# Appendix to: "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

## Appendix A: A Combined Model of Non-Bayesian Updating

In this appendix, we briefly outline models of other biases besides NBLLN that also appear to matter for inference, focusing on the same single-clump context of Section 2. These models help organize our review of the evidence in Appendix B. Although these models are far more cursory and preliminary than our model of NBLLN, we hope that formalizing these alternatives can both crisply differentiate them from NBLLN—clarifying in particular that none of them are the "opposite" of or inconsistent with NBLLN—and also be suggestive of how to develop these other biases along lines we have done with NBLLN.

At first glance, NBLLN appears to be directly at odds with another bias in beliefs, the Law of Small Numbers (LSN): Tversky and Kahneman (1971) formulated the term "Law of Small Numbers" to refer to the idea that people exaggerate the likelihood that small samples will reflect the underlying population. While NBLLN generally leads to under-inference, LSN generates over-inference. However, these two biases are neither logically not psychologically inconsistent. Indeed, we believe it is the combination of the two which has led judgment researchers to posit a bias of "sample-size neglect" in which people overestimate the resemblance of small samples and underestimate the resemblance of large samples as if they simply do not see the relevance of sample size. We re-interpret what appears to be sample-size neglect as a combination of these two biases, and we show how the basic under-inference implications of NBLLN goes through after LSN is accounted for.

According to LSN, people exaggerate how much small samples resemble the population. In Rabin's (2002) model, an agent forms beliefs about $N$ i.i.d. draws that have known rate $\theta$ *as if* signals were drawn without replacement from an "urn" of size $M$ that contains exactly $\theta M$ $a$-signals and $(1-\theta)\,M$ $b$-signals.[1] To make sure that the agent continues to view the draws as random even after many signals have been realized, it can be assumed that $M>2$, and the urn is "renewed" every odd number of draws. In other words, every odd-numbered draw, the signal is drawn from a refilled urn of size $M$, and every even-numbered draw, the signal is drawn from the urn of size $M-1$ that is depleted by the previous signal's draw. An agent who believes in LSN is called Freddy. The parameter $M$ governs the strength of LSN, with Freddy becoming Tommy in the parameter limit $M\to\infty$. In Rabin's model, for a given $\theta$, the parameter $M$, in addition to being an integer larger than 2, must satisfy the constraint that $\theta M$ is an integer. This constraint becomes problematic when combining the model of LSN with our model of NBLLN. In our model of NBLLN, the agent thinks the outcome of a random sample is determined by the subjective rate $\beta$, which is drawn from a distribution with full support on either $(0,1)$ or $[0,1]$ (depending on whether A1 or A1$'$ holds). Hence for any $M$, $\beta M$ will be non-integer-valued with probability 1. For this reason, we propose a variant of Rabin's model that, while still requiring that $M$ is an integer larger than 2, remains well-defined even when $\theta M$ is not an integer.

Our variant of Rabin's model of LSN is identical to Rabin's model except that instead of believing that the "urn" contains $\theta M$ $a$-signals and $(1-\theta)\,M$ $b$-signals, Freddy believes that it contains $\widetilde{A}$ $a$-signals and $\left(M-\widetilde{A}\right)$ $b$-signals, where $\widetilde{A}$ is an integer-valued random variable that equals $j\in\{0,1,...,M\}$ with probability $\binom{M}{j}\theta^{j}\,(1-\theta)^{M-j}$. In words, Freddy thinks that signals are drawn without replacement from an urn of size $M$, but he believes the composition of the urn is random, with the "average" urn containing $\theta M$ $a$-signals and $(1-\theta)\,M$ $b$-signals. When the rate is known to be $\theta$, Freddy believes that the number of $a$-signals in the urn is a binomial random variable with parameters $(\theta,M)$.

NBLLN is the belief that the mean of a random sample converges to a non-trivial distribution, rather than a precise estimate of the mean of the population, in the limit as the sample size gets large. We refer to an agent who believes in both LSN and NBLLN as Barney-Freddy, with beliefs denoted by $f^{\psi M}$. Like Barney, he predicts that the sample is drawn according to a subjective rate that may not equal the true rate. We assume that the subjective rate $\beta\in[0,1]$ is drawn from a density $f^{\psi}_{\text{ß}|\Theta}\,(\beta|\theta)$ determined by $\theta$ and Barneyness parameter $\psi$ that satisfies Assumptions A1-A4 from Section 2. In accordance with LSN, Barney-Freddy thinks that signals are drawn without replacement from an "urn" of size $M$ that contains $\widetilde{A}$ $a$-signals and $\left(M-\widetilde{A}\right)$ $b$-signals, where $\widetilde{A}$ is a binomial random variable with parameters $(\beta,M)$. Hence Barney-Freddy thinks that the "average" urn contains $\beta M$ $a$-signals and $(1-\beta)\,M$ $b$-signals. The integer $M>2$ parameterizes

[1] Rabin and Vayanos (2010) improves on Rabin's (2002) model that we discuss here, generalizing LSN beyond the binomial case and without assuming the signal-generating process is *i.i.d.* We explore this more contrived simple model of LSN here for ease of combining it with NBLLN.

the degree of belief in LSN, and we assume that the urn is "renewed" and $\widetilde{A}$ is re-drawn every odd number of signals. As usual for the model of NBLLN, the same subjective rate $\beta$ applies to the entire clump.

Barney-Freddy both believes in the gambler's fallacy—that is, he expects recent $a$-signals to be followed by $b$'s and vice-versa—and believes that the sample mean of a large population will converge toward a full-support limit distribution. He does not observe the subjective rate $\beta$, but conditional on any given $\beta$, Barney-Freddy expects that if the even-numbered draw $t$ is more likely to be an $a$ signal if the $(t-1)^{st}$ draw was $b$ than if it was $a$. Hence even without knowing $\beta$, Barney-Freddy excessively expects that $a$ and $b$ signals will alternate between an odd draw and an even draw. Consequently, for reasonable calibrations of $\psi$ and $M$, Barney-Freddy's subjective sampling distribution in a finite sample will be too peaked, putting too little weight in the tails. On the other hand, Lemma A1 states that Barney-Freddy's beliefs about a large sample converge to a full-support limit distribution. Like for Barney, Barney-Freddy's limit density will be equal to $f^{\psi}_{\text{ß}|\Theta}(\beta|\theta)$, regardless of the degree of belief in LSN. Intuitively, every odd-even pair of draws will have, on average, proportion $\beta$ of $a$-signals. Hence by the Law of Large Numbers, the sample as a whole will tend toward having proportion $\beta$ of $a$-signals almost surely.

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

$$\lim_{N\to\infty} \sum_{x=\lfloor a_1 N \rfloor}^{\lceil a_2 N \rceil} f^{\psi M}_{S_N|\Theta}(A_s = x|\theta) = F^{\psi}_{\text{ß}|\Theta}(\beta = \alpha_2|\theta) - F^{\psi}_{\text{ß}|\Theta}(\beta = \alpha_1|\theta) > 0.$$

Not only are LSN and NBLLN mutually consistent, but LSN actually *magnifies* NBLLN. For smaller $M$, Barney-Freddy believes that odd-even signal pairs more frequently alternate, and hence his subjective sampling distribution converges to the limit distribution more quickly.

While it is true that Freddiness generates over-inference while Barneyness tends to generate under-inference, there is a clear pattern to when Barney-Freddy over-infers and when he under-infers. For reasonable calibrated values of $\psi$ and $M$, Barney-Freddy will over-infer from small samples. For any values of $\psi$ and $M$, it follows immediately from Lemma A1 that Barney-Freddy will under-infer when the sample size $N$ is sufficiently large.

A well-known bias is base-rate neglect (Kahneman and Tversky, 1973), an underweighting of prior probabilities in drawing inferences. Instead of assuming that an agent updates according to Bayes's Rule applied to his subjective sampling distributions, we can capture base-rate neglect by assuming that the agent draws inferences according to:

$$f_{S_N|\Theta}^{\psi Mb}(\theta_A|s) = \frac{f_{S_N|\Theta}^{\psi M}(s|\theta_A)\, f_\Theta(\theta_A)^b}{f_{S_N|\Theta}^{\psi\phi\gamma}(s|\theta_A)\, f_\Theta(\theta_A)^b + f_{S_N|\Theta}^{\psi\phi\gamma}(s|\theta_B)\, f_\Theta(\theta_B)^b}$$

and

$$f_{S_N|\Theta}^{\psi Mb}(\theta_B|s) = \frac{f_{S_N|\Theta}^{\psi M}(s|\theta_B)\, f_\Theta(\theta_B)^b}{f_{S_N|\Theta}^{\psi M}(s|\theta_A)\, f_\Theta(\theta_A)^b + f_{S_N|\Theta}^{\psi M}(s|\theta_B)\, f_\Theta(\theta_B)^b},$$

where $f_{S_N|\Theta}^{\psi M}(s|\theta_A)$ and $f_{S_N|\Theta}^{\psi M}(s|\theta_B)$ are the subjective sampling distributions of an agent with Barneyness parameter $0 < \psi < \infty$ and Freddiness parameter $M > 2$, and $0 \leq b \leq 1$ parameterizes the degree of base-rate neglect. If $b = 1$, these formulae specialize to Bayes' Rule, where there is no base-rate neglect. If $b = 0$, the agent ignores base rates altogether, treating any prior probabilities as if they were 50-50. This formulation of base-rate neglect has been previously adopted in empirical work by (e.g., Grether, 1980) and concurrently in theoretical work by Bodoh-Creed (2010). Applying it theoretically in dynamic settings raises many of the same conceptual issues as NBLLN; the way in which the agent processes groups of signals will matter a great deal in how his beliefs evolve. For that reason, we believe the framework we have begun to develop in this paper for analyzing dynamic NBLLN may be of use for analyzing dynamic base-rate neglect as well.

A simple explanation for some of the evidence on people's beliefs is "extreme-belief aversion," an aversion to holding beliefs that are close to certainty. Consider a discrete probability density function, $f_X(\cdot)$, that puts positive probability on a set of possible outcomes $x_1, x_2, ..., x_J$. We capture the idea of extreme-belief aversion by defining a mapping from the true probability density, $f_X(\cdot)$, to a subjective probability density that is less extreme,

$$f_X^\xi(x_i) = \frac{.5 + \xi \cdot (f_X(x_i) - .5)}{\sum_{j=1}^J (.5 + \xi \cdot (f_X(x_j) - .5))}.$$

The parameter $0 < \xi < 1$ describes the degree of extreme-belief aversion, with smaller values corresponding to greater bias. If $\xi = 1$, the subjective probabilities coincide with the true probabilities, while if $\xi = 0$, all outcomes $x_1, x_2, ..., x_J$ are treated as equally likely.

One interpretation of the transformed probability, $f_X^\xi(x)$, is that it represents the agent's truly-held beliefs. Another interpretation is that the agent actually holds beliefs $f_X(x)$ but *reports* beliefs that are transformed by the $\xi$ function. While the latter is certainly plausible when Tommy's beliefs are very extreme—we can easily imagine a person saying she is 99% sure when her true belief is .9999—it does not address the evidence that beliefs inferred from betting behavior also exhibit under-inference.

Qualitatively, extreme-belief aversion can explain both the excessively-dispersed subjective sampling distributions and the under-inference evidence that we discuss in Appendix B. An agent with extreme-belief aversion will have a more dispersed subjective sampling distribution than Tommy in large samples by virtue of compressing beliefs away from 0 and 1. An agent with extreme-belief aversion will also under-infer in situations where Tommy's inference would be extreme, as would almost always occur when the sample is large. Extreme-belief aversion taken alone, however, implies that in any two inference problems where Tommy's posterior is the same, people would hold (or at least report) the same belief. There is evidence contradicting this implication, indicating that extreme-belief aversion is not the only deviation from Bayesian belief formation that is going on.

For example, consider the experiment in Griffin and Tversky (1992), where $\theta_A = .6$. Tommy's inference depends only on the difference between the number of $a$-signals and the number of $b$-signals. However, when the sample is 4 $a$'s and 1 $b$, subjects' median belief in favor of $\theta_A$ is .80, while when the sample is 10 $a$'s and 7 $b$'s, subjects' median belief in favor of $\theta_A$ is .60. Tommy's belief would be .77 in both cases, so people are under-inferring from the sample of size 17 and actually slightly over-inferring from the sample of size 5. Consistent with Proposition 5—but inconsistent with extreme-belief aversion being the only bias in beliefs—people infer less from the same difference in $a$ and $b$ signals when the sample is larger. Kahneman and Tversky (1972) and Kraemer and Weber (2004) also report evidence that beliefs are sensitive to sample size, holding constant the difference in the number of $a$ and $b$ signals.

While extreme-belief aversion may help to describe the evidence on people's beliefs, and it may be a confound for other interpretations of biased beliefs, we do not review evidence for extreme-belief aversion, and we know of no evidence that our crude formulation is a close match for people's thinking. Extreme-belief aversion may also lead to internal-inconsistency modeling challenges that we do not address. For example, it seems reasonable to assume that the transformation above could be applied to an agent's sampling distribution or to the agent's inferences, depending on which beliefs are being elicited. In that case, however, subjective sampling distributions and inferences will not in general be linked by Bayes' Rule.

Combining the three biases above with NBLLN gives a complicated model that captures many features that could be applied to predict beliefs and behavior in economic settings. One insight that comes immediately out of the combination is that base-rate neglect—i.e., underweighting priors—is not the opposite of, or contradictory to, the way NBLLN leads people to underweight likelihood information. Indeed, as we have noted in some discussions above about vividness and other biases, NBLLN is, especially in understanding "multi-clump" information processing, likely a contributor to the relevance of other biases. In many information-rich environments where full Bayesians would correctly become very confident independent of their priors, NBLLN is necessary for the question of whether people neglect base rates to be relevant. At the same time, we show above why NBLLN means that unless people completely neglect base rates, people's initial beliefs matter even in the long run. In combination, in fact, NBLLN suggests that it is possible that the real and often important fact that people under-use base rates may be consistent with the possibility that base rates matter more for social and economic phenomena than fully rational models have supposed that they do.

One last bias sits less comfortably with the others, and is harder to integrate. Many experiments make clear that peoples' subjective sampling distributions have flatter tails than our model of NBLLN by itself can explain. We attribute the flatness to "sampling-distribution-tails diminishing sensitivity (SDTDS)," a bias in which people perceive very unlikely outcomes as similar to each other and hence similar in probability. Consider 10 flips of a coin that is biased .8 in favor of heads. We know of no direct evidence but conjecture that most people would judge the likelihood of observing 1 head as very close to the likelihood of observing 0 heads, even though observing 1 head is actually 40 times more likely.

SDTDS can be formalized by assuming that an agent forms beliefs as if the likelihood of sample realizations far from the average are more similar to each other (and to the average) than they truly are. For a sample of size $N$ from a $\theta$-biased coin, let $\sigma$ denote the standard deviation of the sample proportion $\frac{A_N}{N}$ (which equals $\sqrt{\frac{\theta(1-\theta)}{N}}$ for Tommy but not for Barney-Freddy). Let the perceived distance between the realized $\frac{A_N}{N}$ and $\theta$ be $\sigma\gamma\left(\frac{\frac{A_N}{N}-\theta}{\sigma}\right)$, where the twice-differentiable "sample-perception function" $\gamma : (-1,1) \to (-1,1)$ has the properties (1) $\gamma(0) = 0$, (2) $0 < \gamma' < 1$, (3) $\gamma''(x) < 0$ for all $x > 0$, and $\gamma''(x) > 0$ for all $x < 0$, and (4) $\gamma(x) = -\gamma(-x)$ for all $x$.

Property (1) says that the agent perceives a sample proportion of $\theta$ accurately. Property (2) ensures that the agent perceives any other sample proportion as more similar to $\theta$ than it actually is. The key concavity and convexity assumption (3) means that neighboring samples are perceived as more similar to each other the further they are from $\theta$. For this reason, the agent makes little distinction between outcomes that fall far in the tails of his subjective sampling distribution. Property (4) specifies that $\gamma$ is symmetric around 0 so that the misperception is symmetric around $\theta$.

Roughly speaking, a person who exhibits SDTDS with sample-perception function $\gamma$ judges the probability of a sample $s \in S_N$ with proportion $\frac{A_N}{N}$ of $a$-signals as if it were the sample $\Lambda(s)$, which has proportion $\theta + \sigma\gamma\left(\frac{\frac{A_N}{N}-\theta}{\sigma}\right)$ $a$-signals. Formally, $f^{\psi\phi\gamma}_{S_N|\Theta}(s|\theta) = \frac{f^{\psi\phi}_{S_N|\Theta}(\Lambda(s)|\theta)}{\sum_{s' \in S_N} f^{\psi\phi}_{S_N|\Theta}(\Lambda(s')|\theta)}$, where $f^{\psi\phi}_{S_N|\Theta}$ is the subjective sampling distribution for $\psi\phi$-Barney-Freddy, and the denominator is a normalization that ensures that the subjective sampling distribution adds up to 1.[2] While NBLLN by itself leads to subjective sampling distributions that have flat tails, SDTDS is an additional force for flat tails. When the rate $\theta$ is not .5, the mean of the agent's subjective sampling distribution no longer equals the mean of the objective sampling distribution because the "long tail" of the distribution is overweighted. Both of these features—flatter tails than can be accommodated with a reasonably-calibrated model of NBLLN and a mean shifted toward the long tail—are present in the subjective sampling distributions measured by Kahneman and Tversky (1972).

Unlike both our model of NBLLN and the Rabin (2002) model of LSN (and consequently also unlike the model of Barney-Freddy presented in this Appendix), SDTDS predicts under-inference from samples of size 1. In fact, the evidence we review in Appendix B suggests that people *do* under-infer in a sample of size 1. We note, however, that no experiment has disentangled the SDTDS explanation of under-inference from an alternative explanation of extreme-belief aversion. Nonetheless, in inference problems with normally-distributed signals, we suspect that SDTDS does help explain why subjects under-infer in a sample of size 1 when the realized signal is extreme. In contrast, when signals are binomial as in our analysis in the main text, we believe that the psychology of SDTDS — extreme events in the same tail of a distribution being judged as similar in probability — is unlikely to apply because there are no two extreme events in the same tail. Moreover, if SDTDS predicts under-inference from a sample of size 1, then it necessarily also implies incorrect predictions about the likelihood of a single signal realization, an implication that we view as implausible when signals are binomial.

Unfortunately, our formulation of SDTDS cannot be so easily integrated with NBLLN—or with a model combining NBLLN with LSN and base-rate neglect—because it does not arm somebody with a theory of the sequence of signals, only the frequency of signals within a sample. While surely a form of it could be specified that embeds a concrete theory of permutations within the sample (allowing for instance a person to believe all different sequences are equally likely), we do not know that the existing evidence provides a guide, nor do we believe the psychology underlying it translates easily into situations where an agent is cognizant of the ordering of signals. Moreover, an improved model or a better formulation than we have found to interpret existing evidence may be more compatible with the models of other biases than we have supposed.

[2]In general, $\Lambda(s)$ could be a sample outside the support of the objective sampling distribution. For example, it may be a sample with 3.2 $a$'s. The expression above is nonetheless well-defined as long as the density $f^{\psi\phi}_{S_N|\Theta}(\cdot|\theta)$ can be evaluated at that sample, as it can for a binomial objective sampling distribution.

# Appendix B: Experimental Evidence

In this appendix, we report on all papers we could identify with experimental results related to prediction and inference for binomial signals. Table B1 lists the papers, whether their evidence relates to prediction or inference, their experimental subject population, and their incentive structure. Most of the studies we review did not incentivize subjects' responses; we will also discuss how the evidence from the few incentivized experiments relates to the unincentivized studies.

Table B1. Experimental evidence on NBLLN for binomial signals.

| Author(s) | Year | Prediction or Inference? | Subjects | Incentives? |
|---|---|---|---|---|
| Beach, Wise, and Barclay | 1970 | inference | 169 male undergrads | no |
| Camerer | 1987 | inference | 74 undergrads | financial market |
| Chinnis and Peterson | 1968 | inference | 40 male undergrads | no |
| Dave and Wolfe | 2003 | inference | 40 undergrads | BDM for probability |
| DeSwart | 1972 | inference | 21 male undergrads | no |
| DeSwart | 1972 | inference | 18 male undergrads | no |
| Donnell and DuCharme | 1975 | inference | 24 male undergrads | no |
| Gettys and Manley: Study 1 | 1968 | inference | 20 undergrads | no |
| Gettys and Manley: Study 2 | 1968 | inference | 28 undergrads | no |
| Green, Halbert, and Robinson | 1965 | inference | 32 grad students | paid for guess about state |
| Grether | 1980 | inference | 341 undergrads | paid for guess about state |
| Grether: Studies 1 and 2 | 1992 | inference | 97 undergrads | paid for guess about state |
| Grether: Study 3 | 1992 | inference | 55 summer students | BDM for probability |
| Griffin and Tversky: Study 1 | 1992 | inference | 35 undergrads | paid for accurate posterior |
| Griffin and Tversky: Study 2 | 1992 | inference | 40 undergrads | paid for accurate posterior |
| Griffin and Tversky: Study 3 | 1992 | inference | 50 undergrads | no |
| Kahneman and Tversky: prediction | 1972 | prediction | unclear | no |
| Kahneman and Tversky: inference | 1972 | inference | 560 high school students | no |
| Kraemer and Weber | 2004 | inference | 51 students (most grad) | paid for accurate posterior |
| Marks and Clarkson | 1972 | inference | 68 undergrads | no |
| Nelson, Bloomfield, Hales, and Libby: Study 1 | 2001 | inference | 27 MBA students | financial market |
| Peterson and Miller | 1965 | inference | 42 undergrads | no |
| Peterson and Swensson: Study 1 | 1968 | inference | 15 male undergrads | no |
| Peterson and Swensson: Study 2 | 1968 | inference | 18 male undergrads | no |
| Peterson, DuCharme, and Edwards: Study 1 | 1968 | prediction | 41 male undergrads | no |
| Peterson, DuCharme, and Edwards: Study 2 | 1968 | both | 24 male undergrads | no |
| Peterson, Schneider, and Miller | 1965 | inference | 44 undergrads | no |

| Author(s) | Year | Prediction or Inference? | Subjects | Incentives? |
|---|---|---|---|---|
| Phillips and Edwards: Study 1 | 1966 | inference | 5 male undergrads | no |
| Phillips and Edwards: Study 2 | 1966 | inference | 48 male undergrads | paid for accurate posterior |
| Phillips and Edwards: Study 3 | 1966 | inference | 48 male undergrads | no |
| Pitz | 1967 | inference | 28 undergrads | no |
| Sanders | 1968 | inference | 32 undergrads | bets on the state |
| Sasaki and Kawagoe | 2007 | inference | 1033 employees | no |
| Strub | 1969 | inference | 12 male undergrads | paid for guess about state |
| Teigen: Study 1 | 1974 | prediction | 22 undergrads | no |
| Teigen: Study 2 | 1974 | prediction | 73 undergrads | no |
| Wheeler and Beach | 1968 | both | 17 male undergrads | paid for beliefs (prediction), and bets (inference) |

### 0.1 Evidence on Subjective Sampling Distributions

We begin by assessing subjective sampling distributions. We have found only 6 experiments from 4 papers in which researchers explicitly elicited experimental participants' beliefs about the likelihood of each possible sample.[3] None of these elicitations were incentivized. For 5 of these studies—Kahneman and Tversky (1972), Peterson, DuCharme and Edwards's (1968) Study 1, Wheeler and Beach (1968), and both of Teigen's (1974) studies—the data are displayed in the paper, and we have reproduced the graphs in Figures 1, B4, B5, and B1, respectively (we display both of Teigen's studies together), shown below in the order we discuss them.

Among the papers, Kahneman and Tversky (1972) elicited sample-proportion beliefs for the largest sample sizes. As discussed in the Introduction, they find that subjective sampling distributions are "constant in proportions" for $N = 10$, 100, and 1000. There is no noticeable tightening of the distribution even for $N = 1000$; while in fact there is less than a .01 chance of the proportion of heads falling outside the range 45% to 55%, subjects' distributions assign probability .79 to a proportion outside that range.[4]

A straightforward implication of our model of NBLLN is that subjective sampling distributions will be flatter than the objective sampling distributions. In all 7 experiments, with the exception of the $N = 3$ conditions of one experiment, the researchers indeed concluded that subjective sampling distributions are excessively close to uniform.[5]

[3]Cohen and Hansel (1955) also elicited subjective sampling distributions, but we cannot compare their data with our model because they did not tell their subjects the rate that was generating the signals.

[4]At the time of this draft, two of the authors (Benjamin and Rabin, in joint work with Don Moore) have also collected data on people's subjective sampling distributions for $N = 10$ and 1000. We designed the experiment to measure subjective sampling distributions in several different ways to deal with potential confounds such as extreme-belief aversion. When we elicit distributions in the same manner as Kahneman and Tversky, we replicate their results almost exactly. While our preliminary findings support NBLLN for $N = 1000$, we also find that the evidence for NBLLN for $N = 10$—and hence presumably also the evidence about smaller sample sizes reviewed below—is confounded by other explanations.

[5]Unlike the other 6 studies, Teigen (1974) asked subjects about the probability of each possible outcome

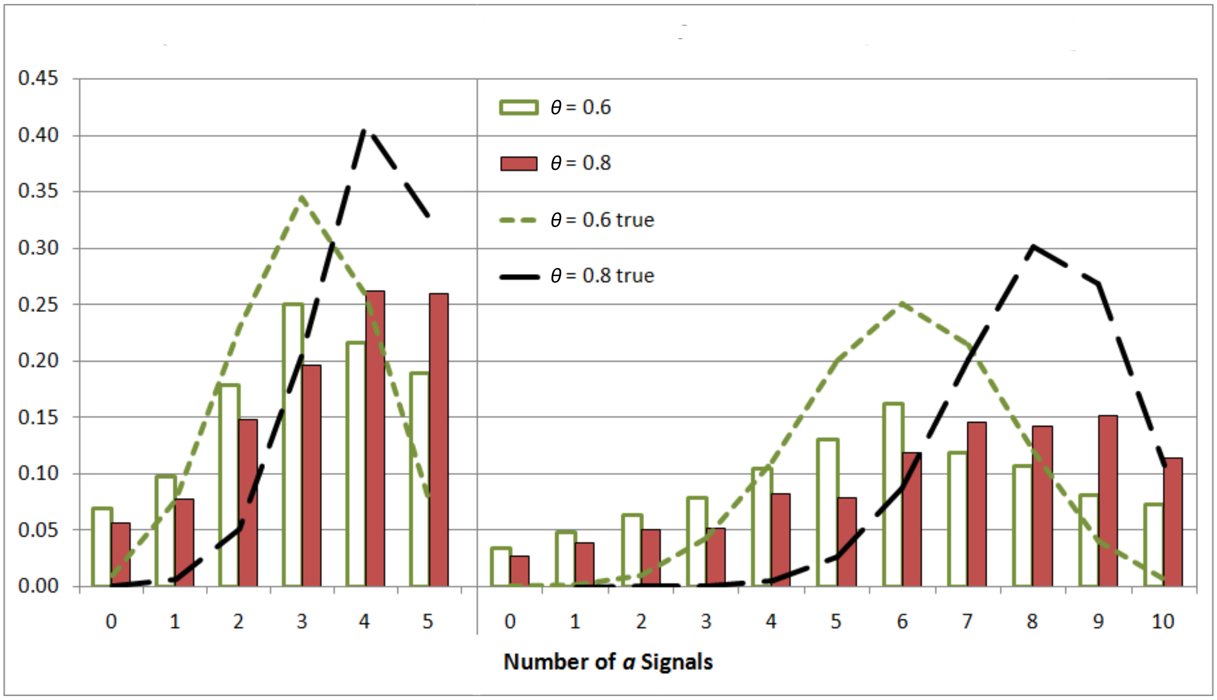


Figure B1: Median probability estimates, N=5 and N=10 (Teigen, 1974)

A feature of our model of NBLLN is that people have correct beliefs about samples of size 1. We know of no evidence on this point, but our own introspection suggests it is virtually a tautology that if a person knows the rate of $a$-signals is $\theta$, the person thinks the probability is $\theta$ that a single draw will be $a$.[6] However, our model of NBLLN implies that for any sample size larger than 1, no matter how small, the subjective sampling distribution will be too close to uniform. Peterson, DuCharme and Edwards's (1968) Study 1 $N = 3$ conditions are the only cases where researchers found subjective sampling distributions that are *not* too flat. For all three rates they studied ($\theta =$ .6, .7, .8), the $N = 3$ subjective sampling distribution nearly coincides with the objective sampling distribution. We interpret this evidence as consistent with the combined effects of NBLLN and LSN, rather than either of the biases considered alone. The combined model we discuss in Appendix A predicts that subjective sampling distributions will be correct for $N = 1$ and too flat for $N$ large. For plausible parameter values, there will be a non-monotonicity for intermediate $N$: due to LSN, the subjective sampling distribution may be too peaked when $N$ is larger than 1 but small. The relative strength of NBLLN grows with $N$, so the subjective sampling distribution may be nearly

separately, without requiring that the probabilities sum to 1. He found that the probabilities summed to greater than 1, and subjects often assigned probabilities that were too high to every outcome. When the subjective sampling distributions were normalized to sum to 1, they were too flat.

