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

Daniel J. Benjamin
Cornell University and NBER

Matthew Rabin
University of California — Berkeley

Collin Raymond*
University of Oxford

March 23, 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, Larry Phillips, Daniel Read, 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 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. Yet evidence indicates that people *also* do not believe in the Law of Large Numbers: they believe that even in very large random samples, proportions might depart significantly from the overall population rate. This paper develops a formal model of this error, which we call "non-belief in the Law of Large Numbers" and abbreviate by NBLLN.[1] Our goal is not to explain the source of this bias, nor to provide a unifying model of it and other biases, but rather to explore NBLLN's implications and assess its potential importance in economic decision-making. We show that NBLLN has a range of economic consequences, including causing too little or too much risk-taking, a lack of demand for information, and a persistence of incorrect beliefs despite large amounts of data. In addition, we identify and explore conceptual challenges with modeling NBLLN that are also likely to arise in modeling other biases in statistical reasoning.

Even though NBLLN has received far less attention from economists than other biases, extensive experimental evidence establishes that people have incorrect beliefs about and inferences from large- and medium-sized samples. An early example is Kahneman and Tversky (1972), who find that subjects seem to think sample proportions reflect a "universal sampling distribution," virtually neglecting sample size. In doing so, subjects vastly exaggerate the probability of unbalanced ratios in large samples. 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 non-belief in the Law of Large Numbers in a simple setting, where a person is trying to predict the distribution of—or make an inference from—a fixed sample 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 "θ-binomial distribution." But Barney, as we model him, believes that large-

[1] NBLLN 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.

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 NBLLN: 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 sample distribution" intuition from Kahneman & Tversky (1972), it also embeds *some* sensitivity to sample sizes, consistent with other evidence (e.g., Griffin and Tversky's 1992 Study 1).[3] Other models would share the basic features of NBLLN 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 if Barney applies Bayesian updating based on his wrong beliefs about the likelihood of different sample realizations, NBLLN 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. In Appendix B, we review and meta-analyze the extensive experimental evidence on inference. Consistent with the general features of our model—and contrary to the widespread impression that overconfidence is the pervasive direction of mistakes in beliefs—this evidence clearly indicates that the typical finding is under-inference, and this under-inference is especially severe in large samples.[4]

In the remainder of the paper, we draw out some of the consequences of NBLLN in a wide range of applications that cover many of the major areas in the economics of uncertainty, such as valuation

[3]Even 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.

[4]In light of this evidence, we do not know why NBLLN has not been widely embraced or emphasized by judgment researchers or behavioral economists. Perhaps 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. In our view, summarizing people as overly conservative or not conservative enough is manifestly the wrong way to parse human judgment. By focusing on the concrete biases at play and highlighting the co-existence of NBLLN with other biases, we hope to make clear that there is no contradiction.

of risky prospects, information acquisition, and optimal stopping. The applications highlight which features of the economic environment determine in which direction (e.g., more risk averse or less) Barney is biased relative to Tommy. Two basic properties of NBLLN appear throughout these analyses and tie them together: Barney under-infers from large samples, and Barney believes just about anything has a real chance of occurring.

Section 3 illustrates some of the basic economic implications of NBLLN, beginning with 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 NBLLN and 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, 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 3 also 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. This refusal must come from a plausible model of risk preferences, such as loss aversion, generating the intrinsic aversion to small risks. But even such an improvement in assumptions about risk attitudes would not generate the observed behavior if a person 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.

Both Sections 2 and 3 analyze a model of NBLLN when there is a single, "given" sample that Barney will observe. Yet information does not always arrive in a single package of signals. A person may hear a series of individual reports from random strangers at cocktail parties about their car experiences, while also reading large-sample statistics of car performance. If Barney pools each of his interlocutors' tales with the statistics from *Consumer Reports*, his inferences will be very different than if he separately updates his beliefs following each cocktail party anecdote, and then treats the *Consumer Reports* data as one big sample. Such cases confront us with a conceptual challenge intrinsic to the very nature of NBLLN: because Barney under-infers more for larger samples than smaller ones, he will infer differently if he lumps observations together versus separately. A model of NBLLN *must* involve a theory of how Barney groups information as a function of how it is presented to him and other features of his decision-making environment. Our dynamic model also makes clear, in turn, that to be generally applicable in economics, any model of biased sample inference must specify what a person believes about her future information processing. With little empirical research to guide us, in Section 4 we discuss and formalize various combinations of assumptions on 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. Different combinations of assumptions may be warranted by different perceptual, framing, and decisionmaking environments. Of special interest for some of our dynamic applications below is the possibility that Barney retrospectively groups signals differently than he prospectively anticipates he will. He may, for instance, plan to separately ask people at cocktail parties about their experiences, and prospectively focus on each conversation as if it is a separate signal; but then in retrospect, he may pool the conversations together as a single sample.

In Section 5, we explore Barney's behavior in various environments involving learning and inference. A number of conclusions follow from the central fact that, even when observing extensive evidence, Barney fails to reach appropriately strong confidence. For example, NBLLN 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. Absent NBLLN, after having processed from *Consumer Reports* a summary of the experiences of thousands of random strangers, hearing one random stranger recount a vivid story about her car experience—even if the story is overweighted a hundredfold—could not plausibly affect an agent's beliefs. Similarly, absent NBLLN, optimistic priors would give way to more realistic self-assessments after a lifetime of experience.

When agents gather information, variants of NBLLN predict not only that people will never figure out the truth when they rationally should, but that their efforts to learn may be enormously costly. Like Tommy, Barney will plan to quit his costly information acquisition once he reaches some threshold of confidence. And, like Tommy, he may stop gathering signals very quickly if information is decisive. But because Barney tends to infer far less from signals than Tommy, when the initial signals are mixed, Barney may continue trying to learn even after many signals. Indeed, if Barney prospectively anticipates separately updating using each arriving signal but actually pools them retrospectively, Barney may become stuck in a "learning trap": he persistently expects to soon be confident enough to stop experimenting, or buying information, but, because he never achieves the confidence he anticipates, continues his costly efforts forever.

In Section 6, we discuss why we think our model is more compelling than alternative possible explanations and modeling approaches—both fully rational and not fully rational—that might seem to accommodate the psychology evidence. Our model, of course, ignores other important departures from Bayesian inference—such as base-rate neglect and belief in the Law of Small Numbers—that seem separable from NBLLN. But it also omits features—including the psychophysics of diminishing sensitivity, as well as unwillingness to hold or express extreme beliefs—that, as alternative sources of under-inference from large samples, are less separable. In Appendix A, in fact, we present a (complicated) formal model embedding some of these other errors along with NBLLN. Guided by this formal model, in Appendix B we attempt to give a fairly exhaustive review of the empirical work on sampling predictions and inference. We believe this review makes clear that our model of NBLLN is capturing a broad empirical reality.

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 $N \in \{1,2,...\}$ by $S_N \equiv \{a,b\}^N$, and we denote an arbitrary clump (of size N) by $s \in S_N$.[5] 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 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)

[5] Note 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. Conserving notation here, in Section 5.4 we use t to index the clumps of signals.

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. When generalizing the model, complicating conceptual challenges arise. Some of our analysis in fact concerns precisely these complications, but we defer discussion of these issues and ways to handle them until Section 4.

