Seeing is believing: Priors, trust, and base rate neglect

Matthew B. Welsh (matthew.welsh@adelaide.edu.au)
Australian School of Petroleum, University of Adelaide, Adelaide SA 5005, Australia

Daniel J. Navarro (daniel.navarro@adelaide.edu.au)
School of Psychology, University of Adelaide, Adelaide SA 5005, Australia

Abstract

Tversky and Kahneman (1974) described an effect they called 'insensitivity to prior probability of outcomes', better known as base rate neglect (Bar-Hillel, 1980). This describes people's tendency to underweight prior information in favor of new data. Probability theory requires these prior probabilities to be taken into account, via Bayes' theorem, when determining an event's posterior probability. The fact that most people fail to do so has been taken as evidence of human irrationality and, by other authors, of a mismatch between our cognitive processes and the questions being asked (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). In contrast to both views, we suggest that simplistic Bayesian updating using given base rates is not always a rational strategy. Instead, we reconsider Bar-Hillel's original relevance theory, and argue that, since base rates differ in their perceived degree of trustworthiness they are, accordingly, rationally discounted by people.

In the closing remarks to *A Philosophical Essay on Probabilities*, Laplace (1814/1951) argues that "the theory of probabilities is at bottom only common sense reduced to calculus; it makes us appreciate with exactitude that which exact minds feel by a sort of instinct without being able ofttimes to give a reason for it". Using probability theory, Bayes' rule provides the mechanism by which a set of prior beliefs can be updated in light of evidence, as follows: given a hypothesis, h, which we believe has some prior probability of being correct $p(h)$, if we then observed some data, x, Bayes' theorem tells us how to find $p(h|x)$, the posterior probability that h is true given that we have now seen x.

$$p(h|x) = p(x|h)p(h)/p(x) \tag{1}$$

As to whether Laplace's claim provides a plausible account of human reasoning, one of the principal sources of discussion is *base rate neglect*, a phenomenon that seems to contradict the assertion that analytic probabilities are merely formalized versions of people's intuitions about chance. The general finding is that, when people are provided with prior information (in the form of a base rate) along with new evidence, they typically weight the evidence provided by data far more heavily than the base rates (Tversky & Kahneman, 1974). This tendency to downgrade the value of the prior relative to the likelihood is taken to imply that: firstly, Bayes' theorem does not provide a complete account of the reasoning employed by people (Villejoubert & Mandel, 2002); and, secondly, that people are therefore suboptimal or biased in their judgments, and may be taken to be acting irrationally. Note, however, that there are two distinct claims here. Clearly, underweighting the base rate information will lead people to make judgments that differ from those provided by a simplistic application of Equation 1. However, the charge of irrationality is a stronger claim, and a more questionable one.

Traditional approaches to the study of human decision making have tended to assume that rational behavior is best operationalized in terms of strict adherence to an externally-provided optimality criterion, such as expected utility. Any deviations from these criteria are deemed to be irrational. However, this approach has been criticized for the assumption that it is always rational to conduct exhaustive calculations, rather than to make a far swifter decision that leads to a satisfactory outcome (Todd & Gigerenzer, 2000). This criticism is based on Simon's (1956) notion of bounded rationality, and indeed some models proposed within this framework can be reconciled very neatly with the classical Bayesian view (e.g., Lee & Cummins, 2004). In a similar vein, in this paper we discuss the question of how one might appropriately weight base rates and novel information in order to make predictions in real environments.

The Existence of Base Rate Neglect

The research on base rate neglect is heavily polarized. The 'heuristics and biases' school of thought argues that base rate neglect is robust – resulting from people's inability to update in a Bayesian manner (Kahneman & Tversky, 1996) – while their opponents argue that the effect disappears under experimental conditions better suited to human cognition (Cosmides & Tooby, 1996; Gigerenzer, 1996). In particular, it is suggested that questions phrased in a frequency format rather than in terms of probabilities are more easily dealt with by people and thus less susceptible to base rate neglect. The reasoning behind this argument is the claim that frequencies of events are easily observed whereas the probability of a single event is intrinsically unobservable (Cosmides & Tooby, 1996). Thus people would be expected to have cognitive abilities suited to counting events and comparing these absolute frequencies rather than determining one-off probabilities.

