Individual differences in base rate neglect responses and susceptibility.

Matthew B. Welsh[1]

Nicholas R. Burns[2]

Paul H. Delfabbro[2]

Steve H. Begg[1]

[1]Australian School of Petroleum, [2]School of Psychology,

University of Adelaide

North Terrace, Adelaide

SA, 5005, AUSTRALIA

Corresponding Author:

Dr Matthew B. Welsh

Ph: +61 8 8303 8024

Fax: +61 8 8303 8030

matthew.welsh@adelaide.edu.au

Emails:

nicholas.burns@adelaide.edu.au

paul.delfabbro@adelaide.edu.au

steve.begg@adelaide.edu.au

Abstract

Base rate neglect refers to the underweighting of base rate information in favor of newer (reliability) data (Bar-Hillel, 1980). Analyses of responses to a base rate neglect problem, however, indicate that individual behavior is more complex than this statement indicates. Participants' responses can be categorized as resulting from intuitive or methodical reasoning (i.e., System 1 or System 2 thinking, Stanovich & West, 2000) and which of these an individual relies on may be related to differences in their cognitive abilities. Specifically, sustained attention, numerical reasoning, and processing speed all predict participant responses to a base rate neglect problem. Decision style measures, including Frederick's (2005) Cognitive Reflection Test, designed to measure people's preference between the two systems, show less clear results. We conclude that individual differences should be considered in order to understand the cognitive processes underlying base rate neglect and that explanations based around the average effect are almost certainly wrong.

Keywords: decision making; base rate neglect; individual differences; cognitive ability; decision style.

1. Introduction

Kahneman and Tversky (1973) first described the effect that came to be called base rate neglect (Bar-Hillel, 1980), which describes people's tendency to underweight base rate information in favor of new data. The fact that many people fail to make use of base rate data (as Bayes theorem requires) has been taken by some as evidence of 'irrationality' and, by others, of a mismatch between our cognitive processes and the questions being asked (Cosmides & Tooby, 1996). Thus, a significant literature has been devoted to the question of base rate neglect and when or if it will occur (see, e.g., Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995; Kahneman & Tversky, 1996; Sloman et al., 2003).

Although of theoretical interest in its own right, base rate neglect, can also have important practical implications. For example, Bayesian updating often forms the basis for probability assessment in industries such as oil and gas exploration. Welsh et al. (2005, 2007) showed that people have significant difficulty in accurately undertaking such tasks and that errors caused by such cognitive biases have serious economic implications. There has, however, been a disconnect between much of the research in the field and the desire of industries to either select personnel who are less susceptible to base rate neglect or create debiasing methods to improve the decision-making of specific individuals.

A large part of this disconnect may result from the fact that, in almost all base rate research, the existence and strength of the effect has been discussed in terms of average or group-level results rather than at the level of individuals, where industry interests lie. However, as recent work has suggested, even where a group seems to show base rate neglect, analyses of individual responses can demonstrate that relatively few individuals show base rate neglect as it is commonly understood (Welsh & Navarro, in press). Given this, an analysis of base rate neglect at the level of the individual is necessary in order to determine what cognitive processes actually underlie the effect and, thus, whom it will effect and how.

1.1. System 1 vs System 2 Thinking

A key distinction drawn in recent years within the judgment and decision making literature is between System 1 and System 2 styles of thinking (Stanovich & West, 2000). System 1 referring to fast, almost effortless, intuition-based judgments, whereas System 2 refers to effortful, reasoned decision making that requires conscious attention be paid to a problem. In relation to statistical problems involving base rate information, for example, it is assumed that decision-making is overly influenced by intuitive or System 1 reasoning. People quickly adopt highly salient and simple solutions and either (a) fail to apply System 2 reasoning altogether or (b) do not have the knowledge or System 2 ‘processing ability’ to find the correct solution via Bayes theorem.

1.2. Individual Differences

Given that individuals within groups demonstrate different approaches to base-rate problems (Welsh & Navarro, in press), a question arises as to whether these variations reflect measurable individual differences. This has not, previously, been examined and evidence from other decision biases has been mixed. Some studies (see, e.g., Stanovich & West, 1998, 2000, 2008) have found small to moderate relationships between measures of intelligence, aptitude or achievement (e.g., Raven’s Progressive Matrices, SAT scores) and people’s susceptibility to biases. Similarly, Peters et al. (2006) found that more numerate subjects were less prone to ratio biases and framing effects, although this effect was not sustained across all experiments. While some effects were consistent with a division between two styles of thinking, others (e.g., how highly numerate participants responded to a choice-bet task) were not (Reyna & Brainerd, 2008).

Similar inconsistencies emerge in studies of cognitive style. Bruine de Bruin et al. (2007) found relationships between decision-making competency and cognitive style. Specifically, people who try to think rationally are less likely to make decisions quickly and perform better on these tasks. Another study by McElroy and Seta (2003) showed that susceptibility to decision-making biases is lower in people with preferences for more systematic or analytical processing styles. However, Peters et al. (1996) found, for some tasks, an interaction between cognitive style and numerical ability, implying that more optimal responses sometimes coincided with more intuitive thinking styles.

1.3. The Present Study

The core of the present study is to undertake a more detailed investigation of the covariates of base rate neglect susceptibility, using the tools provided by modern individual differences testing and theory. In this way, we hope to shed light on why this bias occurs and who is most susceptible to it – thereby informing both cognitive theories of decision making and potential debiasing methods.

Given the inconsistent findings noted in Section 1.2 relating other cognitive biases and individual differences, avenues for methodological refinement and further analysis remain. Most studies so far have used a relatively limited range of cognitive abilities measures, some of which (e.g., Peter's numeracy test) may generate highly attenuated scores in the educated populations often used in these studies. Accordingly, there is a need to examine decision-making quality in relation to a wider range of cognitive ability measures. In particular, there is a need to capture abilities considered central to modern theories of intelligence and cognitive performance (e.g., the Cattell-Horn-Carroll model, McGrew, 2005) and, specifically, those that seem most likely to impact on a the cognition required within a base rate task.