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

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

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. 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 NBLLN 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.[6]

[6]In keeping with this interpretation, Barney does not believe the realized β is a real feature of the coin,

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 $f^{\psi}_{ß|\Theta}(\beta|\theta)$. We refer to $f^{\psi}_{ß|\Theta}$ as Barney's *subjective rate distribution* and assume that it has the following properties:

A1. *For all β and θ, $F^{\psi}_{ß|\Theta}(\beta|\theta)$ is absolutely continuous in β, and $f^{\psi}_{ß|\Theta}(\beta|\theta)$ 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 $f^{\psi}_{ß|\Theta}(\beta|\theta)$ has full support on $[0,1]$.*

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

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

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

Assumptions A1 and A2 are mild and consistent with the experimentally elicited densities of people's beliefs about the distribution of signals in large samples. A1 differs from A1′ in allowing the density functions 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.

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 $f^{\psi}_{ß|\Theta}(\beta|\theta)$—especially in the spirit of the type of diminishing-sensitivity evidence discussed in Appendix A—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.[7] It holds for our main example of the beta distribution and more generally is useful for

and is certainly not an object he makes inferences about. Instead, it is a representation of Barney's subjective uncertainty that the θ will manifest itself in a given sample. By comparison, Acemoglu, Chernozukov, and Yildiz (2009) analyze a model that is formally similar to ours but *is* a model of parameter uncertainty. Similar to our Proposition 2 below, they show that a Bayesian agent fails to learn the state with certainty even after observing an infinite number of signals if he is uncertain about the meaning of signals. In contrast, we assume that Barney has a concrete belief about the meaning of signals, but he thinks that their meaning can vary from one sample to the next. In the dynamic applications we develop below, this interpretation will be much more than an aspiration to get the psychology right but rather an integral part of the formal model. For instance, in situations where Barney must predict further signals after observing his first 100 signals, we assume his expected proportions are still θ, rather than being influenced by the first signals as the parameter-uncertainty interpretation would suggest.

[7] In particular, Griffin and Tversky asked subjects to infer the likelihood that a coin is biased $\theta_A =$

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 says that the mean of Barney's subjective rate distribution is the known objective rate. Although we rely on A4 extensively in the analysis, it is in fact violated in existing data. Below we discuss how it is violated, and in Appendix A, 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

$$f^{\psi}_{S_N|\Theta}(s|\theta) = \int_{\beta\in[0,1]} f_{S_N|\text{ß}}(s|\beta)\, f^{\psi}_{\text{ß}|\Theta}(\beta|\theta)\, d\beta, \tag{1}$$

where $f_{S_N|\text{ß}}(s|\beta)$ is the (correct) probability of observing s *if the rate were* β, and this is averaged over the density of subjective rates, $f^{\psi}_{\text{ß}|\Theta}(\beta|\theta)$. 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.[8]

Lemma 1. *Assume A1-A4. Barney does not believe in LLN: for any* $\theta \in \Theta$ *and interval* $[\alpha_1, \alpha_2] \subseteq [0, 1]$*,*

$$\lim_{N\to\infty} \sum_{x=\lfloor \alpha_1 N\rfloor}^{\lceil \alpha_2 N\rceil} f^{\psi}_{S_N|\Theta}(A_s = x|\theta) = F^{\psi}_{\text{ß}|\Theta}(\beta = \alpha_2|\theta) - F^{\psi}_{\text{ß}|\Theta}(\beta = \alpha_1|\theta) > 0.$$

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.

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 empirical beliefs in studies such as those illustrated in Figure 1, drawn from Kahneman and Tversky (1972). The black, gray, and white bars—which correspond to people's reported beliefs regarding samples of size 10, 100, and 1000, respectively—are virtually the same height. This distribution presumably corresponds to people's large-sample beliefs about sample proportions and thus could be directly

.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.

[8] All proofs are elevated to Appendix C.

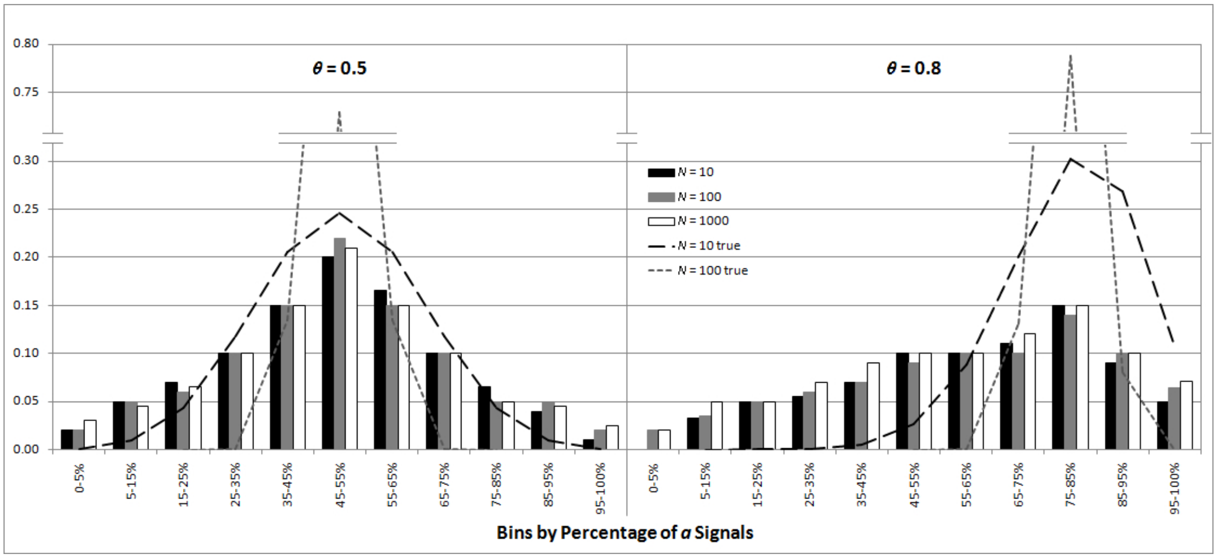

Figure 1: Evidence from Kahneman and Tversky (1972)

assumed to be the density of $\beta|\theta$.

Although Assumptions A1 and A2 are consistent with Figure 1, A4 is not. Beliefs for $\theta = .5$, depicted in the left panel, naturally have mean approximately equal to .5. However, beliefs for $\theta = .8$, depicted in the right panel, have 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 A and B, 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 A 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 $N \in \{1, 2, ...\}$:*

1. $E^{\psi}_{S_N|\Theta}\left(\frac{A_s}{N}|\theta\right) = E_{S_N|\Theta}\left(\frac{A_s}{N}|\theta\right) = \theta$.

2. $F_{S_N|\Theta}(A_s|\theta)$ *second-order stochastically dominates (SOSD)* $F^{\psi}_{S_N|\Theta}(A_s|\theta)$*, and* $Var^{\psi}_{S_N|\Theta}\left(\frac{A_s}{N}|\theta\right) \geq$

$Var_{S_N|\Theta}\left(\frac{A_s}{N}|\theta\right)$ *with strict inequality for* $N > 1$.

3. $Var^{\psi}_{S_N|\Theta}\left(\frac{A_s}{N}|\theta\right)$ *is strictly decreasing in* N.

4. $F^{\psi}_{S_N|\Theta}\left(A_s|\theta'\right)$ *first-order stochastically dominates (FOSD)* $F^{\psi}_{S_N|\Theta}\left(A_s|\theta\right)$ *whenever* $\theta' > \theta$.

Part 1 states that Barney, like Tommy, expects the average proportion of a's in the sample to be θ. An immediate and important corollary of Part 1 is that Barney's beliefs coincide with Tommy's when $N = 1$. 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.