Initial support for this was found by a number of researchers (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995) but subsequent research has found that relative frequencies, such as percentages, give an equal or greater reduction in base rate neglect (Harries & Harvey, 2000; Sloman, Over, Slovak, & Stibel, 2003). Sloman et al. further argue that, rather than a difference in the underlying cognition, the effect described by earlier work results from nesting probabilities to make it clear to participants which values need to be compared and that the same benefit can be achieved in probability formats if the data are presented equivalently.

Even accepting that the method works, however, base rate neglect is generally *reduced* by using frequency formats, not eliminated. Even where people are given direct experience of a sample, rather than merely summary statistics, base rate neglect persists (e.g., Goodie & Fantino, 1999; Gluck & Bower 1988) and an analogous effect has been observed in pigeons (Zentall & Clement, 2002), suggesting that the effect is not simply an artifact of experimental design.

Unstable Base Rates

In some respects, this debate seems somewhat confusing. As the proponents of bounded rationality, Gigerenzer and Todd (1999; Todd & Gigerenzer, 2003) have argued that we should seek to understand cognitive processes in light of the environments in which they are designed to operate. Given the preponderance of data illustrating the robustness of the effect (e.g., Kahneman & Tversky, 1996), rather than attempt to force the effect to disappear, it seems more productive to consider the ecological reasons that might suggest that neglecting base rates is the right thing to do.

The tendency within the literature has been to treat base rates as eternal and unchanging truths given unto people and which, therefore, it is irrational to ignore. This is not, however, generally the case. The value of base rate data is limited, or bounded, in a number of important ways. Goodie and Fantino (1999) touch on this, arguing that people need to be sensitive not just to base rates but also to the how base rates change. In the classic "taxicab" problem they argue that the base rates given for taxicab colors and eyewitness reliability are specific to one place at one time and thus subject to change. Indeed, in advocating a relevance-based account for the base rate neglect effect, Bar-Hillel (1980) examined variants of the taxicab problem in which the source of information for the base rate is varied, and notes corresponding effects on the strength of the phenomenon.

In this paper, we expand on the perceived-relevance view of base rate neglect. Our approach can be motivated by considering the following example, adapted from one commonly used by philosophers and originating in the work of David Hume (1739/1898).

> *Imagine that you have been calculating the proportion (base rate) of white swans amongst the general swan population. You have been across Europe and observed 999 swans – all of which were white. You then take a plane to Australia and continue your survey. Your first observation is of a black swan. You have now observed one thousand swans and have a base rate of 99.9% for white swans. As you plan to continue your survey, what is the probability that the next swan you observe will be white?*

In a naive statistical sense, one would expect the next swan to be white with a 99.9% probability, as the base rate indicates. Rationally, however, you are aware that there is such a thing as regional variation and that your observations from Europe are less likely to be predictive in Australia. Accordingly, a belief in regional variation provides a strong justification for a decision to neglect the base rate.

When deciding how much faith to place in a base rate, at least four environmental factors would appear to be relevant.

- *Location.* Even if you genuinely believed that 99.9% was the true, world-wide base rate of white swans, the existence of regional variations implies that the single black swan observed in Australia *should* be more highly weighted. Therefore, when changing from one location to another, a rational person will discount prior observations against current ones – that is, they will neglect the base rate in favor of new data.

- *Age.* Old data are less likely to be relevant to a new prediction than more current data as base rates change over time. Consider, for example, the proportion of land predators that are dinosaurs. If you were relying on base rates incorporating data from the last 170 million years, you might predict a fairly high proportion of observations. A more current analysis, however, would yield a lower figure. While this is a deliberately extreme example, this "information aging" effect is observed in library curves that track the frequency with which books are borrowed as a function of age, which have a similar shape to human forgetting curves, suggesting that forgetting old information is a rational adaptation to changing environments (Anderson & Schooler, 1991).

- *Source.* In general, people trust the evidence of their own senses to a greater extent than they do that of another person. Thus, a sample that a person has collected themselves is likely to be weighted more heavily than data given to that same person from an outside source. This is, of course, quite rational in that, with outside data, the degree of certainty over its veracity and how it was collected will tend to be lower than that regarding one's own observations.

- *Quantity.* Sample size must also be considered for both the prior and current samples. The former is often ignored in base rate neglect experiments but must be considered as sample size partially determines a base rate's reliability. Empirically, a "base rate" can only be discovered via observation: in truth, base rates are simply prior samples. As a consequence, the decision-maker should consider how much data contributes to the base rate itself.