In this research, measures are, therefore, selected to provide a more refined assessment of: numerical reasoning, as numerical skill is required to solve a base rate problem correctly; processing speed, reflecting how quickly a person can solve a problem and thus how likely they are to give up before solving it; working memory, reflecting the ability to hold and manipulate the necessary pieces of information in mind during attempts to solve the problem; and executive function, which reflects a person's ability to notice and overrule errors in automatic responses – and thus bears a clear similarity to the proposed function of System 2 thinking.

This study is also predicated on the assumption that many existing cognitive measures are not designed to examine a preference for System 1 or 2 thinking (Stanovich & West, 2008). For this reason, a variety of personality/decision style measures were also selected – to account for potential differences in how people prefer to respond to base rate neglect questions. For example, Frederick's (2005) cognitive reflection task (CRT) is examined because this is thought to capture, not only people's ability to undertake systematic processing (the function of IQ and aptitude tests), but also their preference to apply this processing style to tasks. If this is the case, then one would expect that people scoring more highly on the CRT will be less susceptible to biases resulting from the use of System 1 thinking.

As a more general point, a study of this sort offers the potential to ask questions regarding how the two-system theory of decision making research aligns with the empirically-derived facets of cognitive ability and personality that are generally accepted within the field of individual differences. Specifically, in order to demonstrate the utility of the two-system theory, one must first establish that it accounts for something not incorporated within those attributes already established in the individual differences literature – an approach that is uncommon in the judgment and decision making field and which has not previously been applied to base rate neglect.

2. Methodology

2.1. Participants

Participants were 102 university students and members of the general public, recruited via posters and research participation email lists from around the {institution}. The sample had a mean age of 22.5 (SD = 4.89) and consisted of 34 males and 68 females. All participants received $50 for their completion of a three-hour battery of tasks including those described in this paper.

2.2. Materials and Procedure

Prior to coming in to the laboratory, participants completed an online instrument, which gathered demographic details, the decision style tasks and delivered the base rate neglect problem described below. They then were invited to the laboratory to complete the computerized cognitive tests.

2.3. Demographics. In addition to age and sex, participants were asked to provide their highest level of education on a 7-point scale indicating various levels of academic achievement. As a result of the small sizes of some of these categories, responses were recoded on a 3 point scale: 1, have not attended university (n = 22); 2, current university student (n = 58); and 3, university graduate (n = 22).

2.4. Measures

2.4.1. Base Rate Neglect Task. To test for susceptibility to base rate neglect, we used the traditional 'taxi-cab' problem (Bar-Hillel, 1980), giving participants the following information:

> A cab was involved in a hit-run accident at night. Two cab companies, the Green and the Blue operate in the city in which the accident occurred.
>
> You are given the following facts:
> 85% of the cabs in the city are Green and 15% are Blue.
> A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each of the two colours 80% of the time.
>
> What is the probability that the cab involved in the accident was Blue? (%)

This scenario was part of the online instrument and participants entered their response in a text box at the bottom of the page.

2.4.2. Cognitive Ability Measures

2.4.2.1. Numerical Abilities Test (NAT). A short version of the Numerical Abilities scale from the Differential Aptitude Test (Bennett et al., 1989) was prepared, asking 12 rather than 48 questions and restricting time to 9 minutes rather than 36. A participant's score was simply the number of questions they answered correctly. This task was run via a Matlab GUI and conducted in the laboratory.

2.4.2.2. Symbol-Digit Test. A cognitive processing speed measure, this is a computerized test similar to the Digit-Symbol test from the Wechsler IQ tests (see McPherson & Burns, 2005). The measure taken from the test is simply the number of symbols correctly translated within the given time limit.

2.4.2.3. Dot Matrix Task. A computer-administered version of the Dot Matrix Task (Law et al., 1995), a measure of working memory. The task requires participants to verify a series of simple matrix equations while simultaneously remembering dot locations on a 5×5 grid. The test had four levels (i.e., 2, 3, 4, and 5 equation–grid pairs) with four questions per level (total 16 questions). The measure for the task was the total number of dot positions correctly recalled.

2.4.2.4. Sustained Attention to Response Test. The SART (Robertson et al., 1997) is a computer-administered test of executive function. The stimuli consisted of the individual presentation of randomly selected digits between 1 and 9. Each digit was presented for 245 ms and was followed immediately by a 900ms mask. Participants were required to depress the left mouse button for all digits except the digit '3'. The digit '3' was presented randomly either 24 or 25 times out of 225 trials. The measure taken from this was the number of errors of commission (i.e., incorrectly responding to '3s').

2.4.3. Decision Style Measures

2.4.3.1. Cognitive Reflection Task (CRT). As described by Frederick (2005), the CRT is designed to measure a person's level of "cognitive reflection"; that is, how likely they are to engage rational and reflective System II reasoning rather than relying on the fast and intuitive System I reasoning implicated in decision biases (Stanovich & West, 2000). Participants answer three questions and are scored either right or wrong. CRT score is simply the number of questions that a person gets right – from 0 to 3.

In addition to predicting susceptibility to a number of biases, this measure shows a strong relationship with educational level in Frederick's (2005) data. This test was part of the online instrument.

2.4.3.2 Need for Cognition. The 10-item, International Personality Item Pool (IPIP; Goldberg et al., 2006) version of the Cacioppo and Petty (1982) scale. NFC is a personality variable reflecting the extent to which people engage in
and enjoy cognitive activities.

2.4.3.3. DOI- Decision-making Outcomes Inventory. 20-items that examine whether people have made various poor decisions (e.g., purchased things they've not used, been unable to make decisions, forgot birthdays) in the last 10 years. A modification of the measure formulated by Bruine de Bruin et al. (2007), a person's score on this was the number of questions to which they responded affirmatively.

2.4.3.4. Rational Experiential Inventory. A 30-item test of risk-style developed by Epstein et al., (1996). Yields 4 measures: Rational Engagement (RE); Rational Ability (RA); Experiential Engagement (EE); and Experiential Ability (EA). These distinguish between 'engagement' and 'ability' for two different cognitive styles – rational (conscious, analytical) and experiential (holistic, intuitive).