[6] Of course there is a good reason no one has done this experiment: the correct answer about the probability of an $a$-signal is obvious if the experimenter has just said that it is $\theta$. Nonetheless, we discuss below evidence from experiments on inference from a single signal. The correct answer to an inference problem is not obvious, and indeed subjects often do not get the correct answer, tending somewhat to under-infer on average.

correct for some intermediate $N$.

Assumption A4 of our model is that the subjective sampling distribution has the same mean as the objective sampling distribution, regardless of sample size. In contrast, the evidence indicates that when $\theta \neq .5$, the mean of the subjective sampling distribution is generally between the objective mean, $\theta N$, and .5. Kahneman and Tversky (1972) explicitly comment that "the mean is displaced towards the long tail" (p.440), and this pattern is visually evident in all of the figures except the $N = 3$ cases discussed above. From a modeling perspective, we believe it is appropriate for our model to have the counterfactual feature of an accurate mean for the subjective sampling distribution because it allows us to draw out the implications of believing that the limit distribution has full support, without mixing in the implications of an inaccurate mean. Moreover, we speculate that the "displaced mean" is the result of a psychologically distinct bias, "sampling-distribution-tails diminishing sensitivity (SDTSD)," sketched in Appendix A, and our calibrated model of NBLLN generates subjective sampling distributions that come closer to matching the data in Figure 1 when we additionally incorporate SDTSD into the model (calculations not shown).

There is mixed evidence about whether training and feedback affect subjective sampling distributions. Wheeler and Beach (1968) elicited subjective sampling distributions for a sample of size $N = 8$ for rates $\theta = .6$ and .8; these are shown in Figure B5. Next, their subjects were faced with 100 asymmetric binomial inference problems ($\theta_A = .8$ and $\theta_B = .4$). After each problem, the subject was told the true rate for that problem. Subjective sampling distributions were elicited again, the subjects responded to 100 more symmetric binomial inference problems with feedback, and the subjective sampling distributions were elicited a final time.[7] Comparing the initial subjective sampling distributions with the final ones, the final subjective sampling distributions are less flat. For $\theta = .8$, the final subjective sampling distribution is quite close to the objective sampling distribution. For $\theta = .6$, the final subjective sampling distribution is actually more peaked than the objective sampling distribution. On the other hand, Peterson, DuCharme, and Edwards (1968, Study 2) conducted an experiment with four stages: (1) subjects were faced with symmetric binomial inference problems (with no feedback); (2) subjective sampling distributions were elicited for each combination of $N = 3, 5, 8$, and $\theta = .6, .7, .8$; (3) subjects were shown the objective sampling distributions; and (4) subjects were faced with another series of symmetric binomial inference problems. Subjects' responses in the inference problems were similar in stage 4 as in stage 1, suggesting that showing subjects the objective sampling distributions had little effect on beliefs.

[7]These subsequent elicitations are not shown in Figure B5.

## 0.2 Evidence on Inference

We have found 33 studies from 26 papers measuring inferences from samples about which of two equally-likely rates, $\theta_A$ and $\theta_B$, generated the samples. Most of these binomial inference problems are symmetric in the sense that $\theta_A = 1 - \theta_B$. Unless otherwise noted, all the studies we mention are symmetric. We focus first on the studies where the prior probabilities, $f_\Theta(\theta_A)$ and $f_\Theta(\theta_B)$, are equal. Equal priors neutralizes the role of base-rate neglect. We study below how inferences are affected by unequal priors.

To compare the degree of under- or over-inference across studies, note that Bayes' Rule can be written as $\frac{f_{\Theta|S_N}(\theta_A|s)}{f_{\Theta|S_N}(\theta_B|s)} = \frac{f_{S_N|\Theta}(s|\theta_A)f_\Theta(\theta_A)}{f_{S_N|\Theta}(s|\theta_B)f_\Theta(\theta_B)}$. Since the signals are binomial, the rates are symmetric, and the priors are equal, Bayes' Rule can be expressed as $\frac{f_{\Theta|S_N}(\theta_A|s)}{f_{\Theta|S_N}(\theta_B|s)} = \left(\frac{\theta_A}{1-\theta_A}\right)^{\frac{2A_s-N}{N}\times N}$. Taking the natural log twice and rearranging,

$$\ln\ln\frac{f_{\Theta|S_N}(\theta_A|s)}{f_{\Theta|S_N}(\theta_B|s)} - \ln\left(\frac{2A_s-N}{N}\right) - \ln\ln\left(\frac{\theta_A}{1-\theta_A}\right) = \ln N. \tag{1}$$

It is possible in 9 of the papers to identify the value of $\theta_A$ for the inference problem, the actual sample observed by subjects, and subjects' mean or median reported posterior. Using the experimental data, Figure B2 plots the left-hand side of equation (1) against $\ln N$.[8] If the subjects' beliefs were Bayesian, the points should cluster along the identity line (the dashed line in Figure B2). Instead, most points fall below the identity line, indicating that subjects generally infer less in favor of rate$\theta_A$ than a Bayesian would. Moreover, the best-fitting regression line (the solid line in the figure) has a slope smaller than 1, suggesting that in the metric defined by the left-hand side of equation (1), the under-inference is greater for larger $N$.

[8] The left-hand side is well-defined only for inference problems such that $\frac{\theta_A}{1-\theta_A} > 1$ (that is, $\theta_A > .5$) and $\frac{A_s}{N} > \frac{1}{2}$ (that is, over half the realized signals are $a$'s). Hence, as written, equation (1) only applies to such cases. Although this holds for only 63 of the 99 regression observations in Figure B2 and Table B2, we can include additional regression observations by relabeling the rates and sample proportions that we plug into formula (1). In inference problems such that $\frac{\theta_A}{1-\theta_A} > 1$ and $\frac{A_s}{N} < \frac{1}{2}$, we express Bayes' Rule as $\frac{f_{\Theta|S_N}(\theta_B|s)}{f_{\Theta|S_N}(\theta_A|s)} = \left(\frac{\theta_B}{1-\theta_B}\right)^{\frac{2A_s-N}{N}\times N} = \left(\frac{\theta_A}{1-\theta_A}\right)^{\frac{N-2A_s}{N}\times N}$, so equation (1) becomes $\ln\ln\frac{f_{\Theta|S_N}(\theta_A|s)}{f_{\Theta|S_N}(\theta_B|s)} - \ln\left(\frac{N-2A_s}{N}\right) - \ln\ln\left(\frac{\theta_A}{1-\theta_A}\right) = \ln N$. This allows us to use an additional 32 regression observations. Finally, we can use a further 4 regression observations for which $\theta_A < .5$ and $\frac{A_s}{N} < \frac{1}{2}$ by expressing Bayes' Rule as $\frac{f_{\Theta|S_N}(\theta_A|s)}{f_{\Theta|S_N}(\theta_B|s)} = \left(\frac{\theta_A}{1-\theta_A}\right)^{\frac{2A_s-N}{N}\times N} = \left(\frac{1-\theta_A}{\theta_A}\right)^{\frac{N-2A_s}{N}\times N}$, and we can take the log-log of this equation. In the case of 3 inference problems, $\frac{A_s}{N} = \frac{1}{2}$, so the Bayesian posterior ratio is equal to 1, and it is impossible to define what constitutes "over-inference" or "under-inference." Those 3 datapoints are dropped from the Figure B2 and Table B2.

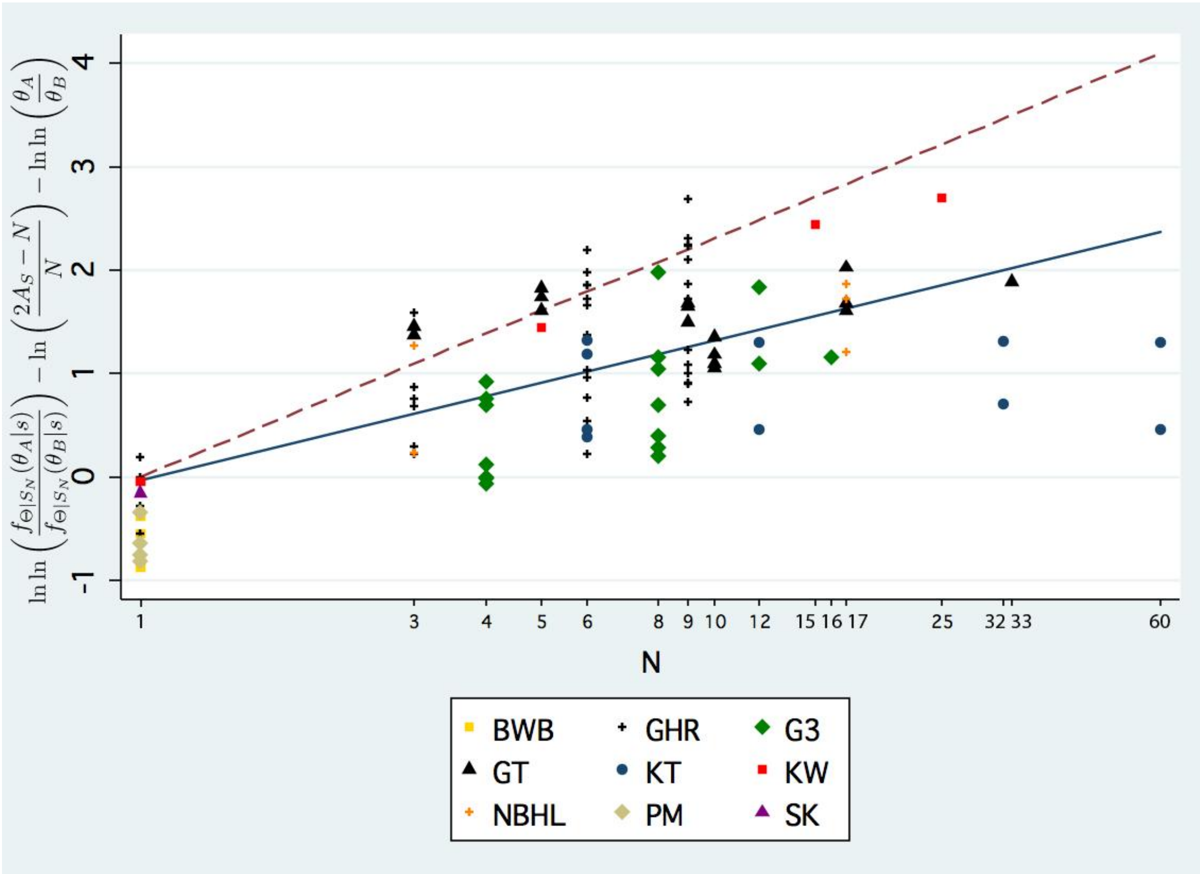


Figure B2: Inference with symmetric rates and equal priors

Notes: The x-axis is depicted on the natural log scale. For all datapoints in the figure, subjects knew that prior probabilities of the two rates were equal. The dotted line represents the null hypothesis of Bayesian updating, and the solid line is the best-fitting regression line from column 2 of Table B2. The includes studies are: BWB = Beach, Wise, and Barclay (1970); GHR = Green, Halbert, and Robinson (1965); G3 = Grether's (1992) Study 3; GT = Griffin and Tversky's (1992) Study 1; KT = Kahneman and Tversky (1972); KW = Kraemer and Weber (2004); NBHL = Nelson, Bloomfield, Hales, and Libby's (2001) Study 1; PM = Peterson and Miller (1965); SK = Sasaki and Kawagoe (2007).

To estimate the degree of under-inference and to probe its robustness, we rewrite (1) as a regression equation:

$$\ln\ln\frac{f_{\Theta|S_N}(\theta_A|s)}{f_{\Theta|S_N}(\theta_B|s)} = \gamma_0 + \gamma_1 \ln N + \gamma_2 \ln\left(\frac{2A_s - N}{N}\right) + \gamma_3 \ln\ln\left(\frac{\theta_A}{1-\theta_A}\right) + \varepsilon. \quad (2)$$

The null hypothesis of Bayesian updating is $\gamma_0 = 0$, $\gamma_1 = \gamma_2 = \gamma_3 = 1$. In Table B2, we estimate versions of equation (2) with several different restrictions on the coefficients and data.

Table B2. Inference with symmetric rates and equal priors.

| | (1) Restriction: $\gamma_1 = \gamma_2 = \gamma_3 = 1$ | (2) Restriction: $\gamma_2 = \gamma_3 = 1$ | (3) Restriction: $\gamma_3 = 1$ | (4) Coeffs unrestricted | (5) Coeffs unrestricted Incentivized only |
|---|---|---|---|---|---|
| $\ln N$ | | 0.587 | 0.470 | 0.485 | 0.675 |
| | | (0.083) | (0.109) | (0.085) | (0.086) |
| $\ln\left(\frac{2A_s-N}{N}\right)$ | | | 0.776 | 0.914 | 0.957 |
| | | | (0.116) | (0.109) | (0.121) |
| $\ln\ln\left(\frac{\theta_A}{1-\theta_A}\right)$ | | | | 0.333 | 0.421 |
| | | | | (0.101) | (0.104) |
| Constant | -0.753 | -0.031 | -0.002 | -0.145 | -0.168 |
| | (0.076) | (0.143) | (0.142) | (0.108) | (0.130) |
| $R^2$ | 0.000 | 0.426 | 0.360 | 0.603 | 0.763 |
| #obs | 99 | 99 | 99 | 99 | 47 |
| #papers | 9 | 9 | 9 | 9 | 3 |

Notes: Results are from OLS regressions, with standard errors in parentheses. The dependent variable is as described in the text. Coefficients for blank entries are restricted to equal 1. The fifth column restricts the data to incentivized experiments.

Column 1 estimates just $\gamma_0$, under the restriction that $\gamma_1 = \gamma_2 = \gamma_3 = 1$. The estimate, $\widehat{\gamma}_0 = -.753$, is significantly smaller than zero, indicating that the average pattern is under-inference.

Column 2 estimates both $\gamma_0$ and $\gamma_1$, while restricting $\gamma_2 = \gamma_3 = 1$. The predicted regression line is plotted as the solid line in Figure B2. The regression confirms that the degree of under-inference is related to sample size; $\widehat{\gamma}_1 = .587$ is significantly smaller than 1 (but greater than 0). Moreover, once the degree of under-inference is allowed to depend on sample size, there is no residual under-inference to be picked up by the constant term: $\widehat{\gamma}_0 = -.031$, which is not statistically distinguishable from zero. Breaking down the data by study (not shown in the table), every study that manipulates sample size, while holding constant other features of the experiment—Peterson, Schneider, and Miller (1965); Pitz (1967); Peterson, DuCharme, and Edwards's (1968) Study 2; Kahneman and Tversky (1972); Griffin and Tversky's (1992) Study 1; Nelson, Bloomfield, Hales, and Libby's (2001) Study 1; Kraemer and Weber (2004)—concludes that there is greater under-inference in larger samples.[9]

Column 3 of Table B2 relaxes the regression model still further, estimating $\gamma_0$, $\gamma_1$, and $\gamma_2$, with the only remaining restriction being $\gamma_3 = 1$. The coefficient on the proportion of $a$ signals, $\widehat{\gamma}_2 = .776$, is just statistically distinguishable from 1, and the estimates $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$ are not substantially different in column 3 compared with column 2.

Column 4 estimates all four coefficients. Qualitatively, the main effect of relaxing the $\gamma_3 = 1$ restriction on the conclusions from column 3 is that the coefficient on the proportion of $a$-signals, $\widehat{\gamma}_2 = .914$, is no longer statistically distinguishable from 1. Hence we cannot reject the hypotheses that $\gamma_0 = 0$ and $\gamma_2 = 1$, but $\widehat{\gamma}_1$ is significantly smaller than 1. However, column 4 makes clear that the extremeness of the rates—the extent to which $\theta_A = 1 - \theta_B$ departs from .5—also matters for the degree of biased inference. The coefficient $\widehat{\gamma}_3 = .333$ is much smaller than 1. This means that subjects under-infer by more the further is $\theta_A$ from .5. Breaking down the data by study (not shown in the table), every study that manipulates $\theta_A$, while holding constant other features of the experiment either concludes that there is greater under-inference for $\theta_A$ further from .5 (Green, Halbert, and Robinson, 1965; Phillips and Edwards's, 1966 Study 1 and 3; Peterson and Miller, 1965; Peterson and Swensson's, 1968 Studies 1 and 2; Peterson, DuCharme, and Edwards's, 1968 Study 2; Sanders, 1968; Donnell and DuCharme, 1975; Kahneman and Tversky, 1972) or finds it without explicitly stating it (Chinnis and Peterson, 1968; Beach, Wise, and Barclay, 1970; Shu and Wu, 2003).[10]

Griffin and Tversky (1992) use the term "discriminability" to describe the phenomenon of under-inference becoming more severe when $\theta_A$ and $\theta_B$ are further apart. In Griffin and Tversky's (1992; Study 3) particularly clear evidence from asymmetric inference problems, subjects were asked to infer the likelihood of rate $\theta_A$ where the rates have equal priors, the sample has size $N = 12$, and number of $a$-signals is $A_s = 7$, 8, 9, or 10. When the rates are close together, $(\theta_A, \theta_B) = (.6, .5)$, the subjects exhibit slight over-inference: a Bayesian's posteriors for these four inference problems would be .54, .64, .72, and .80, respectively, while subjects' median posteriors were .55, .66, .75, and .85. When the rates are further apart, $(\theta_A, \theta_B) = (.6, .25)$, subjects exhibited massive under-inference: whereas a Bayesian's posteriors in these problems would be .95, .98, .998, and .999, respectively, subjects' posteriors were .60, .70, .80, and .90.

[9] For Green, Halbert, and Robinson (1965), we can also reach this conclusion by estimating the regression equation (2) on the data reported just in that paper. There are a number of other studies that manipulate sample size but do not analyze or display the data in a way that makes it clear how sample size affects the degree of bias in inference: Sanders (1968); Peterson and Swensson's (1968) Study 2; Beach, Wise, and Barclay (1970); Marks and Clarkson (1972); and DeSwart (1972a, 1972b).

[10] DeSwart (1972a, 1972b) manipulates how far $\theta_A$ is from .5 but does not analyze or display the data in a way that makes it clear how it affects the degree of under-inference.

Griffin and Tversky's evidence cannot be fully explained by extreme-belief aversion (that simply maps an objective posterior into a less extreme subjective one) because, for example, subjects' posterior of .80 is identical whether $\theta_B = .5$ and 10 $a$-signals were observed or $\theta_B = .25$ and 9 $a$-signals were observed, but the objective posteriors are quite different in those two cases.

Our one-parameter model of NBLLN does predict $\gamma_3 < 1$, i.e., "more severe" under-inference (to be precise, Barneys likelihood ratio being closer to 1 than Tommys likelihood ratio, conditional on any realized sample) when the symmetric rates are further apart. Nonetheless, we suspect that SDTSD also contributes to explaining the flat tails of subjective sampling distributions and the discriminability phenomenon in inference.

Column 5 estimates the same regression as column 4, but with the data restricted to incentivized experiments. There are only three such studies—Green, Halbert, and Robinson (1965), Nelson, Bloomfield, Hales, and Libby (2001), and Kraemer and Weber (2004)—but the results in column 5 are largely similar to column 4, except that both $\widehat{\gamma}_1 = .675$ and $\widehat{\gamma}_3 = .421$ are larger, suggesting greater sensitivity to sample size and to rates when accurate inferences are rewarded. Nonetheless, both coefficients remain far less than 1, indicating substantial biases relative to Bayesian inference.

Training in inference appears to reduce but not eliminate under-inference. When subjects were told after each inference which state actually occurred, they became less biased over time but still under-inferred by the end of the experiment (Phillips and Edwards's, 1966, Study 2; Camerer, 1987). Strub (1969) found under-inference among subjects who had received 114 hours of lecture sessions, demonstrations, problem-solving sessions, and other training in dealing with probabilities, including prior participation in inference experiments. When subjects were told after each inference what the normatively correct inference is, they very quickly learned to report more extreme beliefs, but they do not seem to have learned to draw better inferences. While reporting more extreme beliefs led to more accurate beliefs in problems where their pre-training beliefs were not extreme enough, it led to less accurate beliefs in problems where their pre-training beliefs were accurate (Donnell and DuCharme, 1975).

A few studies have found over-inference. In all cases, $N$ is relatively small, and $\theta_A$ is relatively close to .5.[11] Griffin and Tversky's Study 1 (1992; $\theta_A = .6$) , which compared inference from samples of size 3, 5, 9, 17, and 33, found over-inference for $N = 3$ and 5 and under-inference for the others. Nelson, Bloomfield, Hales, and Libby's Study 1 (2001; $\theta_A = .6$) conducted an experimental asset market where payoffs depended on correct inferences from samples of size 3 and 17. Subjects under-inferred for $N = 17$ and over-inferred for $N = 3$. There is some evidence, though it is weak,

[11] Peterson and Swennson's (1968) Study 1 finds over-inference for $N = 1$ and $\theta_A = .6, .67, .75, .9$ in the first half of their data. In the same inference problems in the second half of their data from Study 1, and in both halves from Study 2, however, they find under-inference.

that over-inference in favor of a particular state occurs when the realized sample exactly matches the expected sample in that state, a phenomenon that has been called "exact representativeness." In an experimental asset market, Camerer (1987; $N = 3$, $\theta_A = .67$) found that the price of a state-contingent asset that pays off if state $A$ is true was too high—indicating over-inference in favor of state $A$—when the observed sample contained exactly 2 $a$-signals and 1 $b$-signal, and symmetrically over-inference in favor of state $B$ when the sample proportions were reversed. In an experimental asset market with asymmetric rates of $\theta_A = .67$ and $\theta_B = .5$, Grether (1980; $N = 6$) similarly found evidence indicating over-inference in favor of state $A$ when the realized sample was 4 $a$'s and 2 $b$'s and over-inference in favor of state $B$ when the realized sample was 3 $a$'s and 3 $b$'s. In a similar experiment, Grether (1992) found less support for "exact representativeness." Neither NBLLN alone, nor NBLLN combined with SDTSD, can explain over-inference.

Such over-inference can, however, be explained by the Law of Small Numbers (LSN). As shown in Appendix A, when NBLLN and LSN are combined in a single model, small $N$ is a necessary condition for over-inference. Moreover, we have argued that SDTSD will tend to generate under-inference when $\theta_A$ and $\theta_B$ are far apart.

A distinctive feature of our theory—a feature that differentiates it from alternative theories of under-inference from large samples discussed in Section 6 and Appendix A—is the prediction that inferences from a sample of size 1 will be correct. Sample-size neglect predicts over-inference for samples of size 1, while extreme-belief aversion predicts under-inference for samples of any size, including 1. There are 11 experiments that measure inference when $N = 1$. Peterson, Schneider, and Miller (1965; $\theta_A = .6$), Dave and Wolfe (2003; $\theta_A = .7$), Peterson and Swennson's Study 2 (1968; $\theta_A = .6, .67, .75, .9$), and Gettys and Manley's (1968) Studies 1 and 2 (which used a variety of asymmetric inference problems) found substantial under-inference. Chinnis and Peterson (1968; $\theta_A = .67, .8$), Kraemer and Weber (2004; $\theta_A = .6$), and Sasaki and Kawagoe (2007; $\theta_A = .67$) found slight under-inference, very close to Bayesian, and Peterson and Swennson's (1968; $\theta_A = .6, .67, .75, .9$) Study 1 found over-inference in the first half of their data and under-inference in the second half. In a mix of symmetric and asymmetric problems, Peterson and Miller (1965) found under-inference for $(\theta_A, \theta_B) = (.83, .17)$, $(\theta_A, \theta_B) = (.71, .2)$, and $(\theta_A, \theta_B) = (.67, .33)$, and over-inference for $(\theta_A, \theta_B) = (.6, .43)$. Green, Halbert, and Robinson (1965; $\theta_A = .6, .8$) found inferences very close to Bayesian when $N = 1$. The evidence is mixed but with more of the studies leaning toward under-inference.[12]

[12] Presumably, the many papers on base-rate neglect also contain evidence on inferences from samples of size 1. Virtually none of them have 50-50 priors, however, so it is difficult to disentangle biased inference from base-rate neglect. We do not review this literature systematically, but to give a flavor of what it may indicate, we examined Bar-Hillel's (1980) seminal paper. Our impression is that the evidence in Bar-Hillel's paper roughly mirrors the evidence from experiments on single-signal inference reviewed above. The

As an aside, we can use the same balls-and-urns experiments to study how priors affect inference and thereby measure the prevalence and extent of base-rate neglect in these studies. Taking the log of Bayes' Rule and rearranging:

$$\ln \frac{f_{\Theta|S_N}(\theta_A|s)}{f_{\Theta|S_N}(\theta_B|s)} - \ln \frac{f_{S_N|\Theta}(s|\theta_A)}{f_{S_N|\Theta}(s|\theta_B)} = \ln \frac{f_\Theta(\theta_A)}{f_\Theta(\theta_B)}. \tag{3}$$

As above, from the published experiments, we obtain for each inference problem the values $\theta_A$, $\theta_B$, $f_\Theta(\theta_A)$, the actual sample observed by subjects, and subjects' mean or median reported posterior. The right-hand side of equation (3) can be readily calculated from $f_\Theta(\theta_A)$, as can the first term on the left-hand side, $\ln \frac{f_{\Theta|S_N}(\theta_A|s)}{f_{\Theta|S_N}(\theta_B|s)}$, from subjects' posterior. In order not to confound base-rate neglect with other biases (that affect inference even when priors are equal), we calculate the second term on the left-hand side, $\ln \frac{f_{S_N|\Theta}(s|\theta_A)}{f_{S_N|\Theta}(s|\theta_B)}$, as the predicted value, $\ln \widehat{\frac{f_{S_N|\Theta}(s|\theta_A)}{f_{S_N|\Theta}(s|\theta_B)}}$, from the previously-estimated regression equation (2). We use only the symmetric-rate data. Figure B3 plots the left-hand side of equation (3) against $\ln \frac{f_\Theta(\theta_A)}{f_\Theta(\theta_B)}$. Regardless of whatever other biases may affect inference, if the subjects correctly incorporate base rates into their inferences, the data should lie along the identity line (the dashed line in Figure B3). However, the best-fitting line (the solid line in the figure) has a slope less than 1, indicating that subjects' inferences are too insensitive to the prior probabilities.

full distribution of subjects' reported posteriors can be eyeballed from histograms reported for Bar-Hillel's Studies 1, 2, 3, 7 and 8, each of which presents an inference problem where a single signal is indicative of the less likely of two states that subjects are given base rates for. We divide the 222 subjects' responses into four categories. Because the signal strength always was in the opposite direction of the base rate, the 33% of subjects whose posteriors equaled the base rate or weaker must have been either under-inferring from the signal or (as is presumably unlikely) "over-using" the base rate. By contrast, 9% of subjects reported posteriors stronger than the signal, almost surely indicating over-use of the signal (since otherwise they must be reversing the base rate). 31% of subjects reported posteriors of exactly the signal strength. Although not logically necessary, we share the presumption of Bar-Hillel and most researchers in this area that these subjects were almost surely simply using the signal strength and ignoring the base rate altogether. The remaining 27% of subjects reported posteriors strictly between the correct Bayesian posteriors and the posteriors that would completely ignore the base rate. It is unclear how many of these subjects were over-using or under-using the signal because anyone under-using the base rate could have been either over-inferring or under-inferring from the signal. From these data taken together, it seems likely that between 9% and 36% of the subjects were over-inferring from the signal, and at least 33% of the subjects were under-using the signal.

