more about low-probability events than expected-utility theory predicts, whereas

NBLLN is about distortion in beliefs. The two theories frequently make different predictions. For example, the degree of belief distortion predicted by NBLLN depends on N , whereas the degree of probability weighting does not. Thus, NBLLN and probability weighting make qualitatively different predictions regarding the comparison between gambles composed of multiple realizations of random variables vs. attitudes towards gambles composed of a single realization of a random variable, but which otherwise have the same distribution over outcomes.

A different possibility from decision weights is that people are averse to holding extreme beliefs—and thus the beliefs themselves are distorted away from very low (or high) probabilities. We refer to this idea as “extreme-belief aversion” and discuss it in detail in Appendix C. As a very brief summary, extreme-belief aversion seems to be a separate phenomenon that cannot by itself explain the evidence that we interpret as NBLLN. Extreme-belief aversion predicts that in any two inference problems where Tommy’s posterior is the same, people would hold the same belief. Evidence such as in Griffin and Tversky (1992), however, indicates a lower willingness to have extreme beliefs when inferring from large samples than when inferring from small samples.

An interpretation proposed for many cognitive biases is “ecological mismatch”: while a person’s thought process leads to biased beliefs for i.i.d. processes studied in the laboratory, the same thought process would generate appropriate beliefs for the typical, real-world random processes people encounter. For example, in the case of under-inference, Winkler and Murphy (1973) posit that people may treat independent signals as if they were positively correlated because their real-world experience is with positively correlated signals. Such positive correlation would generate excessively-dispersed subjective sampling distributions and under-inference but *not* the invariance (or near-invariance) to sample size that Kahneman and Tversky (1972) observed because (under mild regularity assumptions) the Law of Large Numbers still “works” for positively correlated signals (see Hu, Rosalsky, and Volodin, 2008). Moreover, while ecological-mismatch arguments often have merit, we think the argument is unappealing in this context because the bias we call NBLLN is evident in examples with which subjects have a great deal of real-world experience, such as coin-flipping.²⁶

A related point is that even if NBLLN describes people’s beliefs in abstract laboratory experiments, it may be a poor model for economic applications if people form beliefs in an entirely different way in realistic settings. Since the clean evidence comes from abstract experiments, there is no direct evidence on this point. Our view, however, is that a model of people’s intuitive beliefs

²⁶We also note that in the case of the Law of Small Numbers, the opposite ecological-mismatch hypothesis is often proposed: that people ordinarily deal with *negatively*-autocorrelated signals. Typical real-world processes would have to have a fairly complicated form involving short-run negative autocorrelation and long-run positive autocorrelation to rationalize both the Law of Small Numbers and NBLLN.

about random processes should explain people’s beliefs about abstract situations like coin flips—the random processes people have most experience with and for which we can be most confident that subjects correctly understand what they are being asked.

Many have proposed conceptualizing under-inference in large samples as one consequence of the “representativeness heuristic,” according to which people draw inferences based on the degree of similarity between features of a sample and features of a population from which the sample might have been drawn. Indeed, Kahneman and Tversky (1972) present evidence for what we call NBLLN in precisely this context. Although NBLLN certainly seems consistent with representativeness, it is not clear how the logic of representativeness predicts the prototypical case of under-inference: e.g., an agent who observes 600 heads and 400 tails continues to put non-trivial probability on the coin being fair. Representativeness could explain this kind of observation if it is interpreted as inferences based on proportions, combined with the additional assumptions of reasonably accurate inferences in small samples and insensitivity to sample size, but that combination of assumptions essentially amounts to our model.

A natural alternative modeling approach would be to build a theory of “sample-size neglect,” in which, loosely speaking, an agent forms beliefs about a sample of any size as if it were a “medium-sized” sample of, say, size 7. Such a model would imply under-inference for sample sizes larger than 7 and over-inference for sample sizes smaller than 7. This is the formal model one might build to capture Griffin and Tversky’s (1992) verbal theory that people overweight the “strength” of the evidence (extremeness of the proportion of heads) and underweight the “weight” of the evidence (sample size). It is a common conceptualization and one which we found compelling enough to consider as our first (and more parsimonious) approach.

But we have come to the view that NBLLN and LSN are distinct phenomena: LSN is an error intrinsically tied to mistaken beliefs about specific *sequences* of random events, whereas NBLLN is an error specific to *sampling distributions*. Consistent with existing models of LSN (Rabin, 2002; Rabin and Vayanos, 2010), we believe that it is intrinsically linked to the gambler’s fallacy, which is an incorrect belief about sequences. In i.i.d. coin flips, the gambler’s fallacy manifests as a belief that a head becomes less likely than 50% following a streak of heads. If LSN were reversed in large samples—as implied by the sample-size-neglect theory—then people would have to believe that a long streak of heads makes a subsequent head *more* likely. But evidence indicates that the gambler’s fallacy seems to have as much force in large samples as in small samples. For example, Benjamin, Moore, and Rabin (2012) find that people think that the probability of a head following a streak of 9 heads from a fair coin is only 32%.²⁷ Benjamin, Moore, and Rabin also find that

²⁷One might think that the prediction of the sample-size neglect theory—that a head is believed to be more

people’s sampling distributions, which seem to exhibit NBLLN, are inconsistent with their beliefs about sequences.