2.4.3.5. Intellect. A 20-item inventory from IPIP (Goldberg et al., 2006) that combines both the Cattell (1973) and Costa & McRae (1992) approaches to this construct. Intellect measures the facet of the "openness-to-experience" scale from the NEO PI-R that measures a person's openness to new ideas in an intellectual context.

2.4.3.6. Stimulating Instrumental Risk Inventory. 10-item test that yields two measures of risk attitude, one of *Stimulating* and one of *Instrumental Risk Taking* (*SRT* and *IRT*). *SRT* reflects positive arousal and is emotional, short-term and impulsive, while *IRT* reflects negative arousal is cognitive, long term and reflective (Zaleskiewicz, 2001).

2.4.3.7. Rationality. 14-item test from the IPIP (Goldberg et al., 2006). This is a facet from the AB5C personality questionnaire, reflecting a combination of Conscientiousness and Agreeableness (in Big Five terms, Costa & McCrae, 1992). Specifically, high rationality reflects high Conscientiousness and low Agreeableness.

3. Results

3.1. Base Rate Neglect

One-hundred-and-one valid responses were received from the 102 participants. As an initial approach, participants' responses to the base rate neglect problem were compared to the Bayesian solution of 41%. The mean estimate was, in fact, very close to the correct answer (X = 42.8%), which might lead one to argue that the sample displayed relatively little base rate neglect. This is not, however, a good characterization of the data, as suggested by the high variability of the data (SD = 31.9). Therefore a detailed examination of individual responses was undertaken.

3.1.2. Individual Differences in Responding

Inspection of the distribution of estimates in Figure 1 shows that few participants gave estimates close to the Bayesian solution of 41%. The histogram is, instead, dominated by spikes at 80%, 15% and 12%, which together account for 75 of the 101 estimates. Each of these emerging clusters may provide insights into how different people approached the base-rate task. The first and largest group (the spike at 80%) appears easily interpretable. It represents the traditional finding of base rate neglect, where 35 participants have ignored the base rate data and relied solely on the reliability of the eye-witness in making their decision. The second spike, at 15%, comprises participants (*n* = 13) who have relied entirely on the base rate, ignoring the reliability data (plus one who estimated 14.1%).

That is, almost half of the sample simply repeated back one of the two numbers they were given as their estimate. There is also a small spike (*n* = 5) at 41% - consistent with participants capable of calculating the Bayesian solution (or those who had seen the question before). The remaining responses, however, require more interpretation.

Values above the eyewitness reliability (80%) or below the base rate (15%) make little sense within the base rate neglect paradigm in that people are assumed to be incorporating the two values (with some weightings). Estimates should, therefore, lie within the range dictated by these values. Inspection of Figure 1, however, shows that 3 participants made estimates above 80% and 32 made estimates below 15% - dominated by the spike of 27 participants at 12%. This spike is, initially, puzzling because it does not correspond to any number given in the taxi-cab problem and yet attracts more estimates than the base rate. The solution to this puzzle, however, is quite simple – 12% is the *product* of the two percentages given in the question.

That is, we have four clearly distinguishable patterns of responding in our sample: complete base rate neglect; complete base rate reliance; a simple (incorrect) mathematical ‘solution’; and the Bayesian solution. Taken together, these account for 80 of the 101 estimates.

The remaining 21 responses are more difficult to categorize. A simple schema, in keeping with the traditional base rate neglect paradigm, would be to divide the values between the three primary categories of: base rate neglect, base rate overreliance and Bayesian reasoning. Simply sorting values into these categories, however, proves difficult.

Even having accounted for the majority of estimates falling outside the 15-80% range by recognizing the Simple Maths response class, there remain 10 participants who gave estimates outside the range for no immediately obvious reason (We note, however, that two of these ‘errors’ were values of 0.12 and 0.41 – that is, exactly 1/100th of the Simple Maths and Bayesian solutions respectively – which may indicate simple errors in responding or calculation). While there can be rational reasons for people extrapolating beyond the range dictated by a base rate neglect problem (for a discussion of this, see Welsh & Navarro, in press) these vary by individual and cannot be distinguished, in any *ad hoc* sense, from simple errors or misunderstandings of the experimental instructions.

Additionally, there remain 11 participants who gave estimates within the dictated range but not conforming to one of the major patterns. These responses could be interpreted as attempts to integrate the two values given in the problem – with varying weights assigned to each. For these cases, therefore, it could make sense to ‘bin’ these responses with the type they most resemble. That is, values between 15% and 28% could be added to the base rate reliance category, those between 29% and 60% to the Bayesian category and those between 61% and 80% to the base rate neglect category. This does, however, run the risk of, for example, ‘rewarding’ participants who responded with 50% in order to indicate that they had no real idea of the true answer by adding them to the Bayesian group. For the reasons discussed above, the analyses described below were, therefore, undertaken using only the 80 ‘definite’ members of the four, described response categories.

Table 1 shows the response categories and their memberships. The ordering of the response categories also carries information about our expectations regarding the quality of the response and how it might relate to cognitive measures. The Bayesian calculation provides the correct answer and, so, is the *best* response. It also requires the most cogitation to complete. Of the remaining types, there is little to separate base rate neglect and base rate overreliance – as both rely simply on repeating back one of the numbers presented in the problem. Base Rate Neglect is recorded as worse in this case (that is, named group 1) as it leads to larger errors on this specific task. The simple mathematical calculation, however, while providing a worse estimate (i.e., further from the true value) than base rate overreliance, requires more cogitation than just repeating back a number and is likely, therefore, to be more strongly linked to cognitive abilities than the simpler strategies.

That is, our four groups fall naturally into two super-groups: those who used a mathematical solution; and non-mathematical solutions where participants repeated back on of the numbers from the question. In Two-System terms, of course, a mathematical solution requires the use of deliberate System 2 thinking, while repeating back an observed number as an answer seems likely to be a System 1 shortcut. Given uncertainty regarding this division and for additional reasons discussed later (Section4.2), however, we prefer to maintain the distinction between mathematical and non-mathematical, rather than System 1 and System 2 responses.