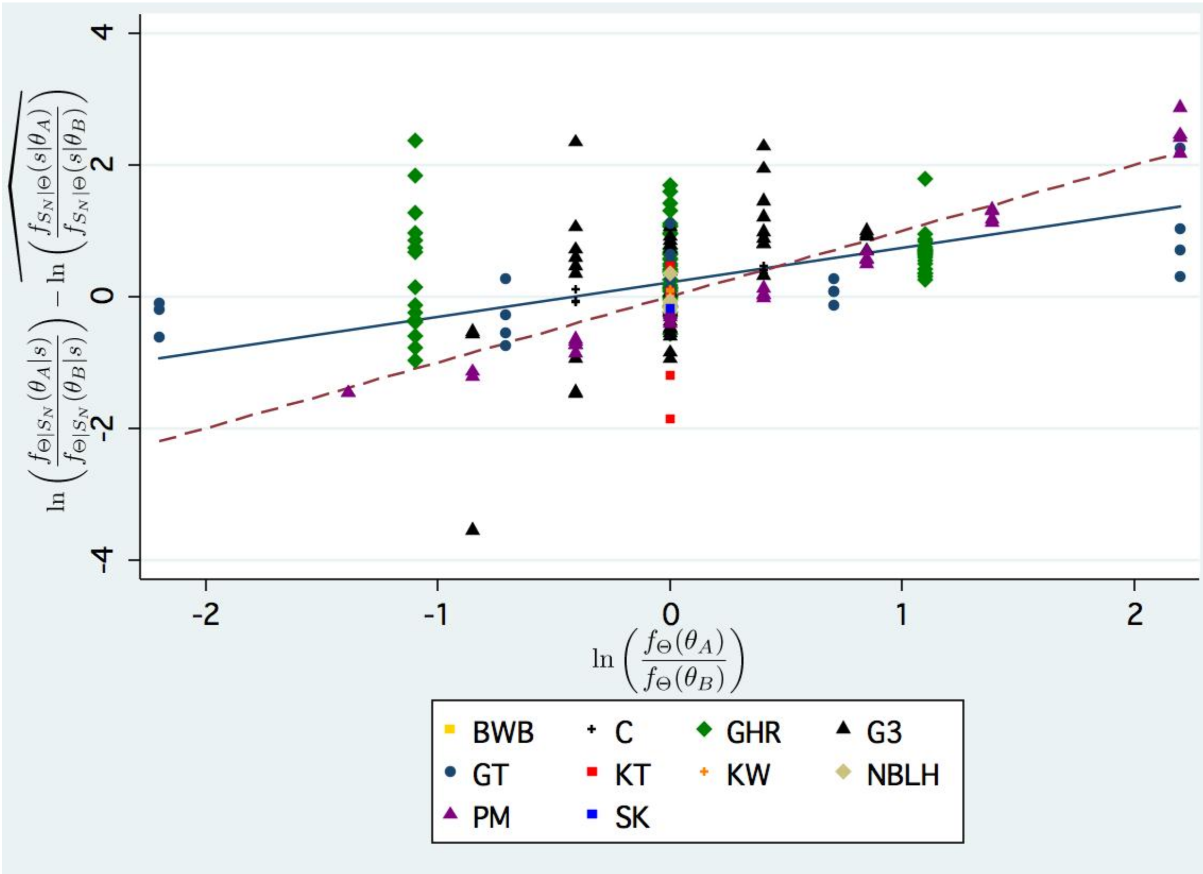


Figure B3: Base-rate neglect in symmetric-rate inference problems.

Notes: The dotted line represents the null hypothesis of Bayesian updating, and the solid line is the best-fitting regression line from column 1 of Table B5. The includes studies are: BWB = Beach, Wise, and Barclay (1970); C = Camerer (1987); GHR = Green, Halbert, and Robinson (1965); G3 = Grether's (1992) Study 3; GT = Griffin and Tversky's (1992) Study 1; KT = Kahneman and Tversky (1972); KW = Kraemer and Weber (2004); NBHL = Nelson, Bloomfield, Hales, and Libby's (2001) Study 1; PM = Peterson and Miller (1965); SK = Sasaki and Kawagoe (2007).

To formally investigate the degree of base-rate neglect, we rewrite (3) as a regression equation:

$$\ln \frac{f_{\Theta|S_N}(\theta_A|s)}{f_{\Theta|S_N}(\theta_B|s)} - \ln \frac{f_{S_N|\Theta}(s|\theta_A)}{f_{S_N|\Theta}(s|\theta_B)} = \varphi_0 + \varphi_1 \ln \frac{f_\Theta(\theta_A)}{f_\Theta(\theta_B)} + \varepsilon. \tag{4}$$

The null hypothesis of Bayesian updating corresponds to $\varphi_0 = 0$ and $\varphi_1 = 1$.

The first column of Table B3 estimates equation (4) on the full sample. While we can reject the null hypothesis that $\varphi_0 = 0$, it is clear from Figure B5 that the deviation is small. On the other hand, the fitted value $\widehat{\varphi}_1 = .601$ is significantly less than 1 (and larger than 0). The second column estimates the equation using only the data with unequal prior probabilities, but the results are similar. Finally, the third column restricts the data to incentivized experiments. While $\widehat{\varphi}_0$ is now essentially zero, $\widehat{\varphi}_1 = .405$ indicates even stronger base-rate neglect in these data.

Table B3. Base-rate neglect.

| | (1) All data | (2) Only unequal priors | (5) Only incentivized |
|---|---|---|---|
| $\ln \frac{f_\Theta(\theta_A)}{f_\Theta(\theta_B)}$ | 0.524 | 0.510 | 0.360 |
| | (0.078) | (0.081) | (0.115) |
| Constant | 0.216 | 0.299 | -0.130 |
| | (0.053) | (0.087) | (0.084) |
| $R^2$ | 0.242 | 0.303 | 0.176 |
| #obs | 209 | 110 | 48 |
| #papers | 10 | 5 | 3 |

Notes: Results are from OLS regressions, with standard errors in parentheses. The dependent variable is as described in the text.

### 0.3 Consistency Between Subjective Sampling Distributions and Inference

A core feature of the fully rational model that our model retains is that people draw inferences that are consistent with Bayes's Rule applied to their subjective sampling distributions. Although surely not a perfect fit, we believe (like other researchers before us) that this feature is approximately right except insofar as people neglect base rates. There is indirect supportive evidence from the qualitative correspondence between the evidence on subjective sampling distributions (reviewed in section B.1) and the evidence on inferences (reviewed in section B.2). There is direct evidence from two studies that measured subjective sampling distributions and inferences for the same subject.

Peterson, DuCharme, and Edwards (1968, Study 2) conducted symmetric inference experiments with every combination of $N = 3$, 5, 8, and $\theta_A = 1 - \theta_B = .6, .7, .8$. Then subjects drew subjective sampling distributions for the nine binomial distributions (shown in Figure B4). Peterson, DuCharme, and Edwards plotted subjects' inferences against what their inferences would be if they applied Bayes's Rule to their subjective sampling distributions. Peterson, DuCharme, and Edwards found that "most points cluster extremely close to the identity line."

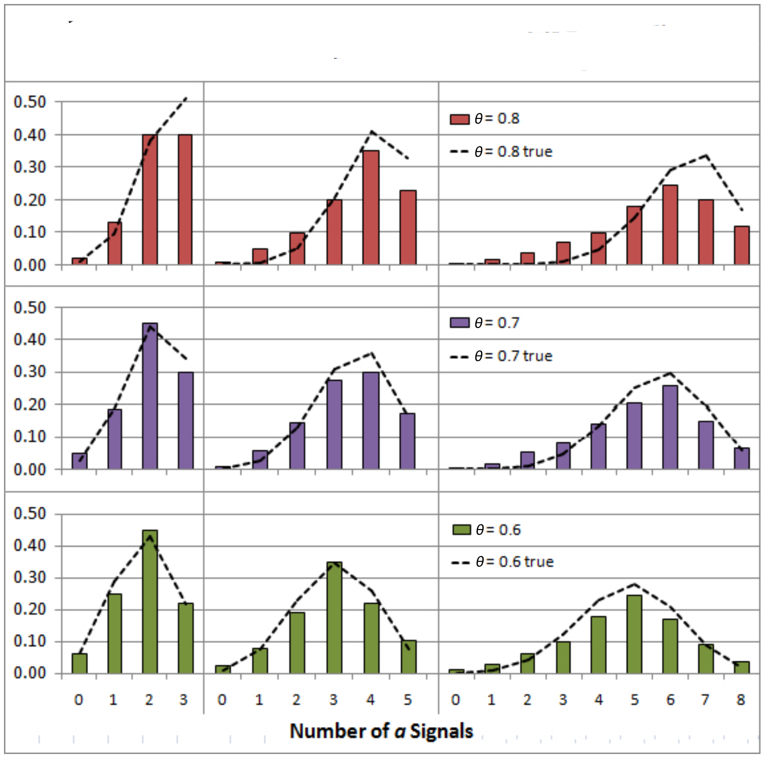


Figure B4: Median probability estimates, N=3, N=5, and N=8 (Peterson DuCharme and Edwards 1968)

Wheeler and Beach (1968) elicited subjects' subjective sampling distributions for a sample of size $N = 8$ for rates $\theta = .6$ and .8 (see Figure B5) and then asked subjects to make bets in an inference task. Wheeler and Beach inferred subjective posteriors from the bets, under the assumption that subjects sought to maximize expected winnings. The correlation between the median subjective log-posterior (calculated from the first 20 inference problems) and the median subjective log-likelihood (calculated from applying Bayes's Rule to the subjective sampling distribution elicited at the very beginning of the experiment) was .90. For individual subjects' data, the median correlation was .85.

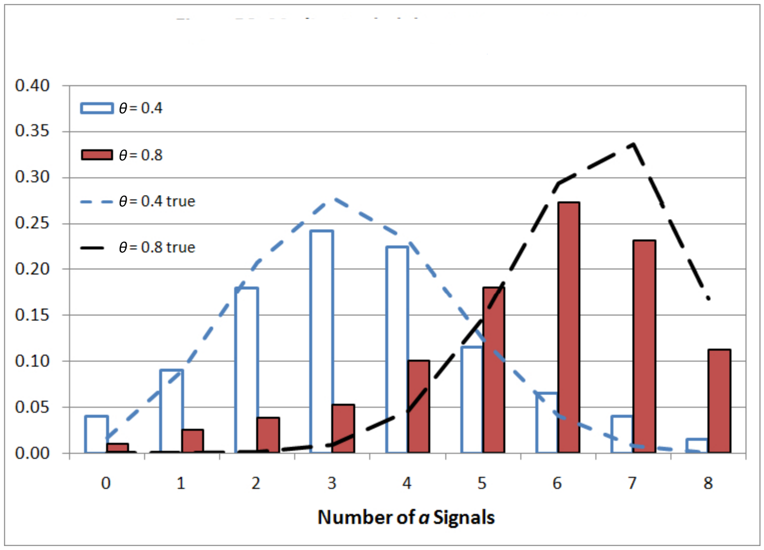


Figure B5: Median probability estimates, N=8 (Wheeler and Beach 1968)

### 0.4 Inference For Sequential Clumps

In Section 4, we lay out various possible dynamic extensions of our model of NBLLN. Unfortunately, there are no experiments that elicit people's beliefs about what they will infer conditional on observing samples in the future. Here we review the few experiments that compare inferences from a sample presented simultaneously with a sample drawn sequentially.

In symmetric inference with $\theta = .6$ and $N = 48$, Peterson, Schneider, and Miller (1965) presented the sample clumped as a single sample of size 48, four samples of size 12, twelve samples of size 4, or forty-eight samples of size 1. Subjects reported their updated beliefs after each clump was shown. The results are somewhat difficult to interpret because the data are averaged across clumps and realizations. Nonetheless, Peterson et al.s finding that under-inference is more severe when clumps contain more signals is consistent with either retrospective-acceptive or retrospective-pooling.

Grether (1992, Study 3) confronted subjects with (incentivized) symmetric, binomial inference problems, with rates $\theta_A = 1 - \theta_B = .2, .3, .4, .6, .7, .8$, and priors $f_\Theta(\theta_A) = .3, .4, .5, .6, .7$. The sample size always began as $N = 4$. In some cases, however, after subjects made their inference, they were asked to make an updated inference after an additional 4 signals were drawn, up to a maximum of 12 signals in total. Although only aggregate statistics are reported in the paper, David Grether sent us the subject-level data. In a few cases, we can learn about how subjects process the signals by comparing their inferences before and after they receive a clump. For example, in one situation, the rates were $\theta_A = 1 - \theta_B = .2$, the prior probabilities of the rates were equal, and the first four signals were all $b$'s. The next four signals were 2 $a$'s and 2 $b$'s. The objective posterior probability of rate $\theta_B$ is the same after all eight as after the first four: .9961. However, the subjects' subjective posterior (that is, the median across subjects) is .95 after the first four signals and .70 after all eight. This pattern of additional uninformative signals causing the subjective posterior to move toward .5 is consistent with retrospective-pooling Barney, not retrospective-acceptive Barney. The same pattern holds in all three other test cases in Grether's data.[13]

In contrast, Kraemer and Weber (2004) present evidence that supports retrospective-acceptive Barney. For an incentivized, symmetric binomial inference problem with $\theta_A = 1 - \theta_B = .6$ and $N = 5$, subjects presented with a sample of 3 $a$'s and 2 $b$'s gave mean posterior probability for rate $\theta_A$ of .585. Other subjects who were instead shown the same signals as two separate samples, one with 3 $a$'s and 0 $b$'s and one with 0 $a$'s and 2 $b$'s, gave mean posterior probability for state A of .56, which is marginally statistically different. Similarly, with $\theta_A = 1 - \theta_B = .6$ and $N = 25$, subjects' mean posterior probability for rate $\theta_A$ was .56 when the sample was 13 $a$'s and 12 $b$'s, but .53 (strongly statistically distinguishable from .56) when two samples of 13 $a$'s and 0 $b$'s, and then 0 $a$'s and 12 $b$'s were presented sequentially. The fact that subjects make different inferences in these two cases is inconsistent with being retrospective-pooling, but not inconsistent with being retrospective-acceptive.[14]

Shu and Wu's (2003) Study 3 appears to be inconsistent with any of the dynamic extensions of our model that we consider. They conduct a symmetric binomial inference problem with samples of size $N = 10$ and three different levels of the rates, $\theta_A = 1 - \theta_B = .6$, .75, or .9. In one condition, subjects observed the 10 signals one at a time before stating a posterior belief. In the other conditions, subjects observed the 10 signals in clumps of 2 signals each or 5 signals each. While for some realizations of the 10-signal sets subjects draw less extreme inferences when the signals arrive in larger clumps—as predicted for retrospective-acceptive Barney—the results on average

[13] In the $2^{nd}$ test case, $f_\Theta(\theta_A) = .4$, $\theta_A = 1 - \theta_B = .3$, and the first eight signals were 1 $a$ and 7 $b$'s. The next four signals were 2 $a$'s and 2 $b$'s. While the objective posterior probability of rate $\theta_B$ is identically .9908, subjects' median subjective posterior fell from .9 to .5. In the $3^{rd}$ test case, $f_\Theta(\theta_A) = .7$, $\theta_A = 1 - \theta_B = .4$, and the first four signals were 2 $a$'s and 2 $b$'s, and the next four signals were 2 $a$'s and 2 $b$'s. The objective posterior probability of rate $\theta_A$ remains .7, but subjects' median subjective posterior fell from .775 to .7. In the $4^{th}$ test case, $f_\Theta(\theta_A) = .3$, $\theta_A = 1 - \theta_B = .7$, and the first four and next four signals were 2 $a$'s and 2 $b$'s. The objective posterior probability of rate $\theta_A$ remains .3, but subjects' median subjective posterior increased from .42 to .51. If we examine subjects' mean posterior probabilities, the pattern is robust in the first three test cases (.87 to .66, .72 to .59, and .73 to .68) and hard to interpret in the fourth (.54 to .58).

[14] Unlike the experimental subjects, however, retrospective-acceptive Barney will under-infer weakly *more* from two clumps than from one. This is because an additional clump provides information about an additional draw of $\beta$ from the distribution, which is weakly more informative about the true rate than an additional signal about a fixed $\beta$. That being said, a combination of retrospective-acceptive Barneyness with SDTSD could explain why the two extreme clumps lead to a weaker overall inference than the single clump.

Kraemer and Weber (2004) also have a third experimental treatment that generates ambiguous evidence. In the $N = 5$ case, when subjects are presented with a sample of 1 $a$ and 1 $b$, followed by a sample of 2 $a$'s and 1 $b$, their mean posterior is .575, in between the other two treatments and not statistically distinguishable from either. Similarly, in the $N = 25$ case, the mean posterior is .555 when subjects are presented with a sample of 6 $a$'s and 6 $b$'s, followed by a sample of 7 $a$'s and 6 $b$'s.

go in the opposite direction.[15] While it may be possible to reconcile Shu and Wu's results with a combination of NBLLN and the dynamics of base-rate neglect, that combined model should be worked out to see if it systematically reverses some of the conclusions we reach in Section 5.

### 0.5 Evidence For Non-Binomial Distributions

While the vast majority of simple inference experiments have been conducted with binomial signals, there are a few studies with other distributions. The results overall are consistent with NBLLN and SDTSD applying beyond binomial subjective sampling distributions.

There are a handful of studies where signals are multinomial. For example, in Beach's (1968) experiment, there were two decks of cards, a Red Deck and a Green Deck. Each card had a letter from A to F written on it. The Red and Green Decks had equal priors, but each deck had different proportions of the lettered cards. The subjects were shown $N = 3$ cards, one card at a time, and reported their subjective probability that the cards were being drawn from the Red Deck, as opposed to the Green Deck, after each draw. The likelihood ratios for the cards ranged from 1:2.5 to 3:1. For example, the likelihood ratio for card F was 1:2. A second group of subjects faced the same inference task with the same likelihood ratios for each card, but with the absolute probabilities scaled down for some cards and scaled up for others. For example, for the first group of subjects, the probability of an F card was .03 for the Red Deck and .06 for the Green Deck; for the second group of subjects, the probability of an F card was .16 for the Red Deck and .32 for the Green Deck. The first main finding is that subjects under-inferred on average. The other finding was that, for a given objective likelihood ratio, Group 1 under-inferred more for cards where Group 1's probabilities for that card were scaled down relative to Group 2's. Our interpretation is that when Group 1's probabilities are scaled down, the observed sample lies further in the tails on the subjective sampling distribution for both decks. SDTSD predicts more extreme under-inference in such cases.

Under-inference is also the general finding in the other multinomial experiments we could find that compared subjects' posteriors with Bayesian posteriors (Phillips, Hays, and Edwards's Study 1, 1966; Dale, 1968; Martin, 1969; Martin and Gettys, 1969; Chapman, 1973). However, there are two exceptions: (1) Phillips, Hays, and Edwards (1966) varied the sample size of signals observed by subjects and, while finding under-inference for $N = 3$, 5, and 9, found essentially Bayesian inference for $N = 1$, and (2) Dale (1968) reported that in one particular trial where the data happened to

[15] Sanders (1968) and Beach, Wise, and Barclay (1970) also compare inferences from simultaneously-presented samples with inferences from sequentially-presented samples, but it is difficult to interpret their findings because the results for the simultaneously-presented samples are averaged across different sample sizes.

exactly match one of the multinomial rates, about 1/8 of the subject over-inferred—the most over-inference he observed on any trial. Martin (1969), Martin and Gettys (1969), and Chapman (1973) also reported that subjects' under-inferred by more when an observed sample warranted a more extreme conclusion. That finding could be due to NBLLN, SDTSD, or both.

DuCharme (1970) conducted a normal-signal inference experiment where subjects observed samples of size 1. He found under-inference when the sample was relatively far in the tails of both distributions, consistent with SDTSD, although he interpreted his results as meaning that people are reluctant to report extreme probabilities. Gustafson, Shukla, Delbecq, and Walster (1973) told subjects the average heights and weights of Midwestern college-age men and women. Subjects were then asked to draw inferences from samples of size 1; specifically, they were asked a series of questions such as, "The observed height of a person is 68 inches. Is the person more likely to be a male or female? How much more likely?" Gustafson et al. found that subjects over-inferred when the objective likelihood ratio was relatively small and under-inferred when the objective likelihood ratio was relatively large. Assuming that subjects believed that the sampling distributions for height and weight were normal distributions, this result means that subjects over-inferred when the sample was relatively close to the men's or women's mean height or weight and under-inferred when the sample was relatively far in the tails of both distributions. In two studies, DuCharme and Peterson (1968) familiarized subjects with normal distributions for male and female heights and then elicited subjects' beliefs that a sample was being drawn from the population of men or of women. Subjects' posteriors were nearly Bayesian when $N = 1$, but subjects under-inferred for samples of size $N = 4$.

Peterson and Phillips (1966) conducted an an experiment where the rate generating binary signals was drawn from a uniform distribution on $[0, 1]$. Subjects observed 48 binary signals and after each signal had to specify a 33% confidence interval for the rate. Subjects' confidence intervals were almost always too wide, indicating that subjects under-inferred from the data about the rate.

**Appendix C: Proofs**

## 0.6 Preliminary and Appendix A Results

According to Whitt's (1979) Theorem 4, A3 implies that for any $N$ and any $\theta_A, \theta_B \in \Theta$, Barney's likelihood ratio $\frac{f^{\psi}_{S_N|\Theta}(s|\theta_A)}{f^{\psi}_{S_N|\Theta}(s|\theta_B)}$ is strictly increasing in the number of $a$-signals in the sample. We will repeatedly use this fact in the proofs.

The preliminary results all apply to the beta-distribution functional form, given by equation (3) in the main text. Let $p_0 \equiv \frac{f_\Theta(\theta_A)}{f_\Theta(\theta_B)}$ denote the agent's prior ratio, let $\Pi^{\psi}_{S_N|\Theta\times\Theta}(s|\theta_A,\theta_B) \equiv \frac{f^{\psi}_{S_N|\Theta}(s|\theta_A)}{f^{\psi}_{S_N|\Theta}(s|\theta_B)}$ denote Barney's likelihood ratio after observing a clump of $N$ signals $s \in S_N$, and let $\Pi^{\psi}_{\Theta\times\Theta|S_N}(\theta_A,\theta_B|s) \equiv \frac{f^{\psi}_{\Theta|S_N}(\theta_A|s)}{f^{\psi}_{\Theta|S_N}(\theta_B|s)} = p_0 \Pi^{\psi}_{S_N|\Theta\times\Theta}(s|\theta_A,\theta_B)$ denote his posterior ratio.

**Lemma $\beta$1.** *Assume Barney has the beta-distribution functional form given by equation (3). Suppose Barney is prospective-acceptive, knows the rate is $\theta \in \Theta$, and will observe a random clump of $N$ signals $s \in S_N$. Then he believes the probability of observing $A_s$ a-signals in the sequence $s$ is:*

$$f^{\psi}_{S_N|\Theta}(s|\theta_A) = \frac{\Gamma(\psi)}{\Gamma(\psi+N)} \frac{\Gamma(\theta\psi + A_s)}{\Gamma(\theta\psi)} \frac{\Gamma((1-\theta)\psi + N - A_s)}{\Gamma((1-\theta)\psi)} \frac{\Gamma(N+1)}{\Gamma(A_s+1)\Gamma(N-A_s+1)}.$$

**Proof:**

$$\begin{aligned}
& f^{\psi}_{S_N|\Theta}(s|\theta_A) \\
= & \int_0^1 f_{S_N|\text{ß}}(s|\beta) f^{\psi}_{\text{ß}|\Theta}(\beta|\theta_A)\, d\beta \\
= & \int_0^1 \frac{\Gamma(N+1)}{\Gamma(A_s+1)\Gamma(A_s-N+1)} \beta^{A_s}(1-\beta)^{N-A_s} \frac{\Gamma(\psi)}{\Gamma(\theta\psi)\Gamma((1-\theta)\psi)} \beta^{\theta\psi-1}(1-\beta)^{(1-\theta)\psi-1}\, d\beta \\
= & \frac{\Gamma(\psi)}{\Gamma(\psi+N)} \frac{\Gamma(\theta\psi+A_s)}{\Gamma(\theta\psi)} \frac{\Gamma((1-\theta)\psi+N-A_s)}{\Gamma((1-\theta)\psi)} \frac{\Gamma(N+1)}{\Gamma(A_s+1)\Gamma(N-A_s+1)} \times \\
& \int_0^1 \frac{\Gamma(\psi+N)}{\Gamma(\theta\psi+A_s)\Gamma((1-\theta)\psi+N-A_s)} \beta^{\theta\psi+A_s-1}(1-\beta)^{(1-\theta)\psi+N-A_s-1}\, d\beta \\
= & \frac{\Gamma(\psi)}{\Gamma(\psi+N)} \frac{\Gamma(\theta\psi+A_s)}{\Gamma(\theta\psi)} \frac{\Gamma((1-\theta)\psi+N-A_s)}{\Gamma((1-\theta)\psi)} \frac{\Gamma(N+1)}{\Gamma(A_s+1)\Gamma(N-A_s+1)},
\end{aligned}$$

where the fourth equality follows because the term being integrated is the pdf of a beta distribution, and the integral of a pdf is equal to 1.

