Supplementary materials to: "Seeing is believing: priors, trust and base rate neglect."

Matthew B. Welsh
Australian School of Petroleum
University of Adelaide

Daniel J. Navarro
School of Psychology
University of Adelaide

This note contains the supplementary materials to:

Welsh, M.B. & Navarro, D.J. (in press). Seeing is believing: priors, trust and base rate neglect *Organizational Behavior and Human Decision Processes.* Accepted February 8th 2012.

Table of Contents

Details of Modeling

Constructing the Simple Estimators

In what follows, we describe how to construct the estimators used in the analysis of Experiment 1A and 1B. We focus in particular on the naive estimator (Equation 2 of the main paper), since the other estimators (Equations 3, 4 and 5) are minor modifications to this. To start with, note that (by definition) the expected value of θ is just the mean of the posterior distribution,

$$E[\theta \mid x_0, x_1, n_0, n_1] = \int_0^1 \theta P(\theta \mid x_0, x_1, n_0, n_1) d\theta$$

and if we treat the sample sizes n_0 and n_1 , as fixed (or, more precisely, as variables that are independent of θ) then by Bayes' rule we can write:

$$P(\theta \mid x_0, x_1, n_0, n_1) \propto P(x_0, x_1 \mid \theta, n_0, n_1) P(\theta)$$

where $P(\theta)$ denotes the prior degree of belief that the decision-maker has in θ, and

P (x_0, x_1|θ, n_0, n_1) is the probability that the two samples would have produced the data x_0 and x_1. Note that for equality (rather than proportionality) to hold, then the expression on the right hand side would be divided by the normalizing constant P (x_0, x_1|n_0, n_1). Under the naive model, we assume that both samples are independent, and moreover are assumed to be equally relevant to the problem of estimating θ, so this becomes:

$$P(\theta \mid x_0, x_1, n_0, n_1) \propto P(x_0 \mid \theta, n_0) P(x_1 \mid \theta, n_1) P(\theta)$$

If the observations in a particular sample are generated independently, with probability θ of possessing the relevant characteristic, then the probability that x of n items possess the characteristic is given by a binomial distribution, in which:

$$P(x \mid n_1, \theta) \propto \theta^x (1-\theta)^{n-x}$$

Under such circumstances, a standard prior to use in a Bayesian analysis is a beta distribution over θ, in which one effectively "pretends" to have observed u previous observations that possess the characteristic and v previous observations that do not.

$$P(\theta) \propto \theta^{u-1} (1-\theta)^{v-1}$$

Note that when $u = v = 1$ we obtain the uniform prior density $P(\theta) = 1$ (to understand why the uniform prior corresponds to the assumption of having seen one case of each possible outcome, see Jaynes 2003). However, we will assume that the learner has seen zero previous observations before the two data sets are collected (i.e., $u = v = 0$), since this produces a simpler estimator. Under these assumptions,

$$\begin{aligned} P(\theta \mid x_0, x_1, n_0, n_1) &\propto \theta^{x_0} (1-\theta)^{n_0 - x_0} \theta^{x_1} (1-\theta)^{n_1 - x_1} \theta^{-1} (1-\theta)^{-1} \\ &= \theta^{x_0 + x_1 - 1} (1-\theta)^{n_0 + n_1 - x_0 - x_1 - 1} \end{aligned}$$

which corresponds to the expression for a beta distribution with parameters $u = x_0 + x_1$ and $v = n_0 + n_1 - x_0 - x_1$. Thus, to construct our estimator, the decision maker should simply report the mean of the corresponding distribution. Since the mean of a beta distribution with parameters *u* and *v* is simply $u/(u+v)$ we obtain the estimator reported in Equation 2:

$$E[\theta \mid x_0, x_1, n_0, n_1] = \frac{x_0 + x_1}{n_0 + n_1}$$

To construct the "skeptical" estimator (i.e., Equation 3), we need to be a little more careful, since there are many different ways in which we could model the influence of factors like, age, location and source of the older sample. The simplest is to suppose that only some proportion t of the observations that make up the older data set is relevant to the *current* problem. If that is the case, then the old sample is effectively only of sample size $t \times n_0$, and the number of those observations that would meet the relevant criterion is reduced to $t \times x_0$. Having made those assumptions, it should be clear that the logic of the original naive estimator then carries through, and indeed this is how we produced all of the modified estimators (Equations 3, 4 and 5). It should be noted, of course, that this method is extremely crude: in truth we should expect the number of "actually relevant" observations in the prior sample to be a random variable rather than a fixed value, and indeed we might expect the effects in question to be much more complex than just "censoring" some proportion of the old data. Nevertheless, in view of the simplicity of the experimental design this model is sufficient for our purposes, and in our view it would serve no useful purpose to complicate the model further.