3.2. Predictors of Base Rate Neglect

3.2.1. Demographics

Participants' age, sex and educational level were recorded for each of the four base rate response groups. These data are displayed in Table 2.

Looking at Table 2, one sees that for two variables there seems to be an observable pattern. Both females and more educated people tend to fall into the 'better' response groups. The Bayesian group, for example, is composed entirely of females who are attending or have graduated from university, whereas the Base Rate Neglect Group is almost 50% male and more than a third of its members have not attended university. Age, by comparison, seems to have no predictable value – perhaps due to homogeneity in our sample.

As both Sex and Education are categorical, chi-squared tests are suggested to determine whether the observed relationships were statistically significant. However, given the very small numbers in some cells, the data were analyzed after dividing participants into "mathematical" (i.e., Simple Maths and Bayesian) and "non-mathematical" (Base Rate Neglect and Overreliance) base rate response groups – that is, System 1 versus System 2 responses from Table 1. This showed no significant association between Sex and response group, $\chi^2(1, N=80) = 2.0$, $p = 0.16$, but indicated a significant result for Education, $\chi^2(3, N=80) = 9.7$, $p = 0.008$, with higher Education predicting an increased likelihood of responding with a mathematically derived solution to the base rate problem.

3.2.2. Cognitive Ability Measures. In total, four cognitive ability tests were used. Three of these were 'positive' tests for which higher scores indicated better performance, whereas the last (the SART) was a 'negative' test that counted errors and thus had lower scores indicate better performance. Results for all four measures are presented in Figure 2.

Inspection of Figure 2 appears to show a relationship between cognitive ability and base rate response group. The Bayesian response group performed best on all four cognitive measures, whereas the Simple Maths group was the second best performed on three of the four and third on the other. The worst performing group in each case was either the Base Rate Neglect group (3 out of 4) or the Base Rate Overreliance group (1 out of 4).

Given the four groups being compared, a one-way ANOVA is initially suggested as a statistical analysis. Given the very small number of people in the Bayesian response group, however, a more robust method is the use of *t*–tests comparing the scores of participants who gave mathematical and non-mathematical solutions – following what was done in Section 3.2.1, above. Table 3 summarizes these tests.

Table 3 indicates that all of the relationships are in the expected directions – with people who gave mathematical solutions having better scores on the cognitive ability variables – although only one, the SART (measuring executive function), is significant is its own right. Examination of the Cohen's *d* values, however, indicates that the SART has a medium effect on response type, while weak effects were obtained for Numerical Ability and Symbol-Digit (processing speed). Finally, the effect of Dot Matrix (working memory) is very weak - although still of comparable strength to those observed, for example, by Frederick (2005) for his CRT measure, self-reported SAT or ACT scores and IQ scores (as measured by the 15 minute Wonderlic Personnel Test; Wonderlic, 1973) – for which correlations with decision making biases (although not base rate neglect) ranged from 0 to a maximum of 0.24.

3.2.3. Decision Style and Personality

The decision style measures relating to intellect and rationality measures – Need for Cognition, the four Rational-Experiential Inventory measures, Intellect and Rationality - on inspection of an initial correlation matrix (see Appendix A), seemed to form a related cluster. In order to simplify analyses, these seven variables were, therefore, subjected to an exploratory factor analysis. A parallel roots analysis (Horn, 1965) of the seven scores indicated a three-factor solution, with loadings for a Promax rotated solution shown in Table 4.

An examination of Table 4 reveals a sensible clustering of the variables – with the Factor 1 variables of Intellect, NfC and RE all reflecting a person's willingness or desire to engage in rational approaches to a problem; Factor 2's EE and EA reflecting aspects of intuition; and Factor 3's RA and Rationality both measuring a person's ability to behave rationally – as opposed to their preference for it. Given this, the factors were renamed: Preference for Rationality (PR); Intuition (IN); and Aptitude for Rationality (AR). These factors were used in subsequent analyses in preference to the seven original variables that loaded on these factors.

Figure 3 shows the mean scores on each of the remaining decision style/personality variables achieved by each of the four base rate response groups.

Looking at Figure 3, there is less evidence for an overall trend than was observed in the cognitive ability data. Given the descriptions of the variables (and factors), one might have expected four – Cognitive Reflection Task (CRT), Preference for Rationality, Aptitude for Rationality and Instrumental Risk - to increase with response group number, while the others – Intuition, Stimulating Risk and the Decision Outcomes Inventory – would decrease. In fact, only for Intuition is there a clear pattern conforming to this expectation.

As was the case for the cognitive abilities, given the very small number of participants in the Bayesian group and differences in variance between the groups, Welch's *t*-tests rather than ANOVAs were conducted, comparing the scores achieved by participants who responded with mathematical versus non-mathematical solutions to the base rate problem. Table 5 summarizes these results.

Table 5 indicates three significant effects: Stimulating Risk, Intuition and Preference for Rationality, in descending order of strength. The first two effects, being measures of a person's tendency towards preferring or using intuitive reasoning and of responding impulsively, respectively, are also in the expected direction, with the groups who responded mathematically having lower scores than those who responded non-mathematically. In both cases, Cohen's *d* indicates that the effect is of medium strength. The only other variable expected to show a negative relationship – that is, where non-mathematical respondees score higher – was the Decision Outcomes Inventory. In fact, this showed no effect at all, with the two groups having exactly the same mean score on this test.

The remaining four variables were all expected to show positive relationships, with the mathematical respondees scoring more highly. Further inspection of Table 5 shows, however, that this effect was only observed for the Instrumental Risk measure, which was a very weak, non-significant predictor of response type. The remaining three measures all showed effects in the opposite direction, with Preference for Rationality reaching significance and having a weak to medium effect on response type while Aptitude for Rationality and CRT showed very weak and near-zero effect sizes, respectively.

4. Discussion

Overall, the results were consistent with the proposition that individual differences in cognitive abilities are related to people's susceptibility to base rate neglect, predicting the types of responses that they are likely to make. Education level – often used as a proxy measure of cognitive ability - was also significantly related to base rate response type. By comparison, the relationships between base rate neglect and decision style/personality measures were much more variable and less convincing. These points are discussed in greater detail in Sections 4.1 and 4.2, below.