In light of the inconsistency between people’s sampling-distribution beliefs and their sequence beliefs, our model of NBLLN cannot be used to make predictions regarding Barney’s beliefs about the likelihood of particular sequences. For example, since Barney believes that 10 heads out of 10 is more likely than it is, one might be tempted to conclude that he believes that the probability of a head following 9 heads is higher than it is (and that he believes the probability of any sequence yielding 5 heads out of 10 is lower than it is). Such an application of the model would yield false predictions. For this reason, none of our applications relied on Barney’s beliefs about sequences.

What is the psychology underlying people’s mistaken beliefs about sampling distributions? While we are not aware of any evidence on this issue, one intuition is that people fail to realize just how many combinations of a and b signals generate proportions close to the population mean. For example, our introspection suggests that people overestimate the likelihood of 90% heads in a sample of 100 largely because they vastly underappreciate how many more ways there are to get 50% heads than 90% heads: $\frac{\binom{100}{50}}{\binom{100}{90}} \approx \frac{10^{29}}{10^{13}} = 10^{16}$. While we know of no attempt to formulate NBLLN along such lines, we believe that such a model would share the main features and predictions of our model. Relative to our model, it would have the advantage that there would be no temptation to use the model to make predictions about the agent’s beliefs about the likelihood of specific sequences. It would have two disadvantages, however. First, in some circumstances, it would counterfactually predict “thin tails.” For example, for $\theta = .8$, suppose the agent believes the probability of h heads out of 10 is $r\left(\binom{10}{h}\right) (.8)^h (.2)^{10-h}$, where $r(\cdot)$ is a distortion function. In order to fit the evidence from Kahneman and Tversky (1972) that people think 5 heads out of 10 is more likely than 10 heads, it must be that $r\left(\binom{10}{5}\right) (.8)^5 (.2)^5 > r\left(\binom{10}{10}\right) (.8)^{10} (.2)^0$, or $\frac{r\left(\binom{10}{5}\right)}{r(1)} > 1024$. But then $\frac{r\left(\binom{10}{5}\right)}{r\left(\binom{10}{0}\right)} > \frac{\binom{10}{5}}{\binom{10}{0}} = 252$, which implies that people must *underestimate* the likelihood of 0 heads out of 10 relative to 5 heads. Second, this model would be harder to work with because the mean of the agent’s subjective sampling distribution would be incorrect.

likely following a short streak of heads but less likely following a long streak—provides a parsimonious account of both the gambler’s fallacy and its apparent opposite, the “hot hand fallacy.” This is a false parsimony, however, because shoe-horning the gambler’s fallacy and the hot hand fallacy into the same psychological mechanism generates counterfactual predictions about when they occur. As noted, the gambler’s fallacy occurs even after a long streak of heads, and as far as we are aware, the hot hand fallacy has never been observed for coins. Instead, the hot hand fallacy is usually understood as occurring in situations where an agent believes that the random process alternates between “hot” and “cold” rates.

7 Concluding Remarks

While the logic of NBLLN unambiguously predicts that people will extract far too little information from large samples, there are strands of literature both within psychology and within economics on “over-confidence” in beliefs. Rather than viewing over-confidence and under-confidence as fundamental biases in themselves, we view both as outcomes to be explained as a function of the information a person is confronted with. Our model of NBLLN highlights a feature of the decision-making environment—namely, sample size—that affects the degree to which an agent will draw too weak an inference from evidence. In Appendix C we combine NBLLN with LSN, which generates a bias toward too strong an inference from evidence, and the overall pattern we predict is: correct inference for samples of size 1, over-inference in small samples larger than 1, and under-inference in large samples. LSN exacerbates people’s tendency to rely on smaller samples.

In Section 3, we used the example of lack of demand for *Consumer Reports* to illustrate a fundamental implication of NBLLN: people underweight the information contained in a large sample. Our example of *Consumer Reports* is borrowed from Nisbett and Ross (1980), however, who used the contrast between pallid statistics and colorful anecdotes to illustrate a different and better-known phenomenon called “vividness bias”: people *overweight* vivid evidence in reaching their judgments, perhaps because of its emotional tug. For example, somebody’s graphic description of the horrors that ensued when her car broke down while trying to pick up her child from school may weigh more heavily on our judgment of which brand of car we should buy than summary statistics based on large samples of data.

