Learning from Social Data

What Should Make Us Uncomfortable about Bayesian Priors in the Social Sciences?

Marika Csapo, UID: 003643895
mcsapo@ucla.edu

February 8, 2015

Abstract

Bayesian analyses are often critiqued on the basis of dubious exchangeability claims regarding the data (as are frequentist analyses regarding claims of independence). Not only must observed data be exchangeable, but prior "data" must be as well, and the observed data must also be exchangeable with the prior data—an assumption typically not explicitly justified by the practitioner. I focus on the last of these assumptions, which is likely to be violated in the social sciences where many of our "observations" are informed by human behavior and, therefore, by the prior beliefs of the humans taking the observed actions. Social or psychological biases that characterize the beliefs of a society will thus inform the observed data, violating the exchangeability assumption. One of the common defensive arguments offered by many Bayesian practitioners is that as long as there is some component of new information in the observed data, repeated observation-updating cycles will still eventually produce a posterior distribution that is highly informative (similar to the idea of consistency in frequentist statistics). In frequentist statistics we have power analyses—a way of estimating how much data we need to get desirable properties from our estimator. This paper develops a model that parameterizes the degree of non-exchangeability between the observed data and the prior data and offers a standard way to calculate how many observations are needed to achieve an arbitrary (parameterized) definition of an "informative" estimator in a single iteration of updating, or the number of updating iterations needed given a fixed observation size n at each iteration. Though I focus on non-exchangeability between observed data and prior data, the same type of analysis may be applied to any non-exchangeability between any data.

Relevant Theoretical Questions

Observed human behaviors that rely on human judgment, including consumption and investment decisions, cooperative behavior, jury trial results, voting behavior, psychological diagnoses, and behavioral responses to risk, among others, are often the source of social data. Kahneman explains that the inputs to human behavior combine essentially autonomic prior beliefs, which may include psychological biases, but which have the convenience of leading to speedy decisions since they do not require much cognitive processing time, with information processing and logic (2011). Social priors are, therefore, non-exchangeable with social data almost by definition. The social priors factor in to the data-generating process itself.

We can think of individuals' prior beliefs as being drawn from a social distribution of prior beliefs, with some mean, and many small, additive sources of individual variation. It is important to note that the researcher is, too, a sociological product of the same society as the data-generating members of society. That is, the same source of prior information informs both the data and the researcher's priors.

We commonly note that our observations ought to be exchangeable with each other, but there is no mathematical reason that our prior observations should be exempt from this standard. Prior observations are treated mathematically as data (whether or not that data is quantitative in its raw form), just as are new observations. From a mathematical perspective, it is not clear which distribution is the prior and which is the "data." Thus, permuting the order in which we observe the prior observations and the new observations should not effect our results. This implies the new data should not be influenced by our prior beliefs if we are to update accurately (or, at least, the updating process should take into account the non-exchangeability). However, we rarely attempt to explicitly justify the implicit assumption that our prior and new observations are exchangeable with each other. The problem of ignoring non-exchangeability between prior and new data is especially worrisome when prior data are both biased and non-exchangeable with the new data.

Typically Bayesians assuage fears over biased priors and non-exchangeability by noting that if we allow our priors to update with each iteration of newly observed data and if the newly observed data contain at least some component of independent information about the world, eventually (after repeated iterations of updating) the posterior distribution will still become highly informative. However, two practical problems exist with this logic. First, we often must make an inference after a single round of updating. For practical purposes, we

still need to know, given the non-exchangeability and potential for bias in our priors, what size n is needed in this single iteration to produce an informative estimate? Furthermore, we may wish to know, given a certain number of new observations on each iteration, how many iterations of new data and updating would be required before the estimate becomes highly informative? Both have clearer practical implications than the claim that if we gathered information infinitely, in the limit we would definitely have a good answer.

New Data Non-exchangeable with Biased Prior

First Period Prior

The central limit theorem implies that if the distribution of social beliefs is made up of a group mean and the sums of many small individual factors the distribution may be adequately characterized by $\mathcal{N}(\alpha, \tau)$.