□

**Lemma $\beta$2.** *Assume Barney has the beta-distribution functional form given by equation (3) and is retrospective pooling. Consider two states of the world, A and B, associated with known rates $\theta_A, \theta_B \in \Theta$ and a prior ratio $p_0 = \frac{f(\theta_A)}{f(\theta_B)} \in (0, \infty)$. Barney's likelihood ratio after observing a clump of $N$ signals $s \in S_N$ which has $A_s$ a-signals is:*

$$\Pi^{\psi}_{S_N|\Theta\times\Theta}(s|\theta_A, \theta_B) = \frac{\Gamma(\theta_A\psi + A_s)}{\Gamma(\theta_B\psi + A_s)} \frac{\Gamma((1-\theta_A)\psi + N - A_s)}{\Gamma((1-\theta_B)\psi + N - A_s)} \frac{\Gamma(\theta_B\psi)}{\Gamma(\theta_A\psi)} \frac{\Gamma((1-\theta_B)\psi)}{\Gamma((1-\theta_A)\psi)}.$$

*Moreover, in the "symmetric inference" case where $\theta_A = 1 - \theta_B$,*

$$\Pi^{\psi}_{S_N|\Theta\times\Theta}(s|\theta_A, \theta_B) = \frac{\Gamma(\theta_A\psi + A_s)}{\Gamma(\theta_B\psi + A_s)} \frac{\Gamma((1-\theta_A)\psi + N - A_s)}{\Gamma((1-\theta_B)\psi + N - A_s)}$$

**Proof:** This is an immediate implication of the previous lemma.

□

**Lemma $\beta$3.** *Assume Barney has the beta-distribution functional form given by equation (3). Consider two states of the world, A and B, associated with known rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A \geq \theta_B$. Suppose Barney has observed $N$ signals, $s \in S_N$, and he observes his $N + 1^{st}$ signal. Then Barney's likelihood ratio is closer to $\Pi^{\psi}_{S_N|\Theta\times\Theta}(s|\theta_A, \theta_B)$ if he pools the $N + 1^{st}$ signal together with the others than if he groups it separately: that is, if the $N + 1^{st}$ signal is an a-signal, then*

$$\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A, \theta_B)\, \Pi^{\psi}_{S_N|\Theta\times\Theta}(s|\theta_A, \theta_B) \geq \Pi^{\psi}_{S_{N+1}|\Theta\times\Theta}(s \cup a|\theta_A, \theta_B),$$

*and if $N + 1^{st}$ signal is a b-signal, then*

$$\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A, \theta_B)\, \Pi^{\psi}_{S_N|\Theta\times\Theta}(s|\theta_A, \theta_B) \leq \Pi^{\psi}_{S_{N+1}|\Theta\times\Theta}(s \cup a|\theta_A, \theta_B).$$

**Proof:** We will prove the claim in the lemma that pertains if the final signal is an $a$-signal; the proof for the other case is analogous. Denote by $A_N$ the number of $a$-signals in $s_N$. If Barney pools the $a$-signal with the others, then his likelihood ratio is

$$\begin{aligned}
& \Pi^{\psi}_{S_{N+1}|\Theta\times\Theta}\left(s\cup a|\theta_A,\theta_B\right) \\
= & \frac{\Gamma\left(\theta_A\psi+A_N+1\right)}{\Gamma\left(\theta_B\psi+A_N+1\right)}\frac{\Gamma\left(\left(1-\theta_A\right)\psi+N+1-A_N-1\right)}{\Gamma\left(\left(1-\theta_B\right)\psi+N+1-A_N-1\right)}\frac{\Gamma\left(\theta_B\psi\right)}{\Gamma\left(\theta_A\psi\right)}\frac{\Gamma\left(\left(1-\theta_B\right)\psi\right)}{\Gamma\left(\left(1-\theta_A\right)\psi\right)} \\
= & \frac{\left(\theta_A\psi+A_N\right)\Gamma\left(\theta_A\psi+A_N\right)}{\left(\theta_B\psi+A_N\right)\Gamma\left(\theta_B\psi+A_N\right)}\frac{\Gamma\left(\left(1-\theta_A\right)\psi+N-A_N\right)}{\Gamma\left(\left(1-\theta_B\right)\psi+N-A_N\right)}\frac{\Gamma\left(\theta_B\psi\right)}{\Gamma\left(\theta_A\psi\right)}\frac{\Gamma\left(\left(1-\theta_B\right)\psi\right)}{\Gamma\left(\left(1-\theta_A\right)\psi\right)} \\
= & \frac{\theta_A\psi+A_N}{\theta_B\psi+A_N}\Pi^{\psi}_{S_N|\Theta\times\Theta}\left(s|\theta_A,\theta_B\right),
\end{aligned}$$

where the second equality uses Lemma $\beta 2$, and the third equality uses the fact that $\Gamma(n+1) = n\Gamma(n)$. If Barney instead separately groups the final signal, then his likelihood ratio is

$$\begin{aligned}
\Pi^{\psi}_{S_1|\Theta\times\Theta}\left(a|\theta_A,\theta_B\right)\Pi^{\psi}_{S_N|\Theta\times\Theta}\left(s|\theta_A,\theta_B\right) &= \frac{\theta_A\psi}{\theta_B\psi}\Pi^{\psi}_{S_N|\Theta\times\Theta}\left(s|\theta_A,\theta_B\right) \\
&= \frac{\theta_A}{\theta_B}\Pi^{\psi}_{S_N|\Theta\times\Theta}\left(s|\theta_A,\theta_B\right).
\end{aligned}$$

By assumption, $\frac{\theta_A}{\theta_B} \geq 1$. Therefore, for all $k > 0$,

$$\frac{\theta_A}{\theta_B} \geq \frac{\theta_A\psi+k}{\theta_B\psi+k} \geq 1.$$

The result follows.

□

**Lemma $\beta$4.** *Assume Barney has the beta-distribution functional form given by equation (3) and is retrospective pooling. Consider two states of the world, A and B, associated with known rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A > \theta_B$ and a prior ratio $p_0 = \frac{f(\theta_A)}{f(\theta_B)} \in (0,\infty)$.*

- *Barney draws the same inferences as Tommy when the observed sequence is ab or ba.*
- *If the sample is all a's or all b's, Barney under-infers relative to Tommy.*
- *For any true state $\theta \in (0,1)$, Barney believes that: as $N \to \infty$, his posterior ratio $\frac{f^{\psi}_{\Theta|S_N}(\theta_A|s)}{f^{\psi}_{\Theta|S_N}(\theta_B|s)}$ will converge in distribution to a random variable that has full support on $(0,\infty)$.*

**Proof:**

- Barney's posterior after one $a$ and one $b$ signal is

$$\begin{aligned}
& p_0 \frac{\Gamma(\theta_A\psi+1)}{\Gamma(\theta_B\psi+1)} \frac{\Gamma((1-\theta_A)\psi+1)}{\Gamma((1-\theta_B)\psi+1)} \frac{\Gamma(\theta_B\psi)}{\Gamma(\theta_A\psi)} \frac{\Gamma((1-\theta_B)\psi)}{\Gamma((1-\theta_A)\psi)} \\
= \; & p_0 \frac{\Gamma(\theta_A\psi)\theta_A\psi}{\Gamma(\theta_B\psi)\theta_B\psi} \frac{\Gamma((1-\theta_A)\psi)(1-\theta_A)\psi}{\Gamma((1-\theta_B)\psi)(1-\theta_B)\psi} \frac{\Gamma(\theta_B\psi)}{\Gamma(\theta_A\psi)} \frac{\Gamma((1-\theta_B)\psi)}{\Gamma((1-\theta_A)\psi)} \\
= \; & p_0 \frac{\theta_A}{\theta_B} \frac{(1-\theta_A)}{(1-\theta_B)},
\end{aligned}$$

which equals Tommy's posterior.

- Barney's posterior after $N$ $a$-signals and 0 $b$-signals is:

$$\begin{aligned}
& p_0 \frac{\Gamma(\theta_A\psi+N)}{\Gamma(\theta_B\psi+N)} \frac{\Gamma((1-\theta_A)\psi)}{\Gamma((1-\theta_B)\psi)} \frac{\Gamma(\theta_B\psi)}{\Gamma(\theta_A\psi)} \frac{\Gamma((1-\theta_B)\psi)}{\Gamma((1-\theta_A)\psi)} \\
= \; & p_0 \frac{\Gamma(\theta_A\psi+N)}{\Gamma(\theta_B\psi+N)} \frac{\Gamma(\theta_B\psi)}{\Gamma(\theta_A\psi)} \\
= \; & p_0 \frac{(\theta_A\psi+N-1)(\theta_A\psi+N-2)...(\theta_A\psi)\Gamma(\theta_A\psi)}{(\theta_B\psi+N-1)(\theta_B\psi+N-2)...(\theta_B\psi)\Gamma(\theta_B\psi)} \frac{\Gamma(\theta_B\psi)}{\Gamma(\theta_A\psi)} \\
= \; & p_0 \frac{(\theta_A\psi+N-1)(\theta_A\psi+N-2)...(\theta_A\psi)}{(\theta_B\psi+N-1)(\theta_B\psi+N-2)...(\theta_B\psi)}.
\end{aligned}$$

Note that the numerator and denominator each has $N$ terms in it. Tommy's posterior ratio is

$$p_0 \frac{(\theta_A)^N}{(\theta_B)^N},$$

so also for Tommy, the numerator and denominator each has $N$ terms. Furthermore, since $\frac{\theta_A}{\theta_B} > 1$, for all $k > 0$, $\frac{\theta_A+k}{\theta_B+k} < \frac{\theta_A}{\theta_B}$. Therefore,

$$p_0 \frac{(\theta_A\psi+N-1)(\theta_A\psi+N-2)...(\theta_A\psi)}{(\theta_B\psi+N-1)(\theta_B\psi+N-2)...(\theta_B\psi)} < p_0 \frac{(\theta_A)^N}{(\theta_B)^N}.$$

Moreover, the likelihood ratio clearly favors state $A$ for both Tommy and Barney. Hence, Barney under-infers relative to Tommy. The case of $N$ $b$-signals and 0 $a$-signals proceeds analogously.

- $F^{\psi}_{\text{ß}|\Theta}(\beta|\theta)$ for our parameterized model is absolutely continuous (since it is simply the cdf of the beta distribution). Therefore, the results of Lemma 1 hold. Lemma 1 implies that Barney's subjective sampling distribution for an infinite sample has pdf $\lim_{N\to\infty} f^{\psi}_{S_N|\Theta}\left(\frac{A_s}{N}=\alpha|\theta\right) = f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta)$ a.s. Therefore, after observing proportion $\alpha$ of $a$-signals in a large sample, Barney anticipates having likelihood ratio $\frac{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_B)}$ a.s. Since the beta distribution satisfies the monotone likelihood ratio ordering (for a fixed $\psi$) as $\theta$ changes, by Whitt (1979), $\frac{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_B)}$ is strictly increasing in $\alpha$.

Furthermore,

$$\frac{f^{\psi}_{ß|\Theta}(\beta=0|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=0|\theta_B)} = \frac{\frac{\Gamma(\psi)}{\Gamma(\theta_A\psi)\Gamma((1-\theta_A)\psi)}0^{\theta_A\psi-1}(1-0)^{(1-\theta_A)\psi-1}}{\frac{\Gamma(\psi)}{\Gamma(\theta_B\psi)\Gamma((1-\theta_B)\psi)}0^{\theta_B\psi-1}(1-0)^{(1-\theta_B)\psi-1}} = \frac{\Gamma(\theta_B\psi)\Gamma((1-\theta_B)\psi)}{\Gamma(\theta_A\psi)\Gamma((1-\theta_A)\psi)}\frac{0^{(\theta_A-\theta_B)\psi}}{1^{(\theta_A-\theta_B)\psi}} = 0$$

and

$$\frac{f^{\psi}_{ß|\Theta}(\beta=1|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=1|\theta_B)} = \frac{\frac{\Gamma(\psi)}{\Gamma(\theta_A\psi)\Gamma((1-\theta_A)\psi)}1^{\theta_A\psi-1}0^{(1-\theta_A)\psi-1}}{\frac{\Gamma(\psi)}{\Gamma(\theta_B\psi)\Gamma((1-\theta_B)\psi)}1^{\theta_B\psi-1}0^{(1-\theta_B)\psi-1}} = \frac{\Gamma(\theta_B\psi)\Gamma((1-\theta_B)\psi)}{\Gamma(\theta_A\psi)\Gamma((1-\theta_A)\psi)}\frac{1^{(\theta_A-\theta_B)\psi}}{0^{(\theta_A-\theta_B)\psi}} = \infty.$$

Because we are using the beta density for the subjective rate distribution, Barney's beliefs put full support on $\alpha \in (0,1)$. Hence Barney thinks his large-sample likelihood ratio will converge in distribution to a random variable whose support is $\left(\frac{f^{\psi}_{ß|\Theta}(\beta=0|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=0|\theta_B)}, \frac{f^{\psi}_{ß|\Theta}(\beta=1|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=1|\theta_B)}\right) = (0,\infty)$. Since Barney's posterior ratio is his likelihood ratio times his prior ratio, the result follows.

□

**Lemma $\beta$5.** *If Barney has the beta-distribution functional form given by equation (3), then his subjective-rate distribution satisfies A1-A4.*

**Proof:** We will prove each property in turn.

- Clearly the full-support property holds since $f^{\psi}_{ß|\Theta}(\beta|\theta) = \frac{\Gamma(\psi)}{\Gamma(\theta\psi)\Gamma((1-\theta)\psi)}\beta^{\theta\psi-1}(1-\beta)^{(1-\theta)\psi-1} > 0$ for all $\beta \in (0,1)$ and $\theta \in (0,1)$. This function is also clearly point-wise continuous in $\theta$ since all the components are point-wise continuous. Note that $f^{\psi}_{ß|\Theta}(\beta|\theta)$ is defined for all $\beta \in [0,1]$, is Lebesgue-integrable with respect to $\beta$ and $F^{\psi}_{ß|\Theta}(\beta|\theta) = 0 + \int_0^{\beta} f^{\psi}_{ß|\Theta}(x|\theta)\,dx$ for all $\beta$ in $[0,1]$. Therefore, $F^{\psi}_{ß|\Theta}(\beta|\theta)$ is absolutely continuous with respect to $\beta$.

- Note that

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

Therefore,

$$\int_0^{1-x} f^{\psi}_{ß|\Theta}(\beta|\theta)\,d\beta=\int_0^{1-x} f^{\psi}_{ß|\Theta}(1-\beta|1-\theta)\,d(1-\beta).$$

The left-hand side is $F^{\psi}_{ß|\Theta}(1-x|\theta)$. Using the change-of-variables $\widetilde{\beta}=1-\beta$, the right-hand side becomes

$$\begin{aligned} -\int_1^x f^{\psi}_{ß|\Theta}\left(\widetilde{\beta}|1-\theta\right)d\widetilde{\beta} &= \int_x^1 f^{\psi}_{ß|\Theta}\left(\widetilde{\beta}|1-\theta\right)d\widetilde{\beta} \\ &= 1-\int_0^x f^{\psi}_{ß|\Theta}\left(\widetilde{\beta}|1-\theta\right)d\widetilde{\beta}. \end{aligned}$$

Hence

$$F^{\psi}_{ß|\Theta}(1-x|\theta)=1-F^{\psi}_{ß|\Theta}(x|1-\theta).$$

- Note that

$$\begin{aligned} \frac{f^{\psi}_{ß|\Theta}(\beta|\theta')}{f^{\psi}_{ß|\Theta}(\beta|\theta)} &= \frac{\frac{\Gamma(\psi)}{\Gamma(\theta'\psi)\Gamma((1-\theta')\psi)}\beta^{\theta'\psi-1}(1-\beta)^{(1-\theta')\psi-1}}{\frac{\Gamma(\psi)}{\Gamma(\theta\psi)\Gamma((1-\theta)\psi)}\beta^{\theta\psi-1}(1-\beta)^{(1-\theta)\psi-1}} \\ &= \frac{\Gamma(\theta\psi)\Gamma((1-\theta)\psi)}{\Gamma(\theta'\psi)\Gamma((1-\theta')\psi)}\left(\frac{\beta}{1-\beta}\right)^{(\theta'-\theta)\psi}. \end{aligned}$$

If $\theta'>\theta$, then this expression is increasing in $\beta$.

- The mean of the beta distribution with parameters $\theta\psi$ and $(1-\theta)\psi$ is $\frac{\theta\psi}{\theta\psi+(1-\theta)\psi}=\theta$.

□

**Lemma A.** *Assume Barney has the beta-distribution functional form given by equation (3). Fix rates $\theta_A,\theta_B\in\Theta$ such that $\theta_A\geq\theta_B$, prior $p_0=\frac{f(\theta_A)}{f(\theta_B)}\in(0,\infty)$ and some $0<\underline{\lambda}<\bar{\lambda}<\infty$. If Barney's posterior ratio after one signal is in $(\underline{\lambda},\bar{\lambda})$,*

$$[p_0\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B),p_0\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)]\subseteq(\underline{\lambda},\bar{\lambda}),$$

*then for any natural number $N$, we can construct a sequence of $N$ signals, $s \in S_N$, such that for every truncation of $s$—i.e., for every sequence comprising only the first $n \leq N$ signals of $s$—the posterior ratio after that truncation, $\Pi^{\psi}_{\Theta\times\Theta|S_n}(\theta_A, \theta_B|s_n)$, is in $(\underline{\lambda}, \bar{\lambda})$.*

**Proof:** We prove this by induction. By hypothesis, the result is true for the truncation $n = 1$. Assume that the likelihood ratio after $N$ signals is in $(\underline{\lambda}, \bar{\lambda})$, and we will show that we can add a signal, and it will still be in $(\underline{\lambda}, \bar{\lambda})$. Suppose that there are $A_N$ $a$-signals and $N - A_N$ $b$-signals. Since

$$[p_0\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A, \theta_B), p_0\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A, \theta_B)] \subseteq (\underline{\lambda}, \bar{\lambda})$$

and $\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A, \theta_B) \leq 1 \leq \Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A, \theta_B)$, it follows that $p_0 \in (\underline{\lambda}, \bar{\lambda})$.

There are two cases. The first is that Barney's posterior ratio after the sequence of $N$ signals is (weakly) less than $p_0$, and so $\Pi^{\psi}_{S_N|\Theta\times\Theta}(s_N|\theta_A, \theta_B) \leq 1$. In this case add an $a$-signal. The new posterior ratio is $p_0\Pi^{\psi}_{S_{N+1}|\Theta\times\Theta}(s_N \cup a|\theta_A, \theta_B)$. Now,

$$\begin{aligned}
\bar{\lambda} \geq p_0\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A, \theta_B) &\geq p_0\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A, \theta_B)\,\Pi^{\psi}_{S_N|\Theta\times\Theta}(s_N|\theta_A, \theta_B) \\
&\geq p_0\Pi^{\psi}_{S_{N+1}|\Theta\times\Theta}(s_N \cup a|\theta_A, \theta_B) \\
&\geq p_0\Pi^{\psi}_{S_N|\Theta\times\Theta}(s_N|\theta_A, \theta_B) \geq \underline{\lambda}
\end{aligned}$$

The first inequality is by hypothesis, the second is due to the fact that $\Pi^{\psi}_{S_N|\Theta\times\Theta}(s_N|\theta_A, \theta_B) \leq 1$, the third inequality is due to Lemma $\beta$3, the fourth inequality holds since the posterior ratio increases when an additional signal is an $a$-signal, and the last inequality is again by hypothesis.

The second case is that the likelihood ratio after $N$ signals is greater than $p_0$, in which case add a $b$-signal, and the argument is analogous.

□

**Lemma B.** *Assume Barney has the beta-distribution functional form given by equation (3). Fix rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A \geq \theta_B$ and prior $p_0 = \frac{f(\theta_A)}{f(\theta_B)} \in (0, \infty)$. If $p_0 \in [\frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)}, \frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B)}]$, then for any $N$, there exists a sequence of signals $s \in S_N$ such that for all truncations of $s$ to its first $i$ signals, denoted $s_i$ for $i = 0, 1, ..., N$, $p_0\Pi^{\psi}_{S_i|\Theta\times\Theta}(s_i|\theta_A, \theta_B) \in [\frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)}, \frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B)}]$.*

**Proof:** The lemma is a direct implication of the following claim: If

$$p_0 \in \left[\frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A, \theta_B)}, \frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A, \theta_B)}\right],$$

then for any $n$, there exists a sequence $\{s_i\}_{i=0}^{n}$, where for all $i = 0....n$: (i) $s_i \in S_i$, (ii) $s_i$ is a subsequence of $s_{i+1}$, and (iii)

$$p_0 \Pi^{\psi}_{S_{i+1}|\Theta\times\Theta}\left(s_i \cup a|\theta_A, \theta_B\right) \geq 1$$

and

$$p_0 \Pi^{\psi}_{S_{i+1}|\Theta\times\Theta}\left(s_i \cup b|\theta_A, \theta_B\right) \leq 1.$$

In words, we can construct a sequence of signals such that Barney's posterior ratio flips around 1 with each signal. We will prove this claim by induction and then show that it implies the lemma.

Clearly for $i = 0$ the claim is true, since the prior ratio is in $\left[\frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)}, \frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B)}\right]$. Now we will assume the statement is true up to $i$ and prove it is true for $i + 1$.

There are two cases. In the first case, assume that after $i$ signals, it is the $a$-signal that flips the posterior ratio around 1: $p_0 \Pi^{\psi}_{S_i|\Theta\times\Theta}\left(s_i|\theta_A, \theta_B\right) \leq 1$ and $p_0 \Pi^{\psi}_{S_{i+1}|\Theta\times\Theta}\left(s_i \cup a|\theta_A, \theta_B\right) \geq 1$ (since adding an $a$-signal must increase the posterior ratio). Assume that the claim is not true for $i + 1$: that is, there is no set of $i + 1$ signals for which $p_0 \Pi^{\psi}_{S_{i+2}|\Theta\times\Theta}\left(s_{i+1} \cup a|\theta_A, \theta_B\right) \geq 1$ and $p_0 \Pi^{\psi}_{S_{i+2}|\Theta\times\Theta}\left(s_{i+1} \cup b|\theta_A, \theta_B\right) \leq 1$. In particular, taking the set of $n$ signals plus an $a$-signal,

$$p_0 \Pi^{\psi}_{S_{i+2}|\Theta\times\Theta}\left(s_i \cup a \cup a|\theta_A, \theta_B\right) \geq 1$$

and

$$p_0 \Pi^{\psi}_{S_{i+2}|\Theta\times\Theta}\left(s_i \cup a \cup b|\theta_A, \theta_B\right) \geq 1.$$

Because the claim is true up to $n$ signals, however, we know that an additional $b$-signal must flip the posterior ratio below 1,

$$p_0 \Pi^{\psi}_{S_{i+1}|\Theta\times\Theta}\left(s_i \cup b|\theta_A, \theta_B\right) \leq 1.$$

But then we have identified a set of $i + 1$ signals for which the statement is true, namely $s_i \cup b$: we know that $s_i \cup b \cup b$ must also generate a posterior ratio below 1 (since adding another $b$-signal must decrease the likelihood ratio); and $s_i \cup b \cup a$ generates the same posterior ratio as $s_i \cup a \cup b$, which we know is above 1. So we have a contradiction. Therefore, either $s_i \cup a$ or $s_i \cup b$ must satisfy the statement. The proof for a $b$-signal proceeds analogously.

To see that the statement implies the lemma, assume WLOG that $p_0 \Pi^{\psi}_{S_i|\Theta\times\Theta}\left(s_i|\theta_A, \theta_B\right) \leq 1$, and that $p_0 \Pi^{\psi}_{S_{i+1}|\Theta\times\Theta}\left(s_i \cup a|\theta_A, \theta_B\right) \geq 1$. Since

$$p_0 \Pi^{\psi}_{S_{i+1}|\Theta\times\Theta}\left(s_i \cup a|\theta_A, \theta_B\right) \leq p_0 \Pi^{\psi}_{S_i|\Theta\times\Theta}\left(s_i|\theta_A, \theta_B\right) \Pi^{\psi}_{S_1|\Theta\times\Theta}\left(a|\theta_A, \theta_B\right)$$

(due to Lemma $\beta 3$), it follows that

$$1 \leq p_0 \Pi^{\psi}_{S_i|\Theta\times\Theta}\left(s_i|\theta_A, \theta_B\right) \Pi^{\psi}_{S_1|\Theta\times\Theta}\left(a|\theta_A, \theta_B\right).$$

Hence

$$\frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}\left(a|\theta_A, \theta_B\right)} \leq p_0 \Pi^{\psi}_{S_i|\Theta\times\Theta}\left(s_i|\theta_A, \theta_B\right) \leq 1 \leq \frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}\left(b|\theta_A, \theta_B\right)}.$$

The proof for the case that a $b$-signal flips the posterior ratio below 1 proceeds analogously.

□

**Lemma C.** *Assume Barney has the beta-distribution functional form given by equation (3). Fix rates $\theta_A, \theta_B \in \Theta$ such that $\theta_A \geq \theta_B$, prior ratio $p_0 = \frac{f_\Theta(\theta_A)}{f_\Theta(\theta_B)}$, and the true state $i \in \{A, B\}$ (with corresponding true rate $\theta_i$). Denote Barney's limit likelihood ratio conditional on the true state as $\Pi^{\psi}_{S_\infty|\Theta\times\Theta}\left(s_\infty|\theta_A, \theta_B, \Omega = i\right)$. Suppose Barney is retrospective-pooling. For any $0 < \underline{\lambda} < \bar{\lambda} < \infty$ satisfying the following two statements:*

1. $p_0 \Pi^{\psi}_{S_\infty|\Theta\times\Theta}\left(s_\infty|\theta_A, \theta_B, \Omega = i\right) \in (\underline{\lambda}, \bar{\lambda})$, *and*
2. $\left[p_0 \Pi^{\psi}_{S_1|\Theta\times\Theta}\left(b|\theta_A, \theta_B\right), p_0 \Pi^{\psi}_{S_1|\Theta\times\Theta}\left(a|\theta_A, \theta_B\right)\right] \subseteq (\underline{\lambda}, \bar{\lambda})$ *or* $p_0 \in \left[\frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)}, \frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B)}\right] \subseteq (\underline{\lambda}, \bar{\lambda})$,

*then there exists a set of infinite sequences $\hat{S}_\infty$ that are realized with strictly positive probability, such that for any truncation $s_i$, $i \in \{1, 2, ...\}$, of any $s_\infty \in \hat{S}_\infty$,*

$$\Pi^{\psi}_{\Theta\times\Theta|S_i}\left(\theta_A, \theta_B|s_i\right) \in (\underline{\lambda}, \bar{\lambda}).$$

**Proof:** The proof proceeds in three steps:

1. Show that the prior ratio is in $(\underline{\lambda}, \bar{\lambda})$. This is true under either of the assumptions in condition 2 of the lemma.