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 B) 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.[9] Consequently, Barney's beliefs after observing a particular clump $s \in S_N$ are $f^{\psi}_{\Theta|S_N}(\theta_A|s) = \frac{f^{\psi}_{S_N|\Theta}(s|\theta_A) f_\Theta(\theta_A)}{\sum_{\theta\in\Theta} f^{\psi}_{S_N|\Theta}(s|\theta) f_\Theta(\theta)}$ and $f^{\psi}_{\Theta|S_N}(\theta_B|s) = \frac{f^{\psi}_{S_N|\Theta}(s|\theta_B) f_\Theta(\theta_B)}{\sum_{\theta\in\Theta} f^{\psi}_{S_N|\Theta}(s|\theta) f_\Theta(\theta)}$.

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:

[9]Therefore, 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 A discusses and formalizes how to combine NBLLN with base-rate neglect.

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 $N \to \infty$, Barney's posterior ratio converges almost surely (with respect to the true probability distribution over events) to a positive, finite number:*

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

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.

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, because, given his wrong model of the data-generating process, Barney's model of the world at any given moment is Bayesian, 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: for any $N \geq 1$, $E^{\psi}_{\Theta|S_N}(\theta|s) = E^{\psi}_{\Theta}(\theta)$. In fact, however, Barney's actual beliefs do not have this martingale property. As per Proposition 2, Barney's posterior beliefs will converge to a point mass that is not equal to the mean of his prior beliefs.

We next turn to inference in finite samples. Because Barney's subjective sampling distribution is correct when the sample size is 1, he will draw correct inferences 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 $N \in \{1, 2, ...\}$ 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 B 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, we impose the functional form of the beta distribution:

$$f^{\psi}_{\text{ß}|\Theta}(\beta|\theta) = \beta^{\theta\psi-1}(1-\beta)^{(1-\theta)\psi-1}\frac{\Gamma(\psi)}{\Gamma(\theta\psi)\Gamma((1-\theta)\psi)}, \tag{3}$$

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 $x > 0$.[10] The properties A1-A4 are satisfied (see Appendix C, Lemma $\beta 5$), 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.[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 $\psi \to \infty$.

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-

[10] The more common way of writing this beta density is $f^{\psi}_{\text{ß}|\Theta}(\beta|\theta) = \beta^{\theta\psi-1}(1-\beta)^{(1-\theta)\psi-1}\frac{(\psi-1)!}{(\theta\psi-1)!((1-\theta)\psi-1)!}$. 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, $\Gamma(x) = (x-1)!$.

[11] The 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 NBLLN. These are presented in Lemma $\beta 4$ in Appendix C. 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.

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 NBLLN 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 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: $\left| f^{\psi}_{\Theta|S_N}(\theta_A|s) - f_{\Theta}(\theta_A) \right|$. 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 $N \in \{1, 2, ...\}$ 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.*

[12] For 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.

[13] We have simulated Barney's and Tommy's expected change in beliefs for a range of parameter values: for each of $\psi \in \{1, 2, ..., 30\}$ 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.

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 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 Static Applications

This section explores the implications of the simple model from the last section. First, 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. We then turn to how NBLLN's basic implications for an agent's subjective sampling distribution play out in various gambling and investment environments. Finally, we build on that analysis to examine the value of information in an environment of choice under risk.

3.1 Lack of Demand for Large Samples

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

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

also overestimates the probability of an extreme outcome that would allow for stronger inferences. Therefore, perhaps counterintuitively, if Barney and Tommy can both choose the total number of signals to purchase, Barney can actually end up being *more* certain about the state of the world.

For a *large* sample, however, Barney unambiguously has a lower willingness to pay than Tommy does. Tommy correctly anticipates 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.

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 knowing the state is valuable: $u(\mu_A, A) > u(\mu_B, A)$ and $u(\mu_A, B) < u(\mu_B, B)$. Furthermore, 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. Public health announcements are more effective if they feature vivid anecdotes rather than statistics, and the car purchaser who actually consults *Consumer Reports* is the exception rather than the rule. In some cases, such as restaurant satisfaction ratings, people may correctly expect large samples not to be very useful because preferences are heterogeneous. When preference heterogeneity is less of an issue, however, e.g., the frequency of car battery failure, the lack of demand for statistical data is a major "dog that didn't bark" clue that implicates NBLLN.[15]

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

[15] In principle, people may distrust large datasets because of selection issues—worrying (say) that those getting vaccines are less prone to autism than those not and finding statistics that vaccines do not increase autism unpersuasive. Yet 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.

immediate social network, or anecdotes from strangers.

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 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. Furthermore, fix payoffs* $u(\mu, \omega)$ *so that knowing the state is valuable:* $u(\mu_A, A) > u(\mu_B, A)$ *and* $u(\mu_A, B) < u(\mu_B, B)$*. 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.[16] In section 5.3 below, 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.

[16]As such, even the limited number of cases where decisionmakers seem to hunger for large datasets is not proof of instances where NBLLN is not having an impact.

3.2 Perceived Aggregate Risk

We now consider the implications of NBLLN for risky investments and gambles, in which the agent's payoff depends on the outcome of a series of random draws. The comparison between Barney and Tommy 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 that outcome is more likely than Tommy does.

A leading special case is where the particular outcome is an extreme sample proportion, which in general Barney thinks is more likely to occur than Tommy does. For example, suppose the agent is deciding whether to play a lottery game where he picks N numbers, he wins a prize if all his picks match the numbers chosen randomly by the lottery, and the probability that any given number he picks is chosen by the lottery is θ. If $N > 1$, then Barney believes his chance of matching all the numbers is higher than it is, and hence his 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 gets a payoff 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. Since Barney believes his chance of getting at least one job offer is lower than it is, Barney's 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.[17] 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 $w(A_S)$.

Proposition 8. *Assume A1-A4. Fix a risky gamble (θ, N). If $u(w(A_S))$ is a concave (resp.,*

[17] Because the outcome space is discrete, in this subsection, as well as the following one, we use the standard definition that $u(w(A_S))$ is a convex (resp., concave, linear) function of A_S if $u(w(A_S - 1)) + u(w(A_S + 1)) - 2u(w(A_S)) \geq$ (resp., $\leq, =$) 0.

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 2, 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 will often 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.

3.3 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, not 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.[18] Somebody with a simple, piecewise-linear loss-averse utility function

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

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.[19] 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% of participants said they would turn down a single scaled-down Samuelson type bet (win \$100 or

[18] Samuelson 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.

[19] Although 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.

lose \$50), but fully 25% also reject the 100-times repeated gamble.[20]

NBLLN helps explain why many people turn down these gambles.[21] 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 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.[22,23]

Formally, 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

[20] In 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 B.

[21] Our 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.

[22] Note 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.

[23] Benartzi 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%.

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 $\widehat{\lambda} > 1$ such that: if $\lambda < \widehat{\lambda}$, then there is some N' sufficiently large such that Barney will accept N repetitions of the gamble for all $N > N'$; and if $\lambda \geq \widehat{\lambda}$, 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.[24]

Using the one-parameter functional form for Barney, calibrations suggest that NBLLN 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.[25]

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 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 NBLLN in the context of long-term investing: university employees vastly overestimated the probability that equities would lose money

[24] Preferences 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, NBLLN predicts rejection precisely due to mistaken beliefs about the *combined* outcomes of the gambles. Furthermore, NBLLN predicts risk-seeking behavior in settings where these other models would not, e.g., in the lottery example in the previous subsection.