Unobserved Data-Generating Process

Suppose the socially-influenced data-generating process is a linear combination of our prior beliefs, $\alpha + \nu_i$, and some independent contribution, μ plus noise, z_i. That is,

$$x_i = \delta(\alpha + \nu_i) + (1 - \delta)(\mu + z_i),$$

where

$$Z \sim \mathcal{N}(0, \sigma),$$

and where,

$$\nu \sim \mathcal{N}(0, \tau),$$

and where $\delta \in (0, 1)$ and both ν and Z are independently and identically distributed. Note that by design there is no information about the world contained in our pre-updated (first period) prior beliefs, α. We are interested in uncovering μ, the independent systematic contribution of the world, but our observed data are influenced by both this independent component and our prior beliefs. Based on the data-generating process above, we can think of $X = (x_1, ..., x_n)$, given a certain value of the parameter, μ, as the weighted sum of two

independent normal distributions.

$$P(X|\mu) \sim \mathcal{N}\left(\delta\alpha + (1-\delta)\mu, \sqrt{\delta^2\tau^2 + (1-\delta)^2\sigma^2}\right) \tag{1}$$

If we were aware of the true data-generating process we could model the likelihood function in this way to account for the fact that our observed data, X, are in part informed by our prior. This would allow us to back out the unique and systematic contribution of the world (μ) to our observed data. That is, we could manufacture a model in which the exchangeability assumption would hold and the only bias in our posterior would come from the biased prior. In fact, in discussions of biased priors, this type of exchangeability is commonly assumed, and so leads to different and perhaps less grave implications regarding biased priors than if we allow for the possibility of mis-modeled non-exchangeability.

Naive Likelihood

When we overlook the question of exchangeability between our new data and our prior beliefs and we do not use the backout method, we allow the bias in our prior beliefs to carry over into our posterior via both the prior and the data (the degree to which it carries over through the data in one period depends on δ). Unaware of the exchangeability violation, we construct the naive likelihood of the data as a normal distribution with mean, μ. For now, for mathematical ease, we will assume that the standard deviation is known to be the quantity given by $\sqrt{\delta^2\tau^2 + (1-\delta)^2\sigma^2}$. Therefore, the likelihood function is proportional to

$$P(X|\mu) \sim \mathcal{N}(\mu, \sqrt{\delta^2\tau^2 + (1-\delta)^2\sigma^2}).$$

Posterior Belief about μ After Observing the Data

Using this naive likelihood function, we can construct our posterior exploiting normal-normal conjugacy. Our posterior belief regarding the updated probability distribution of μ is proportional to the joint likelihood of all the observed data, conditional on μ, multiplied by our prior belief regarding the probability of μ.

$$\begin{aligned} P(\mu|X) &\propto P(X|\mu)P(\mu) \\ &\propto N(\alpha_1, \tau_1) \end{aligned}$$

$$\alpha_1 = \frac{\left(\frac{\alpha}{\tau^2} + \frac{n\bar{X}}{\delta^2\tau^2 + (1-\delta)^2\sigma^2}\right)}{\left(\frac{1}{\tau^2} + \frac{n}{\delta^2\tau^2 + (1-\delta)^2\sigma^2}\right)} \tag{2}$$

$$\tau_1 = \frac{1}{\sqrt{\frac{1}{\tau^2} + \frac{n}{\delta^2\tau^2 + (1-\delta)^2\sigma^2}}} \tag{3}$$

Note that $E[\bar{X}] = \delta\alpha + (1-\delta)\mu$ and that $\bar{X}$ is also the (naive) frequentist maximum likelihood estimate.

Quantifying the Bias in Our Point Estimates

Both our ML estimate of the parameter value and our Bayes estimate will be biased due to the fact that the data are influenced by human psychological biases but we are not modeling them as such. We can quantify the bias produced by these two estimates in one period as the absolute distance between the expectation of our estimator and μ, the part of the data that is neither informed by noise nor by systematic psychological bias.