Experiment 1

As an initial examination, we consider the impact of varying the location, age, source and quantity of the data that provides a base rate. The approach is similar to, but more systematic than, the variations considered by Bar-Hillel (1980).

Method

Participants. Participants were twenty university students and members of the general public, 10 males and 10 females, with a mean age of 30.4 (SD = 12.1). Each was paid for their participation with a $10 bookstore voucher.

Experimental Design. The scenarios used in our experiment were designed to maximize the extent to which people recognize a need to combine two sources of information, by explicitly placing the base rate data on a scale commensurate with a second source of evidence. To do so, both sources of evidence are described as samples of data (prior sample and new sample) that need to be taken into account. In this experiment we chose to examine the effect of varying sample size, while combining the source, age and location variables into a general cover story. Under the "high trust" cover story, the prior sample was described as recent data, collected by the participant, in the same location. Under the "low trust" cover story, the data was old, collected by someone else, and in a different location. Sample sizes were varied for both the prior data (20 or 200 data points) and for the new data (4, 8 or 12 data points). Moreover, the implied base rate could be either 25% or 75% (with the new data implying the alternate). With all factors fully crossed, this gave 24 (2x3x2x2) conditions in total.

All of the scenarios used variations on the same cover story: that the participant was part of a survey team exploring an alien planet and reporting on the proportion of some native life form or natural event that met a particular criterion. In every case, the participant was given a prior sample and then told what they had observed. Finally they were asked for an estimate combining both sets of information to be included in their report. For example:

> *You are currently classifying predators according to whether they pose a threat to humans. Your team, working at this location recently collected 200 observations and found that 50 (25%) of them met this criterion. This week, you have made another 4 observations, of which 3 (75%) met the above criterion. What proportion of predators in the area do you estimate pose a threat to humans?*

This example shows a prior sample size of 200 with a base rate of 25%. The current sample has a size of 4 and a rate of 75%. The prior is trustworthy in that it is described as recent, local and self-collected. Twenty-four scenarios were created so each participant would see a scenario in each condition.

Procedure. All scenarios were incorporated into a GUI and presented in random order. Participants sat at the computer and read the introductory cover story before proceeding to the first randomly determined scenario. During each scenario, all of the information remained visible on the screen until the participant had entered a predicted rate of future occurrence. No time limit was imposed and most participants completed the 24 scenarios within an hour.

Descriptive Model. To analyze the data, we will adopt a heavily simplified model for how a "rational" decision-maker might solve this kind of induction problem. Suppose the participant makes the assumption that the observed data reflect some unknown Bernoulli probability θ, and reports the expected value for θ given the data. In this situation, a straightforward choice of prior might be a Beta distribution with parameters estimated from the prior sample.[1] So, for instance, in the example given in the method section, the observer might specify a Beta(50,150) prior. In general, if n_0 denotes the number of observations that make up this prior and x_0 is the number of those observations that meet the criterion, then the prior is Beta($x_0, n_0 - x_0$), and the expected prior value for θ is given by the base rate, $E[\theta|x_0, n_0] = x_0/n_0 = r_0$. In the event that the new data are assumed to have exactly the same distribution as the prior data, then $x_1|n_1 \sim$ Binomial(θ, n_1). Given this, the posterior distribution over θ is Beta($x_0 + x_1$,$n_0 + n_1 - x_0 - x_1$), and the observer would report the obvious choice, $E[\theta|x_0, x_1, n_0, n_1] = (x_0+x_1)/(n_0+n_1)$.[2]

This model assumes that the observer assigns equal weight to all data. However, this is highly unlikely. Firstly, the scenarios encourage participants to assume that the prior sample may be less closely related to the quantity of interest θ than the new data. Accordingly, if each prior datum is "worth" only t new data, then a natural description of the prior is a Beta($tx_0, tn_0 - tx_0$).[3] Updating in the usual manner, we might expect the participant to report the value,

$$E[\theta|r_0, r_1, n_0, n_1, t] = \frac{tr_0 n_0 + r_1 n_1}{tn_0 + n_1}, \quad (2)$$

where $r_0 = x_0/n_0$ denotes the base rate, and $r_1 = x_1/n_1$ denotes the sample rate. In this experiment, we vary the way people weight r_0 against r_1 in two distinct ways. By altering the description applied to the prior sample, we expect to see a direct change in the value of t. This is a direct "cover story" manipulation, and is expected to result in some explicit downgrading of the usefulness of prior sample.