4.1. Cognitive Abilities and Base Rate Neglect

The measure of sustained attention and executive control (SART) was a significant predictor of response type on the base rate neglect problem, which makes sense in terms of a Two System approach to decision making as executive functioning reflects a person's ability to monitor their decisions for errors – and would thus be an excellent candidate for provoking a switch between intuitive System 1 and reasoned System 2 styles of thinking.

By comparison, numerical ability (NAT) and processing speed (Symbol-Digit) showed consistent but non-significant, weak effects. Similarly, working memory (Dot Matrix), had an effect in the expected direction but was even more weakly predictive of responses to the base rate problem, which may reflect decreasing relevance to the task at hand of these cognitive abilities. In summary, therefore, it appears that, following the executive functioning measures (SART), numerical ability, with clear face validity in this type of task, yielded the largest effect, followed by processing speed, which may have influenced how long a participant would take to reach the solution and thus how likely they are to pursue such reasoning. By contrast, working memory very likely had comparatively little effect beyond its shared variance with these other cognitive abilities.

These effects are, in most cases, as strong or stronger than those commonly reported by researchers who have used more general measures of cognitive ability (see, e.g., Bruine de Bruin et al., 2007; Frederick, 2005; Stanovich & West, 2008), suggesting that looking at cognitive abilities in finer detail than has, previously, been done provides greater insight into the relationship between bias susceptibility and cognitive ability. At the same time, the results also indicate that task performance on base rate neglect problems needs to be more carefully considered. Although adoption of the Bayesian solution is often considered indicative of System 2 processing, our results also show that variations in systematic processing also need to be considered. As also shown by Peters et al. (1996) and discussed by Reyna and Brainerd (2008), numerate people can achieve erroneous solutions using mathematical reasoning. Thus, as we have shown, adoption of a clearly systematic, mathematical (System 2) solution can lead to a *worse* estimate than the (System 1 based) strategy of simply accepting the base rate (depending, of course, on the particular probabilities in the updating task). An assessment of performance solely confined to a group-level comparison of Bayesian and other solutions may have missed this observation.

4.2. Decision Style, Personality and Base Rate Neglect

In contrast to the clear results of the cognitive ability measures, the use of decision-style/personality measures led to a number of unexpected results. For example, somewhat surprisingly, we did not confirm the prediction that the CRT would predict responses on the base-rate problem as it has been shown to do with some other biases (Frederick, 2005).

One possibility is that the CRT actually measures something other than 'cognitive reflection'. For example, in our data (Table A1), CRT correlated with numerical ability at 0.44 – reflecting the fact that the CRT questions are all numerical in nature - but the rank-order correlation between the CRT and SART was only -0.06, which seems very low given the parallels between System 2 thinking and executive functioning discussed above (Section 1.3).

A second, although related, explanation relates to the whether there are differences between base rate and other bias tasks. Although there is evidence in the literature to suggest relationships between performance on different bias tasks (Bruine de Bruin et al., 2007; Reyna & Brainerd, 2008), there are also factors which can militate against this effect. For example, the CRT is designed to measure a person's *preference* for engaging System 2 but is relatively silent on what happens next. This is because the CRT questions, once System 2 has been engaged, are trivially easy – unlike a base rate neglect problem.

This same explanation makes sense for the unexpected results of our constructed Preference for Rationality measure. Specifically, while we can identify those people who definitely engaged System 2, we cannot rule out the possibility (or even likelihood) that the groups we identified as responding in a System 1-like fashion contain participants who did, in fact, engage System 2 but then could not find the correct answer and thus *reverted* to their initial impression. Thus, the difficulty of the base rate neglect problem could hide people's preference for System 2 reasoning, with the result that the CRT would *not* predict performance - as we observed.

That Preference for Rationality significantly predicted response type but in the opposite direction (as, did Aptitude for Rationality to a lesser, non-significant extent) seems more surprising. However, considered in light of the above argument, this too can be made sense of. Specifically, while someone may have a preference or aptitude for rationality, this will not necessarily enable them to construct, on the spot, the necessary understanding of probability theory required to produce a Bayesian solution but *may* result in their realizing that the simple mathematical solution is incorrect. This would lead to the situation, observed in our data, where people with higher Preference and Aptitude for Rationality are *less* likely to respond with the Simple Maths solution and thus less likely, overall, to display an obvious System 2 response (as the Simple Maths group is the majority of the mathematical response group).

By comparison with this, the results for the measures relating to intuition and impulsivity (our constructed Intuition scale and the Stimulating Risk scale of the Stimulating Instrumental Risk Inventory, respectively) were fairly straightforward, with more intuitive and more impulsive people showing a significant tendency to respond in a System 1-like manner. That is, responding with a non-mathematical answer.

Instrumental Risk, which reflects cognitive, long-term and reflective approaches to risk, also showed a very weak effect in the expected direction. This makes sense if the 'risk' being considered is the risk of being wrong and there is some tendency for people with higher Instrumental Risk to, therefore, apply more effort to the problem to avoid this undesired outcome.

Finally, the Decision Outcomes Inventory scales failed to predict people's base rate neglect task responses at all. The reason for this is harder to isolate – although the fact that the effect is driven by the Simple Maths response group suggests a similar cause to that discussed for Preference for Rationality, above. That is, people with a history of making better decisions may have been more likely to spot that the Simple Maths solution was wrong and thus revert to an (equally wrong) alternative.

4.3. Caveats

The interpretation of the findings in this study needs to be qualified by several caveats. The overall sample size combined with the rarity of Bayesian solutions meant that only five participants produced such a response - so that the study was clearly under-powered to make comparisons between the four response groups identified in Section 3.1.2. This lack of power was compounded by the (we believe necessary) decision to exclude 21 participants from the analyses as a result of their response not being clearly classifiable.