[25] Incorporating some of the biases missing from our model of NBLLN (such as the sampling-distribution-tails diminishing sensitivity bias (SDTDS) described in Appendix A) that would generate even fatter tails in subjects' subjective sampling distribution helps reduce the required level of loss aversion even more—although still not to a reasonable level of λ such as 2.25. Indeed, even if an individual exhibited the most extreme form of NBLLN *and* SDTDS (and possibly also probability weighting), putting equal weight on every possible outcome of the repeated gamble, the required level of loss aversion would be around 4 (because only 1/3 of the outcomes are losses). This exercise implies that some other bias is also implicated in turning down the repeated bet. Whatever it is might also explain why some subjects turn down the aggregate bet even when it is presented in histogram form.

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 NBLLN 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 NBLLN 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 NBLLN.

3.4 Risky Gambles and the Value of Information

An important implication of Barney's distorted predictions is that he does not believe that the rate pins down the distribution of good and bad outcomes in an investment as much as it does. This in turn means that if there is uncertainty about the rate, then Barney's willingness to pay to reduce this uncertainty may differ from Tommy's. In this subsection, we explore the implications of NBLLN for the value of information about a risky gamble.

For simplicity, we return to the set-up with classical risk preferences, $u(w)$, in which utility depends only on final wealth. As before, the agent is deciding whether or not to make a risky investment (θ, N) whose monetary payoff, $w(A_S)$, is increasing in the number of a-signals. Here, however, the agent is uncertain about whether or not a mutual fund he is considering investing in has a talented manager. If the manager is talented, the rate of a-signals is $\theta = \theta_A$; if untalented, it is $\theta_B < \theta_A$. The agent's alternative to the risky investment is a safe asset that generates known wealth w_0. The agent has a prior over whether the manager is talented but can reduce his uncertainty by incurring a cost to consult with an investment adviser. We aim to compare Barney's willingness to pay for this information with Tommy's.

Unlike in Section 3.1, where we compared how Barney and Tommy value different sample sizes, here we model the information as a single signal. Doing so allows us to hold constant how much the agent learns from the signal; we focus on how Barney and Tommy value information differently due to how it might affect their actions differently.

We suppose that consulting the investment adviser reveals a signal $\sigma \in \{L, H\}$. A high signal H is more likely if the manager is talented: $f_{\Sigma|\Theta}(\sigma = H|\theta = \theta_A) > f_{\Sigma|\Theta}(\sigma = H|\theta = \theta_B)$ and $f_{\Sigma|\Theta}(\sigma = L|\theta) = 1 - f_{\Sigma|\Theta}(\sigma = H|\theta)$.

A critical factor for the comparison between Barney's and Tommy's demand for information is the shape of utility as a function of the number of good draws, $u(w(A_S))$. As discussed in Section 3.2, this utility depends both on risk preferences and on how the number of a-signals translates into monetary payoffs. If $u\left(w\left(A_S\right)\right)$ is linear, then only the expected number of a-signals matters, which is the same for Barney and Tommy. In that case, Barney's willingness to pay for the signal σ is exactly the same as Tommy's. Proposition 10 compares Barney's willingness to pay with Tommy's, focusing on the case where $u\left(w\left(A_S\right)\right)$ is concave. (If $u\left(w\left(A_S\right)\right)$ is convex, the logic is analogous, and all the conclusions are simply reversed.)

Proposition 10. *Assume A1-A4. Suppose an agent with initial wealth w_0 can choose whether or not to take a risky gamble (θ, N) whose monetary payoff, $w\left(A_S\right)$, is increasing in A_S. The agent does not know whether $\theta = \theta_B$ or $\theta = \theta_A > \theta_B$, and has priors $f_\Theta\left(\theta_A\right) = 1 - f_\Theta\left(\theta_B\right) \in (0,1)$.*

Suppose $u\left(w\left(A_S\right)\right)$ is a concave function of A_S. In that case, if given the prior Barney invests, then so does Tommy. Moreover:

1. *If given the prior neither Tommy nor Barney invests, then Tommy's willingness to pay for a signal σ is higher than Barney's.*

2. *If given the prior both Tommy and Barney invest, then Barney's willingness to pay for a signal σ is higher than Tommy's.*

3. *If given the prior Tommy invests while Barney does not, then Barney's willingness to pay for a signal σ is higher than Tommy's if and only if*

$$\begin{aligned} &\sum_{\sigma \in \Sigma_{\mathbb{I}}^{\psi}} f_\Sigma(\sigma) \left\{ E_{\Theta|\Sigma}\left[E^{\psi}_{S_N|\Theta}\left[u(w(A_S))|\theta\right] |\sigma \right] - u(w_0) \right\} \\ \geq &\sum_{\sigma \in \Sigma_{\mathbb{S}}} f_\Sigma(\sigma) \left\{ u(w_0) - E_{\Theta|\Sigma}\left[E_{S_N|\Theta}\left[u(w(A_S))|\theta\right] |\sigma \right] \right\}, \end{aligned}$$

where $\Sigma_{\mathbb{I}}^{\psi} \subseteq \{L, H\}$ is the set of signals such that given his posterior Barney would invest; and $\Sigma_{\mathbb{S}} \subseteq \{L, H\}$ is the set of signals such that given his posterior Tommy would not invest.

Alternatively, if $u\left(w\left(A_S\right)\right)$ is a convex function of A_S, then all of the conclusions in the previous paragraph hold with "Barney" and "Tommy" switched.

A key piece of the logic underlying the proposition is that observing the signal has value only to the extent that it would cause an agent to switch actions, relative to what the agent would do in the absence of observing the signal. The three parts of the proposition address each of the possible cases of whether Tommy and Barney would undertake the risky investment given the prior. The formulation of these cases in terms of parameter restrictions is contained in the proof.

The impossibility of the fourth potential configuration—that Barney would invest but Tommy would not—follows directly from Proposition 8: for both possible θ's, because Barney's beliefs are a mean-preserving spread of Tommy's and $u(w(A_S))$ is concave, the gamble is at least as attractive to Tommy as it is to Barney. Indeed, this same observation also plays a role in the logic underlying the first and second parts of Proposition 10. The first part of Proposition 10 addresses the case where, given the prior, neither Tommy nor Barney would invest. In that case, the signal has value to the extent that it might cause an agent to switch actions and undertake the investment. Since the investment is more attractive to Tommy (in either state), any signal realization that leads both agents to invest has greater value to Tommy. Moreover, a weaker signal is required for Tommy to switch his action, so the set of signal realizations that are valuable is at least as large for Tommy. Due to both of these reasons, Tommy's willingness to pay for the signal is higher.

The second part supposes that, given the prior, both Tommy and Barney would invest. Now, the same logic operates in reverse: the signal has value to the extent that it might cause the agent to switch to the safe option. Since Barney finds it less attractive than Tommy to undertake the investment (in either state), any signal realization that leads both agents to switch to the safe option is more valuable to Barney, and the set of signal realizations that are valuable is at least as large for Barney. Thus, Barney's willingness to pay for the signal is higher.

The third case is where given the prior, Tommy would invest, but Barney would not. Barney's willingness to pay for the signal is greater than Tommy's exactly when Barney's expected gain from switching his action times the probability that he observes a signal that causes him to change his action exceeds Tommy's expected gain from switching his action times the probability that he observes a signal that causes him to change his action; the inequality is the formal statement of this condition.

4 NBLLN 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 NBLLN, 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 thinking about this issue, and—in order to facilitate studying the range of possible consequences of NBLLN in dynamic decision-making environments as well as to provide guidance for future experimental work—we formulate plausible alternative assumptions.[26] We review in Appendix B the scant and somewhat contradictory evidence we can find about which assumptions are appropriate; since there is so little

[26]Moreover, the implications of other non-Bayesian models of judgment biases—such as base-rate neglect—similarly exhibit sensitivity to how data are framed. As such, the range of approaches we outline here may prove useful for studying those other biases.

evidence, our proposals are tentative. We keep our discussion in this section informal, relegating formal definitions to Appendix D.