The second manipulation involves sample size, and is somewhat more complex, since sample size is already built into the naive model predictions. By altering the ratio n_0/n_1, we would expect some reweighting of the two estimates. However, in view of the widely studied "insensitivity to sample size" effect (e.g., Kahneman & Tversky, 1974), the subjective "value" of a particular sample size is unlikely to be the same as its actual value. Nevertheless, following Sedlmeier and Gigerenzer (1997), we might reasonably expect that people's behavior will accord with Bernoulli's (1713) statement of the so-called empirical law of large numbers: "even the stupidest of men, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one's goal" (see Stigler, 1986, p.65). For the moment, then, we make the assumption that the subjective value $\tilde{n}$ is related to the objective value n via some unknown monotonic increasing function $\tilde{n} = f(n)$. Given this, we model the participants' judgments by assuming that they will report the value of θ to be expected when one applies Bayes' theorem to the *subjective* sample values, with

[1] This is equivalent to starting with a non-informative Haldane prior over θ, assuming that $x_0|n_0 \sim$ Binomial(θ, n_0), and then updating belief about θ via Bayes' theorem (see Jaynes, 2003).

[2] For people of a less Bayesian persuasion, it is worth noting that this is also equivalent to a decision-maker reporting the maximum likelihood estimate for a pooled sample.

[3] The likelihood in this case involves a trivial generalization of the Binomial distribution, straightforward to derive but not discussed here for space considerations.

some constant effect expected to arise due to the cover story:

$$E[\theta|r_0, r_1, \tilde{n}_0, \tilde{n}_1, t] = \frac{tr_0\tilde{n}_0 + r_1\tilde{n}_1}{t\tilde{n}_0 + \tilde{n}_1}. \quad (3)$$

In order to fit the data from the 24 conditions, we fit 4 values for t (high and low trust for both base rates), and 4 values for subjective sample size. Assuming that $f(20) = 20$, we estimate the values for $\tilde{n}$ that correspond to $f(4)$, $f(8)$, $f(12)$ and $f(200)$, which are expected to be more-or-less invariant across experimental conditions. Note that, since 8 parameters are used to fit 24 data points, there is a sense in which this model is more descriptive than explanatory. However, it will transpire that $f(\cdot)$ has a very regular form, allowing these parameters to be fixed in a sensible fashion, and leaving only the explicit trust parameter t as truly 'free'.

Results

Since the simplified framework discussed here makes no provision for extrapolation (i.e., participants perceiving a trend), only those 14 participants whose data show no evidence of extrapolation (i.e., all 24 judgments lie in the range [25, 75]) are considered in this initial investigation. Figures 1 and 2 show the mean estimates for the underlying probability given by these participants in all 24 conditions. The triangles show empirical data for the "high trust" cover story, and the circles show data for the "low trust" cover story. The dashed line shows the predictions made by the simplistic Bayesian solution (Equation 2). Overall, there is a clear base rate neglect effect: the empirical predictions tend to be shifted away from the Bayesian solution towards the current rate (i.e., above it in Figure 1 and below it in Figure 2). In total, data for 23 of the 24 conditions are shifted in this direction (one-tailed sign test gives $p \approx 1.5 \times 10^{-6}$). More important, however, is the fact that trustworthiness is having a clear effect. In all 12 cases, the mean predictions made by participants in high trust scenarios are closer to the Bayesian solution than estimates made in otherwise equivalent low trust scenarios (one-tailed sign test gives $p \approx 2.4 \times 10^{-4}$).

A finer grain of analysis is possible by fitting the model. Parameter estimates for t and $\tilde{n}$ were obtained by minimizing sum squared errors. Figure 3 shows the recovered parameter estimates for the subjective sample size parameters, $\tilde{n}$. Comparison with the solid line makes clear that $\tilde{n} \propto \log n$: in this task, subjective impressions of sample size rise logarithmically with the actual sample size. This logarithmic relationship is in agreement with both the classic Weber-Fechner law, and with other data suggesting that the mental representation of magnitude is approximately logarithmic (e.g., Shepard, Kilpatric & Cunningham 1975; Dehaene 2003).