The size of the bias in our MLE may be expressed as, $\left|E[\bar{X}] - \mu\right|$. The expectation of our frequentist maximum likelihood estimator is $E[\bar{X}] = \delta\alpha + (1-\delta)\mu$. Expanding and re-arranging terms, we get

$$\begin{aligned} E[\bar{X}] - \mu &= [\delta\alpha + (1-\delta)\mu] - \mu \\ &= \delta\alpha + \mu - \delta\mu - \mu \\ &= \delta(\alpha - \mu). \end{aligned}$$

Similarly, we may characterize the bias in our naive Bayes estimate as $\left|\alpha k + E[\bar{X}](1-k) - \mu\right|$, where $k = \left(\frac{1}{\tau^2}\right) / \left(\frac{1}{\tau^2} + \frac{n}{\delta^2\tau^2 + (1-\delta)^2\sigma^2}\right)$. Expanding and re-arranging terms, we get

$$\begin{aligned}
\alpha k + E[\bar{X}](1-k) - \mu &= \alpha k + [\delta\alpha + (1-\delta)\mu](1-k) - \mu \\
&= \alpha k + \delta\alpha - \delta\alpha k + (1-\delta)\mu - (1-\delta)\mu k - \mu \\
&= \alpha(\delta + k - \delta k) + \mu - \mu(\delta + k - \delta k) - \mu \\
&= [\delta + k(1-\delta)](\alpha - \mu).
\end{aligned}$$

Since we know that δ lies in the interval $(0, 1)$ and that k lies in the interval $(0, 1)$ we can conclude that $[\delta + k(1-\delta)] > \delta$. That is, the size of the bias in the Bayes estimator (after a single iteration of updating) is larger than the size of the bias in the ML estimator after one time period.

Iterative Updating

We can recycle the notation above when thinking about the second, or any subsequent, iteration by making one change and indexing α and α_1 by their iteration number, $\alpha^{\{t\}}$ and $\alpha_1^{\{t\}}$, where t denotes the time period. Since we are updating our prior beliefs with each iteration, in $t = 2$ we will have $\alpha^{\{2\}} = \alpha_1^{\{1\}}$. That is, our posterior mean from $t = 1$ becomes our new prior mean for $t = 2$, and so on. Note that this assumes social learning is occurring (the social prior distribution is being updated with each iteration). This is a saccharinely optimistic assumption and, importantly, the question of whether the researcher is a Bayesian or not has little-to-no bearing upon it—I will return to this issue in the final discussion. Then, the parameterization of the posterior distribution defined through Equations 2 and 3 also holds for the posterior in periods $t > 1$ (though the definition of α evolves with each iteration).

We can see in Equation 2 that α_1 is the inverse-variance-weighted mean of α and the data. This implies that the size of the posterior bias, $|\alpha_1^{\{t\}} - \mu|$ is smaller than the size of the prior bias, $|\alpha^{\{t\}} - \mu|$, as we would expect with Bayesian updating since there is some independent informational content in the data. In fact, we can get a sense for the rate at which the bias decays over one time period (for a given n and δ). The proportion of the prior bias that remains in the posterior is given by

$$\frac{\alpha_1^{\{t\}} - \mu}{\alpha^{\{t\}} - \mu}.$$

Thus, the rate of decay in the bias over one period is

$$1 - \frac{\alpha_1^{\{t\}} - \mu}{\alpha^{\{t\}} - \mu}.$$

We know from above that $\alpha_1^{\{t\}} - \mu = [\delta + k(1-\delta)]\,(\alpha^{\{t\}} - \mu)$. Therefore, the bias decay rate for one iteration is

$$1 - [\delta + k(1-\delta)]\,.$$

This bias decay rate is a function (either implicitly, explicitly or both) of δ, n, τ and σ.

Power Analysis Analog

Single Iteration Inference

We often must make an inference after a single round of updating unless the inquiry is purely theoretical or the data simulated. If biased social priors partly inform our data, but we use the naive likelihood function, is there an n sufficiently large to come up with a reasonably good estimate in a single iteration of updating? If so, what size n is needed in this single iteration to produce an informative estimate?

By quantifying the bias in our Bayes and ML estimates after one iteration, we know that when the priors are biased, the Bayesian approach performs more poorly than the frequentist strategy (equivalent to assuming no prior knowledge). It is also clear, however, that in the limit, as $n \to \infty$, the difference between the two methods goes to zero. As such, the implications of the choice of whether to be a Bayesian researcher or a frequentist researcher become trivial in the presence of large data.