NBLLN has two important implications for the study of vividness bias. First, NBLLN is a confound for evidence that has been used to establish its existence. Indeed, in their 1982 review paper, still considered authoritative, Taylor and Thompson (1982) find that the empirical support for vividness bias is surprisingly weak. Of special note, they also observe it rests almost entirely on evidence of the *comparative* over-use of vivid information relative to statistical information. Clearly, such comparative over-use could instead be due to under-use of the statistical data.

Second, even assuming an anecdote *is* genuinely over-used due to its vividness, NBLLN is nonetheless *also* needed in order to explain how the anecdote could outweigh a large sample of statistical information. Absent NBLLN, to outweigh *Consumer Reports*, the story-teller’s car experience would have to be overweighted by a factor of many thousands, but we think it is clearly implausible that vividness bias (if it exists!) could be that strong. By drastically attenuating the persuasiveness of the statistical data, NBLLN enables the overweighted sample of a single anecdote to prevail.

As we show formally in Appendix B, NBLLN also serves as an “enabling bias” for misbeliefs

people have about themselves. In particular, NBLLN implies that people will fail to learn the truth despite a lifetime of feedback about the outcomes from their own actions. Consequently, while people have optimistic priors about their own abilities and preferences for reasons unrelated to NBLLN, NBLLN explains how people remain uncertain about their own type and why optimistic priors do not give way to more realistic self-assessments. For example, NBLLN may explain why people persist in being overoptimistic about their ability on tasks that they regularly engage in. NBLLN may also explain why people remain uncertain regarding their own altruistic preferences—an otherwise-puzzling lack of knowledge that is a crucial ingredient for self-signaling to help explain altruistic behavior.

NBLLN can also matter directly in many economic applications we have not explored in this paper. Consider, for example, persuasion. Barney is again uncertain whether to buy a Volvo or a Lada. A Lada saleswoman has observed N signals about which car is better. She must reveal all the signals but can choose how to clump the signals when revealing them to Barney. Due to the strategic nature of the interaction, additional assumptions are needed. To illustrate, suppose that Barney is retrospective-acceptive but unaware of his own NBLLN—and hence does not realize that the clumping of signals will affect his beliefs. In that case, Barney will not draw any inferences about the state from the saleswoman’s behavior. The saleswoman can therefore maximally move Barney’s beliefs in favor of the Lada by clumping all the pro-Volvo signals together and separating out each pro-Lada signal. For any two distinct rates and any priors, if N is sufficiently large, then the saleswoman can make Barney arbitrarily confident that the Lada is superior.

Finally, we note that, while our model of NBLLN is defined only when the signals are i.i.d. and binomial, the intuition applies more broadly. There are some natural approaches to modeling NBLLN for non-i.i.d. signal sequences. Consider a binomial random process defined by a mapping from any initial rate, θ_0 , and any history of t observed signals, h_t , into a rate that the $(t + 1)^{\text{st}}$ signal will be an a -signal, $\theta(\theta_0, h_t)$. When Barney knows the initial rate is θ_0 , he forms his beliefs as if the initial rate were β , a random variable drawn from distribution $f_{\beta|\Theta}^\psi(\beta|\theta_0)$. For the first signal in a group, he believes that the probability of an a -signal is β , and for the $(t + 1)^{\text{th}}$ signal within that group, he believes that the probability of an a -signal is $\theta(\beta, h_t)$. This modeling approach can be applied not only when the signals truly are non-i.i.d., but also when an agent falsely believes they are non-i.i.d., due to another psychological bias, such as in LSN.

We also believe that there are natural extensions of our modeling approach to non-binomial cases. Suppose, for example, that the signals are normally distributed i.i.d. with known mean μ and variance σ^2 . We can imagine a cousin of Barney believes instead that signals are generated by a two-stage process, where a subjective mean ν is drawn from some distribution centered at μ , and

then the signals are drawn from a normal distribution with mean ν and variance σ^2 . Effectively, Barney thinks that common noise ν is added to every signal in a given sample. While Tommy believes that the mean of a large random sample of signals will converge to a point mass at μ , Barney's cousin believes it will converge to the density of ν . We could assume that the density of ν corresponds to the empirical large-sample beliefs, or for analytical tractability, we could assume that ν follows the conjugate prior distribution for the normal distribution, which is itself a normal distribution.