A secondary concern is whether our sample is non-representative. This question arises from the fact that the overall level of base rate neglect observed in our sample is negligible, with the average estimate being almost exactly the Bayesian solution. However, such an interpretation is, we would argue, inappropriate given the results herein and previous evidence (Welsh & Navarro, in press) that has demonstrated that base rate neglect primarily occurs as the result of the proportion of individuals adopting different response strategies rather than any internal weighting of probabilities (an effect also observed in Bar-Hillel's ,1980, data). Understood in this light, base rate neglect cannot be examined at the group level through a simple averaging and must, instead, be regarded as occurring only where individual responses ignore base rate data.

That said, a comparison of our results with Bar-Hillel (1980) does suggest that our sample differs from the original sample asked to answer the taxi-cab problem. Specifically, Bar-Hillel had no participants correctly provide the Bayesian solution and neither was there the large spike at 12% that we have observed in our data. This suggests that our sample is more mathematically inclined than Bar-Hillel's and or that a greater proportion of people have now seen problems of this sort and thus realize that a mathematical solution is necessary. In either case, however, this does not undermine the results here – instead just suggesting an even greater need for analyses of individual differences in base rate studies.

This examination of Bar-Hillel's (1980) results also, however, suggests that the number of participants required to get an adequate representation of each response type within a sample may be prohibitively large. Even with our more ‘aware’ sample, less than 5% of people responded with the Bayesian solution, meaning that to get a group of sufficient size to find the relatively weak effects commonly observed linking cognitive ability and bias susceptibility, may require lengthy cognitive testing on many hundreds or even thousands of participants.

Finally, only a relatively limited range of individual differences were included in the study. In terms of cognitive abilities, we included measures of executive functioning and three of the ten specific cognitive abilities identified by the CHC theory of intelligence (McGrew, 2005). It is, therefore, possible that inclusion of additional abilities would further refine our results. Similarly, the decision-style/personality measures we included, while selected for what we believe are sound, theoretical reasons, exclude many facets of the Big 5 personality measures (Costa & McCrae, 1992) that might, potentially, impact on decision-making and base rate neglect specifically. For example, one candidate is extraversion, which has previously been linked to susceptibility to other biases (Eroglu & Croxton, 2010), perhaps due ot its capturing of higher impulsivity?

4.4. Future Research

Although this study extended the range of cognitive abilities examined over previous studies, there remain many abilities that could be predictive of susceptibility to specific biases. For example, as Bruine de Bruin et al. (2007) suggest, there may be value in conducting comparisons of abilities that coincide with the *gc* and *gf* components of the CHC model. Ideally, of course, base rate neglect and other cognitive biases would be compared with all of the specific intelligences from the CHC model and all facets of the Big 5 personality measures.

Another line of research would be to examine the relationships between the wide range of biases discussed under the ‘heuristics and biases’ umbrella. While taxonomies do exist, these have largely (but see Bruine de Bruin et al., 2007) been created *a priori* based on researchers’ perceptions regarding the underlying processes leading to the bias (see, e.g., Tversky & Kahneman, 1974). There exists, therefore, an opportunity to test these taxonomies against observed susceptibilities to different biases.

A third issue is to ensure that participants are recruited from a wide range of ability levels and education backgrounds and, consistent with item-response theory, that tasks with varying levels of difficulty are chosen for the studies.

A more specific hypothesis that might follow from the research presented here, for example, would be that those people who tend to respond by simply repeating one of the numbers presented in the base rate neglect problem might be the same people as would be most susceptible to the anchoring effect – where people use given numbers as the basis for their responses (Tversky & Kahneman, 1974).

4.5. Conclusions

Participants' cognitive abilities – specifically their executive functioning but also, to decreasing extents, their numerical reasoning, processing speed and short term memory – and level of education all showed some predictive power in terms of whether participants would respond with a non-mathematical or mathematical solution to the base rate neglect problem. Of particular interest is that fact that, in the SART measure of executive function, we have a predictive variable that corresponds quite closely to important aspects of the proposed System 2 reasoning style involving the recognition and overruling of automatic errors.

The decision style measures, however, showed a less clear relationship. This, we argue, is because a preference for responding in a certain way does not, in the case of a complex probability updating task (like a base rate neglect task), actually enable the person to solve the problem, with the result that people with a preference to respond rationally can still fail to solve the problem and become indistinguishable in terms of their responses from people who prefer to respond intuitively. This leads to the observation that the division of people's responses into System 1 and System 2 types is significantly harder than it seems at first glance, with the overlap between this distinction and the mathematical/non-mathematical distinction that we draw herein being far from perfect.

Overall, the results presented here support the idea that a deeper understanding of cognitive biases can be achieved through the application of individual differences research; both in terms of paying more attention to the types of responses that individual participants make and through a consideration of the psychometric qualities of individuals. Specifically, we would argue that an individual differences approach to base rate neglect will be far more fruitful than looking for explanations of the average effect – as the average describes the behavior of almost no participants at all. This will assist not only in advancing cognitive theories of decision making but also in the development of regimes to avoid or reduce bias within industries where cognitive biases can have multi-million dollar impacts.

Acknowledgments

{authors} thank ExxonMobil and Santos for their funding of the {institution}. The authors thank { } for her assistance with data collection.

Appendix A

Table A1 shows the correlations between all of the cognitive ability, decision style/personality and demographic predictor variables collected for our analyses.

References

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

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1989). *Differential Aptitude Test*. Marrickville, NSW: Psychological Corporation.

Bruine de Bruin, W., Parker, A.M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology, 92*, 938-956.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131.doi:10.1037/0022-3514.42.1.116

Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies.* Cambridge, UK: Cambridge University Press.

Cattell, R.B. (1973). *Personality and mood by questionnaire*. San Francisco: Jossey-Bass.

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.

Costa, P. T. J., & McCrae, R. R. (1992). *NEO PI-R Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in Intuitive-Experiential and Analytical-Rational Thinking Styles. *Journal of Personality and Social Psychology, 71*(2), 390-405. doi:10.1037/0022-3514.71.2.390

Eroglu, C. & Croxton, K.L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting, 26,* 116-133.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25-42.

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

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Goodie, A.S., & Crooks, C.L. (2004). Time pressure effects on performance of a base-rate task. *Journal of General Psychology, 131*, 18-25.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237-251.