   - Since $\Pi^{\psi}_{S_1|\Theta\times\Theta}\left(b|\theta_A, \theta_B\right) \leq 1$ and $\Pi^{\psi}_{S_1|\Theta\times\Theta}\left(a|\theta_A, \theta_B\right) \geq 1$, the assumption

     $$\left[p_0 \Pi^{\psi}_{S_1|\Theta\times\Theta}\left(b|\theta_A, \theta_B\right), p_0 \Pi^{\psi}_{S_1|\Theta\times\Theta}\left(a|\theta_A, \theta_B\right)\right] \subseteq (\underline{\lambda}, \bar{\lambda})$$

     implies $p_0 \in (\underline{\lambda}, \bar{\lambda})$.
   - Alternatively, we directly assume that $p_0 \in \left[\frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)}, \frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B)}\right] \subseteq (\underline{\lambda}, \bar{\lambda})$.

- Note: If $\theta_A = \theta_B$, then Barney's posterior equals his prior for any sequence of signals. Since $p_0 \in (\underline{\lambda}, \bar{\lambda})$, the conclusion of the lemma follows trivially. Hence, for the remainder of the proof, assume $\theta_A > \theta_B$.

2. Show that for any finite $N$, we can construct a sequence of signals such that the posterior ratio is always in $(\underline{\lambda}, \bar{\lambda})$. There are two ways we can guarantee this:

   - If $[p_0 \Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A, \theta_B), p_0 \Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A, \theta_B)] \subseteq (\underline{\lambda}, \bar{\lambda})$, then it is true by Lemma A.
   - Alternatively, if we assume $p_0 \in \left[\frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)}, \frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B)}\right] \subseteq (\underline{\lambda}, \bar{\lambda})$, then it is true by Lemma B.

3. Show that there exists a large enough number of signals $\hat{N}$ such that if the posterior ratio after $\hat{N}$ signals is in $(\underline{\lambda}, \bar{\lambda})$, then with positive probability the posterior ratio will always be inside $(\underline{\lambda}, \bar{\lambda})$. The rest of the proof demonstrates this. The conclusion of the lemma then follows from combining steps 1-3, taking the sequence in step 2 to have $\hat{N}$ signals, which (because it is a finite sequence) has positive probability.

Let $g(r) \equiv \Pi^{\psi}_{S_\infty|\Theta\times\Theta}(s_\infty|\theta_A, \theta_B, r)$ denote Barney's likelihood ratio in an infinite sample where the proportion of $a$-signals is $r$. This function will allow us to map from an observed proportion $r$ to the posterior ratio that it would imply for an infinite sample, $p_0 g(r)$. Even in dealing with finite samples, it will be useful to work with the $g(\cdot)$ function in describing the mapping from the proportion of $a$-signals to Barney's posterior ratio. In a finite sample, however, $p_0 g(r)$ will only approximate Barney's posterior ratio, but we can bound the approximation error in a large enough sample; formally, using the Law of Large Numbers, for any $r$, $\varepsilon > 0$, and $\nu > 0$, there exists an $N_{r,\epsilon,\nu}$ such that if $N \geq N_{r,\epsilon,\nu}$, then with probability at least $1-\nu$, the posterior ratio after observing a proportion $r$ $a$-signals out of $N$ is within $\epsilon$ of $p_0 g(r)$. Since all we will need for step 3 is that the posterior ratio remains within $(\underline{\lambda}, \bar{\lambda})$ with positive probability, the $\nu$ we pick does not matter. So fix $\nu$.

Since $p_0 g(r)$ is only an approximation of Barney's posterior ratio for a finite sample, we need to impose tighter bounds than $(\underline{\lambda}, \bar{\lambda})$ on where $p_0 g(r)$ can wander in a finite sample in order to ensure that Barney's posterior ratio always remains within $(\underline{\lambda}, \bar{\lambda})$. To that end, fixing $\epsilon$, we will now pick two other numbers, $\underline{\tau} < \bar{\tau}$, such that:

- $\underline{\tau} > \underline{\lambda} + \epsilon$, and $\bar{\tau} < \bar{\lambda} - \epsilon$;

- the limit posterior ratio conditional on the true state, $p_0 \Pi^{\psi}_{S_\infty|\Theta\times\Theta}(s_\infty|\theta_A, \theta_B, \Omega = i)$, is in $(\underline{\tau}, \bar{\tau})$; and

- $g^{-1}(\underline{\tau})$ and $g^{-1}(\bar{\tau})$ are both rational.

Note that $\underline{\tau}$ and $\bar{\tau}$ satisfying the above properties exist because $g$ is continuous and monotonic in $r$.

We now map these bounds on the posterior ratio onto the implied bounds on the proportion of $a$-signals. Let $\underline{r}$ denote the proportion of $a$-signals such that $p_0 g(\underline{r}) = \underline{\tau}$; and $\bar{r}$, the proportion of $a$-signals such that $p_0 g(\bar{r}) = \bar{\tau}$. By assumption both $\underline{r}$ and $\bar{r}$ are rational. Denote the proportion of realized $a$-signals after $N$ signals by $r_N$.

We claim that for any particular $\underline{r}$ and $\bar{r}$ such that the true rate $\theta_i \in (\underline{r}, \bar{r})$, there is a large enough number of signals $\widetilde{N}$ such that if there have already been $\widetilde{N}$ observed with a proportion $a$-signals $r_{\widetilde{N}} \in (\underline{r}, \bar{r})$, then with positive probability, $r_n \in (\underline{r}, \bar{r})$ for all numbers of signals $n \geq \widetilde{N}$. This is because by Chebyshev's inequality, for any $N$ and any $\delta > 0$, $P(|r_N - \theta_i| < \delta) \geq 1 - \frac{\theta_i(1-\theta_i)}{N^2\delta^2}$. Thus, the probability that the proportion of $a$-signals always stays within the bounds $\underline{r}, \bar{r}$ is

$$P\left(|r_n - \theta_i| < \delta \; \forall n \geq \widetilde{N} \;\text{ s.t. }\; \left|r_{\widetilde{N}} - \theta_i\right| < \delta\right) \geq \Pi_{n=\widetilde{N}}^{\infty}\left(1 - \frac{\theta_i(1-\theta_i)}{n^2\delta^2}\right).$$

Note that for any $x > 0$ and $y < 1$, it is true that $y(1-x) > y - x$, and so we can bound the right-hand side of the previous equation by iteratively applying this inequality:

$$\Pi_{n=\widetilde{N}}^{\infty}\left(1 - \frac{\theta_i(1-\theta_i)}{n^2\delta^2}\right) > 1 - \sum_{n=\widetilde{N}}^{\infty} \frac{\theta_i(1-\theta_i)}{n^2\delta^2}.$$

Hence there is positive probability that the proportion of $a$-signals always stays within $(\underline{r}, \bar{r})$. While not relevant for this proof, we point out for reference in proofs below that this infinite series is a known convergent sum; so as $\widetilde{N} \to \infty$, the right-hand side converges to 1.

Therefore, for any $\epsilon$, we can choose a large enough number of signals, $\hat{N} = \max\{\widetilde{N}, N_{\underline{r},\epsilon,\nu}, N_{\bar{r},\epsilon,\nu}\}$, so that if the proportion of $a$-signals after $\hat{N}$ signals is within $(\underline{r}, \bar{r})$, then with positive probability the proportion always remains within $(\underline{r}, \bar{r})$. Therefore, with positive probability $p_0 g(r)$ always remains within $(\underline{\tau}, \bar{\tau})$. Since $\hat{N} \geq N_{\underline{r},\epsilon,\nu}$ and $\hat{N} \geq N_{\bar{r},\epsilon,\nu}$, whenever $p_0 g(r)$ is within $(\underline{\tau}, \bar{\tau})$, Barney's posterior ratio is within $(\underline{\lambda}, \bar{\lambda})$.

□

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

$$\lim_{N\to\infty} \sum_{x=\lfloor a_1 N \rfloor}^{\lceil a_2 N \rceil} f^{\psi M}_{S_N|\Theta}\left(A_s = x|\theta\right) = F^{\psi}_{ß|\Theta}\left(\beta = \alpha_2|\theta\right) - F^{\psi}_{ß|\Theta}\left(\beta = \alpha_1|\theta\right) > 0.$$

**Proof:** First, we claim that conditional on a particular subjective-rate $\beta$, the agent believes that the distribution of percentage of $a$-signals converges to a point mass on $\beta$. To see this, fix $\beta$, and suppose Barney-Freddy has observed $N$ signals. We begin with the case where $N$ is an even number (i.e., the agent has observed $\frac{N}{2}$ pairs of signals, where recall that the "urn" is renewed after each pair of signals). Considering each pair of signals as a single signal that has 4 possible values and a mean value of $2\beta$, LLN implies that Barney-Freddy believes that the distribution of percentage of $a$-signals converges to a point mass on $\beta$. This proves the claim for the case where $N$ is an even number. Now, notice that for any $[\alpha_1, \alpha_2] \subseteq [0,1]$ and $\varepsilon > 0$, we can find a large enough $\hat{N}$ such that for all $N \geq \hat{N}$, if $\frac{A_{s_N}}{N} \in (\alpha_1, \alpha_2)$, then $\frac{A_{s_N}}{N+1} \in (\alpha_1 - \varepsilon, \alpha_2 + \varepsilon)$ and $\frac{A_{s_N}+1}{N+1} \in (\alpha_1 - \varepsilon, \alpha_2 + \varepsilon)$. This observation, combined with the claim for even number $N$, proves the claim for odd number $N$.

The remainder of the proof is identical to the proof of Lemma 1 below.

□

### 0.7 Main Text Results

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

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

**Proof:** We begin by proving a key fact. Consider some rational $\alpha \in [0,1]$ (which will represent the percentage of $a$-signals) and an increasing sequence of integers $m_1, m_2, ...$ (which will represent sample sizes) such that $\alpha m_j$ (which will represent numbers of $a$-signals) is an integer for all $j = 1, 2, ....$ Using the definition of Barneyness,

$$\begin{aligned}
\lim_{j\to\infty} F^{\psi}_{S_{m_j}|\Theta}(\alpha m_j|\theta) &= \lim_{j\to\infty} \int_0^1 F_{S_{m_j}|\text{ß}}(\alpha m_j|\beta)\, f^{\psi}_{\text{ß}|\Theta}(\beta|\theta)\, d\beta \\
&= \int_0^1 \lim_{j\to\infty} F_{S_{m_j}|\text{ß}}(\alpha m_j|\beta)\, f^{\psi}_{\text{ß}|\Theta}(\beta|\theta)\, d\beta \\
&= \int_0^\alpha \lim_{j\to\infty} F_{S_{m_j}|\text{ß}}(\alpha m_j|\beta)\, f^{\psi}_{\text{ß}|\Theta}(\beta|\theta)\, d\beta + \int_\alpha^1 \lim_{j\to\infty} F_{S_{m_j}|\text{ß}}(\alpha m_j|\beta)\, f^{\psi}_{\text{ß}|\Theta}(\beta|\theta)\, d\beta \\
&= \int_0^\alpha (1)\, f^{\psi}_{\text{ß}|\Theta}(\beta|\theta)\, d\beta + \int_\alpha^1 (0)\, f^{\psi}_{\text{ß}|\Theta}(\beta|\theta)\, d\beta \\
&= F^{\psi}_{\text{ß}|\Theta}(\beta = \alpha|\theta),
\end{aligned}$$

where the second equality follows from the absolute continuity of $F^{\psi}_{ß|\Theta}(\beta|\theta)$, and the fourth follows from the Law of Large Numbers.

Now, by the above fact and the definition of the cdf, for any $[a_1, a_2] \subseteq [0, 1]$ with $a_1, a_2$ rational,

$$\begin{aligned}\lim_{N\to\infty}\sum_{x=\lfloor a_1 N\rfloor}^{\lceil a_2 N\rceil} f^{\psi}_{S_N|\Theta}(A_s = x|\theta) &= \lim_{N\to\infty}(F^{\psi}_{S_N|\Theta}(\lceil a_2 N\rceil|\theta) - F^{\psi}_{S_N|\Theta}(\lfloor a_1 N\rfloor|\theta)) \\ &= F^{\psi}_{ß|\Theta}(\beta = a_2|\theta) - F^{\psi}_{ß|\Theta}(\beta = a_1|\theta).\end{aligned}$$

Consider a sequence of pairs of rational numbers, $\left(a_1^{(1)}, a_2^{(1)}\right), \left(a_1^{(2)}, a_2^{(2)}\right), ...$ that converges to the pair of real numbers $(\alpha_1, \alpha_2)$. Taking the limit of the above equality along this sequence gives

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

This is greater than 0 by the full-support assumption in A1.

□

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

**Proof:**

1. Since $A_s$ is a binomial random variable with rate $\theta$, $E_{S_N|\Theta}\left(\frac{A_s}{N}|\theta\right) = \theta$. Using the Law of Iterated Expectations, $E^{\psi}_{S_N|\Theta}\left(\frac{A_s}{N}|\theta\right) = E^{\psi}_{ß|\Theta}\left[E_{S_N|ß}\left(\frac{A_s}{N}|\beta\right)|\theta\right] = E^{\psi}_{ß|\Theta}[\beta|\theta] = \theta$, where the last equality follows from A4.
2. To keep notation compact, let $T$ denote a random variable whose distribution is binomial distribution with rate $\theta$. Let $Y$ denote the random variable whose distribution is $F^{\psi}_{S_N|\Theta}(A_s|\theta)$, i.e., the random variable induced by taking a binomial draw using rate $\beta$ and then integrating

over all possible $\beta$'s using $f^{\psi}_{\mathbb{B}|\Theta}(\beta|\theta)$. We will show that $Y$ is a mean preserving spread of the binomial distribution with rate $\theta$, which implies second-order stochastic dominance.

For $Y$ to be a mean preserving spread of $T$, it must be the case the $Y = T + Z$, where $Z$ is a random variable with conditional mean $E[Z|T] = 0$. We will construct $Z$.

Recall that for any random variable $V$, its moment generating function $M_V(t) \equiv E\left[e^{tV}\right]$ (when it exists) completely characterizes the distribution of $V$ and has the following useful properties: $E[V] = \frac{d}{dt}M_V(t)\big|_{t=0}$; for random variables $V, V'$ and $V''$, $V = V' + V''$ if and only if $M_V(t) = M_{V'}(t)M_{V''}(t)$; and for random variables $V$ and $V'$, $M_V(t) = E\left[M_{V|V'}(t)|V'\right]$.

Since the random variable $Y|\beta$ has a binomial distribution with parameter $\beta$, its moment generating function is $M_{Y|\beta}(t) = 1 - \beta + \beta e^t$. Therefore,

$$M_Y(t) = E\left[M_{Y|\beta}(t)|\beta\right] = \int_0^1 (1 - \beta + \beta e^t) f^{\psi}_{\mathbb{B}|\Theta}(\beta|\theta)\, d\beta,$$

which clearly exists since $M_{Y|\beta}(t)$ exists for each $\beta \in [0,1]$. If there is some $Z$, then we know it must have moment generating function

$$M_Z(t) = \frac{M_Y(t)}{M_T(t)} = \frac{\int_0^1 (1 - \beta + \beta e^t) f^{\psi}_{\mathbb{B}|\Theta}(\beta|\theta)\, d\beta}{1 - \theta + \theta e^t}.$$

(With this moment generating function, the random variable $T + Z$ will have the same distribution as $Y$.) Hence, we will have proved that $F^{\psi}_{S_N|\Theta}(A_s|\theta)$ SOSD $F_{S_N|\Theta}(A_s|\theta)$ once we verify that $E[Z|T] = 0$. By construction, the random variable $Z$ is independent of $T$ (since $Z$'s moment generating function, and hence its distribution, does not depend on the realization of $T$). Therefore, $E[Z|T] = E[Z]$. Finally,

$$\begin{aligned} E[Z] &= \left.\frac{d}{dt}M_Z(t)\right|_{t=0} \\ &= \left.\frac{(1-\theta+\theta e^t)\int_0^1 (\beta e^t) f^{\psi}_{\mathbb{B}|\Theta}(\beta|\theta)\, d\beta - \theta e^t \int_0^1 (1-\beta+\beta e^t) f^{\psi}_{\mathbb{B}|\Theta}(\beta|\theta)\, d\beta}{(1-\theta+\theta e^t)^2}\right|_{t=0} \\ &= \frac{\int_0^1 \beta f^{\psi}_{\mathbb{B}|\Theta}(\beta|\theta)\, d\beta - \theta \int_0^1 f^{\psi}_{\mathbb{B}|\Theta}(\beta|\theta)\, d\beta}{(1-\theta+\theta)^2} = \theta - \theta = 0. \end{aligned}$$

This proves the first claim.

Using the Law of Total Variance,

$$\begin{aligned} Var^{\psi}_{S_N|\Theta}\left(\frac{A_s}{N}|\theta\right) &= E^{\psi}_{ß|\Theta}\left[Var_{S_N|ß}\left(\frac{A_s}{N}|\beta\right)|\theta\right] + Var^{\psi}_{ß|\Theta}\left[E_{S_N|ß}\left(\frac{A_s}{N}|\beta\right)|\theta\right] \\ &= E^{\psi}_{ß|\Theta}\left[\frac{\beta(1-\beta)}{N}|\theta\right] + Var^{\psi}_{ß|\Theta}[\beta|\theta] \\ &= \frac{\theta}{N} - \frac{Var^{\psi}_{ß|\Theta}[\beta|\theta]+\theta^2}{N} + Var^{\psi}_{ß|\Theta}[\beta|\theta] \\ &= \frac{\theta(1-\theta)}{N} + \frac{N-1}{N}Var^{\psi}_{ß|\Theta}[\beta|\theta], \end{aligned}$$

where the third equality uses $E^{\psi}_{ß|\Theta}\left(\beta^2|\theta\right) = Var^{\psi}_{ß|\Theta}[\beta|\theta] + \left[E^{\psi}_{ß|\Theta}(\beta|\theta)\right]^2$. Since $Var_{S_N|\Theta}\left(\frac{A_s}{N}|\theta\right) = \frac{\theta(1-\theta)}{N}$, the result immediately follows.

3. From part 2,
$$Var^{\psi}_{S_N|\Theta}\left(\frac{A_s}{N}|\theta\right) = \frac{\theta(1-\theta)}{N} + \frac{N-1}{N}Var^{\psi}_{ß|\Theta}[\beta|\theta].$$
Note that the highest variance of any distribution that has support on $[0,1]$ and a mean of $\theta$ can have is $\theta(1-\theta)$; this is the variance of a distribution with mass $(1-\theta)$ on 0 and mass $\theta$ on 1. Any shifting of weight that preserves the mean must move positive mass closer to the mean. Since $f^{\psi}_{ß|\Theta}$ has full support (by A1) and mean $\theta$ (by A4), $Var^{\psi}_{ß|\Theta}[\beta|\theta] < \theta(1-\theta)$. The result immediately follows.

4.

$$\begin{aligned} F^{\psi}_{S_N|\Theta}(A_s|\theta) &= \sum_{i=0}^{A_s}\int_0^1 \frac{N!}{(N-i)!i!}\beta^i(1-\beta)^{N-i} f^{\psi}_{ß|\Theta}(\beta|\theta)\,d\beta \\ &= \int_0^1\left[\sum_{i=0}^{A_s}\frac{N!}{(N-i)!i!}\beta^i(1-\beta)^{N-i}\right] f^{\psi}_{ß|\Theta}(\beta|\theta)\,d\beta \\ &= \int_0^1 F_{S_N|ß}(A_s|\beta)\, f^{\psi}_{ß|\Theta}(\beta|\theta)\,d\beta, \end{aligned}$$

where $F_{S_N|ß}(A_s|\beta)$ is the cdf of the binomial distribution conditional on a rate $\beta$. By A3, $F^{\psi}_{ß|\Theta}(\beta|\theta')$ FOSD $F^{\psi}_{ß|\Theta}(\beta|\theta)$ whenever $\theta' > \theta$, and therefore $\int_0^1 g(\beta)\, f^{\psi}_{ß|\Theta}(\beta|\theta')\,d\beta > \int_0^1 g(\beta)\, f^{\psi}_{ß|\Theta}(\beta|\theta)\,d\beta$ for any function $g(\beta)$ that is decreasing in $\beta$. Since $F_{S_N|ß}(A_S|\beta)$ is decreasing in $\beta$, the result follows.

□

**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:*

$$\frac{f^{\psi}_{\Theta|S_N}(\theta_A|s)}{f^{\psi}_{\Theta|S_N}(\theta_B|s)} \rightarrow_{a.s.} \frac{f^{\psi}_{ß|\Theta}(\beta=\theta|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=\theta|\theta_B)}\frac{f(\theta_A)}{f(\theta_B)}. \tag{5}$$

**Proof:** Using the Law of Large Numbers, and the true probability distribution over realizations of infinite samples (i.e., events):

$$\lim_{N\to\infty}\frac{f^{\psi}_{\Theta|S_N}(\theta_A|s)}{f^{\psi}_{\Theta|S_N}(\theta_B|s)} = \left(\lim_{N\to\infty}\frac{f^{\psi}_{S_N|\Theta}(s|\theta_A)}{f^{\psi}_{S_N|\Theta}(s|\theta_B)}\right)\frac{f(\theta_A)}{f(\theta_B)} = \frac{f^{\psi}_{ß|\Theta}(\beta=\theta|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=\theta|\theta_B)}\frac{f(\theta_A)}{f(\theta_B)} \text{ a.s.}$$

□

**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 $\theta_A$, then his limit posterior probability that the rate is $\theta_A$ is 1. In contrast, before having observed any data, Barney believes: if the rate is $\theta_A$, then his limit posterior probability that the rate is $\theta_A$ 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 $[0,1]$.*

**Proof:** The proof for Tommy is entirely standard and follows directly from the Law of Large Numbers. For Barney, recall that Lemma 1 implies that Barney's subjective sampling distribution for an infinite sample is $\lim_{N\to\infty} f^{\psi}_{S_N|\Theta}\left(\frac{A_s}{N}=\alpha|\theta\right) = f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta)$ a.s. Therefore, after observing proportion $\alpha$ of $a$-signals in a large sample, Barney anticipates having posterior ratio arbitrarily close to $\lim_{N\to\infty}\frac{f^{\psi}_{\Theta|S_N}\left(\theta_A|\frac{A_s}{N}=\alpha\right)}{f^{\psi}_{\Theta|S_N}\left(\theta_B|\frac{A_s}{N}=\alpha\right)} = \frac{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_A)f(\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_B)f(\theta_B)}$. Barney believes that this large-sample posterior ratio is a random variable because, according to A1, Barney's beliefs put full support on $\alpha \in (0,1)$.

By A3 and Whitt (1979), $\frac{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_A)f(\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_B)f(\theta_B)}$ is strictly increasing in $\alpha$. It follows that

$$\inf_{\alpha\in(0,1)}\frac{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_A)f(\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_B)f(\theta_B)} = \lim_{\alpha\to 0}\frac{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_A)f(\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_B)f(\theta_B)} < 1$$

and

$$\sup_{\alpha\in(0,1)}\frac{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_A)f(\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_B)f(\theta_B)} = \lim_{\alpha\to 1}\frac{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_A)f(\theta_A)}{f^{\psi}_{ß|\Theta}(\beta=\alpha|\theta_B)f(\theta_B)} > 1.$$

Furthermore, by the Intermediate Value Theorem, the posterior ratio must take on each value in

$$\left(\inf_{\alpha\in(0,1)} \frac{f^{\psi}_{\text{ß}|\Theta}\left(\beta=\alpha|\theta_A\right) f\left(\theta_A\right)}{f^{\psi}_{\text{ß}|\Theta}\left(\beta=\alpha|\theta_B\right) f\left(\theta_B\right)}, \sup_{\alpha\in(0,1)} \frac{f^{\psi}_{\text{ß}|\Theta}\left(\beta=\alpha|\theta_A\right) f\left(\theta_A\right)}{f^{\psi}_{\text{ß}|\Theta}\left(\beta=\alpha|\theta_B\right) f\left(\theta_B\right)}\right)$$

for some $\alpha \in (0,1)$. It follows that, as $\alpha$ ranges from 0 from 1, the limit posterior probability that the rate is $\theta_A$ takes values in a nondegenerate interval in $[0,1]$. By A1, Barney assigns positive density to each possible realization of $\alpha \in (0,1)$, and so all possible posteriors in that interval have a positive density of being realized. Under A1′,

$$\lim_{N\to\infty} \frac{f^{\psi}_{\Theta|S_N}\left(\theta_A|\frac{A_s}{N}=0\right)}{f^{\psi}_{\Theta|S_N}\left(\theta_B|\frac{A_s}{N}=0\right)} = \frac{f^{\psi}_{\text{ß}|\Theta}\left(\beta=0|\theta_A\right) f\left(\theta_A\right)}{f^{\psi}_{\text{ß}|\Theta}\left(\beta=0|\theta_B\right) f\left(\theta_B\right)}$$

and

$$\lim_{N\to\infty} \frac{f^{\psi}_{\Theta|S_N}\left(\theta_A|\frac{A_s}{N}=1\right)}{f^{\psi}_{\Theta|S_N}\left(\theta_B|\frac{A_s}{N}=1\right)} = \frac{f^{\psi}_{\text{ß}|\Theta}\left(\beta=1|\theta_A\right) f\left(\theta_A\right)}{f^{\psi}_{\text{ß}|\Theta}\left(\beta=1|\theta_B\right) f\left(\theta_B\right)}$$

are simply real numbers strictly greater than 0 and less than 1. The result follows.

□

**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 $\frac{A_s}{N} = \frac{1}{2}$.*

**Proof:** The first claim can be seen from the fact that for a single sample, the subjective sampling distribution of Tommy matches that of Barney.

To see that the second claim is true, suppose that Barney observes exactly half $a$ and $b$ signals, $k$ of each signal (so that $N - k = k$). Barney's likelihood ratio is

$$\begin{aligned}
\frac{\int_0^1 \frac{N!}{k!k!}\beta^k(1-\beta)^k f^{\psi}_{\text{ß}|\Theta}(\beta|\theta_A)d\beta}{\int_0^1 \frac{N!}{k!k!}\beta^k(1-\beta)^k f^{\psi}_{\text{ß}|\Theta}(\beta|\theta_B)d\beta} &= \frac{\int_0^1 \frac{N!}{k!k!}\beta^k(1-\beta)^k f^{\psi}_{\text{ß}|\Theta}(\beta|\theta_A)d\beta}{\int_0^1 \frac{N!}{k!k!}\beta^k(1-\beta)^k f^{\psi}_{\text{ß}|\Theta}(\beta|1-\theta_A)d\beta} \\
&= \frac{\int_0^1 \frac{N!}{k!k!}\beta^k(1-\beta)^k f^{\psi}_{\text{ß}|\Theta}(\beta|\theta_A)d\beta}{\int_0^1 \frac{N!}{k!k!}\beta^k(1-\beta)^k f^{\psi}_{\text{ß}|\Theta}(1-\beta|\theta_A)d\beta} \\
&= \frac{\int_0^1 \frac{N!}{k!k!}\beta^k(1-\beta)^k f^{\psi}_{\text{ß}|\Theta}(\beta|\theta_A)d\beta}{\int_0^1 \frac{N!}{k!k!}(1-\beta)^k\beta^k f^{\psi}_{\text{ß}|\Theta}(\beta|\theta_A)d\beta} \\
&= 1,
\end{aligned}$$

where the first equality comes from substituting $\theta_B = 1 - \theta_A$, and the second equality follows from A2. This directly implies that Barney's likehood ratio is equal to 1 if and only if Tommy's is, since the likelihood ratio is strictly increasing in the number of $a$-signals for both Barney and Tommy if $\theta_A > \frac{1}{2}$, and equal to 1 for any sample for both Barney and Tommy if $\theta_A = \frac{1}{2}$.