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 to have this be an exogenous assumption that is not *per se* related to NBLLN, and ideally is 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.

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.[27]

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 at a future date. More generally, we impose the

[27]While 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 Section 6, 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.

following constraint on any model of NBLLN: 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 NBLLN from the single-clump model in Section 2 generalizes to every 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 to process all 11,000 signals together as a single group after already observing the first 10,000 signals, 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 $f^{\psi}_{\text{ß}|\Theta}(\beta|\theta = .5)$ 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 $f^{\psi}_{\text{ß}|\Theta}(\beta|\theta)$ 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 that one would employ if one interpreted it as a fully-rational model with uncertainty about the parameter β. (Barney's predictions regarding a single clump and a single decision node, as in Sections 2 and 3, are of course covered by the prospective-acceptive case.)

Which processing assumptions are the most psychologically plausible? We hypothesize that in many situations, Barney will expect to process data more finely in the future than he actually will do retrospectively.[28] For example, if Barney plans to keep talking to friends one by one until he feels confident, he might think ahead with attention to each separate signal, focusing on how he will update from current beliefs after his next conversation. But then in retrospect, after he has talked to his next friend, he may quite naturally treat that friend's information symmetrically with all the previous conversations and take stock of his current information by thinking together about all the advice he has received. Accordingly, in some of our applications, we explore the implications

[28]We emphasize, however, that we are hypothesizing about what is typical, and we can imagine alternative possibilities in certain situations. For example, Barney might be retrospective-acceptive but pool signals he has not yet observed: before he talks to the 10 friends (say) he plans to talk to, he may not attend to the time separation of the information and not realize that he will update his beliefs story-by-story as he goes along. But retrospectively, he may be retrospective-acceptive, distinguishing colorful details of his friends' stories.

of Barney being retrospective-pooling and prospective-acceptive.

As an implication of all the other ways he is rational, Tommy always processes information the way that he expects to process information. We call this property *processing-consistency*. Despite his irrationality, Barney shares this property if his retrospective and prospective thought processes coincide. In particular, if Barney is retrospective-acceptive and prospective-acceptive, then he is processing-consistent and accurately forecasts what his own future beliefs will be after he observes a sequence of signals. In contrast, if Barney is retrospective-pooling and prospective-acceptive—as we have argued is often plausible—then he is not processing-consistent. As a result, he may behave in a time-inconsistent way, e.g., expecting to learn a lot from purchasing a large number of signals, but remaining uncertain after observing the signals and therefore preferring to purchase yet more signals. This time-inconsistency will play a role in some of the applications we study.

Besides lack of evidence, an additional and major reason the hypotheses in this section are tentative is that we have completely sidestepped the issue of when and how Barney might "think through" his beliefs more fully. Even if Barney is not processing-consistent, our *model* of Barney's beliefs is internally consistent. Barney's *beliefs* themselves, however, may not be internally consistent, and this raises additional conceptual and practical issues in applying a model of NBLLN.[29] For example, a teacher could elicit Barney's belief about the likelihood that a first signal will be a, the likelihood that a second signal will be a conditional on the first signal being a, and the likelihood that a sequence of two signals will be aa; then the teacher could point out that the product of the first two does not equal the second. In fact, even our coherence constraint above could fail depending on the questions a teacher asked Barney. For example, suppose Barney expects to observe 10,000 signals from *Consumer Reports*, then make some payoff-relevant decision, and then observe another 1,000 signals. If a teacher asks Barney to forecast all 11,000 signals, then Barney would presumably do so according to the single-clump model with $N = 11,000$—in violation of our modeling constraint. Even in the absence of a "teacher," Barney might ask himself such questions.[30] While we flag these issues, and we think they are important subjects for future research,

[29]The internal inconsistency we highlight here does not arise in "false-model Bayesian" models of biased beliefs such as Barberis, Shleifer and Vishny (1998), Rabin (2002), and Rabin and Vayanos (2010). In these models, biases are formulated as agents holding the wrong theory as to the statistical structure of the world, but as being fully Bayesian in their interpretation of data within that structure. So long as all events that are possible in the true world are also possible in the agents' imagined world, no internal inconsistency can arise. (Though even in these models, it may be very likely that the agent will observe a sequence of signals that he perceives to be very unlikely.) Processing-consistent variants of Barney likewise reduce to a "false-model Bayesian" theory. The processing-*in*consistent variants of NBLLN, however, assume an intrinsically non-Bayesian thought process. The modeling challenges associated with internally-inconsistent beliefs are not specific to NBLLN and will arise in any model of belief formation that is fundamentally non-Bayesian.

[30]Importantly, a decision Barney faces might itself naturally cause Barney to ask himself such questions. For example, imagine Barney is making a decision whose payoff depends on whether the state is A or B. He could purchase one signal, and then decide whether to purchase a second signal, or he could purchase two signals all at once at a discount. When deciding what to do, it seems natural that Barney would ask

we proceed in subsequent sections with the assumption that Barney does not think through the inconsistencies in his own beliefs.

5 Dynamic Applications

This section explores some implications of NBLLN for economic environments where the grouping of signals matters. We begin by briefly revisiting the lack of demand for large samples, this time focusing on how NBLLN leads to the relative overweighting of a set of individual datapoints. We then turn to settings involving dynamic inference, applying variants of our model to information acquisition, learning, and experimentation. While we highlight how NBLLN has different implications depending on specific features of these environments, we note that two themes run through all these analyses: Barney believes that anything can happen, and he under-infers from large samples.

5.1 Enabling "Vividness Bias"

In section 3.1, 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 phenomenon, "vividness bias": people *overweight* vivid evidence in reaching their judgments. 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 in our judgment of which brand of car we should buy than summary statistics based on large samples of data.

Vividness bias is probably better known among researchers than NBLLN. Here we note here two relevant implications of NBLLN. First, NBLLN is a confound for evidence that has been used to establish the existence of vividness bias. 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

himself what he would conclude after observing each of the three possible outcomes: he observes 1 signal, 1 signal followed by 1 signal, and 2 signals together. Having explicitly asked himself about these possibilities, it seems odd that Barney would—as assumed if Barney is prospective-acceptive—expect to conclude less from the 2 signals together than the 2 signals individually.

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 could be that strong (especially given the murkiness of the evidence for its existence). In the presence of NBLLN, the anecdote could overwhelm the statistics even if vividness bias is much weaker.

5.2 Sequential Information Acquisition

To turn to examples of agents acquiring information, consider an agent who is uncertain about which of two possible states of the world, $\omega \in \{A, B\}$, is true. State A has prior probability $0 < f_\Theta(\theta_A) < 1$, and state B has prior probability $f_\Theta(\theta_B) = 1 - f_\Theta(\theta_A)$. In state A, the probability of an a-signal is θ_A, and the probability of a b-signal is $1 - \theta_A$. In state B, the probability of an a-signal is $\theta_B < \theta_A$, and the probability of a b-signal is $1 - \theta_B$. Depending on the particular application, the agent can take actions, or observe outcomes and signals that can inform him about the state of the world.