As shown in the previous section, we can characterize the bias in our Bayes estimate as $[\delta + k(1-\delta)](\alpha - \mu)$. The question we would like to ask is, in a single iteration, what size n is needed to make the absolute bias drop below an arbitrary acceptable cutoff point, ϵ? That is, we'd like to know what size n gives us

$$[\delta + k(1-\delta)]\,|\alpha - \mu| \leq \epsilon.$$

Substituting the equation for k and solving for n, we get the following.

$$
\begin{aligned}
\left[\frac{\delta\left(\frac{1}{\tau^2}+\frac{n}{\delta^2\tau^2+(1-\delta)^2\sigma^2}\right)+\frac{1-\delta}{\tau^2}}{\frac{1}{\tau^2}+\frac{n}{\delta^2\tau^2+(1-\delta)^2\sigma^2}}\right]|\alpha-\mu| &\leq \epsilon \\
\frac{\frac{\delta}{\tau^2}+\frac{\delta n}{\delta^2\tau^2+(1-\delta)^2\sigma^2}+\frac{1-\delta}{\tau^2}}{\frac{1}{\tau^2}+\frac{n}{\delta^2\tau^2+(1-\delta)^2\sigma^2}} &\leq \frac{\epsilon}{|\alpha-\mu|} \\
\frac{\delta^2\tau^2+(1-\delta)^2\sigma^2+\tau^2\delta n}{\tau^2(\delta^2\tau^2+(1-\delta)^2\sigma^2)} &\leq \frac{\epsilon}{|\alpha-\mu|}\left(\frac{\delta^2\tau^2+(1-\delta)^2\sigma^2+\tau^2 n}{\tau^2(\delta^2\tau^2+(1-\delta)^2\sigma^2)}\right) \\
\tau^2\delta n-\left(\frac{\epsilon}{|\alpha-\mu|}\right)\tau^2 n &\leq \left(\frac{\epsilon}{|\alpha-\mu|}\right)\left(\delta^2\tau^2+(1-\delta)^2\sigma^2\right)-\left(\delta^2\tau^2+(1-\delta)^2\sigma^2\right) \\
\tau^2\delta n\,|\alpha-\mu|-\epsilon\tau^2 n &\leq \epsilon\left[\delta^2\tau^2+(1-\delta)^2\sigma^2\right]-|\alpha-\mu|\left[\delta^2\tau^2+(1-\delta)^2\sigma^2\right] \\
n\tau^2\left(\delta\,|\alpha-\mu|-\epsilon\right) &\leq \left[\delta^2\tau^2+(1-\delta)^2\sigma^2\right](\epsilon-|\alpha-\mu|)
\end{aligned}
$$

At this point, we would like to isolate n by dividing both sides of the inequality by $\tau^2\left(\delta\,|\alpha-\mu|-\epsilon\right)$. However, we do not know whether the quantity $\delta\,|\alpha-\mu|-\epsilon$ is greater than or less than 0 and this will determine the direction of the inequality. Therefore, I separate the results into the two possible scenarios.

First, suppose $\delta\,|\alpha-\mu|<\epsilon$. Then,

$$n \geq \frac{[\delta^2\tau^2+(1-\delta)^2\sigma^2](\epsilon-|\alpha-\mu|)}{\tau^2(\delta|\alpha-\mu|-\epsilon)}.$$

We know that $\delta\,|\alpha-\mu|-\epsilon<0$, so the denominator must be negative. This scenario further subdivides into two cases, depending on whether $\epsilon-|\alpha-\mu|$ is less than zero.

(a) If $\epsilon-|\alpha-\mu|>0$, then the entire expression to the right of the inequality is negative (and is zero if $\epsilon-|\alpha-\mu|=0$). The conclusion, then, is that if $\epsilon\geq|\alpha-\mu|$, n needs only to be larger or equal to some negative number in order to get an estimate with bias within the tolerable range. This makes sense, as long as we recognize that n cannot be smaller than zero, so for this range, $n=0$. No data is needed to update the prior because the Bayes' estimate from the prior (even without updating using data) is α, and the bias in α by definition falls within the tolerable range.

(b) In the more interesting scenario, $\epsilon-|\alpha-\mu|<0$. In this case, $\epsilon-|\alpha-\mu|<0$. Then, both the numerator and the denominator of the expression to the right of the inequality are negative, so the whole expression is a positive number and n must be at least that great to get the bias in the (posterior) Bayes' estimate to be within the tolerable range

after one iteration.