□

**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 $\theta_A$ or $\theta_B$, for $\psi$ 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 $\psi$, 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.*

**Proof:** We will prove the first claim in two steps. First, we shall show that a sufficient condition is that Tommy's subjective sampling distribution is Blackwell-sufficient for Barney's. Second, we will show that this sufficient condition in fact holds for $\psi$ sufficiently small.

First, suppose $F_{S_N|\Theta}(A_s|\theta_i)$ is Blackwell-sufficient for $F^\psi_{S_N|\Theta}(A_s|\theta_i)$ for $i \in \{A, B\}$. Using Blackwell (1951, 1953), note that $F_{S_N|\Theta}(A_s|\theta_i)$ is Blackwell-sufficient for $F^\psi_{S_N|\Theta}(A_s|\theta_i)$ if and only if for any continuous, convex function $g$,

$$\begin{aligned} & \sum_{k=0}^{N} f_{S_N|\Theta}(A_s = k|\theta_i)\, g\left(\frac{f_{S_N|\Theta}(A_s = k|\theta_i)}{f_{S_N|\Theta}(A_s = k|\theta_{-i})}\right) \\ \geq & \sum_{k=0}^{N} f^\psi_{S_N|\Theta}(A_s = k|\theta_i)\, g\left(\frac{f^\psi_{S_N|\Theta}(A_s = k|\theta_i)}{f^\psi_{S_N|\Theta}(A_s = k|\theta_{-i})}\right) \end{aligned}$$

for $i \in \{A, B\}$ and $-i$ being the other state. Taking $g(x) = |p_0 - p_0 x|$, this implies

$$\begin{aligned} & \sum_{k=0}^{N} f_{S_N|\Theta}(A_s = k|\theta_i) \left| p_0 - p_0 \frac{f_{S_N|\Theta}(A_s = k|\theta_i)}{f_{S_N|\Theta}(A_s = k|\theta_{-i})} \right| \\ \geq & \sum_{k=0}^{N} f^\psi_{S_N|\Theta}(A_s = k|\theta_i) \left| p_0 - p_0 \frac{f^\psi_{S_N|\Theta}(A_s = k|\theta_i)}{f^\psi_{S_N|\Theta}(A_s = k|\theta_{-i})} \right| \end{aligned}$$

for $i \in \{A, B\}$.

From Proposition 1, we know that $f^{\psi}_{S_N|\Theta}(A_s|\theta_i)$ is a mean-preserving spread of $f_{S_N|\Theta}(A_s|\theta_i)$. Since $g(x) = |p_0 - p_0 x|$ is a continuous convex function, it follows that

$$\begin{aligned} & \sum_{k=0}^{N} f^{\psi}_{S_N|\Theta}(A_s = k|\theta_i) \left| p_0 - p_0 \frac{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_i)}{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_{-i})} \right| \\ \geq \; & \sum_{k=0}^{N} f_{S_N|\Theta}(A_s = k|\theta_i) \left| p_0 - p_0 \frac{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_i)}{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_{-i})} \right| \end{aligned}$$

for $i \in \{A, B\}$. Chaining the two inequalities,

$$\begin{aligned} & \sum_{k=0}^{N} f_{S_N|\Theta}(A_s = k|\theta_i) \left| p_0 - p_0 \frac{f_{S_N|\Theta}(A_s = k|\theta_i)}{f_{S_N|\Theta}(A_s = k|\theta_{-i})} \right| \\ \geq \; & \sum_{k=0}^{N} f_{S_N|\Theta}(A_s = k|\theta_i) \left| p_0 - p_0 \frac{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_i)}{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_{-i})} \right|. \end{aligned}$$

Aggregating over states weighted by the priors,

$$\begin{aligned} & f(\theta_A) \left( \sum_{k=0}^{N} f_{S_N|\Theta}(A_s = k|\theta_A) \left| p_0 - p_0 \frac{f_{S_N|\Theta}(A_s = k|\theta_A)}{f_{S_N|\Theta}(A_s = k|\theta_B)} \right| \right) \\ & + f(\theta_B) \left( \sum_{k=0}^{N} f_{S_N|\Theta}(A_s = k|\theta_B) \left| p_0 - p_0 \frac{f_{S_N|\Theta}(A_s = k|\theta_B)}{f_{S_N|\Theta}(A_s = k|\theta_A)} \right| \right) \\ \geq \; & f(\theta_A) \left( \sum_{k=0}^{N} f_{S_N|\Theta}(A_s = k|\theta_A) \left| p_0 - p_0 \frac{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_A)}{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_B)} \right| \right) \\ & + f(\theta_B) \left( \sum_{k=0}^{N} f_{S_N|\Theta}(A_s = k|\theta_B) \left| p_0 - p_0 \frac{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_B)}{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_A)} \right| \right). \end{aligned}$$

This is what we were seeking to prove.

Now we will show that for $i \in \{A, B\}$ and $N > 1$, $F_{S_N|\Theta}(A_s|\theta_i)$ is Blackwell-sufficient for $F^{\psi}_{S_N|\Theta}(A_s|\theta_i)$ if $\psi$ is sufficiently small. Since Barney's subjective sampling distribution is the same as Tommy's for $N = 1$, $F_{S_1|\Theta}(A_s|\theta_i)$ is Blackwell-equivalent to $F^{\psi}_{S_1|\Theta}(A_s|\theta_i)$ (regardless of the value of $\psi$). Note that as $\psi \to 0$, Barney's subjective sampling distribution converges to point masses at 0 and 1. Hence in the limit $\psi \to 0$, for any $N$, Barney expects to observe all $a$'s or all $b$'s; therefore, in the limit $\psi \to 0$, $F^{\psi}_{S_N|\Theta}(A_s|\theta_i)$ is Blackwell-equivalent to $F^{\psi}_{S_1|\Theta}(A_s|\theta_i)$. In contrast, for Tommy, $F_{S_N|\Theta}(A_s|\theta_i)$ for $N > 1$ is strictly Blackwell-sufficient for $F_{S_1|\Theta}(A_s|\theta_i)$ (e.g., Shaked and Tong, 1990). Therefore, for any $N > 1$, in the limit $\psi \to 0$, $F_{S_N|\Theta}(A_s|\theta_i)$ is Blackwell-sufficient for $F^{\psi}_{S_N|\Theta}(A_s|\theta_i)$. Using the necessary and sufficient inequality for Blackwell-sufficiency: for any $N > 1$ and any continuous, convex function $g$, in the limit $\psi \to 0$,

$$\sum_{k=0}^{N} f_{S_N|\Theta}(A_s = k|\theta_i)\, g\left(\frac{f_{S_N|\Theta}(A_s = k|\theta_i)}{f_{S_N|\Theta}(A_s = k|\theta_{-i})}\right)$$
$$> \sum_{k=0}^{N} f^{\psi}_{S_N|\Theta}(A_s = k|\theta_i)\, g\left(\frac{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_i)}{f^{\psi}_{S_N|\Theta}(A_s = k|\theta_{-i})}\right).$$

Note that the right-hand side of this inequality is continuous in $\psi$. Therefore, the strict inequality holds for any sufficiently small $\psi > 0$, and so $F_{S_N|\Theta}(A_s|\theta_i)$ is Blackwell-sufficient for $F^{\psi}_{S_N|\Theta}(A_s|\theta_i)$ for any sufficiently small $\psi > 0$.

To see that the second claim in the proposition is true, consider some sample of size $N$, with $A_s$ realizations of heads, and assume that $\theta_A > \frac{1}{2}$ and $A_s > N - A_s$. (The other cases have analogous proofs.) Tommy's likelihood ratio is

$$\frac{\frac{N!}{\theta_A!(N-\theta_A)!}\theta_A^{A_s}(1-\theta_A)^{N-A_s}}{\frac{N!}{\theta_A!(N-\theta_A)!}\theta_A^{N-A_s}(1-\theta_A)^{A_s}} = \left(\frac{\theta_A}{1-\theta_A}\right)^{2A_s-N},$$

which has $2A_s - N$ terms multiplied together. Barney's likelihood ratio is

$$\frac{\Gamma(\theta_A\psi + A_s)}{\Gamma((1-\theta_A)\psi + A_s)}\frac{\Gamma((1-\theta_A)\psi + N - A_s)}{\Gamma(\theta_A\psi + N - A_s)} = \frac{(\theta_A\psi + A_s - 1)...(\theta_A\psi + N - A_s)}{((1-\theta_A)\psi + A_s - 1)...((1-\theta_A)\psi + N - A_s)}.$$

This also has $2A_s - N$ terms. Since by assumption $\frac{\theta_A\psi}{(1-\theta_A)\psi} = \frac{\theta_A}{1-\theta_A} > 1$, then for any $k > 0$,

$$\frac{\theta_A}{1-\theta_A} = \frac{\theta_A\psi}{(1-\theta_A)\psi} > \frac{\theta_A\psi + k}{(1-\theta_A)\psi + k} > 1,$$

which implies that

$$\left(\frac{\theta_A}{1-\theta_A}\right)^{2A_s-N} > \frac{(\theta_A\psi + A_s - 1)...(\theta_A\psi + N - A_s)}{((1-\theta_A)\psi + A_s - 1)...((1-\theta_A)\psi + N - A_s)} > 1.$$

Hence Barney under-infers relative to Tommy.

Finally, consider two samples, one of size $N$ with $A_s$ $a$-signals, and the other of size $N' > N$ with $A'_s$ $a$-signals. Suppose both samples have the same difference, $d$, between the number of $a$ and $b$ signals: $d = A_s - (N - A_s) = 2A_s - N$ and $d = 2A'_s - N'$. Since Tommy's likelihood ratio is $\left(\frac{\theta_A}{1-\theta_A}\right)^d$, it is the same for both samples. To consider Barney's inference, assume that $\theta_A > \frac{1}{2}$ and $d > 0$. (The other cases have analogous proofs.) Barney's likelihood ratio, stated above, has $d$

terms for both samples. Since by assumption $\frac{\theta_A \psi}{(1-\theta_A)\psi} = \frac{\theta_A}{1-\theta_A} > 1$, then for any $k' > k > 0$,

$$\frac{\theta_A \psi + k}{(1-\theta_A)\psi + k} > \frac{\theta_A \psi + k'}{(1-\theta_A)\psi + k'} > 1.$$

Note that $A'_s > A_s$. It follows that each term in Barney's likelihood ratio for the sample of size $N$ is larger than the corresponding term in Barney's likelihood ratio for the sample of size $N'$ (e.g., for the first term, $\frac{\theta_A \psi + A_s - 1}{(1-\theta_A)\psi + A_s - 1} > \frac{\theta_A \psi + A'_s - 1}{(1-\theta_A)\psi + A'_s - 1}$). Hence while Barney infers in favor of state $A$ in both cases (like Tommy does), Barney's change in beliefs is smaller from the larger sample.

□

**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.*

**Proof:** First, the fact that in the absence of *Consumer Reports*, Tommy will ask a friend immediately implies the same about Barney, since they infer the same from a single signal.

We will prove the proposition in two steps. First we will show that both Tommy and Barney follow a threshold rule, purchasing *Consumer Reports* if and only if $c_r$ is less than some threshold. Then we will show that Barney's threshold $c'_r$ is less than Tommy's threshold $c''_r$.

Let $q$ denote an arbitrary posterior ratio. Denote the set of possible $q$'s that could occur after asking a friend as $Q_1$ (the same for Tommy and Barney), Tommy's set of possible $q$'s after reading *Consumer Reports* as $Q_N$, and Barney's set of possible $q$'s after reading *Consumer Reports* as $Q_N^\psi$. Denote the expected value of taking the decision after asking a friend—which is the same for Tommy and Barney—as $V(1) = E_s[\max_\mu E_q[u(\mu,\omega)|q \in Q_1]|s \in S_1]$. (Note that this does not include the cost of asking a friend.) Denote the *subjective* expected value of taking the decision, after reading *Consumer Reports*, as $V(N) = E_s[\max_\mu E_q[u(\mu,\omega)|q \in Q_N]|s \in S_N]$ for Tommy and as $V^\psi(N) = E_s^\psi[\max_\mu E_q[u(\mu,\omega)|q \in Q_N^\psi]|s \in S_N]$ for Barney. Denote the value of knowing the state for sure, i.e., having full information, as $V(fi)$. Since the agents would like to condition their actions, $V(1) < V(fi)$.

For all $N' > N \geq 1$, a sample size of $N'$ is strictly Blackwell-more-informative than a sample size of $N$. Therefore, for each of Tommy and Barney, the subjectively-anticipated distribution of posteriors after $N'$ signals is a strict mean-preserving spread of the subjectively-anticipated distribution of posteriors after $N$ signals. Because each of the two actions is better in one of the two states, $V(N') > V(N)$ and $V^{\psi}(N') > V^{\psi}(N)$; that is, $V(N)$ and $V^{\psi}(N)$ are both strictly increasing in $N$.

Since for any $N > 1$, $V(N) > V(1)$ and $V^{\psi}(N) > V(1)$, both Tommy and Barney follow a threshold rule.

Denote the agents' value of an infinite number of signals as $V(\infty) = \lim_{N\to\infty} V(N)$ and $V^{\psi}(\infty) = \lim_{N\to\infty} V^{\psi}(N)$. Since Tommy expects to learn the state almost surely after an infinite number of signals, $V(\infty) = V(fi)$. Since Barney believes his posterior will place positive weight on both states even in an infinite sample, and since Barney would like to condition his action on the state, $V^{\psi}(\infty) < V(fi)$. Therefore, there exists $N'$ sufficiently large that for all $N > N'$, $V(N) > V^{\psi}(N)$. In that case, Barney's threshold $c_r'$ is less than Tommy's threshold $c_r''$.

□

**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 $\mu_A$ and $\mu_B$, 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.*

**Proof:** As in the previous proof, denote the *subjective* expected value of taking the decision, after observing $N$ signals as $V(N)$ for Tommy and as $V^{\psi}(N)$ for Barney. Denote the subjective expected value of taking the decision without observing any signals as $V(\emptyset)$. Denote the value of knowing the state for sure, i.e., having full information, as $V(fi)$.

By construction $V(fi) > V(\emptyset)$. Denote $V(fi) - V(\emptyset) = c_r^*$. Tommy believes that as the number of signals goes to $\infty$, his posterior converges almost surely to placing a weight of 1 on the true state. Hence $V(N)$ converges to $V(fi)$. Therefore, the proposition holds for Tommy.

Now, note that for both Barney and Tommy, there exists a threshold posterior ratio,

$$\tau = \frac{\frac{u(\mu_A,B)-u(\mu_B,B)}{u(\mu_B,A)+u(\mu_A,B)-u(\mu_B,B)-u(\mu_A,A)}}{1-\frac{u(\mu_A,B)-u(\mu_B,B)}{u(\mu_B,A)+u(\mu_A,B)-u(\mu_B,B)-u(\mu_A,A)}},$$

such that if the posterior ratio is less than $\tau$, the agent takes action $\mu_B$; if the posterior ratio is greater than $\tau$ the agent takes action $\mu_A$; and if the posterior ratio is equal to $\tau$, the agent is indifferent between either action. Let $p_0 = \frac{f_\Theta(\theta_A)}{1-f_\Theta(\theta_A)}$ be the prior ratio. Also note that (due to A1$'$) Barney believes that there exist $0 < \underline{L} < \overline{L} < \infty$ such that as $N$ goes to $\infty$, his likelihood ratio will converge to a random variable $L_\infty$ that has support over a bounded interval that is a subset of $[\underline{L}, \overline{L}]$. This means that his subjective posterior will converge to $p_0 L_\infty$ that has support on $[p_0\underline{L}, p_0\overline{L}]$.

To prove $(i)$, notice that we can always find a value $f^*$ large enough such that for all $f_\Theta(\theta_A) > f^*$, $p_0\underline{L} > \tau$. This means that regardless of how many signals Barney observes, the value of the information contained in those signals is 0 (because Barney will take action $\mu_A$ no matter what). In this case, Barney will not purchase *Consumer Reports* for any positive price, regardless of the number of signals.

Next we prove $(ii)$. Without loss of generality assume $p_0 > \tau$. Because the subjective-rate distribution is point-wise continuous in $\theta$, for any $0 < \epsilon < 1$, we can find $\theta_A(\epsilon)$ and $\theta_B(\epsilon)$ close enough together such that $\epsilon < \frac{f^\psi_{\mathbb{B}|\Theta}(\beta=0|\theta_A)}{f^\psi_{\mathbb{B}|\Theta}(\beta=0|\theta_B)} < 1$. Recall that $\frac{f^\psi_{\mathbb{B}|\Theta}(\beta=0|\theta_A)}{f^\psi_{\mathbb{B}|\Theta}(\beta=0|\theta_B)} = \underline{L}$, the minimum possible limit likelihood ratio for Barney. Therefore, $\epsilon < \underline{L} < 1$. Furthermore, there exists an $\epsilon^*$ close enough to 1 such that for all $\epsilon^* < \epsilon < 1$, $p_0\epsilon > \tau$. Therefore, for all $\epsilon^* < \epsilon < 1$, for rates $\theta_A(\epsilon)$ and $\theta_B(\epsilon)$, $p_0\underline{L} > \tau$. As above, this implies that Barney will not purchase *Consumer Reports* for any positive price, regardless of the number of signals.

□

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

**Proof.** By Proposition 1, Barney's beliefs about the distribution of $A_S$ is a mean-preserving spread of Tommy's. We next define the continuous function $\hat{u} : \mathbb{R}_+ \to \mathbb{R}$:

$$\hat{u}(x) = \begin{cases} u(w(x)) & \text{if } x \in \{0,1,2,...\} \\ u(w(\lfloor x \rfloor)) + (x - \lfloor x \rfloor)\left[u(w(\lceil x \rceil)) - u(w(\lfloor x \rfloor))\right] & \text{otherwise.} \end{cases}$$

(This function is simply $u(w(x))$ when $x$ is an integer and the linear interpolation of $u(w(x))$ when $x$ is not an integer.) If $u(w(A_S))$ is a concave (resp., convex) function of $A_S$—that is, if $u(w(A_S - 1)) + u(w(A_S + 1)) - 2u(w(A_S)) \geq 0$ (resp., $\leq$) for all $A_S$—then $\hat{u}(x)$ is a concave (resp., convex) function of $x$. The result immediately follows from standard results from the theory of choice under risk. □

**Proposition 9.** *Assume A1-A4. Suppose Barney and Tommy have simple, piecewise-linear loss-averse preferences as specified in (4). Fix any gamble* $(\theta, h, t)$*, paying off* $h > 0$ *with probability* $\theta$ *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* $N > N''$*.*

**Proof:** WLOG let the reference point be $w_0$, and fix $\lambda \geq 1$. Denoting $G(z)$ as the distribution of monetary outcomes $z$, the expected utility of any lottery is:

$$\begin{aligned} & \int_0^\infty (w_0 + z)\, dG(z) + \int_{-\infty}^0 (w_0 + \lambda z)\, dG(z) \\ = \quad & w_0 + \int_0^\infty z dG(z) + \lambda \int_{-\infty}^0 z dG(z)\,. \end{aligned}$$

Clearly, this is better than the option of refusing the lottery if and only if

$$\int_0^\infty z dG(z) > -\lambda \int_{-\infty}^0 z dG(z) \Leftrightarrow \frac{\int_0^\infty z dG(z)}{-\int_{-\infty}^0 z dG(z)} < \lambda.$$

The monetary outcome of an $N$-times repeated gamble is $z = A_s h - (N - A_s)t = -Nt + A_s(h + t) = N(-t + \frac{A_s}{N}(h+t))$. Hence the agent earns a positive payoff from the gamble if and only if $A_s \geq \left\lceil \frac{t}{h+t} N \right\rceil$. Substituting in, we find that for Tommy,

$$\frac{\int_0^\infty z dG(z)}{-\int_{-\infty}^0 z dG(z)} = \frac{\sum\limits_{A_s = \lceil \frac{t}{h+t} N \rceil}^{N} (-t + \frac{A_s}{N}(h+t)) f_{S_N|\Theta}(A_s|\theta)}{-\sum\limits_{A_s=0}^{\lceil \frac{t}{h+t} N \rceil - 1} (-t + \frac{A_s}{N}(h+t)) f_{S_N|\Theta}(A_s|\theta)},$$

where $\theta$ is the known rate of the good outcome (and the $N$'s in the numerator and denominator cancel out).

Since the bet is better-than-fair, the probability of losing money goes to 0 as $N \to \infty$. Therefore, as $N \to \infty$, the denominator goes to 0. As $N \to \infty$, $\frac{A_s}{N}$ converges almost surely to $\theta$, so the numerator goes to $-t + \theta(h+t) \geq 0$. Hence the ratio of the numerator to the denominator goes to infinity. This implies that Tommy always accepts the $N$-times repeated gamble for $N$ sufficiently large.

Alternatively, for Barney,

$$\frac{\int_0^\infty zdG(z)}{-\int_{-\infty}^0 zdG(z)} = \frac{\sum_{A_s=\lceil \frac{t}{h+t}N\rceil}^{N} \left(-t+\frac{A_s}{N}(h+t)\right) f^\psi_{S_N|\Theta}(A_s|\theta)}{-\sum_{A_s=0}^{\lceil \frac{t}{h+t}N\rceil -1} \left(-t+\frac{A_s}{N}(h+t)\right) f^\psi_{S_N|\Theta}(A_s|\theta)}$$

$$\to \frac{\int_{\frac{t}{h+t}}^1 \left(-t+\beta(h+t)\right) f^\psi_{\text{ß}|\Theta}(\beta|\theta)\, d\beta}{-\int_0^{\frac{t}{h+t}} \left(-t+\beta(h+t)\right) f^\psi_{\text{ß}|\Theta}(\beta|\theta)\, d\beta}$$

almost surely as $N \to \infty$. This is a finite, positive number.

□

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

*Suppose $u(w(A_S))$ 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 $\sigma$ is higher than Barney's.*

2. *If given the prior both Tommy and Barney invest, then Barney's willingness-to-pay for a signal $\sigma$ 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 $\sigma$ is higher than Tommy's if and only if*

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

*where $\Sigma^\psi_{\mathbb{I}} \subseteq \{L, H\}$ 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 $u(w(A_S))$ is a convex function of $A_S$, then all of the conclusions in the previous paragraph hold with "Barney" and "Tommy" switched.*

**Proof.** We prove each part in turn assuming $u$ is concave: that is, $u(w(A_S-1))+u(w(A_S+1))-2u(w(A_S)) \geq 0$ for all $A_S$. (The proof for convex $u$ is analogous.) Proposition 8 implies that conditional on either the good state ($\theta_A$) or the bad state ($\theta_B$), if Barney invests, then so does Tommy. To see this, observe that Tommy invests if and only if

$$E_{S_N|\Theta}\left[u(w(A_S))|\theta=\theta_A\right]f_\Theta\left(\theta_A\right)+E_{S_N|\Theta}\left[u(w(A_S))|\theta=\theta_B\right]\left(1-f_\Theta\left(\theta_A\right)\right)\geq u(w_0).$$

Similarly, Barney invests if and only if

$$E^{\psi}_{S_N|\Theta}\left[u(w(A_S))|\theta=\theta_A\right]f_\Theta\left(\theta_A\right)+E^{\psi}_{S_N|\Theta}\left[u(w(A_S))|\theta=\theta_B\right]\left(1-f_\Theta\left(\theta_A\right)\right)\geq u(w_0).$$

From Proposition 8, we know that both left-hand-side terms are smaller for Barney than the analogous terms are for Tommy. Thus, if given the prior Barney invests, then so does Tommy.

Before proving Parts 1-3 of the proposition, we characterize behavior given the posterior (after observing the signal $\sigma$). Observe that, given the realization of the signal, Tommy invests if and only if

$$E_{\Theta|\Sigma}\left[E_{S_N|\Theta}\left[u(w(A_S))|\theta\right]|\sigma\right]\geq u(w_0).$$

Define $\Sigma_{\mathbb{I}}$, the set of signal realizations such that Tommy invests, as the realizations of $\sigma$ such that this inequality holds. The set of signal realizations such that Tommy invests in the safe asset is the complement, $\Sigma_{\mathbb{S}}\equiv\Sigma\backslash\Sigma_{\mathbb{I}}$. Thus, Tommy's expected utility if he observes the signal $\sigma$ is

$$\sum_{\sigma\in\Sigma_{\mathbb{I}}}f_\Sigma\left(\sigma\right)E_{\Theta|\Sigma}\left[E_{S_N|\Theta}\left[u(w(A_S))|\theta\right]|\sigma\right]+\sum_{\sigma\in\Sigma_{\mathbb{S}}}f_\Sigma\left(\sigma\right)u(w_0),$$

which can be expressed as

$$u(w_0)+\sum_{\sigma\in\Sigma_{\mathbb{I}}}f_\Sigma\left(\sigma\right)\left\{E_{\Theta|\Sigma}\left[E_{S_N|\Theta}\left[u(w(A_S))|\theta\right]|\sigma\right]-u(w_0)\right\}. \tag{6}$$

Analogously, given the realization of the signal, Barney invests if and only if

$$E^{\psi}_{\Theta|\Sigma}\left[E^{\psi}_{S_N|\Theta}\left[u(w(A_S))|\theta\right]|\sigma\right]\geq u(w_0).$$

$\Sigma^{\psi}_{\mathbb{I}}$ is defined as the values of $\sigma$ such that this inequality holds, and $\Sigma^{\psi}_{\mathbb{S}}\equiv\Sigma\backslash\Sigma^{\psi}_{\mathbb{I}}$. Thus, Barney's expected utility if he observes the signal $\sigma$ is

$$\sum_{\sigma\in\Sigma^{\psi}_{\mathbb{I}}}f_\Sigma\left(\sigma\right)E^{\psi}_{\Theta|\Sigma}\left[E^{\psi}_{S_N|\Theta}\left[u(w(A_S))|\theta\right]|\sigma\right]+\sum_{\sigma\in\Sigma^{\psi}_{\mathbb{S}}}f_\Sigma\left(\sigma\right)u(w_0),$$

where note that we write Barney's probability distribution over signal realizations, $f_\Sigma(\sigma)$, as the same as Tommy's because the signal $\sigma$ has a sample size of 1 and because Barney and Tommy have the same prior. This expression for Barney's expected utility after observing the signal can be expressed as

$$u(w_0) + \sum_{\sigma \in \Sigma_{\mathbb{I}}^{\psi}} f_\Sigma(\sigma) \left\{ E^{\psi}_{\Theta|\Sigma} \left[ E^{\psi}_{S_N|\Theta} \left[ u(w(A_S)) | \theta \right] | \sigma \right] - u(w_0) \right\}. \tag{7}$$

We now prove Part 1 of the proposition. Suppose Tommy does not invest given the prior. Using the observation above, we know that Barney also does not invest given the prior. Thus, given their prior, Barney and Tommy expect to get the same expected utility from not observing the signal: $u(w_0)$. Tommy's gain from observing the signal is the gain from changing his action times the probability that he receives a signal that would cause him to change his action:

$$\sum_{\sigma \in \Sigma_{\mathbb{I}}} f_\Sigma(\sigma) \left\{ E_{\Theta|\Sigma} \left[ E_{S_N|\Theta} \left[ u(w(A_S)) | \theta \right] | \sigma \right] - u(w_0) \right\}. \tag{8}$$

Analogously, Barney's gain from observing the signal is

$$\sum_{\sigma \in \Sigma_{\mathbb{I}}^{\psi}} f_\Sigma(\sigma) \left\{ E^{\psi}_{\Theta|\Sigma} \left[ E^{\psi}_{S_N|\Theta} \left[ u(w(A_S)) | \theta \right] | \sigma \right] - u(w_0) \right\}. \tag{9}$$

By definition of $\Sigma_{\mathbb{I}}$ and $\Sigma_{\mathbb{I}}^{\psi}$, each term in both sums is positive. Since Barney's value of investing in either state is lower than Tommy's: for any $\sigma \in \Sigma$,

$$E_{\Theta|\Sigma} \left[ E_{S_N|\Theta} \left[ u(w(A_S)) | \theta \right] | \sigma \right] > E^{\psi}_{\Theta|\Sigma} \left[ E^{\psi}_{S_N|\Theta} \left[ u(w(A_S)) | \theta \right] | \sigma \right].$$

This fact has two important implications. First, for any $\sigma \in \Sigma_{\mathbb{I}} \cap \Sigma_{\mathbb{I}}^{\psi}$ (i.e., any signal realization such that both Tommy and Barney invest), each term for $\sigma$ in the sum in (8) is larger than the corresponding term in the sum in (9). Second, since it takes a higher posterior (on the good state) for Barney to switch to investing, relative to Tommy, the set of posteriors that cause Barney to switch is a subset of those that cause Tommy to switch: $\Sigma_{\mathbb{I}}^{\psi} \subseteq \Sigma_{\mathbb{I}}$. Therefore, Tommy's expected utility from observing the signal exceeds Barney's.