In this subsection, we explore sequential information acquisition. Imagine that Barney is (still) trying to decide whether to buy a Volvo or a Lada. He polls one friend at a time, asking which car is better. Conditional on the responses that he receives, he can decide to ask more friends, or to stop and choose a car. Formally, each period $t = 1, 2, ...$, the agent can choose to purchase a single signal at cost $c > 0$ or take an action $\mu \in \{\mu_A, \mu_B\}$.[31] If the agent takes an action, he gets payoff $u(\mu, \omega)$, which equals 1 if the action matches the state ω and 0 otherwise, and the agent faces no further decisions. If the agent decides to purchase an additional signal, he sees the realization of the signal, and he proceeds to the next period. The agent lives forever and seeks to maximize the expected action payoff minus expected signal-purchase costs; if the agent purchases κ signals and then takes action μ, his utility is $u(\mu, \omega) - \kappa c$. We assume no discounting, so that the only reason an agent would stop acquiring information before being absolutely certain is the cost c of obtaining an additional signal.

For Tommy, the characterization of optimal behavior is well-known (e.g., Wald, 1947). Each time Tommy purchases a signal, he updates his posterior beliefs. His optimal behavior is characterized by two probabilities, ν_l and ν_h, with $0 < \nu_l < \nu_h < 1$. If and only if the posterior probability

[31] Here and later, we restrict the agent to purchasing information a single realization at a time. For Tommy, this restriction entails no loss of generality. If we allowed Barney to choose how many signals he could purchase each period, however, Barney would have to think about what he would learn from purchasing two signals sequentially in order to compare it to purchasing two signals simultaneously. Modeling this thought process raises challenges—the same as those mentioned in footnote 30—that we sidestep in this paper. Nonetheless, we think the results and intuitions about Barney we develop below will generalize, as long as Barney's thought process does not eliminate his NBLLN.

of state A exceeds ν_h, he stops and takes action μ_A; if and only if it goes below ν_l, he stops and takes action μ_B. Tommy continues to purchase signals as long as his posterior beliefs remains between ν_l and ν_h. But because his posterior ratio is a martingale process, Tommy will eventually feel strongly enough to take an action almost surely. Importantly, as $c \to 0$, $\nu_l \to 0$ and $\nu_h \to 1$, and so as information becomes cheaper, the agent requires more extreme beliefs to stop purchasing information.

Because we assume that the signals arrive one at a time and that Barney is prospective-acceptive, Barney expects to group the signals as samples of size 1. Since Barney *expects* to behave exactly like Tommy, his policy is the same as Tommy's, with the *same* thresholds ν_l and ν_h determining when he stops and takes an action. If Barney is retrospective-acceptive, his beliefs and behavior will be identical to Tommy's.

If he is retrospective-pooling, however, Barney's behavior can differ qualitatively from Tommy's. In this case, the impact of an additional signal on his posterior beliefs is smaller than for Tommy. Moreover, the marginal impact will approach zero as his sample of observed signals grows. This is because Barney's inference becomes more and more driven by the proportion of a-signals, which is less affected by an additional signal in a larger sample. However, Barney *believes* that an additional signal will matter, regardless of the sample size he has already observed. As a result, Barney can become stuck in a *learning trap*, in which he purchases signals forever, but they will never change his confidence in the state of the world enough for him to stop. The first part of Proposition 11 shows that such a learning trap can occur and becomes more likely the more signals he has already observed:

Proposition 11. *Assume Barney has the beta-distribution functional form given by equation (3). Fix payoffs $u(\mu, \omega)$, rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A > \theta_B$ and prior $f_\Theta(\theta_A) = 1 - f_\Theta(\theta_B) \in (0,1)$. Suppose Barney is prospective-acceptive and retrospective-pooling.*

1. *For all $\overline{p} < 1$, there exists $\bar{c} > 0$ such that for all $c \leq \bar{c}$, Barney buys an infinite number of signals with probability $p > \overline{p}$. Furthermore, suppose that Barney, before buying any signals, has a positive probability of buying an infinite number of signals. Then for any $\varepsilon > 0$, there exists $N_\varepsilon > 2$ such that if Barney buys an additional signal after having already bought N_ε signals, the probability of Barney buying a finite number of signals from then on is less than ε.*

2. *Suppose $\theta_A = 1 - \theta_B$. Suppose Barney is willing to buy an additional signal when his posterior probability (of state A) is equal to q, and suppose Barney's posterior is q after observing N signals. If Barney's posterior probability of state A is q after observing $N' > N$ signals, then*

the probability that Barney will buy an infinite number of signals is weakly higher after he has observed the N' signals than it was after the N signals.

3. *Again, suppose $\theta_A = 1 - \theta_B$ and the prior $f_\Theta(\theta_A) \geq .5$. For any $\varepsilon > 0$, there exists $N > 2$ such that if Barney chooses an action after buying at least N signals, then the likelihood ratio of Barney having taken the action that does not match the state to the action that matches the state is less than ε.*

If Barney ends up in a learning trap, then his welfare is unboundedly negative. Because a small c tempts him to wait longer, the probability of a learning trap becomes arbitrarily close to 1 as the signal cost c becomes arbitrarily small.[32]

The second part of the proposition states that, for a given c, Barney is more likely to get caught in a learning trap the more signals he has already observed, holding constant his posterior belief. For example, if the rates θ_A and θ_B are symmetric, then an equal number of a and b signals does not change Barney's beliefs; hence Barney is more likely to end up purchasing an infinite number of signals after having observed *ababab* than he was before he observed any signals.[33,34]

This result points to a more general implication of NBLLN that emerges across a range of dynamic applications: it makes a bigger difference to his eventual beliefs for Barney than for Tommy whether he happens to observe strong evidence of the true state early or late in his learning process. The basic logic of NBLLN implies that Barney finds observing strong, early evidence, such as the group of signals *aa*, more persuasive that A is the true state than a group such as *abababaa*, even

[32]Although we prove the results of Proposition 11, and Proposition 12 below, using our parameterized model of Barney, we believe the results in both propositions extend to any distribution satisfying A1-A4. Furthermore, if A1 is replaced with A1$'$, then there exist situations in which Barney will purchase an infinite number of signals (in dynamic information acquisition) or take a sub-optimal action in every period (in experimentation) with probability 1—unlike in the current propositions, where these possibilities can occur but always with probability strictly less than 1. That is because, under A1$'$, Barney's likelihood ratio is bounded away from zero and infinity. Hence Barney could be in a situation where no infinite sequence of signal realizations would affect his action (even though he perpertually believes there exist sequences that would).

[33]Furthermore, elaborating on this second part: there exists a $\hat{N}$ such that for all $N' > \hat{N}$, if Barney's posterior probability of state A is q after observing N' signals, then the probability that Barney will buy an infinite number of signals is *strictly* higher after he has observed the N' signals than after the N signals.

[34]We have found several experiments that set up a dynamic information-purchase setting with a payoff structure similar to the model in the text and that compare subjects' behavior with a Bayesian benchmark, which is calculated assuming expected-value maximization. Tversky and Edwards (1966), Pitz (1968), Wendt (1969), and Hershman and Levine (1970) found that subjects purchased too much information. In contrast, Fried and Peterson (1969) and Pitz and Barrett (1969) found that subjects purchased too little information. Moreover, Pitz and Barrett found that when the already-observed sample size was larger, holding constant the objective strength of evidence, subjects bought *fewer* additional signals. Also contrary to our model's prediction, Sanders and Ter Linden (1967) Studies 1-3 found that, when the already-observed sample size was larger, subjects stopped acquiring information at a point where the objective evidence was weaker. In Sanders and Ter Linden's experiments, however, the signals arrived at a rate of 2, 5, or 10 signals per second, which is so fast that the nature of the inference task is likely quite different than in other studies.

though these two groups are, objectively, similarly strong signals about the state (indeed, exactly equally strong when $1-\theta_A = \theta_B$). As a result, if Barney observes strong evidence early on, he may stop trying to learn about the state after only a few signals, while if he observes ambiguous data early on, he may continue trying to learn even after many signals.