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

Law, D.J., Morrin, K.A., Pellegrino, J.W. (1995). Training effects and working memory contributions to skill acquisition in a complex coordination task. *Learning and Individual Differences, 7,* 207–234.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp.136–182). New York: Guilford.

McElroy, T., & Seta, J.J. (2003). Framing effects: An analytic-holistic perspective. *Journal of Experimental Social Psychology, 39*, 610-617.

McPherson, J.L & Burns, N.R. (2005). A speeded coding task using a computer based mouse response. *Behavior Research Methods, Instruments, & Computers, 37,* 538-544.

Peters, E., Vastfjall, D., Slovic, P., Mertz, C.K., Mozzoco, K., & Dickert, S. (2006). Numeracy and decision-making. *Psychological Science, 17,* 406-413.

Reyna, V.F., & Brainerd, C.J. (2008). Numeracy, ratio bias and denominator neglect in judgments of risk and probability. *Learning and Individual Differences, 18*, 89-107.

Robertson, I.H., Manly, T., Andrade, J., Baddeley, B.T., Yiend, J. (1997). 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35,* 747–758.

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).

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General,* 127(2), 161-188.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral and Brain Sciences*, 23(645-726).

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, 94(4), 672-695.

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

Welsh, M. B., & Navarro, D. J. (in press). Seeing is believing: priors, trust and base rate neglect. *Organizational Behavior and Human Decision Processes*. Accepted April 3rd 2012. http://dx.doi.org/10.1016/j.obhdp.2012.04.001

Welsh, M. B., Bratvold, R. B., & Begg, S. H. (2005). SPE 96423 - Cognitive biases in the petroleum industry: impact and remediation. *Proceedings of the Society of Petroleum Engineers 81st Annual Technical Conference and Exhibition*.

Welsh, M. B., Begg, S. H., & Bratvold, R. B. (2007). SPE 110765: Modeling the economic impact of cognitive biases on oil and gas decisions. *Proceedings of the Society of Petroleum Engineers 83rd Annual Technical Conference and Exhibition*.

Wonderlic, E. F. (1973). *Wonderlic Personnel Test*. Los Angeles: Western Psychological Services.

Zaleskiewicz, T. (2001). Beyond risk seeking and risk aversion: Personality and the dual nature of economic risk taking. *European Journal of Personality*, *15*(S1), S105-S122. doi:10.1002/per.426

Table 1. Response category memberships and corresponding ‘super-group’.

Response Category	# Members	Super-group
1. Base Rate Neglect	35	Non-mathematical
2. Base Rate Overreliance	13	Non-mathematical
3. Simple Math. Calculation	27	Mathematical
4. Bayesian Calculation	5	Mathematical

Table 2. Comparisons of demographic measures by base rate response group.

	Base Rate Response Group			
	1. Neglect	2. Overreliance	3. Simple Math	4. Bayesian
n	35	13	27	5
Sex: % Male	48.6%	30.8%	33.3%	0.0%
Age: *M* (*SD*)	21.6 (*3.3*)	21.7 (*3.0*)	24.0 (*6.0*)	20.6 (*1.4*)
Education:				
- 1. % No Uni	37.1%	23.1%	7.4%	0.0%
- 2. % Current Uni	42.9%	46.2%	74.1%	80.0%
- 3. % Graduate	20.0%	30.8%	18.5%	20.0%

Note: the education group percentages for Overreliance sum to more than 100% due to rounding.

Table 3. Summary of Welch's t-tests for cognitive ability variables comparing participants using mathematical and non-mathematical strategies in base rate neglect problem.

Variable	M_1 *(SD)*	M_2 *(SD)*	*t*	df	*p* (2-tailed)	Cohen's *d*
Numerical Ability	6.9 (2.5)	7.7 (1.8)	1.56	77	.123	0.34
Symbol-Digit	85.0 (18.0)	90.5 (16.2)	1.42	71	.159	0.32
Dot Matrix	43.4 (7.2)	44.9 (6.7)	0.99	70	.325	0.22
SART	**7.2 (9.6)**	**3.1 (3.1)**	**-2.80**	**60**	**.007**	**-0.54**

Note: Significant results are emboldened. Welch's *t*-tests were used given concerns about equality of variance between the groups. M_1 is mean of the non-mathematical, System 1, response group. M_2 is that of the mathematical, System 2, group.

Table 4. Factor loadings of 7 decision-style variables.

	Factor 1	Factor 2	Factor 3
Intellect	**0.99**	0.06	-.03
Need for Cognition (NfC)	**0.87**	-0.01	0.06
REI –Rational Engagement (RE)	**0.67**	-0.04	0.11
REI – Experiential Engagement (EE)	0.13	**0.90**	-0.22
REI – Experiential Ability (EA)	-0.12	**0.83**	0.23
REI – Rational Ability (RA)	-0.05	-0.09	**0.86**
Rationality	0.24	0.06	**0.62**

Note: Primary factor loadings are emboldened.

Table 5. Summary of Welch's *t*-tests for decision style variables comparing participants using mathematical and non-mathematical strategies in base rate neglect problem.

Variable	M_1 *(SD)*	M_2 *(SD)*	t	d.f.	p (2-tailed)	Cohen's d
CRT	1.33 (1.15)	1.28 (0.92)	-0.22	75	0.824	-0.05
Pref. Rationality	**0.23 (0.99)**	**-0.20 (0.90)**	**-2.00**	**71**	**0.049**	**-0.45**
Apt. Rationality	0.16 (0.97)	-0.07 (0.93)	-1.03	69	0.307	-0.23
Intuition	**0.36 (1.47)**	**-0.33 (1.32)**	**-2.20**	**71**	**0.031**	**-0.49**
Stim. Risk	**22.9 (5.0)**	**20.2 (4.4)**	**-2.51**	**67**	**0.014**	**-0.57**
Instr. Risk	18.6 (3.4)	19.0 (2.9)	0.58	71	0.567	0.13
Decision O.I.	12.2 (4.1)	12.2 (9.3)	0.00	39	1.000	0.00

Note: Significant results are emboldened. Welch's *t*-tests were used given concerns about equality of variance between the groups. M_1 is mean of the non-mathematical, System 1, response group. M_2 is that of the mathematical, System 2, group.