Turning to Part 2, suppose both Tommy and Barney invest given the prior. Now, Tommy's gain from observing the signal is the gain from switching his action to not investing times the probability he observes a signal that would cause him to switch:

$$\sum_{\sigma \in \Sigma_{\mathbb{S}}} f_{\Sigma}(\sigma) \left\{ u(w_0) - E_{\Theta|\Sigma} \left[ E_{S_N|\Theta} \left[ u(w(A_S)) | \theta \right] | \sigma \right] \right\}.$$

Analogously, Barney's gain from observing the signal is

$$\sum_{\sigma \in \Sigma_{\mathbb{S}}^{\psi}} f_{\Sigma}(\sigma) \left\{ u(w_0) - E_{\Theta|\Sigma}^{\psi} \left[ E_{S_N|\Theta}^{\psi} \left[ u(w(A_S)) | \theta \right] | \sigma \right] \right\}.$$

By definition of $\Sigma_{\mathbb{S}}$ and $\Sigma_{\mathbb{S}}^{\psi}$, each term in both sums is positive. Since Barney's value of investing in either state is lower than Tommy's, an argument analogous to the argument in the previous paragraph shows that Barney's gain from observing the signal exceeds Tommy's.

Finally, we prove Part 3 of the proposition. Suppose that, given the prior, Tommy invests but Barney does not. As in the proof of Part 1, Barney's gain from observing the signal is

$$\sum_{\sigma \in \Sigma_{\mathbb{I}}^{\psi}} f_{\Sigma}(\sigma) \left\{ E_{\Theta|\Sigma}^{\psi} \left[ E_{S_N|\Theta}^{\psi} \left[ u(w(A_S)) | \theta \right] | \sigma \right] - u(w_0) \right\}.$$

As in the proof of Part 2, Tommy's gain from observing the signal is

$$\sum_{\sigma \in \Sigma_{\mathbb{S}}} f_{\Sigma}(\sigma) \left\{ u(w_0) - E_{\Theta|\Sigma} \left[ E_{S_N|\Theta} \left[ u(w(A_S)) | \theta \right] | \sigma \right] \right\}.$$

The result immediately follows. □

**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 $N_{\varepsilon}$ signals, the probability of Barney buying a finite number of signals from then on is less than $\varepsilon$.*

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 $\varepsilon$.*

**Proof:** Since Barney is prospective-acceptive and gets one signal at a time, he conceives of his problem as in the classic sequential information acquisition setting of Wald (1947). Therefore, we can characterize his optimal policy using two thresholds in the posterior ratio. Denote the lower and upper thresholds by $\underline{\lambda}$ and $\bar{\lambda}$, respectively. These thresholds are functions of the cost of signals and the payoffs for each action in each state (but not of the current posterior ratio). So long as Barney's current posterior ratio is in $(\underline{\lambda}, \bar{\lambda})$, he will continue to acquire information. If the posterior ratio ever is in the region $(0, \underline{\lambda}] \cup [\bar{\lambda}, 1)$, Barney will stop and take an action.

To prove the first part, assume WLOG that state $A$ is the true state. Since Barney is retrospective-pooling, for any two values $m$ and $n$ such that $0 < m < n < \infty$, we can find a $c'$ small enough so that if $c \leq c'$, then $\underline{\lambda} < m < n < \bar{\lambda}$. This is because, as in Wald (1947), the upper (lower) threshold is strictly increasing (decreasing) in $c$ without bound.

Since $0 < p_0 < \infty$ and Barney's limit likelihood ratio $\Pi^{\psi}_{S_\infty|\Theta\times\Theta}(s_\infty|\theta_A, \theta_B, \Omega = i)$ is finite, his limit posterior ratio is bounded away from 0 and $\infty$:

$$0 < p_0 \Pi^{\psi}_{S_\infty|\Theta\times\Theta}(s_\infty|\theta_A, \theta_B, \Omega = i) < \infty.$$

Moreover, since the posterior ratios after just a single $a$-signal or a single $b$-signal are bounded away from 0 and $\infty$, we can find a small enough $\hat{c} > 0$ so that both the following statements are true for all $c < \hat{c}$:

1. $[p_0 \Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A, \theta_B), p_0 \Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A, \theta_B)] \subseteq (\underline{\lambda}, \bar{\lambda})$
2. $p_0 \Pi^{\psi}_{S_\infty|\Theta\times\Theta}(s_\infty|\theta_A, \theta_B, \Omega = i) \in (\underline{\lambda}, \bar{\lambda})$.

Therefore, by the proof of Lemma C, if $c < c'' \equiv \min\{c', \hat{c}\}$, then the probability that Barney purchases an infinite number of signals, conditional on still purchasing signals after $N$ signals have already been purchased, is increasing in $N$ and converges to 1 as $N$ goes to infinity.

Fix $\overline{p} < 1$. By the above argument, for any $c < c''$, we find an $\overline{N}$ such that the probability that Barney purchases an infinite number of signals, conditional on Barney still purchasing signals after $\overline{N}$ signals have already been purchased, is larger than $\overline{p}$. Now, there exist $0 < m < n < \infty$ sufficiently far apart such that the probability equals 1 that Barney's posterior ratio lies the region $(m, n)$ after he observes $\overline{N}$ signals. We can find $c'''$ sufficiently small that for all $c \leq c'''$, $\underline{\lambda} < m < n < \bar{\lambda}$. It follows that if $c < \bar{c} \equiv \min\{c'', c'''\}$, then Barney purchases an infinite number of signals with probability larger than $\overline{p}$.

Turning to the "Furthermore" claim, for Barney to stop purchasing signals, his posterior ratio must leave
$\left[\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B), \Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)\right]$. Since Barney has a positive probability of purchasing an infinite number of signals, from the proof of Lemma C, we know that if he still wants to purchase an additional signal after having already purchased $\hat{N}$ signals, the probability of the posterior ratio leaving $\left[\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B), \Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)\right]$ is less than $1 - \sum_{i=\hat{N}}^{\infty} \frac{p(1-p)}{i^2\delta^2}$, where $p = \Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B)$ and $\delta$ is the distance between the limit proportion of $a$-signals and the closest proportion that induces Barney to stop purchasing signals. This expression is an upper bound on the probability of Barney purchasing a finite number of signals, if he still wants to purchase an additional signal after seeing $\hat{N}$ signals. As noted in the proof of Lemma C, this upper bound converges to 0 as $\hat{N} \to \infty$; hence, for any $\epsilon > 0$, we can find a $N_{\varepsilon}$ such that if Barney still wants to purchase an additional signal after seeing $N_{\varepsilon}$ signals, then the probability of him purchasing a finite number of signals is less than $\epsilon$.

Turning to the second part of the proposition, denote the situations where there is a current posterior ratio of $q$, which is a belief where the agent still strictly wants to experiment, after $N$ and $N' > N$ signals as $q_N$ and $q_{N'}$, respectively. Assume that given $q_{N'}$, a sequence of $m$ signals, denoted $s_m$, causes the agent to stop purchasing signals and (WLOG) take action $a$, and that no truncation of $s_m$ causes the agent to stop purchasing signals. We will show that given initial situation $q_N$, the sequence $s_m$ also causes the agent to stop purchasing signals. This proves that the probability of the agent stopping is weakly higher given $q_N$ than $q_{N'}$.

If any truncation of $s_m$ causes the agent to stop given $q_N$, then we are done, so assume not.

Now we shall show that a sufficient condition for part *(a)* is that the posterior induced by $k$ $a$-signals in a row is less in favor of state $A$ when starting with $q_{N'}$ than when starting with $q_N$. To see this, reorder $s_m$ so that it begins with $a$ and $b$ signals paired off in alternating orders. Note that this reordering does not change the posterior after $s_m$, only the path of posterior ratios. Since by assumption the agent ends up favoring action $a$, there must be more $a$-signals than $b$-signals, so the reordered $s_m$ must end with some set of $a$-signals in a row. Denote this number as $k$. Note that after every pair of signals, $a, b$, the likelihood ratio must still be $q$, since the signal structure is symmetric. So therefore, starting with either $q_N$ or $q_{N'}$, after $m - k$ signals (the pairs of $a$ and

$b$ signals) of the reordered $s_m$, the likelihood ratio is still $q$. Hence we are now simply comparing the effect of $k$ $a$-signals in a row, given $N+m-k$ preceding signals versus $N'+m-k$ preceding signals.

Note that given $q_{N'}$, there must have been more preceding $a$-signals than given $q_N$. Assume there were $r'$ and $r$ $a$-signals, respectively. Hence in the sequence $q_{N'} \cup s_m$, there is a total of $N'+m$ signals, and the number of $a$-signals is $r'$ (the number of $a$-signals in the first $N'$) plus $\frac{m-k}{2}$ (the number of $a$-signals in the alternating set) plus $k$ (the number of $a$-signals in a row at the end). Therefore, using Lemma $\beta 2$, the agent's posterior ratio after $q_{N'} \cup s_m$ is

$$\begin{aligned}
& \Pi^{\psi}_{\Theta\times\Theta|S_{N'+m}}(\theta_A,\theta_B|q_{N'}\cup s_m) \\
= \;& p_0 \frac{\Gamma\left(\theta_A\psi + r' + \frac{m-k}{2} + k\right)}{\Gamma\left(\theta_B\psi + r' + \frac{m-k}{2} + k\right)} \frac{\Gamma\left((1-\theta_A)\psi + N' - r' + \frac{m-k}{2}\right)}{\Gamma\left((1-\theta_B)\psi + N' - r' + \frac{m-k}{2}\right)} \\
= \;& p_0 \frac{\Gamma\left(\theta_A\psi + r' + \frac{m-k}{2}\right)}{\Gamma\left(\theta_B\psi + r' + \frac{m-k}{2}\right)} \frac{\Gamma\left((1-\theta_A)\psi + N' - r' + \frac{m-k}{2}\right)}{\Gamma\left((1-\theta_B)\psi + N' - r' + \frac{m-k}{2}\right)} \Pi_{i=1}^{k} \frac{\theta_A\psi + r' + \frac{m-k}{2} + i}{(1-\theta_A)\psi + r' + \frac{m-k}{2} + i} \\
= \;& q\Pi_{i=1}^{k} \frac{\theta_A\psi + r' + \frac{m-k}{2} + i}{(1-\theta_A)\psi + r' + \frac{m-k}{2} + i}.
\end{aligned}$$

In the sequence $q_N \cup s_m$, there is a total of $N+m$ signals, and the number of $a$-signals is $r$ (the number of $a$-signals in the first $N$) plus $\frac{m-k}{2}$ (the number of $a$-signals in the alternating set) plus $k$ (the number of $a$-signals in a row at the end). Therefore, analogously, the agent's posterior ratio after $q_N \cup s_m$ is

$$\Pi^{\psi}_{\Theta\times\Theta|S_{N'+m}}(\theta_A,\theta_B|q_N\cup s_m) = q\Pi_{i=1}^{k} \frac{\theta_A\psi + r + \frac{m-k}{2} + i}{(1-\theta_A)\psi + r + \frac{m-k}{2} + i}.$$

Notice that for all $i$,

$$1 < \frac{\theta_A\psi + r' + \frac{m-k}{2} + i}{(1-\theta_A)\psi + r' + \frac{m-k}{2} + i} < \frac{\theta_A\psi + r + \frac{m-k}{2} + i}{(1-\theta_A)\psi + r + \frac{m-k}{2} + i}$$

since $r' > r$ and both fractions are larger than 1. Therefore, since the agent stops purchasing signals given his posterior in situation $q_{N'}$, he must stop given his posterior in situation $q_N$, which favors state $A$ even more strongly.

Now we turn to the third part of the proposition. Denote the prior ratio on state $A$ by $p = \frac{f_\Theta(\theta_A)}{f_\Theta(\theta_B)}$. Denote the probability of Barney taking the correct action—WLOG, action $\mu_A$—after purchasing $N$ signals (given parameters $\theta_A, \theta_B, \psi, p$) as $V(A, \theta_A, \theta_B, \psi, N, p)$. Denote the probability of Barney taking the incorrect action $\mu_B$ as $V(B, \theta_A, \theta_B, \psi, N, p)$. For brevity, we will refer to these as

$V(A,N,p)$ and $V(B,N,p)$, respectively. We will show that, for any $\nu > 0$, there exists an $N_\nu > 2$ such that if Barney still wants to purchase a signal after having already observed $N_\nu$ signals, then for all $N \geq N_\nu$, $\frac{V(B,N,p)}{V(A,N,p)} < \nu$. We will do this in several steps:

1. First, we will construct $V(i,N,p)$ in a way that will make it manageable to work with.
2. Second, we will fix equal priors $p = 1$, and we will show that, for any $\nu > 0$, there exists $N_\nu > 2$ such that for all $N \geq N_\nu$, $\frac{V(B,N,1)}{V(A,N,1)} < \nu$.
3. Third, we will show that for any priors $p > 1$ that strictly favor the true state and for all $N \geq N_\nu$, $\frac{V(B,N,p)}{V(A,N,p)} < \frac{V(B,N,1)}{V(A,N,1)}$. This is because relative to the equal priors case, there is a uniform shift of probability from likelihoods that favor Barney stopping and taking an incorrect action to a stopping and taking the correct action.

Now we shall do each step in order:

For the first step of the argument, because the signals are symmetric, we can define

$$\lambda \equiv \frac{1}{\Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)} = \Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B).$$

The set of posterior ratios such that Barney still wants to purchase an additional signal, after having seen $N$ signals, is denoted $Q(N) \subset [\lambda, \frac{1}{\lambda}]$, a set with a finite number of elements. Denote a typical element by $q$.

We will construct $V$ from a more elementary function: $\phi(A,q,N,M)$, the probability of Barney stopping purchasing signals and taking the $\mu_A$ action in the next $M$ signals, when the current posterior ratio is $q$, and $N$ signals have already been observed. Given $p$ and $N$, there is a one-to-one mapping between the posterior ratio $q$ and the "number of $a$ signals" previously observed, which may not be a natural number for a particular $q$. Therefore we will consider the number of $a$ and $b$ signals as fixed at the levels implied by $q$.

After $N$ signals have been observed, Barney stops purchasing signals and takes action $\mu_A$ only if the number of $a$-signals exceeds the number of $b$-signals by at least some number. Abusing notation slightly, denote by $s^j_M$ an exact sequence of $a$ and $b$ signals from a set of $M$ signals, with ordering $j$ indexed from 0 to $2^M - 1$ (so $s^j_M$ could contain any number of signals from 0 up to $M$ signals). Denote the number of $a$-signals in $s^j_M$ as $A(s^j_M)$. Define $a(s^j_M, N, q) \in \{0,1\}$ as an indicator of whether Barney would have stopped purchasing signals and taken action $\mu_A$ with the exact ordering of signals in $s^j_M$, given that the posterior ratio was $q$ after the first $N$ signals. Now we can construct the probability of Barney stopping purchasing signals and taking action $\mu_A$ within the next $M$ signals, given that the posterior ratio was $q$ after the first $N$ signals:

$$\phi(A, q, N, M) = \sum_{j=1}^{2^M - 1} a(s_M^j, N, q)\theta_A^{A(s_M^j)}(1 - \theta_A)^{M - A(s_M^j)},$$

a polynomial in $\theta_A$ and $1 - \theta_A$.

To gain some intuition for $\phi$, consider $q = 1$. Because the rates are symmetric, if we consider a set of $j$ signals that causes Barney to stop purchasing signals and take the $\mu_A$ action, and we replace all the $\theta_A$'s with $(1 - \theta_A)$'s, and vice-versa, then that set of signals will cause Barney to stop purchasing signals and take the $\mu_B$ action. Similarly, if we have a set of $j$ signals that has not caused Barney to stop purchasing signals, then the same set of signals, but replacing $\theta_A$'s with $(1 - \theta_A)$'s and vice-versa, will also not cause Barney to stop purchasing signals. Therefore, $\phi(B, 1, N, M)$ is the same function as $\phi(A, 1, N, M)$ but with all the $\theta_A$'s replaced with $(1 - \theta_A)$'s and vice-versa. Now suppose either $q > 1$ or $q < 1$. Now $\phi(B, q, N, M)$ is not simply $\phi(A, q, N, M)$ with the rates switched. Because rates are symmetric, however, if we define $q' = \frac{1}{q}$, then $\phi(B, q', N, M)$ is the same function as $\phi(A, q, N, M)$ but with $\theta_A$ and $1 - \theta_A$ switched.

Define $\rho(N)$ as the minimum number of $a$-signals in a row that would need to occur after an initial set of $N$ signals to put the posterior ratio above $\frac{1}{\lambda}$ (the upper bound of the interval $Q(N)$, as defined above), given that the posterior ratio was $\lambda$ (the lower bound of the interval $Q(N)$) after the first $N$ signals. $\rho(N)$ will be a useful quantity because, regardless of the posterior ratio after $N$ observations, it is the number of additional signals after which Barney stopping purchasing signals and taking either action could always possibly occur. Note that $\rho(N)$ is increasing in $N$.

Define $x(q, N, p) < N$ as the number of $a$-signals out of $N$ such that $x(q, N, p)$ $a$-signals and $N - x(q, N, p)$ $b$-signals induces posterior ratio of $q$ when the prior is $p$.

For each $x(q, N, p)$, we can construct the set of possible sequences that generates $x$ $a$-signals and $N - x$ $b$-signals without causing Barney to stop purchasing signals at any point in the sequence. Denote the set of these sequences as $X(q, N, p)$. Each one of these sequences must have the same final number of $a$ and $b$ signals, but it must be the case that the difference between the number of $a$-signals and $b$-signals cannot have been too large at any point in the sequence, or else Barney would have stopped purchasing signals. Let $g(q, N, p) = \gamma(q, N, p)\theta_A^{x(q,N,p)}(1 - \theta_A)^{N - x(q,N,p)}$ denote the probability of the set of sequences that leads to $x$ $a$-signals and $N - x$ $b$-signals without Barney ever having stopped purchasing signals, where $\gamma$ is the number of sequences in the set.

Define $r_N(i)$ for $i = 0, 1, ...$ recursively: $r_N(0) = N$ and $r_N(i) = r_N(i-1) + \rho(r_N(i-1))$ for $i > 0$.

We now construct $V(A,N,p)$ as follows:

$$V(A,N,p) = \sum_{i=1}^{\infty} \sum_{q \in Q(r_N(i-1))} \phi(A,q,r_N(i-1),\rho(r_N(i-1)))g(q,r_N(i-1),p).$$

We define $V(B,N,p)$ analogously.

To understand this definition of $V$, each term is the probability of a Barney stopping purchasing signals and taking the correct action within the next $r_N(i)$ signals, conditional on having observed $r_N(i-1)$ signals without stopping purchasing signals, starting with a prior ratio of $p$.

For the second step of the argument, now we will fix the prior ratio at 1 and show that for any $\nu > 0$, there exists $N_\nu > 2$ such that for all $N \geq N_\nu$, $\frac{V(B,N,1)}{V(A,N,1)} < \nu$.

We will begin by proving some key claims:

**Claim 10.A** *Let $a_i, b_i > 0$ for $i = 1, 2...$, such that $\frac{a_i}{b_i} \leq \frac{a_{i+1}}{b_{i+1}}$. Then $\frac{a_1}{b_1} \leq \frac{\sum_{i=1}^N a_i}{\sum_{i=1}^N a_i} \leq \frac{a_N}{b_N}$.*

To see this is true, note that for positive $a, b, c, d$, if $0 < \frac{a}{b} < \frac{c}{d}$, then $\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$. Induction then shows that $\frac{\sum_{i=1}^N a_i}{\sum_{i=1}^N b_i}$ is bounded by the maximum and minimum $\frac{a_i}{b_i}$.

**Claim 10.B** *For any $\nu > 0$, there exists $N_\nu > 2$ such that for all $N \geq N_\nu$,*

$$\frac{\sum_{q \in Q(N)} \phi(B,q,N,\rho(N))g(q,N,p)}{\sum_{q \in Q(N)} \phi(A,q,N,\rho(N))g(q,N,p)} < \nu.$$

Together, Claims 10.A and 10.B imply the result we are seeking. Claim 10.A, combined with the fact that the probability of Barney stopping and taking an action is shrinking as the number of signals gets large, implies that

$$\begin{aligned}\frac{V(B,N,p)}{V(A,N,p)} &= \frac{\sum_{i=1}^{\infty} \sum_{q \in Q(N)} \phi(B,q,r_N(i-1),r_N(i))g(q,r_N(i-1),p)}{\sum_{i=1}^{\infty} \sum_{q \in Q(N)} \phi(A,q,r_N(i-1),r_N(i))g(q,r_N(i-1),p)} \\ &\leq \frac{\sum_{q \in Q(N)} \phi(B,q,N,\rho(N))g(q,N,p)}{\sum_{q \in Q(N)} \phi(A,q,N,\rho(N))g(q,N,p)}.\end{aligned}$$

Claim 10.B then says that given any $\nu > 0$, for large enough $N$, this ratio is $< \nu$.

A sufficient condition for Claim 10.B is:

**Claim 10.C** *For any $\nu > 0$, there exists $N_\nu > 2$ such that for all $N \geq N_\nu$ and $q \in Q(N)$,*

$$\frac{\phi(B,q,N,\rho(N))g(q,N,1) + \phi(B,q',N,\rho(N))g(q',N,1)}{\phi(A,q,N,\rho(N))g(q,N,1) + \phi(A,q',N,\rho(N))g(q',N,1)} < \nu.$$

To see that Claim 10.C implies Claim 10.B, notice that $\frac{\sum_{q\in Q(N)}\phi(B,q,N,\rho(N))g(q,N,p)}{\sum_{q\in Q(N)}\phi(A,q,N,\rho(N))g(q,N,p)}$ is

$$\begin{aligned}
&\leq \max_{q,q'\in Q(N)\text{ s.t. } q'=\frac{1}{q}} \frac{\phi(B,q,N,\rho(N))g(q,N,1)+\phi(B,q',N,\rho(N))g(q',N,1)}{\phi(A,q,N,\rho(N))g(q,N,1)+\phi(A,q',N,\rho(N))g(q',N,1)}\\
&< \nu
\end{aligned}$$

where the first inequality follows from Claim 10.A and the second by Claim 10.C.

Therefore, we will have completed the second step of the argument once we prove Claim 10.C. We substitute:

$$\begin{aligned}
&\phi(A,q,N,\rho(N))g(q,N,1)\\
=\;& g(q,N,1)\sum_{j=1}^{2^{\rho(N)}-1} a(s^j_{\rho(N)},N,q)\theta_A^{A(s^j_{\rho(N)})}(1-\theta_A)^{\rho(N)-A(s^j_{\rho(N)})}\\
=\;& \gamma(q,N,1)\theta_A^{x(q,N,1)}(1-\theta_A)^{N-x(q,N,1)}\sum_{j=1}^{2^{\rho(N)}-1} a(s^j_{\rho(N)},N,q)\theta_A^{A(s^j_{\rho(N)})}(1-\theta_A)^{\rho(N)-A(s^j_{\rho(N)})}.
\end{aligned}$$

Note that some of the sequences of signals starting with a posterior of $q$ will cause Barney to stop purchasing signals before $\rho(N)$ signals. Imagine two sequences begin with the same initial $z$ signals but differ in the signals that occur afterward. If the $z^{\text{th}}$ signal causes Barney to stop purchasing signals, then these subsequent signals are irrelevant, and we can aggregate the two sequences together. Let $\zeta(q,\rho(N),N,z,i)$ denote the number of $a$-signals needed out of an additional $\rho(N)$ signals, starting with a posterior of $q$ after $N$ signals, for Barney to stop purchasing signals and take action $i$ after exactly $z$ signals, and Barney did not stop purchasing signals earlier. Let the lowercase letter $\varsigma$ count the number of permutations of such sequences. Continuing to substitute:

$$\begin{aligned}
&\phi(A,q,N,\rho(N))g(q,N,1)\\
=\;& \gamma(q,N,1)\theta_A^{x(q,N,1)}(1-\theta_A)^{N-x(q,N,1)}\sum_{z=1}^{\rho(N)} \varsigma(q,\rho(N),N,z,A)\theta_A^{\zeta(q,z,N,A)}(1-\theta_A)^{z-\zeta(q,z,N,A)}.
\end{aligned}$$

Now, define

$$\begin{aligned}
b_z(q',N) &\equiv \gamma(q',N,1)\theta_A^{x(q',N,1)}(1-\theta_A)^{N-x(q',N,1)}\varsigma(q',\rho(N),N,z,B)(1-\theta_A)^{z-\zeta(q',z,N,B)}\theta_A^{\zeta(q',z,N,B)}\\
a_z(q',N) &\equiv \gamma(q,N,1)\theta_A^{x(q,N,1)}(1-\theta_A)^{N-x(q,N,1)}\varsigma(q,\rho(N),N,z,A)(1-\theta_A)^{z-\zeta(q,z,N,A)}\theta_A^{\zeta(q,z,N,A)}.
\end{aligned}$$

We can write

$$\frac{\phi(B,q,N,\rho(N))g(q,N,1)+\phi(B,q',N,\rho(N))g(q',N,1)}{\phi(A,q,N,\rho(N))g(q,N,1)+\phi(A,q',N,\rho(N))g(q',N,1)}=\frac{\sum_{z=1}^{\rho(N)}b_z(q,N)+\sum_{z=1}^{\rho(N)}b_z(q',N)}{\sum_{z=1}^{\rho(N)}a_z(q,N)+\sum_{z=1}^{\rho(N)}a_z(q',N)}$$

Starting from a prior ratio of 1, because the signals are symmetric, if $x$ $a$-signals and $N-x$ $b$-signals leaves an agent with a posterior of $q$, then $x$ $b$-signals and $N-x$ $a$-signals leaves an agent with a posterior of $q'$. Therefore $x(q',N,1)=N-x(q,N,1)$. Similarly, $\gamma(q',N,1)=\gamma(q,N,1)$. Also, starting with a posterior of $q$, if $x$ additional $a$-signals and $N-x$ additional $b$-signals leaves Barney wanting to take stop purchasing signals and take action $\mu_A$, then starting with a posterior of $q'$, $x$ additional $b$-signals and $N-x$ additional $a$-signals leaves Barney wanting to stop purchasing signals and take action $\mu_B$. Therefore $\zeta(q',z,N,B)=z-\zeta(q,z,N,A)$ and $\varsigma(q',\rho(N),N,z,B)=\varsigma(q,\rho(N),N,z,A)$.