This insight also relates to the third part of Proposition 11 (which we prove for the case where Barney's prior weakly favors the true state, although we believe it holds more generally). It states that if Barney *does* eventually stop purchasing signals, then the probability that he chooses the correct action converges to 1 as the number of signals increases. This is *not* true for Tommy. Tommy stops when the difference between the number of a-signals and b-signals exits some region, and the probability of observing any given difference is independent of the total number of signals observed. In contrast, Barney stops purchasing signals when the *proportion* of a-signals exits some region, and as the number of signals increases, this proportion becomes arbitrarily more likely to cross the threshold that favors the true state than to cross the threshold that favors the false state.

Perhaps surprisingly, an outsider observer who observes the agent purchasing a large number of signals (but does not observe the realizations) should be more confident in betting that the agent took the correct action if the agent was Barney rather than Tommy. Tommy would only have purchased a large number of signals if the evidence were ambiguous up until the very end. Barney, in contrast, having observed a large number of signals, only takes an action when the cumulative evidence from many signals is overwhelming.

5.3 Experimentation and Learning About Oneself

Rather than purchasing signals about the quality of the car, Barney could instead take them for test drives, perhaps by renting them. Here, instead of an explicit cost, the cost of information acquisition is the cost of waiting to purchase the correct car. If the Volvo is the better car, then Barney is losing out every day he drives the Lada. For the same reasons that retrospective-pooling Barney could get caught in a learning trap in a dynamic information-acquisition setting, in such an experimentation setting he could end up remaining forever uncertain about the state. In our working paper (Benjamin, Rabin, and Raymond, 2012), we show that Barney could forever take an action that provides a suboptimal flow payoff in the mistaken expectation that the action will eventually provide useful information.

The failure to learn the truth from a great deal of feedback about the outcomes from one's own actions has important ramifications for people's beliefs about themselves. While there are reasons unrelated to NBLLN for why people have optimistic priors about their own abilities and preferences, we believe that NBLLN acts an "enabling bias" that explains how, despite a lifetime of experience with themselves, people remain uncertain about their own type, and 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.

We formalize these ideas in a simple, two-action experimentation environment. The agent's type is the state of the world, $\omega \in \{A, B\}$, about which the agent is uncertain. In each of an infinite number of periods $t = 1, 2...$, the agent takes an action $\mu_t \in \{A, B\}$. After taking an action, the agent receives either a high payoff, $u^H(\mu_t)$, or a low payoff, $u^L(\mu_t) \leq u^H(\mu_t)$. The agent receives the payoff $u^H(\mu_t)$ with probability $\theta_\omega \in (.5, 1)$ and $u^L(\mu_t)$ with probability $1 - \theta_\omega$, and hence this outcome serves not only as a payoff but also as a signal about the state. If the agent is of type A, she earns a higher expected payoff from taking action A, while if she is of type B, she earns a higher expected payoff from taking action B; formally, $\theta_A(u^H(A) - u^H(B)) \geq (1 - \theta_A)(u^L(B) - u^L(A))$ and $\theta_B(u^H(B) - u^H(A)) \geq (1 - \theta_B)(u^L(A) - u^L(B))$. The agent discounts the future at rate $0 < \delta < 1$.

The first part of Proposition 12 examines the case where both actions are informative. As is well known, Tommy learns the true state and eventually takes his best action in every period. For Barney, if he is retrospective-pooling, then—following the logic of Proposition 2—his beliefs will converge to a limit posterior at which he will remain uncertain about the state. Moreover, the stronger his prior in favor of one of the states, the stronger his limit posterior in favor of that state. If Barney's prior is strong enough, then even if it is incorrect, it may drive his actions in every period. This may describe situations where people begin with strong, but possibly incorrect, intuitions about their own abilities or preferences, and these intuitions are never fully corrected by experience.

Proposition 12. *Assume Barney has the beta-distribution functional form given by equation (3). Fix payoff functions $u^H(\mu_t)$ and $u^L(\mu_t)$, rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A > \theta_B$ and discount factor δ. Suppose Barney is prospective-acceptive and retrospective-pooling.*

1. *Suppose $u^H(A) > u^L(A)$ and $u^H(B) > u^L(B)$. Without loss of generality, suppose the state is $\omega = B$. For all priors $f_\Theta(\theta_A) = 1 - f_\Theta(\theta_B) \in (0, 1)$, Tommy's belief that the state is B converges to 1 almost surely, and Tommy's action converges to B almost surely. Barney's belief that the state is B converges almost surely to a number in the interval $(0, 1)$, and this limit posterior is increasing in $f_\Theta(\theta_B)$. Moreover, if $f_\Theta(\theta_B)$ is sufficiently small, then there is positive probability that Barney takes action A in every period.*

2. *Suppose* $u^H(A) > u^L(A)$ *and* $u^H(B) = u^L(B)$. *For all priors* $f_\Theta(\theta_A) = 1 - f_\Theta(\theta_B) \in (0,1)$, *if the state is* B, *then almost surely at some finite* T, *Tommy will take action* B *for all periods* $t \geq T$. *For Barney, regardless of the state, there exists* $0 < \underline{p} < 1$ *such that for any prior* $f_\Theta(\theta_A) \geq \underline{p}$, *there is positive probability that Barney takes action* A *in every period. This probability is increasing in* $f_\Theta(\theta_A)$ *but is always strictly less than 1.*

The second part of the proposition considers a special case of the experimentation environment, a one-armed bandit problem that can be used to study learning about one's own self-control. In state A, the agent can successfully exert self-control when faced with temptation, while in state B, the agent cannot resist temptation. In each period, the "risky" action, A, is to expose himself to temptation (e.g., buying potato chips at the supermarket to eat at home). The "safe" (or "commitment") action, B, is to avoid the temptation (e.g., not buying the chips). The risky action has a higher payoff than the safe action in state A but a lower payoff in state B. A special feature of this setting is that if the agent chooses the safe action, thereby avoiding the tempting situation, he does not get any information about his own self-control.

This setting is a simplified version of the Planner-Doer model of learning self-control that Ali (2011) analyzes for the case of Tommy. Almost surely at some finite T, a Tommy without self-control will—regardless of his prior—take the safe action for all periods $t \geq T$. While it is possible that a Tommy who has self-control will always take the safe action and therefore never learn that he has self-control, the only Tommys who will take the risky action in the long run are Tommys with self-control. Most of us have the intuition that, despite repeatedly exposing themselves to temptation and succumbing to it, people remain perpetually optimistic about their own self-control—and yet this result for Tommy is the other way around!

In contrast, Proposition 12 shows that if Barney begins with a sufficiently strong belief in his own self-control ability, then it is possible that he will continue to take the risky action even in the long run, in the face of overwhelming evidence that he actually lacks self-control. Moreover, the likelihood that this occurs is increasing in his ex ante optimism.