Now, suppose $\delta\,|\alpha-\mu| \geq \epsilon$. This results in the following inequality for n,

$$n \leq \frac{[\delta^2\tau^2 + (1-\delta)^2\sigma^2](\epsilon - |\alpha-\mu|)}{\tau^2(\delta|\alpha-\mu| - \epsilon)}.$$

Since $\delta\,|\alpha-\mu| \geq \epsilon$, the denominator is positive. Note that since $\epsilon - \delta|\alpha-\mu| \leq 0$ and δ is in the interval $(0,1)$, we may conclude that $\epsilon - |\alpha-\mu| \leq 0$, as well. Therefore, the numerator is negative and the whole expression to the right of the inequality evaluates to a negative number. The conclusion is that n must be less than or equal to some negative number, which is nonsensical. This results because, in this range, there is no size n that is sufficient to get the (posterior) Bayes' estimate within the tolerable range of bias.

Therefore, we have identified three plausible ranges for ϵ. When $\delta|\alpha-\mu| \geq \epsilon$, no size n is sufficient to produce a Bayes' estimate that falls within the tolerable range in a single iteration, given that we have overlooked the exchangeability violation. While it is still true that in the limit, infinite iterations of data observation and updating would deliver a good estimate, in this case it is simply not possible to collect enough data to overwhelm the nonexchangeability problem in a single iteration. When $\epsilon \geq |\alpha-\mu|$, any size n is sufficient. And when $\delta|\alpha-\mu| < \epsilon < |\alpha-\mu|$, we may calculate the n that will deliver a posterior Bayes' estimate within the tolerable range of bias, using the inequality, $n \geq \frac{[\delta^2\tau^2 + (1-\delta)^2\sigma^2](\epsilon - |\alpha-\mu|)}{\tau^2(\delta|\alpha-\mu| - \epsilon)}$. Recall, however, that when we can only count on a single iteration, we still perform better with respect to the point estimate using a frequentist analysis.

Find number of iterations, t, given n, to get within ϵ of truth

Using the general bias equation found above, we can say that the size of the bias after a single iteration of updating occurs is

$$[\delta + k(1-\delta)]\,(\alpha^{\{1\}} - \mu).$$

Furthermore, we know that any iteration, t, reduces the bias from the previous iteration's posterior mean by

$$\frac{\alpha_1^{\{t\}} - \mu}{\alpha^{\{t\}} - \mu}.$$

We can simplify this expression, given the definition of $\alpha_1^{\{t\}}$ as follows.

$$\begin{aligned}
\frac{\left(\frac{\frac{\alpha^{\{t\}}}{\tau^2}+\frac{n\bar{X}}{\delta^2\tau^2+(1-\delta)^2\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\delta^2\tau^2+(1-\delta)^2\sigma^2}}\right)-\mu}{\alpha^{\{t\}}-\mu} &= \frac{\frac{\alpha^{\{t\}}-\mu}{\tau^2}+\frac{n(\bar{X}-\mu)}{\delta^2\tau^2+(1-\delta)^2\sigma^2}}{\frac{\alpha^{\{t\}}-\mu}{\tau^2}+\frac{n(\alpha^{\{t\}}-\mu)}{\delta^2\tau^2+(1-\delta)^2\sigma^2}} \\
&= \frac{(\alpha^{\{t\}}-\mu)[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n(\bar{X}-\mu)}{(\alpha^{\{t\}}-\mu)[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n(\alpha^{\{t\}}-\mu)}
\end{aligned}$$

At this point, since $\bar{X}$ has a sampling distribution, I elect to focus the analysis on the *expected* number of iterations required. Although, if one were more risk averse, we could easily substitute a different number, say $E(\bar{X}) \pm 1.96\ sd(\bar{X})$, instead of the expectation of $\bar{X}$ in order to calculate the number of iterations required to diminish the bias to some other arbitrarily defined level. Substituting the expectation for $\bar{X}$, we then get

$$\begin{aligned}
& \frac{(\alpha^{\{t\}}-\mu)[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n[\delta\alpha^{\{t\}}+(1-\delta)\mu-\mu]}{(\alpha^{\{t\}}-\mu)[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n(\alpha^{\{t\}}-\mu)} \\
= & \frac{(\alpha^{\{t\}}-\mu)[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n\delta(\alpha^{\{t\}}-\mu)}{(\alpha^{\{t\}}-\mu)[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n(\alpha^{\{t\}}-\mu)} \\
= & \frac{[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n\delta}{[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n}.
\end{aligned}$$