The implied trust statistics t for the cover story, shown in Table 1, are more complex. Most importantly but not surprisingly, in both the 25% base rate conditions and the 75% base rate conditions, the estimated value for t is much higher when the cover story suggests high trust as opposed to low trust. Parameter estimates for low trust suggest that a prior datum is worth only 1/4 of a new datum, in subjective (i.e., log) terms. When the base rate is 25%, the high trust parameter is approximately 1, suggesting that the only effect in this condition is the logarithmic scaling of subjective sample size effect shown in Figure 3. The inferred value of 1.4 for

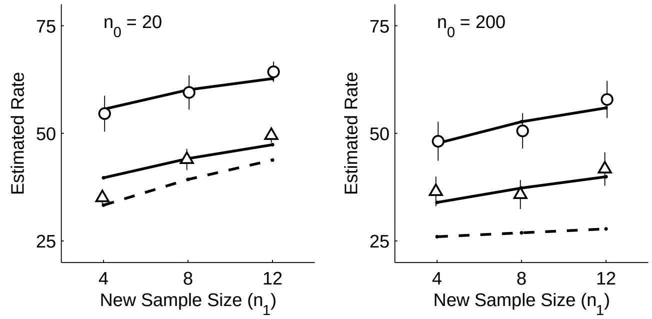

Figure 1: Participants estimated rates for the various conditions in which the base rate was 25%. Triangles denote data from conditions involving the "high trust" cover story, and the circles show the "low trust" condition data. The thin lines are standard error bars. The dashed lines show the predictions of the naive Bayesian model (Equation 2), while the solid lines show the predictions made by the descriptive model.

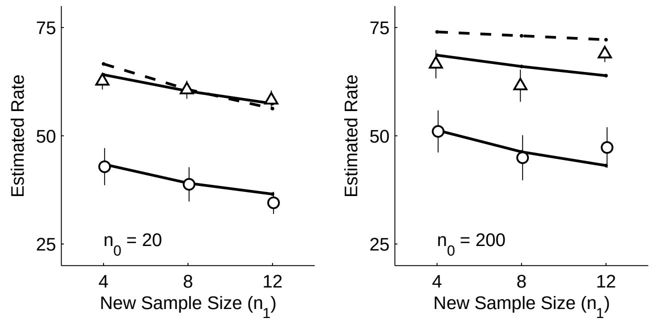

Figure 2: Participants estimated rates for the various conditions in which the base rate was 75%. The format of the plot is the same as for Figure 1.

the 75% base rate and high trust is odd, since it implies that a prior subjective datum is treated as being worth more than one subjective new datum. In view of this, and the fact that the corresponding empirical data for these conditions (solid line at the top of Figure 2) do not show strong evidence of base rate neglect, suggests that this case may be somewhat different to the others.

Discussion

The results paint a relatively clear and somewhat intriguing view of base rate neglect. To a large extent, the base rates implied by larger samples are weighted more heavily than for small samples, in keeping with the so-called empirical law of large numbers (Sedlmeier & Gigerenzer, 1997). In that sense, people can be seen to adapt to the trustworthiness of the data in a very sensible fashion. That said, a kind of "insensitivity" to sample size is observed, since the subjective value rises nearly logarithmically with sample size, rather than linearly. Altering the cover story to devalue the base rate has a large effect on trust, lowering the subjective value of the base rate by three quarters in both the 25% and 75% conditions.

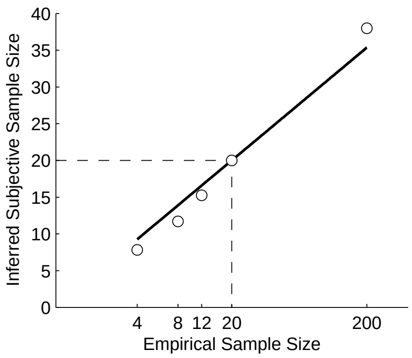

Figure 3: Subjective sample sizes inferred from participants' probability judgments follow an approximately logarithmic function.