Details of the Individual Subjects Analysis

In this section we discuss the analysis used to look at the responses made by individual participants. As noted in the main text, in the individual subject analysis we estimated values of t_a, t_l and t_s for each participant, using a uniform prior on t and reporting the posterior mode (equivalent to maximum likelihood estimation). Thus, the scatterplots in Figure 7 plot the model prediction for $E[\theta]$ against the response x made by the participant, where each datum is shown in the "75% baserate, 25% new data" format. A simple descriptive measure of agreement is provided by the correlation coefficient between $E[\theta]$ and x (i.e., regression with intercept fixed at 0 and slope 1), but to determine whether the model fit is adequate, we need to be a little more careful. Since, as noted in the text, people have a tendency to "round to the nearest multiple of 5 or 10", a conservative approach would be to declare that when both x and $E[\theta]$ lie in the interval [50, 75], then the model has correctly predicted that the old sample (base rate) contributes more to the human judgment than the new sample (an "O+" datum). Similarly, if both x and $E[\theta]$ lie in the interval [25, 50] the model has predicted correctly that the new sample is more important (an "N+" datum). Conversely, if x lies in [75, 50] but $E[\theta]$ does not, then the model has failed to predict that the participant relies more on the old sample (an "O−" datum). Vice versa, if x is in [25, 50] and $E[\theta]$ is not, we have an N− datum. If x = 50, the participant has weighted the two samples equally (a "50" datum), which could be either model-consistent or model-inconsistent. Finally, if $x < 25$ or $x > 75$, the participant has extrapolated beyond the range of the two samples (an "EX" datum), and the model cannot predict the response. We can then count the frequency of each type of datum for each participant (see Table A1), and use this to quantitatively choose between one of seven different explanations, based on the relative frequency of N+, N−, O+ and O− (for simplicity, we treat 50 and EX data as ambiguous and so exclude those data for the present purposes). The seven accounts are as follows:

- NULL EFFECT. In the null effect case, we assume that all four events (N+, N−, O+ and O−) are equally likely with probability 1/4.

- RANDOM EFFECT. In the random effect case, we assume that all four events have different probabilities, but with no particular pattern (that is, we assume a uniform prior on the probabilities).

- NEGLECT. For the neglect account, we assume that a classic base rate neglect effect occurs, and moreover that this effect is inconsistent with the trust model. Neglect implies that the data are more likely to be N than O, and model inconsistency implies that they are no more likely to be + than –. In short, we assume P (N +) = P (N –) > P (O+) = P (O–).

- NAIVE. The logical complement of the neglect account is a naive-Bayesian who sides with the old sample in a manner that that is inconsistent with the trust model That is, P (O+) = P (O–) > P (N +) = P (N –).

- SKEPTIC. The skeptic account corresponds to a participant who sometimes sides with the new sample and sometimes sides with the old sample, but tends to do so in a manner that is consistent with the trust model. That is: P (N +) = P (O+) > P (N –) = P (O–).

- NEW. An important submodel to consider is the intersection of neglect and skeptic: namely a person who sides with the new sample most of the time, in a manner that is consistent with the trust model. In this case we assume P (N +) > P (N –) = P (O+) = P (O–).

- OLD. The last account to consider is the reverse of the last one, in which the participant generally gives responses consistent with the old sample, but also consistent with the model. That is: P (O+) > P (N +) = P (N –) = P (O–).

Formally, the null effect model assigns probability to an observed data set *X* as follows:

$$P(X|\text{null}) = 4^{-N} \tag{6}$$

The random effect model is a standard Dirichlet-multinomial model, in which the marginal likelihood is:

$$P(X|\text{random}) = \frac{n_1!\ n_2!\ n_3!\ n_4!}{(N+3)!} \tag{7}$$

where n_1, n_2, n_3 and n_4 denote the numbers of observations in each cell. For the neglect model, the naive model and the skeptic model, note that we have two “high probability” cells and two “low probability cells (so we denote them the “2-2” models). If N_h denotes the number of observations that fall in a high probability category and N_l is the number that have low probability, the marginal probability is:

$$P(X|2\text{-}2) = \int_{1/2}^{1} \left(\frac{\varphi}{2}\right)^{N_h} \left(\frac{1-\varphi}{2}\right)^{N_l} 2 \, d\varphi$$