Conveniently, the bias ratio is not a function of α and, therefore, is fixed across iterations. This makes it mathematically easy to compound over time. The remaining bias in the Bayes' estimate is, thus, given by the following equation.

$$[\delta+k(1-\delta)](\alpha^{\{1\}}-\mu)\left(\frac{[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n\delta}{[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n}\right)^{(t-1)}$$

We can now use this equation to calculate the number of iterations, given a fixed sample size, n, across iterations, that are necessary to reduce the bias in our point estimate below some arbitrary quantity, ϵ. This occurs when the following condition is satisfied.

$$[\delta+k(1-\delta)](\alpha^{\{1\}}-\mu)\left(\frac{[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n\delta}{[\delta^2\tau^2+(1-\delta)^2\sigma^2]+\tau^2 n}\right)^{(t-1)} \leq \epsilon$$

Therefore, we have the following inequality for t.[1]

$$t \geq 1 + \log_{\frac{\delta^2\tau^2+(1-\delta)^2\sigma^2+\tau^2 n\delta}{\delta^2\tau^2+(1-\delta)^2\sigma^2+\tau^2 n}} \frac{\epsilon}{[\delta + k(1-\delta)]|\alpha^{\{1\}} - \mu|}$$

This implies that it is always possible to find some $t < \infty$ sufficient to produce an informative posterior, as would suggest our Bayesian intuition. When $\epsilon > |\alpha^{\{1\}} - \mu|$, then the equation returns some $t \leq 1$, which implies no updating is necessary beyond the first iteration to obtain a point estimate within the tolerable range.

Final Discussion

This exploration has generated practical guidelines, analogous to power analyses, for situations in which it is plausible that socio-psychological biases inform both the behaviors of the humans generating behavioral data and the prior beliefs of the researcher analyzing those data. This experiment helps to make salient several relevant methodological and philosophical points regarding Bayesian learning. This final discussion will summarize and flesh out several key findings of this research.

First, it is important to consider the plausibility of the exchangeability assumption with respect to prior data—even if that data is non-quantitative in its raw form. Though most of us are in the practice of considering questions of exchangeability with respect to our newly observed data, the question of whether the prior data observations are exchangeable (among themselves as well as with the newly observed data) is not regularly considered, at least not in practice. The tendency to forgo an explicit justification of the exchangeability assumption with respect to priors is especially common when the prior distribution is informed by non-quantitative data or beliefs. The tendency to overlook the importance of exchangeability between new and prior data is therefore, perhaps a symptom of a more general methodological pathology—the tendency not to think of our prior "data" as data, though they are mathematically indistinguishable in the way they enter into Bayes' rule. Prior data is data and therefore exchangeability need apply. This article highlights the potential perils of this type of oversight and deals with practical solutions. I find that sometimes it is impossible to collect enough data to resolve the problem in one iteration if we fail to recognize the non-exchangeability of the data. For cases where it is possible to gather sufficient data to get a good estimate (with respect to managing bias) in a single iteration, I offer a formula

[1]Note that the direction of the inequality switches because the base of the logarithm is necessarily a number in the interval $(0, 1)$. Given that the base is always a positive fraction less than one, the log in this case is a strictly decreasing function.

for calculating the necessary data size.