Table 1: Estimated trust statistics for the low and high trustworthiness conditions, as a function of the underlying base rate.

	high trust story	low trust story
25% base rate	0.94	0.25
75% base rate	1.41	0.23

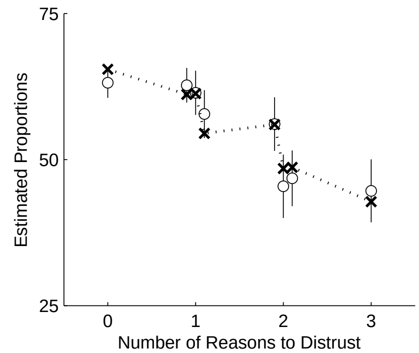

Figure 4: Actual and predicted values for participants' estimates for the underlying rate in experiment 2. Empirical values are shown by white circles with standard error bars shown. Model predictions are shown with crosses.

Table 2: Estimated effect on trust for each element of the cover story.

	location	age	source
t	0.34	0.63	0.62

Experiment 2

Method

Experiment 2 aimed to expand on the three factors that contributed to the cover story in Experiment 1. The design of the experiment was the same as for experiment 1, and was in fact conducted simultaneously with the first experiment using the same 20 participants, with the various conditions intermixed with those used in the first study. In this case, the "base rate" was fixed at 75% using a sample of size 20 (i.e., 15 hits), and the new data are based on a sample size of 4 with a single hit, suggesting a rate of 25%. An independent effect model takes the same format as Equation 3, but with separate terms for the effect of location t_l, age of data t_a and source of the data t_s. For reasons that will become clear shortly, we fix the high value for trust at 1 in each case (e.g., when you collect the data yourself), and simply estimate the low value. Similarly, we again extract $\tilde{n}$ from the raw data.

Results

The basic pattern of results is shown in Figure 4. As more reasons to distrust the prior data (distant location, old data, collected by someone else) are added to the cover story, participants' ratings move away from the base rate and closer to the new data. Moreover, a model that assumes that each manipulation has a constant effect on trust provides a very close fit to the data. As shown in Table 2, each manipulation has a substantial effect. Changing the age or source of the data lowers trust to 2/3, while changing the location lowered trust to 1/3. Fitting the subjective value of the new data, we obtained $\tilde{n} = 4.72$, suggesting that in this case participants' subjective understanding of sample size was almost perfectly calibrated.

The overall pattern of results is highly consistent with results from corresponding conditions in Experiment 1 (i.e., those with 75% base rate, prior sample of 20 and current of 4): if all parameter values are multiplied by 1.41 (the high trust value found for these conditions in Experiment 1), we obtain $\tilde{n} = 6.61$ for the subjective sample size, which is fairly close to the value of 7.82 found in Experiment 1. Similarly, the low trust value of 0.23 from Experiment 1 is close to prediction from Experiment 2 of: $1.41 \times 0.34 \times 0.63 \times 0.62 = 0.18$.

Discussion

Experiment 2 sheds a great deal of light on the data from Experiment 1. Firstly, it is clear that all three elements to the cover story play a role in determining the level of trust that people assign to a particular source of data. Secondly, it helps provide some insight into the otherwise puzzling "1.4 prior trust" parameter estimate that arises in one of the conditions. Specifically, while the majority of the judgments in Experiment 1 reflect a combination of samples based on logarithmic scaling of subjective sample size, it appears that in some scenarios with high trust, participants do weight samples in an exact (i.e., linear) fashion.

General Discussion

Taken together, the two experiments provide an interesting perspective on base rate neglect. Overwhelmingly, they support the view that, although the base rate neglect effect is genuine, it is not the case that people are simply failing to take into account all of the available information. On the contrary, there is evidence to suggest that people are in fact discounting the prior information as being less relevant than current information. Four distinct factors are shown to influence the extent to which people devalue the prior information: geographic distance, age of the data, the source of the information, and the amount of data upon which the base rate is constructed. Sample size, the last of the four factors, was investigated in some detail, and the results suggest that people are highly sensitive to changes in sample size. This is particularly interesting in that, quite incidentally to the overall goals for the study, it suggests that many of the results pertaining to "insensitivity to sample size" might be best explained by supposing that people use an internal logarithmic representation to scale the sample size, in agreement with other studies that have looked at the psychological representation of magnitude.