$$= 2^{1-N} \int_{0}^{1/2} (1-u)^{N_h} u^{N_l} \, d\varphi$$

$$= 2^{1-N} \sum_{k=N_l+1}^{N+1} \frac{N_h! N_l!}{k!(N+1-k)!} \frac{1}{2^{N+1}}$$

$$= 4^{-N} \sum_{k=N_l+1}^{N+1} \frac{N_h! N_l!}{k!(N+1-k)!} \tag{8}$$

For the new and old models, the approach is similar, but in these cases there is one high probability cell and three low probability cells. For these "3-1" models,

$$P(X|3\text{-}1) = \int_{1/4}^{1} \varphi^{N_h} \left(\frac{1-\varphi}{3}\right)^{N_l} (4/3) \, d\varphi$$

$$= \frac{4}{3^{N_l+1}} \int_{0}^{3/4} (1-u)^{N_h} u^{N_l} \, du$$

$$= \frac{4}{3^{N_l+1}} \sum_{k=N_l+1}^{N+1} \frac{N_h! N_l!}{k!(N+1-k)!} \left(\frac{3}{4}\right)^{k} \left(\frac{1}{4}\right)^{N+1-k}$$

$$= 4^{-N} 3^{-N_l+1} \sum_{k=N_l+1}^{N+1} \frac{N_h! N_l! 3^k}{k!(N+1-k)!} \tag{9}$$

Applying Bayes' rule gives the posterior probabilities associated with the *i*th model,

$$P(M_i|X) = \frac{P(X|M_i)P(M_i)}{\sum_{j=1}^{7} P(X|M_j)P(M_j)} \tag{10}$$

where we assume $P(M_j) = 1/7$ for all j. (It should be noted that since the model fitting process attempts to move the data into the two + cells, and there are three free parameters that are fit in the process, an optimal selection procedure should correct for this by penalizing those accounts that allow N + or O+ to be higher probability – this is trivial to do via methods such as BIC, but such methods assume that the fit is optimized with respect to the model in question, which is not the case here: all seven of these models make use of the trust model estimates, which will tend to push data into the N + and O+ sectors, but are not optimized for that precise task. While it is quite possible to extend the analysis to do this sensibly, we are loathe to introduce even more statistical analysis into this paper).

When we apply this analysis to the data shown in Table A1, it turns out that there is an extremely clear winner for 12 of the 20 participants (posterior probability > .94 for one of the seven models - remembering that the prior probability of each model was ~0.14). For 7 of the remaining 8 participants, the best model is at least 1.76 (and up to 8) times as likely as the next best model. In only one case (participant 19) is there a near-tie, where the assigned model, 'Old', has a 0.39 probability and the next best model, 'Naïve', has a 0.36 probability

Overall, there are 9 participants for whom the "skeptic" explanation is best, with a further 4 given the "new" label and one labeled "old" (remembering, of course, that this last individual could almost as well be categorized as "naïve"). That is, one of the model-consistent explanations proves best in 14 cases. In 5 cases, we observe a model-inconsistent "naive" result and, finally, we observe one case of a "random" effect. Overall, there are clear but sensible individual differences, with the majority of participants behaving in accordance with the trust model, but a non-trivial minority of people behaving in a manner according with a naive application of Bayes Theorem.

Table S1

Data, preferred models and posterior probabilities - indicating how likely it is that this model provides the best explanation for the participant's data – of the 7 models considered here.

ID	N+	O+	O-	N-	50	EX	Best Model	Post. Prob.
3	12	9	0	3	8	0	Skeptic	0.94
8	16	14	0	2	0	0	Skeptic	0.99
10	12	8	2	1	7	2	Skeptic	0.94
11	15	15	1	0	1	0	Skeptic	0.99
17	16	13	0	3	0	0	Skeptic	0.97
18	15	12	1	2	2	0	Skeptic	0.99
4	9	6	4	2	1	10	Skeptic	0.49
16	16	8	0	4	4	0	Skeptic	0.70
14	6	6	3	2	11	4	Skeptic	0.46
1	5	12	0	5	10	0	New	0.53
6	2	22	1	3	4	0	New	0.99
12	1	10	2	6	11	2	New	0.72
15	0	18	1	2	11	0	New	0.98
19	16	5	0	7	4	0	Old	0.39
5	6	5	4	15	0	2	Naive	0.74
7	17	0	0	11	4	0	Naive	0.99
9	10	1	1	10	10	0	Naive	0.99
13	19	0	0	13	0	0	Naive	0.99
20	16	1	0	13	0	2	Naive	0.99
2	5	5	4	3	15	0	Random	0.37