6 Concluding Remarks

In a range of important economic settings where we think NBLLN matters, we have drawn out how its consequences depend on specific features of the economic environment. Yet for some other applications where NBLLN may matter, different or additional assumptions will be needed to close the model. Consider, for example, observational learning. Besides learning from his own experience or from gathered information, two situations we focused on in Section 5, an agent could also choose which car to buy by observing which car his neighbors have bought. In the theoretical literature, it is assumed that an infinite sequence of Tommys with common prior beliefs about the state each in turn observes a single signal along with the history of previous actions and then chooses his own action. Each receives a payoff equal to 1 if his action matches the state and 0 otherwise. Much of this work has emphasized that with probability 1, after some date each agent ignores his private signal, and a "herd" forms, with all agents thereafter choosing the same action (e.g., Banerjee, 1992; Bikhchandani, Hirshleifer and Welch, 1992). Although no agent directly observes more than one signal, NBLLN will influence observational learning if agents apply their non-belief to the signals they infer others are getting. In this setting, we think that the most psychologically plausible retrospective signal-processing assumption is that Barney groups all previous agents' actions together and treats his own private signal, which is more vivid, as separate. In addition to such a grouping assumption, formally deriving the implications of NBLLN in this setting requires an extension of the framework introduced in Section 4 to address what Barney's theory is as to how *other* agents draw inferences.

Our working paper (Benjamin, Rabin, and Raymond, 2012) contains a discussion of some possible ways of handling this issue, along with a formal analysis of observational learning. The results parallel those for dynamic information acquisition from Section 5. Intuitively, if Barney groups together previous agents' actions, then NBLLN implies that he will infer less from their actions than he should. Therefore, he will more often rely on his own signal, making it slower for a herd to *begin*, especially if early evidence is mixed. Moreover, since Barney needs to see more agents following their own signal before ignoring signals than does Tommy, when a herd *does* occur, it is more likely to be on the correct action. Finally, because Barney's learning is limited even in an infinite sample, a qualitatively different kind of behavior is possible. If agents do not herd quickly enough, they will instead form what we call an "eddy": following some period, every agent chooses the action corresponding to his own signal.

Another strategic setting in which NBLLN can matter a great deal is 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 and can choose how to clump these signals when revealing them to Barney. If the saleswoman must reveal all the signals, and if Barney is retrospective-acceptive but unaware of his own NBLLN—and hence does not realize that the clumping of signals will affect his beliefs—then Barney will not draw any inferences about the state from the saleswoman's behavior. In that case, she can 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 salewoman can make Barney arbitrarily confident that the Lada is superior.

There are conceptualizations of the tendency to under-infer from large samples that differ from that embedded in our model. One 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* NBLLN (because the Law of Large Numbers still "works" for positively correlated signals under mild regularity conditions; 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.[35]

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

[35]We 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.

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 inherently linked to the gambler's fallacy, the incorrect belief that in i.i.d. coin flips, a head becomes less likely than 50% following a streak of heads. Moreover, the gambler's fallacy has as much force in large samples as in small samples. For example, Benjamin, Moore, and Rabin (2012) found that people think that the probability of a head following a streak of 9 heads from a fair coin is only 32%. If NBLLN were (like LSN) linked to beliefs about sequences of random events, then people would have to believe that a long streak of heads makes a subsequent head *more* likely.[36] Instead, evidence and intuition suggest that NBLLN is not due to *any* belief people have about the likelihood of particular sequences of random events.

Indeed, an important drawback of our model is that it is clearly wrong if used to make predictions regarding Barney's beliefs about the likelihood of particular sequences. It would predict, for example, that Barney overestimates the likelihood of aa and bb relative to ab and ba. For this reason, none of our applications relied on Barney's beliefs about sequences.

In light of the above, we have come to the view that the most psychologically compelling alternative approach to modeling NBLLN would attempt to capture people's failure to realize just how many combinations of a and b signals generate proportions close to the population mean. 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 the bias would not affect agents' predictions about the likelihood of specific sequences. It would have the disadvantages that it would generate "thin tails" (e.g., for $\theta = .8$, the likelihood of 0 heads out of 10 would be underestimated relative to 5 heads) and would be harder to work with because the mean of the agent's subjective sampling distribution would be incorrect.

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

[36] One could argue that the sample-size neglect theory, when linked to beliefs about sequences of random events, 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.

decisionmaking environment—namely, sample size—that affects the degree to which an agent will draw too weak an inference from evidence. In Appendix A we combine NBLLN with LSN, which generates a bias toward over-confidence in inferences, 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 will exacerbate people's tendency to rely on smaller samples. In dynamic settings, LSN will make it more likely for an agent to stop acquiring information after just a few signals—but, if the initial evidence does not cause the agent to stop, the agent will still draw inferences based on sample proportions and may get caught in a learning trap.

Our model of NBLLN is defined only when the signals are i.i.d. and binomial. 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 $f^{\psi}_{\text{ß}|\Theta}(\beta|\theta_0)$. 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 (as in LSN; see Appendix A).

Although we have developed our model for binomial signals, we 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. 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.

Ali, S. Nageeb, "Learning Self-Control," *Quarterly Journal of Economics*, 2011, *126*, 857–893.

Banerjee, Abhijit V., "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 1992, *107* (3), 797–817.

Barberis, Nicholas, Andrei Schleifer, and Robert Vishny, "A model of investor sentiment," *Journal of Financial Economics*, 1998, *49* (3), 307–343.

_ , 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.

Benjamin, Daniel, Don Moore, and Matthew Rabin, "Biased Beliefs about Random Samples: Evidence from an Integrated. Experiment," 2012. Working paper.

_ , Matthew Rabin, and Collin Raymond, "A Model of Non-Belief in the Law of Large Numbers," 2012. Working paper.

Bernatzi, Shlomo and Richard Thaler, "Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments," *Management Science*, 1999, *45* (3), 364–381.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Econmoy*, 1992, *100* (5), 992–1026.

Dillenberger, David, "Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior," *Econometrica*, 2010, *78* (6), 19732004.

Edwards, Ward, "Conservatism in human information processing," in B. Kleinmuntz, ed., *Formal representation of human judgment*, Wiley, 1968.

Fried, Lisbeth S. and Cameron R. Peterson, "Information Seeking: Optimal Versus Fixed Stopping," *Journal of Experimental Psychology*, 1969, *80* (3), 525–529.

Griffin, Dale and Amos Tversky, "The Weighing of Evidence and the Determinants of Confidence," *Cognitive Psychology*, July 1992, *24* (3), 411–435.

Hershman, Ramon L. and J.R. Levine, "Deviations from Optimum Information-Purchase Strategies in Human Decision-Making," *Organizational Behavior And Human Performance*, July 1970, *5* (4), 313–329.

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 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), 909936.

Nisbett, Richard E. and Lee Ross, *Human inference: Strategies and shortcomings of social judgment*, Englewood Cliffs, NJ: Prentice Hall, 1980.

Pitz, Gordon F., "Information Seeking When Available Information Is Limited," *Journal of Experimental Psychology*, 1968, *76* (1), 25–34.

_ **and Helen Reinhold Barrett**, "Information Purchase In A Decision Task Following The Presentation Of Free Information," *Journal of Experimental Psychology*, 1969, *82* (3), 410–414.

Rabin, Matthew, "Diminishing Marginal Utility of Wealth Cannot Explain Risk Aversion," in Daniel Kahneman and Amos Tversky, eds., *Choices, Values, and Frames*, Cambridge University Press, 2000, pp. 202–208.

_ , "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 2000, *68* (5), 1281–1292.

_ , "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.

Sanders, A.F. and W. Ter Linden, "Decision Making During Paced Arrival of Probabilistic Information," *Acta Psychologica*, 1967, *27*, 170–177.

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.

_ **and Ward Edwards**, "Information Versus Rewards in Binary Choices," *Journal of Experimental Psychology*, 1966, *71* (5), 680–683.

Wald, Abraham, *Sequential Analysis*, New York, NY: Wiley, 1947.

Wendt, Dirk, "Value of Information for Decisions," *Journal of Mathematical Psychology*, October 1969, *6* (3), 430–443.

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.