Unfortunately, such oversights are likely to be fairly common in the social sciences where many of our data are human responses to socialized perceptions. It is therefore worth identifying contexts in which this situation is especially likely to arise. Certain contexts, in particular, informational contexts, may make our data-generating actors more or less prone to relying on prior beliefs to inform their behaviors or decisions. When a behavior requires a judgment to be made with relatively little information about the subject (such as in a jury trial or a psychological diagnosis), individuals will automatically tend to use the little bit of information they (feel they) have to fill in the blanks about the subject in a way that is consistent with the details they know, as they understand them (Zaller 1992). Low levels of information are important not just because they can lead to reliance on psychological biases to inform the "blanks," but also because they can result in unreasonable levels of certainty about these conclusions. Kahneman says "you cannot help dealing with the limited information you have as if it were all there is to know...Paradoxically, it is easier to construct a coherent story when you know little" (2011). When we receive little information from the world about a subject, we may find our prior beliefs constitute most of the story (since we have used them to fill in the blanks). Therefore, δ will likely be larger when information to supplement prior beliefs is limited. On the other end of the spectrum, it also seems that information overload can lead a person to give up on sorting out the facts, instead reverting to autonomic shortcuts including psychological biases—as may be true, for example, of voting behavior in elections with a high number of parties or competitors (Cunow 2014).

Second, rather than rely on limiting assumptions, we can calculate how many observations or iterations would be needed to make a reasonable inference, even assuming unfavorable prior exchangeability conditions. The justification we normally use to feel comfortable with potentially-biased Bayesian priors is that as we observe more and more new (exchangeable) data, the weight of our biased prior diminishes. In the limit, the original prior from the first time period has no influence. Thus, even if our prior were completely misinformed, there is no aggregate or long-run bias in our methodology. However, when the data themselves are informed in part by our biased priors (and are therefore non-exchangeable), the medium-run implications of our methodology become quite important. Even if in the limit we can expect reasonable posterior estimates, this gives us little notion of how close our immediate reality comes to the limit and, for practical purposes, tells us little about the quality of our inferences. We should have some notion of what number of observations or updating iterations are needed to arrive at a reasonably informative estimate. This paper provides a method to

assist the researcher in making this judgment.

Third, the most compelling argument for prescriptive use of Bayesian methodology in academia seems to come from the claim that in spite of any prior bias, the process of iterative updating will eventually lead to highly informative estimates and it is therefore a natural framework for thinking about knowledge accumulation. Unfortunately, when social priors inform the data, the fact that the researcher identifies as a Bayesian is no guarantee that social learning will occur. That is, the researcher cannot guarantee that society is Bayesian and cannot guarantee diffusion of results. Without this assumption, in the presence of socially-biased behavioral data, the researcher cannot rely on these limiting claims about Bayesian updating. If the researcher alone updates prior beliefs, but the social prior does not update concurrently, the biased social priors will continue to effect the data at the same rate across iterations. Researcher learning will still occur, but rather than limiting toward μ, the independent contribution of the world, the researcher's posterior Bayes estimate will limit toward $E[\bar{X}] = \delta\alpha^{\{1\}} + (1-\delta)\mu$, the naive maximum likelihood estimate.

This implies that biased social beliefs that affect social data provide an identical challenge for both the Bayesian and the frequentist researcher. Regardless of the researcher's philosophical approach to empirics, both the frequentist and Bayesian researcher live in the same world with respect to social learning. In the absence of social learning, the quality of the Bayesian researcher's point estimates are no less model dependent than are the frequentist's. This is consistent with the oft-noted fact that Bayes' rule is a nested function of the likelihood function. Furthermore, this paper shows that given biased prior beliefs when Bayesian social learning does occur, the ML estimate is actually less biased than the Bayes estimate in any observation time period, t (though as data accumulate the size of this difference goes to zero). Therefore, when considering the expected methodological advantages of Bayesian updating, it is important to acknowledge that any researcher, Bayesian or not, is subject to the same questions of model specification and the realities of social learning. When the data are informed in part by biased social priors and social learning does not occur, these conditions commute the ostensible limiting benefits of updating.

Thus, the question of whether social learning (updating social priors in light of new information) occurs in a Bayesian fashion becomes very important to our understanding of Bayesian methodology in the social sciences, or where ever social data are concerned. This paper has assumed the social prior distribution is independently and identically distributed about its mean, which is an assumption unlikely to be met in a social world. In reality if one

individual (say, a researcher) is drawn from that social distribution and then undergoes an updating of their prior beliefs and is reasserted into the social distribution post-update, the act of reasserting is likely to have some sort of social spillover effect; some social diffusion may occur. However, updated individuals exposed to social interaction are also likely to suffer from social contagion in the other direction, and so it is not clear what the net effects would be. Thus, understanding the dynamics of social learning has important implications for our ability to engage in academic inference.