

## Supplemental Material

Hawkins, G. E., Hayes, B. K., Donkin, C., Pasqualino, M., & Newell, B. R. (in press). A

Bayesian latent mixture model analysis shows that informative samples reduce base rate neglect. *Decision*.

### Section A: A latent mixture model for base rate neglect

In this section we provide a detailed explanation of the latent mixture model we used in our data analysis. Our Bayesian model assumes there are two populations of respondents in the mammogram problem – *low* or *high estimators*. Low estimators will generally be closer to the normative solution for the mammogram problem. Hence, the proportion of low-estimate responders in each experimental condition is the primary dependent measure. Figure S1 depicts our Bayesian model using standard graphical model notation, where latent and observed variables are represented with open and shaded nodes, respectively, and circular and square nodes represent continuous and discrete variables, respectively. Rectangular plates represent independent replications over participants within a condition ( $n = 25$  per cell for the description and posterior conditions,  $n = 50$  per cell in the prior conditions) and over conditions within the experiment ( $m = 5$ ).

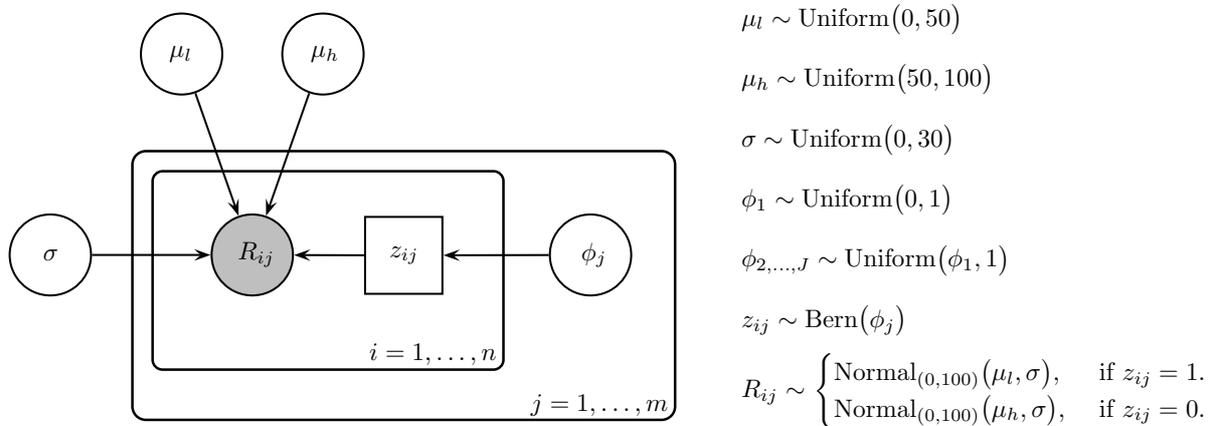


Figure S1. The Bayesian graphical model used to analyze probability estimates in the experiment.

The model assumes that response  $R_{ij}$  of participant  $i$  in condition  $j$  comes from one of two normal distributions (truncated at 0 and 100, to respect the response scale given to participants). The means of the normal distributions are  $\mu_l$  and  $\mu_h$ , the population mean of the low-estimate and high-estimate distributions, respectively. The standard deviation of both normal distributions is  $\sigma$ .<sup>1</sup> Whether  $R_{ij}$  is sampled from  $\mu_l$  or  $\mu_h$  is determined by  $z_{ij}$ , a Bernoulli trial that takes the value of 0 or 1 on the basis of  $\phi_j$ , the probability of an estimate arising from the distribution of low estimators in condition  $j$  (the probability of an estimate arising from the distribution of high estimators is  $1-\phi_j$ , by definition). All prior settings on parameters were uninformative subject to one constraint, as given on the right of Figure S1. The prior distribution on the description condition ( $j = 1$ ) was uniform on the rate scale, and our analysis focused on one-sided tests that examined the evidence that the experimental conditions ( $j = 2-5$ ) had a greater proportion of responses from the low estimator distribution than the description condition. Our prior assumption here is equivalent to examining the extent to which performance was improved in the experimental conditions, and not entertaining the hypothesis that they might hurt performance. We implemented the graphical model using Markov chain Monte Carlo (MCMC) methods in the R statistical programming environment (R Development Core Team, 2014) and Just Another Gibbs Sampler (JAGS, Plummer, 2003), using the *R2jags* package (Su & Yajima, 2014). We took 21,000 samples from the posterior distributions of the parameters from each of three chains with a burn-in period of 1,000 samples, for a total of 60,000 samples from the posterior distributions of the parameters.

We can examine the appropriateness of the Bayesian mixture model for our data with the posterior predictive distribution. Posterior predictives tell us what data to expect from the model once it has been updated by the observed data, shown in the lower row of Figure 2

---

<sup>1</sup> We also investigated a model that estimated a separate value of  $\sigma$  for each population distribution. The posterior distributions for  $\sigma_l$  and  $\sigma_h$  were similar so we simplified the model to assume a common  $\sigma$ . This has the added benefit of more heavily constraining the posterior distributions of other parameters in the model.

from the main text. The distribution follows the trends observed in data (upper row of Figure 2): a good fit to the median, inter-quartile range, and distribution shape. The histogram in the lower right panel of Figure 2 shows the posterior predictive distribution of probability estimates aggregated over conditions, and follows all major trends in the corresponding data histogram.