An intriguing possibility is that the proposed logarithmic law for subjective sample size may not apply in all cases. As indicated, the parameter estimates obtained for Experiment 2 are in close agreement with the relevant ones for Experiment 1. However, when the "high trust" levels are fixed to unity, as in Experiment 2, not only do the "low trust" levels scale accordingly and once again suggest a strong discounting effect, but the sample sizes now appear to scale *linearly* (the same does not occur for Experiment 1, since the vast majority of conditions appear to reflect the logarithmic scaling law). One explanation for the disparity may be that in some (presumably very high-trust) situations, people mentally represent the data in terms of a single pooled sample, rather than as two distinct samples. In this pooled case, the relative weight of the prior data and new samples would combine linearly. It is possible that this distinction between "pooling" and "comparing" data sets could explain the effect, but without further data it is difficult to do more than speculate.

Three final caveats are in order. Firstly, the proposed account only covers the weighting of two data sources. A complete account should extend the approach to deal with extrapolation. Secondly, no claim is made that discounting is always conscious: people may very well have an intuitive preference to rely on more recent data, for instance, but still be willing to admit (post-experiment) that they "should" have used Bayes' theorem. An intuitive (and appropriate) distrust of base rates not being inconsistent with the ability to follow the logic of Bayesian updating. Finally, it should be noted that since our experimental design used within-subject comparisons, it is heavy-handed in terms of the extent to which participants are made aware of the potential variations in reliability for different sources of data. Thus, although it is clear that there are situations in which people are extremely good at incorporating or discounting prior data in a sensible fashion, it is not as clear how generally this holds. In particular, further research is required to determine whether examples without any markers of trustworthiness, such as traditional base rate neglect experiments, lead people to trust or distrust the presented base rates. Until this question has been resolved, however, it seems premature to base any charge of human irrationality on previous base rate findings.

Acknowledgments

MBW was supported by ExxonMobil and Santos, through their funding of the CIBP at the Australian School of Petroleum. DJN was supported by ARC grant DP-0773794. The authors wish to thank Anastasia Ejova for collecting the data, and Steve Begg and Nancy Briggs for helpful discussions.

References

Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396-408.

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

Bernoulli, J. (1713). *Ars Conjectandi*, Basilea: Thurnisius.

Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*, 1-73.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences, 7*(4),145-147.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103*(3), 592-596.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*(4), 684-704.

Gigerenzer, G., & Todd, P. M. (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Gluck, M. A. & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 227-247.

Goodie, A. S., & Fantino, E. (1999). What does and does not alleviate base-rate neglect under direct experience. *Journal of Behavioral Decision Making, 12*, 307-335.

Harries, C. & Harvey, N. (2000). Are absolute frequencies, relative frequencies, or both effective in reducing cognitive biases. *Journal of Behavioral Decision Making, 13*, 431-444.

Hume, D. (1739/1898). *A Treatise of Human Nature*. London: Ward Lock.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press

Kahneman, D., Slovic, P. & Tversky, A. (1982). Judgment Under Uncertainty. Cambridge, UK: Cambridge University Press.

Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*(3), 582-591.

Laplace, P. S. (1814/1951). *Essai Philosophique sur les Probabilites* (F. W. Truscott & F. L. Emory, Trans.). New York: Dover Publications.

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the 'take the best' and 'rational' models. *Psychonomic Bulletin & Review, 11*(2), 343-352.

Sedlmeier, P. & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10*, 33-51.

Shepard, R. N., Kilpatric, D. W. & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology, 7*, 82-138.

Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review, 63*, 129-138.

Sloman, S. A., Over, D., Slovak, L. & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91*, 296-309.

Stigler, S. M. (1986) *The History of Statistics* Cambridge, MA: Harvard University Press.

Todd, P. M. & Gigerenzer, G. (2000). Simple heuristics that make us smart. *Behavioral and Brain Sciences, 23*(5), 727-741.

Todd, P. M. & Gigerenzer, G. (2003). Bounding rationality to the world. *Journal of Economic Psychology, 24*, 143-165.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124-1131.

Tversky, A. & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.

Villejoubert, G. & Mandel, D. R. (2002). The inverse fallacy: an account of deviations from Bayes Theorem and the additivity principle. *Memory and Cognition, 30*(2), 171-178.

Zentall, T. R., & Clement, T. S. (2002). Memory mechanisms in pigeons: evidence of base-rate neglect. *Journal of Experimental Psychology: Animal Behavior Processes, 28*(1), 111-115.