References

- Acemoglu, Daron, Victor Chernozukov, and Muhamet Yildiz**, “Learning and Disagreement in an Uncertain World,” 2009. Working paper.
- Barberis, Nicholas**, “The Psychology of Tail Events: Progress and Challenges,” *American Economic Review Papers and Proceedings*, 2013, *103* (3), 611–616.
- , **Ming Huang, and Richard H. Thaler**, “Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing,” *American Economic Review*, 2006, *96* (4), 1069–1090.
- Benartzi, Shlomo and Richard Thaler**, “Myopic Loss Aversion and the Equity Premium Puzzle,” *The Quarterly Journal of Economics*, 1995, *110* (1), 73–92.
- **and** – , “Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments,” *Management Science*, 1999, *45* (3), 364–381.
- Benjamin, Daniel, Don Moore, and Matthew Rabin**, “Biased Beliefs about Random Samples: Evidence from an Integrated. Experiment,” 2012. Working paper.
- Dillenberger, David**, “Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior,” *Econometrica*, 2010, *78* (6), 1973–2004.
- Edwards, Ward**, “Conservatism in human information processing,” in B. Kleinmuntz, ed., *Formal representation of human judgment*, Wiley, 1968.
- Griffin, Dale and Amos Tversky**, “The Weighing of Evidence and the Determinants of Confidence,” *Cognitive Psychology*, July 1992, *24* (3), 411–435.
- Hu, Tien-Chung, Andrew Rosalsky, and Andrei Volodin**, “On convergence properties of sums of dependent random variables under second moment and covariance restrictions,” *Statistics & Probability Letters*, 2008, *78* (14), 1999–2005.
- Kahneman, Daniel and Amos Tversky**, “Subjective Probability: A Judgment of Representativeness,” *Cognitive Psychology*, July 1972, *3* (3), 430–454.
- **and** – , “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 1979, *47* (2), 263–91.
- **and Dan Lovallo**, “Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking,” *Management Science*, 1993, *39* (1), 17–31.

- Keren, Gideon**, “Calibration and probability judgments: Conceptual and methodological issues,” *Acta Psychologica*, 1991, *77* (3), 217–273.
- **and Willem A. Wagenaar**, “Violation of Utility Theory in Unique and Repeated Gambles,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1987, *13* (3), 387–391.
- Klos, Alexander, Elke U. Weber, and Martin Weber**, “Investment decisions and time horizon: Risk perception and risk behavior in repeated gambles,” March 2005. Working Paper.
- Koszegi, B. and M. Rabin**, “Reference-Dependent Consumption Plans,” *American Economic Review*, 2009, *99* (3), 909–936.
- Nisbett, Richard E. and Lee Ross**, *Human inference: Strategies and shortcomings of social judgment*, Englewood Cliffs, NJ: Prentice Hall, 1980.
- Rabin, Matthew**, “Risk Aversion and Expected-Utility Theory: A Calibration Theorem,” *Econometrica*, 2000a, *68* (5), 1281–1292.
- , “Diminishing Marginal Utility of Wealth Cannot Explain Risk Aversion,” in Daniel Kahneman and Amos Tversky, eds., *Choices, Values, and Frames*, Cambridge University Press, 2000b, pp. 202–208.
- , “Inference by Believers in the Law of Small Numbers,” *The Quarterly Journal of Economics*, August 2002, *117* (3), 775–816.
- **and Dmitri Vayanos**, “The Gambler’s and Hot-Hand Fallacies: Theory and Applications,” *Review of Economic Studies*, 2010, *77* (2), 730–778.
- **and Georg Weizsacker**, “Narrow Bracketing and Dominated Choices,” *American Economic Review*, 2009, *99* (4), 1508–1543.
- **and Richard H. Thaler**, “Anomalies: Risk Aversion,” *The Journal of Economic Perspectives*, 2001, *15* (1), 219–232.
- Read, Daniel, George Loewenstein, and Matthew Rabin**, “Choice Bracketing,” *Journal of Risk and Uncertainty*, 1999, *19* (1-3), 171–197.
- Redelmeier, Donald A. and Amos Tversky**, “On the Framing of Multiple Prospects,” *Psychological Science*, 1992, *3* (3), 191–193.
- Samuelson, Paul A.**, “Risk and Uncertainty: A Fallacy of Large Numbers,” *Scientia*, 1963, *98*, 108–113.

Taylor, Shelley E. and Suzanne C. Thompson, “Stalking the Elusive ‘Vividness’ Effect,” *Psychological Review*, 1982, *89* (2), 155–181.

Tversky, Amos and Daniel Kahneman, “Belief in the Law of Small Numbers,” *Psychological Bulletin*, 1971, *76* (2), 105–110.

– **and** –, “Rational Choice and the Framing of Decisions,” *The Journal of Business*, 1986, *59* (4), S251–S278.

– **and** –, “Loss Aversion in a Riskless Choice: A Reference-Dependent Model,” *The Quarterly Journal of Economics*, 1991, *106* (4), 1039–1061.

Winkler, Robert L. and Allan H. Murphy, “Experiments in the Laboratory and the Real World,” *Organizational Behavior And Human Performance*, October 1973, *10* (2), 252–270.