### **Parameter Estimates**

Posterior distributions of parameters give the feasibility of values of model parameters – the proportion of low estimators in each condition – after having observed the data. We summarize the posterior distributions with the 95% highest density interval (HDI; Kruschke, 2011) – the smallest interval required to contain 95% of the posterior density for a parameter. The 95% HDI for the population mean of the high estimators was 72.7–80.1%, which is broadly consistent with high estimators neglecting the prior probability and heavily weighting the likelihood statistic in the mammogram problem (i.e., the ‘hit rate’,  $p(M|C) = .8$ ) or taking the difference between the two likelihood statistics (hit rate and false alarm rate,  $p(M|C) - p(M|-C) = .8 - .15 = .65$ ). The 95% HDI for the population mean of the low estimators was 0–3.1%, which contains the prior probability of breast cancer ( $p(C) = .01$ ), but not the normative solution ( $p(C|M) \approx .051$ ), suggesting that few participants provided the correct (normative) solution to the task. This was the case in data: only 10% of participants provided an approximately normatively correct response between 4–6%. The 95% HDI for the common standard deviation of the low estimator and high estimator distributions was 11.7–15.1%.

## Section B: Procedure for Bayesian hypothesis tests

In this section we provide a detailed explanation of the Savage-Dickey density ratio test used in the main text. Our hypothesis tests considered the difference between two posterior distributions of  $\phi$ , denoted here as  $\delta$ . For example, Figure 4 from the main text shows the posterior distribution of  $\delta$  for the comparison between the description and posterior sampling conditions as the heavy solid black line, and the prior distribution on this comparison is represented with a thin solid black line. This posterior density is generally greater than 0, indicating that  $\phi$  was generally larger in the posterior sampling condition than the description condition (i.e., larger proportion of low estimators). This is also seen in the upper left panels of Figure 3 in the main text.

Our primary analysis in the experiment used order constrained, one-sided hypothesis tests to compare the level of improvement in the posterior and prior conditions and the sampling and sample summary conditions relative to the description condition. We explicitly used one-sided hypothesis tests because (a) previous research has given no reason to expect that any of our experimental manipulations would hinder performance relative to the baseline description condition, and (b) our aim was to investigate how we can improve performance in base rate neglect tasks, which is a directional test. The Savage-Dickey density ratio uses the prior and posterior distributions of  $\delta$  to compare the two models that correspond to classical one-tailed hypothesis tests: the null hypothesis that there is no difference between the posterior distributions of  $\phi$  for an experimental condition and the description condition,  $H_0 : \delta = 0$ , and the alternative hypothesis of a positive difference between the two posterior distributions,  $H_1 : \delta > 0$ .

We placed an uninformative prior distribution on the unit interval for  $\phi$  in the description condition. Though prior work suggests that we should expect relatively small values of  $\phi$ , since people typically perform poorly in this task, we used a uniform prior to

ensure that our conclusions are as conservative as possible. The prior distribution on the four experimental conditions ensured that  $\phi$  was larger than the description condition, consistent with our one-sided hypothesis tests. The prior distribution for the difference between the experimental conditions and the description condition,  $p(\delta)$ , gives the prior distribution for  $H_I$ , which we obtained through sampling in our model-based analysis, shown as the thin solid black line in Figure 4. This prior distribution has zero density when  $\delta < 0$ , peaked at  $\delta = 0$ , and declines following a convex function tapering to zero density at  $\delta = 1$ .

Our second set of comparisons examined the effect of information format – prior sampling versus prior sample summary and posterior sampling versus posterior sample summary – with two-sided tests to allow for the possibility of increased or decreased performance in the sampling compared to sample summary conditions. We implemented a two-sided test to reflect our lack of prior knowledge for this comparison. The prior distribution for the two-sided hypothesis test is the difference between two of the prior distributions shown in Figure 4 (thin solid black lines), which is a peaked function around the point of no difference in the proportion of low estimators between conditions ( $\delta = 0$ ) that symmetrically tapers to zero density at -1 and 1. For the two-sided tests the null hypothesis states there is no difference between the posterior distributions of two conditions,  $H_0 : \delta = 0$ , and the alternative hypothesis states there is a difference between the two posterior distributions,  $H_I : \delta \neq 0$ .

The Savage-Dickey density ratio is given as the ratio of the density of the posterior to the prior distributions at the point value of relevance to the null hypothesis (i.e.,  $\delta = 0$ , for both one- and two-sided hypothesis tests). The point values of interest for the one-sided tests can be seen in Figure 4 as the density (i.e., value on the y-axis) where each of the difference distributions falls at  $\delta = 0$  (i.e., zero-point on the x-axis). For example, the Bayes factor for the posterior sampling condition versus the description condition is approximately  $.19/7.3 \approx$

.026. This gives the relative odds that the data were generated by the null hypothesis relative to the alternative hypothesis.

This ratio is a Bayes factor indicating support for the null versus alternative hypothesis, where  $BF_{01} > 1$  indicates support for  $H_0$  and  $BF_{01} < 1$  supports  $H_1$ . Following Lee & Wagenmakers (2014), we used the logspline non-parametric density estimator from the *polspline* package in R (Kooperberg, 2013) for plotting of all posterior distributions and obtaining prior and posterior density estimates for hypothesis testing.

## References

- Kooperberg, C. (2013). *polspline: Polynomial spline routines* [Computer software manual]. Available from <http://CRAN.R-project.org/package=polspline> (R package version 1.1.9).
- Kruschke, J. K. (2011). *Doing Bayesian data analysis*. Academic Press.
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press: New York.
- Plummer, M. (2003). JAGS: A program for the analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.
- R Development Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria. (ISBN 3-900051-07-0).
- Su, Y.-S., & Yajima, M. (2014). *R2jags: A package for running jags from R*. [Computer software manual]. Available from <http://CRAN.R-project.org/package=R2jags> (R package version 0.04-03).