Using these relationships,

$$\begin{aligned}\frac{b_z(q',N)}{a_z(q',N)} &= \frac{\gamma(q',N,1)\theta_A^{x(q',N,1)}(1-\theta_A)^{N-x(q',N,1)}\varsigma(q',\rho(N),N,z,B)(1-\theta_A)^{z-\zeta(q',z,N,B)}\theta_A^{\zeta(q',z,N,B)}}{\gamma(q,N,1)\theta_A^{x(q,N,1)}(1-\theta_A)^{N-x(q,N,1)}\varsigma(q,\rho(N),N,z,A)(1-\theta_A)^{z-\zeta(q,z,N,A)}\theta_A^{\zeta(q,z,N,A)}}\\ &= \frac{\gamma(q,N,1)\theta_A^{N-x(q,N,1)}(1-\theta_A)^{x(q,N,1)}\varsigma(q,\rho(N),N,z,A)(1-\theta_A)^{\zeta(q,z,N,A)}\theta_A^{z-\zeta(q,z,N,A)}}{\gamma(q,N,1)\theta_A^{x(q,N,1)}(1-\theta_A)^{N-x(q,N,1)}\varsigma(q,\rho(N),N,z,A)(1-\theta_A)^{z-\zeta(q,z,N,A)}\theta_A^{\zeta(q,z,N,A)}},\end{aligned}$$

which is well-defined as long as $\gamma(q,N,1)\varsigma(q,\rho(N),N,z,A)\neq 0$. (If $\gamma(q,N,1)\varsigma(q,\rho(N),N,z,A)=0$, then $b_z$ drops out from $V(B,N,p)$, and $a_z$ drops out from $V(A,N,p)$. Hence these terms can be ignored.) Canceling and collecting terms:

$$\frac{b_z(q',N)}{a_z(q',N)}=\frac{(1-\theta_A)^{x(q,N,1)+\zeta(q,z,N,A)}\,(\theta_A)^{N+z-x(q,N,1)-\zeta(q,z,N,A)}}{(\theta_A)^{x(q,N,1)+\zeta(q,z,N,A)}\,(1-\theta_A)^{N+z-x(q,N,1)-\zeta(q,z,N,A)}}$$

Using Claim 10.A ,

$$\frac{\sum_z b_z(q',N)}{\sum_z a_z(q',N)}<\max_{z\ s.t.\ \gamma(q,N,1)\varsigma(q',\rho(N),N,z,B)\neq 0}\frac{(1-\theta_A)^{x(q,N,1)+\zeta(q,z,N,A)}\,(\theta_A)^{N+z-x(q,N,1)-\zeta(q,z,N,A)}}{(\theta_A)^{x(q,N,1)+\zeta(q,z,N,A)}\,(1-\theta_A)^{N+z-x(q,N,1)-\zeta(q,z,N,A)}}.$$

The numerator is the probability that, starting with a prior ratio of 1, there will be a sequence of $a$ and $b$ signals, $N+z$ in total, that lands the agent just inside the zone where the agent stops purchasing signals and takes action $\mu_B$. The denominator is the probability that, starting with a prior ratio of 1, there will be a sequence of $a$ and $b$ signals, $N+z$ in total, that lands the agent just inside the zone where the agent stops purchasing signals and takes action $\mu_A$. This expression can be rearranged to yield

$$\left(\frac{1-\theta_A}{\theta_A}\right)^{2x(q,N,1)+2\zeta(q,z,N,A)-N-z},$$

where the exponent, $2x(q,N,1)+2\zeta(q,z,N,A)-N-z \equiv \delta(N+z)$, is the minimum difference between the number of $a$ and $b$ signals out of $N+z$ signals in total at which Barney will stop purchasing signals; if there are $\delta(N+z)$ more $a$ than $b$ signals, then Barney will take action $\mu_A$; and if there are $\delta(N+z)$ more $b$ than $a$ signals, then Barney will take action $\mu_B$. Notice that $\delta(N+z)$ is not a function of $q$. Because Barney's likelihood ratio asymptotically depends on the proportion of $a$ signals, not the difference between the number of $a$ and $b$ signals, $\lim_{N+z\to\infty}\delta(N+z)=\infty$. Furthermore, since $\rho(N)$ is growing in $N$, then it follows that for any $\epsilon>0$ we can find an $N'$ large enough such that for all $q \in \left(\Pi^{\psi}_{S_1|\Theta\times\Theta}(b|\theta_A,\theta_B), \Pi^{\psi}_{S_1|\Theta\times\Theta}(a|\theta_A,\theta_B)\right)$ (and so for any $q\in Q$ for any $N$) and all $N>N'$,

$$\left(\frac{1-\theta_A}{\theta_A}\right)^{2x(q,N,1)+2\zeta(q,z,N,A)-N-z},$$

is less than $\epsilon$. It then follows directly from the previous claims that the upper bound on $\frac{V(B,N,p)}{V(A,N,p)}$, $\left(\frac{1-\theta_A}{\theta_A}\right)^{\delta(N+z)}$, goes to 0 as $N\to\infty$.

Therefore, although $Q$ varies with $N$, it must be the case that for any $q$ in $Q$, the limit result holds. That is, the convergence is uniform over $q\in Q$.

For the third step of the argument, now assume that the prior rate favors $A$: $p>1$. Note that nothing in the construction of the $\phi$ function changes. However, because $p$ is now closer to $q>1$ than it is to $q'<1$, $\frac{g(q,N,p)}{g(q',N,p)} < \frac{g(q,N,1)}{g(q',N,1)}$. Using this fact, and the result proven above for the $p=1$ case: for any $\nu>0$, there exists $N_\nu>2$ such that for all $N\geq N_\nu$, for any $q\in Q(N)$,

$$\frac{\phi(B,q,N,\rho(N))g(q,N,p)}{\phi(A,q',N,\rho(N))g(q',N,p)} < \frac{\phi(B,q,N,\rho(N))g(q,N,1)}{\phi(A,q',N,\rho(N))g(q',N,1)} < \nu,$$

and

$$\frac{\phi(B,q',N,\rho(N))g(q',N,p)}{\phi(A,q,N,\rho(N))g(q,N,p)} < \frac{\phi(B,q',N,\rho(N))g(q',N,1)}{\phi(A,q,N,\rho(N))g(q,N,1)} < \nu.$$

We argued above that Claim 10.A implies

$$\frac{V(B,N,p)}{V(A,N,p)} \leq \frac{\sum_{q\in Q(N)}\phi(B,q,N,\rho(N))g(q,N,p)}{\sum_{q\in Q(N)}\phi(A,q,N,\rho(N))g(q,N,p)}.$$

Using Claim 10.A again, along with the above inequalities, gives the result:

$$\frac{V(B,N,p)}{V(A,N,p)} < \nu.$$

□

**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 $\delta$. 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.*

**Proof:** We begin with the first part of the proposition. Since the high and low payoffs differ for both actions, both actions are informative. If the agent gets the high payoff, consider it an $a$ signal; the low payoff, a $b$ signal. In this case, standard results imply that as the number of signals goes to infinity Tommy almost surely learns the true state. Furthermore, when Tommy is sufficiently confident about the state, he takes the action with the high payoff.

We know that as the number of signals Barney receives goes to infinity, his posterior ratio converges to the finite number

$$\frac{f^{\psi}_{ß|\Theta}(\beta = \theta|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta = \theta|\theta_B)} \frac{f(\theta_A)}{f(\theta_B)}$$

almost surely (where $\theta$ is the rate in the true state). This is increasing in the prior probability of state $A$.

Furthermore, because Barney is prospective-acceptive, we know that for the same current belief as Tommy, his current action will be the same. Hence Barney will only take action $B$ if his current posterior ratio of state $A$ to state $B$ is below some threshold $\bar{\lambda}$. Let $\bar{p}$ denote the largest prior ratio $\frac{f(\theta_A)}{f(\theta_B)}$ such that

$$1 > \frac{f^{\psi}_{ß|\Theta}(\beta = \theta|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta = \theta|\theta_B)} \frac{f(\theta_A)}{f(\theta_B)} \geq \bar{\lambda}$$

and

$$1 > \frac{f(\theta_A)}{f(\theta_B)} \geq \bar{\lambda}.$$

For all prior ratios $p$ such that $1 > p > \bar{p}$,

$$1 > \frac{f^{\psi}_{ß|\Theta}(\beta = \theta|\theta_A)}{f^{\psi}_{ß|\Theta}(\beta = \theta|\theta_B)} p > \bar{\lambda}$$

and

$$1 > p > \bar{\lambda}.$$

By Lemma C, given prior $p$, with positive probability Barney's posterior ratio always stays within $(\bar{\lambda}, 1)$, and so he always takes action $A$.

Now we prove the second part of the proposition. Once again, because Barney is prospective-acceptive, given the same current belief as Tommy, he will choose the same action. Standard results imply that either agent takes action $B$ if his current posterior ratio is below some threshold $\bar{\lambda}$; otherwise he takes action $A$ (and is indifferent at $\bar{\lambda}$). In the present setting, whenever an agent takes action $A$, he receives a signal, but whenever he takes action $B$, he receives no signal. In this environment, the result concerning Tommy is entirely standard. For Barney, define $\bar{p}$ as above, and by the same argument as above, for all prior ratios $p$ such that $1 > p > \bar{p}$, with positive probability Barney's posterior ratio always stays within $(\bar{\lambda}, 1)$, and so he always takes action $A$.

The probability of always taking action $A$ is increasing in the prior. To see this, consider priors $p', p''$ such that $p'' \geq p' > \underline{p}$. Let $\hat{S}_\infty(p')$ denote the set of sequences that lead to Barney taking the $A$ action every period given prior $p'$. Consider any particular sequence $s \in \hat{S}_\infty(p')$. Let $s_N$ be the truncation of $s$ after $N$ signals. Note that $p' \frac{f^{\psi}_{\Theta|S_N}(\theta_A|s_N)}{f^{\psi}_{\Theta|S_N}(\theta_B|s_N)} \in [\bar{\lambda}, 1)$ for all $N = 1, 2, ...$ by construction. Since $p'' \geq p'$,

$$p'' \frac{f^{\psi}_{\Theta|S_N}(\theta_A|s_N)}{f^{\psi}_{\Theta|S_N}(\theta_B|s_N)} \geq p' \frac{f^{\psi}_{\Theta|S_N}(\theta_A|s_N)}{f^{\psi}_{\Theta|S_N}(\theta_B|s_N)}$$

for all $N$, and so $p'' \frac{f^{\psi}_{\Theta|S_N}(\theta_A|s_N)}{f^{\psi}_{\Theta|S_N}(\theta_B|s_N)} \in [\bar{\lambda}, 1)$ for all $N$. Therefore given prior $p''$, Barney will also always take action $A$ for all sequences in $\hat{S}_\infty(p')$. Therefore the probability of always taking action $A$ given prior $p''$ must be weakly higher than given prior $p'$.

The probability of always taking action $A$ is always less than 1 because the limit likelihood ratio for Barney of state $A$ to state $B$ if he observes 0 $a$-signals is 0. Therefore, regardless of the prior, if Barney receives enough $b$ signals in a row initially, his posterior ratio will fall below $\bar{\lambda}$.

$\square$

## Appendix D: Formal Description of the Multiple-Sample Model

To formalize the information-processing assumptions from Section 4, consider a decision problem in which up to $1 \leq T \leq \infty$ binary signals will be realized in total. To greatly simplify notation and clarify presentation, we assume that how the environment or his own thinking leads him to group signals is independent of the realizations of those signals and of earlier decisions.[16] We define $T+3$ partitions, each represented by a set of time periods, of the set of signals that fully characterize our predictions. We conceive of the first three of these partitions as embodying physical, informational, and perceptual assumptions rather than assumptions about Barney's statistical reasoning. The first partition we define by the set of dates at which the agent knows he must make decisions, $D \subseteq \{0, 1, ..., T\}$. If $\tau \in D$, the agent knows that he made or will make a decision after observing $\tau$ signals but before observing $\tau + 1$ signals. The agent's payoff may depend on any or all of the decisions he makes, the signals that are realized, and the underlying state with which the signals are correlated. This opportunity-for-decisions partition, $D$, is of course specified in every economic model of decisionmaking.

The second partition is characterized by a set $C \subseteq \{0, 1, ..., T-1\}$ of dates where a new clump begins, such that if and only if $\tau \in C$, then signal $\tau$ is in a different clump than signal $\tau - 1$. We assume $0 \in C$. For example, if 11,000 signals arrives as a clump of 10,000 signals followed by a clump of 1,000 signals, then $T = 11,000$ and $C = \{0, 10, 001\}$. There is one obvious basic restriction on clumping that we must make:

**Clumping Assumption 0.** $D \subseteq C$.

This assumption states that at any history where the agent makes a decision, subsequent signals arrive as a separate clump than previous signals. This restriction is inherent in the notion of clumps because the signals cannot have "arrived" together (at least in the relevant sense of the agent's knowledge of their realizations) if the agent knows the realizations of only some of the signals. We treat this as a coherence assumption and always impose it.

Another coherence assumption on $C$ is that Barney will not treat signals differently when he does not see the signals distinctly at all. If it is the case, unbeknownst to Barney, that 38 of the people in *Consumer Reports* statistics of 10,000 car owners have the last name Smith, we assume

[16] This is a substantive restriction, and we know of examples where it seems unrealistic, but we do not know how relaxing it will improve insights. Importantly, we are not assuming that *whether* Barney gathers more information is independent of what he has learned; our examples in Section 5 revolve around exactly such a decision by Barney. In those examples, Barney faces a decision after each signal whether to pay to observe another signal. To accommodate those examples in the framework of this section, it can be assumed that whenever Barney chooses not to observe further signals, decisions that "occur" at those future signals do not affect his payoff.

that Barney cannot (even if thusly motivated) treat Smiths as one sample and non-Smiths as another.[17] Let $I \subseteq \{0, 1, ..., T-1\}$ be a set of dates that defines the third partition, a partition of signals into equivalence classes of indistinguishable signals (so that if $\tau \in I$, then the signal at time $\tau$ is distinguishable from the signal at time $\tau - 1$). Clearly it must be the case that $D \subseteq I$, but we also impose a second restriction on clumping:

**Clumping Assumption 1.** $C \subseteq I$.

Although economic models of decisionmaking do not traditionally specify a clumping partition, our aspiration is have the clumping partition be an exogenous assumption that is not *per se* related to NBLLN, and ideally is pinned down by observable characteristics of a situation. In the interest of minimizing the number of assumptions that have to be specified anew for each new economic model (and hence limit degrees of freedom that might reduce the usefulness of the model), it would be especially attractive to pin down some rule for "clumping" that ties it to $D$ or $I$. The two obvious candidates are:

**Clumping Assumption 2(a).** $C = D$.
**Clumping Assumption 2(b).** $C = I$.

While the three partitions characterized by $D$, $C$, and $I$ reflect the physical and perceptual environment facing Barney, the remaining $T$ partitions embed Barney's NBLLN psychology of how he separates out data. To capture the important possibility that Barney's grouping might differ from different time perspectives, we assume that the agent may process the signals differently at different dates. At each $t = 0, 1, ..., T$, there is a set $P_t \subseteq \{0, 1, ..., T-1\}$ of dates where a new group begins. By letting each $P_t$ contain elements both less than $t$ and greater than $t$, each partition specifies both retrospective and prospective grouping of signals at that point in time. For all $\tau \leq t$, if $\tau \in P_t$, then after having observed $t$ signals, the agent processes signal $\tau$ as being in a different group than signal $\tau - 1$; while if $\tau \notin P_t$, then after having observed $t$ signals, the agent processes signal $\tau$ as being in the same group as signal $\tau - 1$. And for all $\tau > t$, $\tau \in P_t$ means Barney *anticipates* separating out signal $\tau$ from signal $\tau - 1$. Whether he actually does so after observing signal $\tau$ is determined by $\{P_\tau, P_{\tau+1}, ...\}$. We assume that $0 \in P_t$ for all $t$.

Our main modeling constraint, which rules out prospective-pooling whenever Barney will face a subsequent decision node, is formalized as:

[17] Conceivably, Barney could decide to label an indistinguishable group any way he wants, such as ordering them from 1 to 10,000, and then either perceptually or psychologically distinguish the signals based on this labeling. It seems a safe assumption that he will not do so.

**Processing Assumption 0.** For any $t = 0, 1, ..., T$, if $\tau \in D$ and $\tau + 1 \geq t$, then $\tau + 1 \in P_t$.

This assumption states that at any date where the agent makes a decision, he processes signals before and after that date as being in separate groups—and that before that date, he knows he will do so. We consider this to be a modeling coherence assumption because it ensures that Barney's NBLLN from the single-clump model in Section 2 generalizes to every decision node in the multiple-clump model.

We additionally impose the coherence assumption that Barney cannot process in separate groups signals that he cannot distinguish from each other.

**Processing Assumption 1.** For any $t = 0, 1, ..., T$, $I \subseteq P_t$.

In words, at every date, Barneys processing partition must be a coarsening of his indistinguishability partition.

We now formalize some ways that Barney might form beliefs retrospectively and prospectively. Because $D$ is the set of nodes where Barney's beliefs are payoff-relevant, these definitions focus the assumptions on the $P_t$'s where $t \in D$:

- **Retrospective-Pooling**: If $t \in D$ and $0 < \tau \leq t$, then $\tau \notin P_t$.
- **Retrospective-Acceptive**: If $t \in D$ and $0 < \tau \leq t$, then $\tau \in P_t \Leftrightarrow \tau \in C$.
- **Prospective-Acceptive**: If $t \in D$ and $\tau > t$, then $\tau \in P_t \Leftrightarrow \tau \in C$.

We omit defining "prospective-pooling" Barney because it is ruled out by Processing Assumption 0 in any decision problem with more than one decision node; at an earlier decision node, Barney cannot expect to pool together future signals that come before and after a future decision node. An agent is **processing-consistent** if he always processes information the way that he expects to process information: $P_t = P_{t'}$ for all $t, t' \in D$.

## References


**Bar-Hillel, Maya**, "The base-rate fallacy in probability judgments," *Acta Psychologica*, 1980, *44* (3), 211–233.

**Beach, Lee Roy**, "Probability Magnitudes and Conservative Revision of Subjective Probabilities," *Journal of Experimental Psychology*, 1968, *77* (1), 57–63.

_ , **James A. Wise, and Scott Barclay**, "Sample Proportions and Subjective Probability Revisions," *Organizational Behavior and Human Performance*, March 1970, *5* (2), 183–190.

**Blackwell, D.**, "Comparison of Experiments," in "Proceedings of the Second Berkeley Symposium on Mathematics, Statistics and Probability," University of California Press, 1951, pp. 93–102.

_ , "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 1953, *24*, 265–272.

**Bodoh-Creed, Aaron L.**, "The Economics of Base Rate Neglect," June 2010. Working paper.

**Camerer, Colin F.**, "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence," *The American Economic Review*, December 1987, *77* (5), 981–997.

**Chapman, Richard C.**, "Prior Probability Bias in Information Seeking and Opinion Revision," *American Journal of Psychology*, 1973, *86* (2), 269–282.

**Chinnis, James O. Jr. and Cameron R. Peterson**, "Inference About a Nonstationary Process," *Journal of Experimental Psychology*, 1968, *77* (4), 620–625.

**Cohen, John and C. E. M. Hansel**, "The Idea of a Distribution," *British Journal of Psychology*, 1955, *46* (2), 111–121.

**Dale, H. C. A.**, "Weighing Evidence: An Attempt to Assess the Efficiency of the Human Operator," *Ergonomics*, May 1968, *11* (3), 215–230.

**Dave, Chetan and Katherine W. Wolfe**, "On Confirmation Bias and Deviations From Bayesian Updating," March 2003. Working paper.

**DeSwart, J.H.**, "A Comment on Marks and Clarkson's Explanation of Conservatism," *Acta Psychologica*, November 1972, *36* (5), 417–419.

**Donnell, Michael L. and Wesley M. Du Charme**, "The Effect of Bayesian Feedback on Learning in an Odds Estimation Task," *Organizational Behavior and Human Performance*, December 1975, *14* (3), 305–313.

**DuCharme, Wesley M.**, "Response Bias Explanation of Conservative Human Inference," *Journal of Experimental Psychology*, 1970, *85* (1), 66–74.

_ **and Cameron R. Peterson**, "Intuitive Inference About Normally Distributed Populations," *Journal of Experimental Psychology*, 1968, *78* (2), 269–275.

**Gettys, Charles F. and Charles W. Manley**, "The probability of an event and estimates of posterior probability based upon its occurence," *Psychonomic Science*, 1968, *11* (2), 47–48.

**Green, Paul E., Michael H. Halbert, and Patrick J. Robinson**, "An Experiment in Probability Estimation," *Journal of Marketing Research*, August 1965, *2* (3), 266–273.

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

_ , "Testing Bayes rule and the representativeness heuristic: Some experimental evidence," *Journal of Economic Behavior and Organization*, January 1992, *17* (1), 31–57.

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

**Gustafson, David H., Ramesh K. Shukla, Andre Delbecq, and G. William Walster**, "A Comparative Study of Differences in Subjective Likelihood Estimates Made by Individuals, Interacting Groups, Delphi Groups, and Nominal Groups," *Organizational Behavior And Human Performance*, April 1973, *9* (2), 280–291.

**Kahneman, Daniel and Amos Tversky**, "Subjective Probability: A Judgment of Representativeness," *Cognitive Psychology*, July 1972, *3* (3), 430–454.

_ **and** _ , "On the psychology of prediction," *Psychological Review*, 1973, *80* (4), 237–251.

**Kraemer, Carlo and Martin Weber**, "How Do People Take into Account Weight, Strength, and Quality of Segregated vs. Aggregated Data? Experimental Evidence," *The Journal of Risk and Uncertainty*, September 2004, *29* (2), 113–142.

**Marks, D.F. and J.K. Clarkson**, "An Explanation of Conservatism in the Bookbag-and-Pokerchips Situation," *Acta Psychologica*, April 1972, *36* (2), 145–160.

**Martin, David W.**, "Data Conflict in a Multinomial Decision Task," *Journal of Experimental Psychology*, 1969, *82* (1), 4–8.

_ **and Charles F. Gettys**, "Feedback and Response Mode in Performing a Bayesian Decision Task," *Journal of Applied Psychology*, 1969, *53* (5), 413–418.

**Nelson, Mark W., Robert Bloomfield, Jeffrey W. Hales, and Robert Libby**, "The Effect of Information Strength and Weight on Behavior in Financial Markets," *Organizational Behavior and Human Decision Processes*, November 2001, *86* (2), 168–196.

**Peterson, C. R. and L.D. Phillips**, "Revision of Continuous Subjective Probability Distributions," *IEEE Transactions on Human Factors in Electronics*, March 1966, *HFE-7* (1), 19–22.

**Peterson, Cameron R. and Alan J. Miller**, "Sensitivity of Subjective Probability Revision," *Journal of Experimental Psychology*, 1965, *70* (1), 117–121.

_ **and Richard G. Swensson**, "Intuitive Statistical Inferences about Diffuse Hypotheses," *Organizational Behavior And Human Performance*, February 1968, *3* (1), 1–11.

_ **, Robert J. Schneider, and Alan J. Miller**, "Sample Size and The Revision of Subjective Probabilities," *Journal of Experimental Psychology*, 1965, *69* (5), 522–527.

_ **, Wesley M. DuCharme, and Ward Edwards**, "Sampling Distributions and Probability Revisions," *Journal of Experimental Psychology*, 1968, *76* (2), 236–243.

**Phillips, Lawrence D. and Ward Edwards**, "Conservatism in a Simple Probability Inference Task," *Journal of Experimental Psychology*, 1966, *72* (3), 346–354.

**Phillips, L.D., W.L. Hays, and W. Edwards**, "Conservatism in Complex Probabilistic Inference," *IEEE Transactions on Human Factors in Electronics*, March 1966, *HFE-7* (1), 7–18.

**Pitz, Gordon F.**, "Sample size, likelihood, and confidence in a decision," *Psychonomic Science*, 1967, *8* (6), 257–258.

**Rabin, Matthew**, "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.

**Sanders, A.F.**, "Choice Among Bets and Revision of Opinion," *Acta Psychologica*, 1968, *28*, 76–83.

**Sasaki, Shunichiro and Toshiji Kawagoe**, "Belief Updating in Individual and Social Learning: A Field Experiment on the Internet," May 2007. Working Paper.

**Shaked, M. and Y L. Tong**, "Comparison of experiments for a class of positively dependent random variables," *Canadian Journal of Statistics*, 1990, *18*, 79–86.

**Shu, Suzanne and George Wu**, "Belief Bracketing: Can Partitioning Information Change Consumer Judgments?," October 2003. Working paper.

**Strub, Michael H.**, "Experience and Prior Probability in a Complex Decision Task," *Journal of Applied Psychology*, 1969, *53* (2), 112–117.

**Swart, J.H. De**, "Effects of Diagnosticity and Prior Odds on Conservatism in a Bookbag-And-Pokerchip Situation," *Acta Psychologica*, February 1972, *36* (1), 16–31.

**Teigen, Karl Halvor**, "Subjective sampling distributions and the additivity of estimates," *Scandinavian Journal of Psychology*, September 1974, *15* (1), 50–55.

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

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

**Wheeler, Gloria and Lee Roy Beach**, "Subjective Sampling Distributions and Conservatism," *Organizational Behavior and Human Performance*, February 1968, *3* (1), 36–46.

**Whitt, Ward**, "A Note on the Influence of the Sample on the Posterior Distribution," *Journal of the American Statistical Association*, 1979, *74* (366), 424–426.