Table A1. Pearson correlations between all predictor variables.

Variables	Sex*	Age	Ed.	NAT	SD	DM	SART	CRT	SR	IR
Exp. Ability	.471	.372	.647	.190	.311	**.027**	.445	.626	.181	.573
Exp. Engage	.782	.957	.159	.167	.907	**.028**	.371	.577	.093	.107
R. Ability	**.034**	.523	**.030**	**.013**	.081	**.024**	.896	**.001**	.417	.227
Rat. Engage.	.211	.607	.631	.888	.428	.213	.477	**.030**	.882	.874
Rationality	.567	.707	.798	.478	.963	.100	.726	.060	.767	.540
Intellect.	.464	.704	.147	.821	.825	.528	.905	.055	.540	.820
Need f Cog.	.507	.932	.524	.953	.531	.447	.937	**.016**	.628	.715
DOI	.314	.784	.658	.640	.339	.438	.344	.963	.085	.621
Instr. Risk	**.041**	**.030**	.373	**.038**	.153	.171	.782	.144	**<.001**	-
Stim. Risk	**<.001**	**.010**	**.018**	.379	.177	1.00	**.044**	.215	-	**0.38**
CRT	**.019**	.124	**.014**	**<.001**	**.011**	**<.001**	.681	-	-0.13	0.15
SART	**.006**	.217	**.043**	.264	.519	.285	-	0.04	**0.20**	0.03
Dot Matrix	.296	.173	**.018**	**<.001**	**<.001**	-	-0.11	**0.34**	0.00	0.14
Symbol Digit	.407	.089	**.028**	**<.001**	-	**0.55**	-0.06	**0.25**	-0.14	0.14
NAT	.176	.080	**.001**	-	**0.38**	**0.47**	-0.11	**0.44**	-0.09	**0.21**
Education *	.527	**.008**	-	**0.34**	**0.22**	**0.23**	**-0.20**	**0.24**	**-0.24**	0.09
Age	.763	-	**0.26**	0.17	-0.17	-0.14	-0.12	0.15	**-0.26**	**-0.22**
Sex*	-	0.03	0.06	0.13	-0.08	0.10	**0.27**	**0.23**	**0.34**	**0.20**

DOI	NfC	Int.	Rat.	RE	RA	EE	EA
.397	**<.001**	**<.001**	**<.001**	**.007**	.061	**<.001**	-
.470	**.007**	.080	**<.001**	.061	.690	-	**0.73**
.065	**<.001**	**<.001**	**<.001**	**<.001**	-	-0.04	0.19
.077	**<.001**	**<.001**	**<.001**	-	**0.45**	0.19	**0.26**
.213	**<.001**	**<.001**	-	**0.72**	**0.50**	**0.38**	**0.40**
.064	**<.001**	-	**0.70**	**0.53**	**0.63**	0.17	**0.34**
.280	-	**0.66**	**0.89**	**0.68**	**0.49**	**0.26**	**0.35**
-	-0.11	-0.18	-0.12	-0.18	-0.18	-0.07	-0.08
0.05	0.04	-0.02	-0.06	0.02	0.12	-0.16	-0.06
0.17	0.05	-0.06	0.03	0.01	-0.08	0.17	0.13
0.00	**0.24**	0.19	0.19	**0.21**	**0.33**	-0.06	0.05
0.09	-0.01	0.01	0.04	-0.07	-0.01	0.09	0.08
0.08	-0.08	-0.06	-0.16	-0.12	**0.22**	**-0.22**	**-0.22**
-0.10	0.06	0.02	0.00	-0.08	0.17	-0.01	-0.10
0.05	-0.01	0.02	-0.07	-0.01	**0.25**	-0.14	-0.13
-0.04	0.06	0.14	-0.03	-0.05	**0.22**	-0.14	-0.05
0.03	-0.01	0.04	-0.04	0.05	0.06	0.01	0.09
0.10	0.07	0.07	0.06	0.12	**0.21**	-0.03	0.07

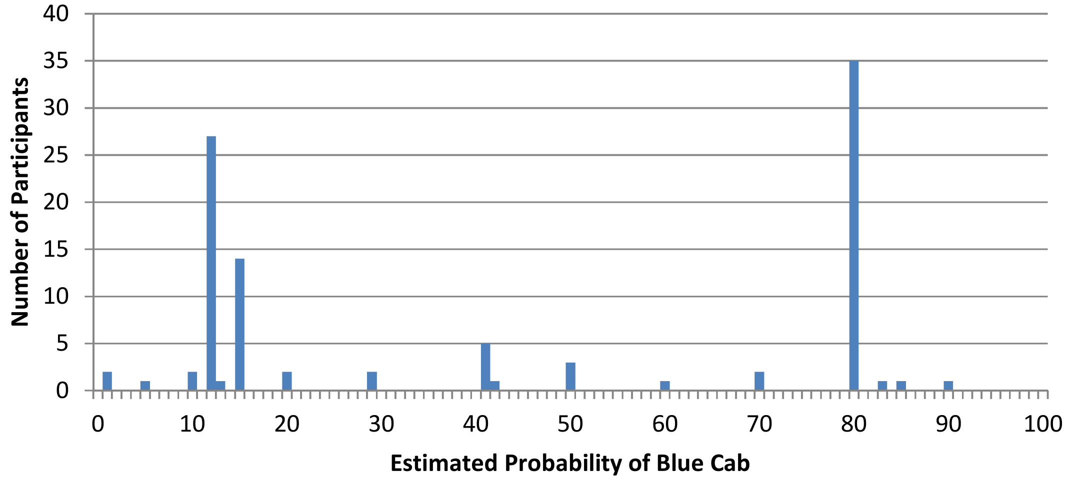
Note: Values in the bottom triangle are r-statistics (N=102, except for SR and IR, N=100); the top triangle displays corresponding p-values (two-tailed); significant results are emboldened. * - Sex and Education are included for completeness, despite being categorical.

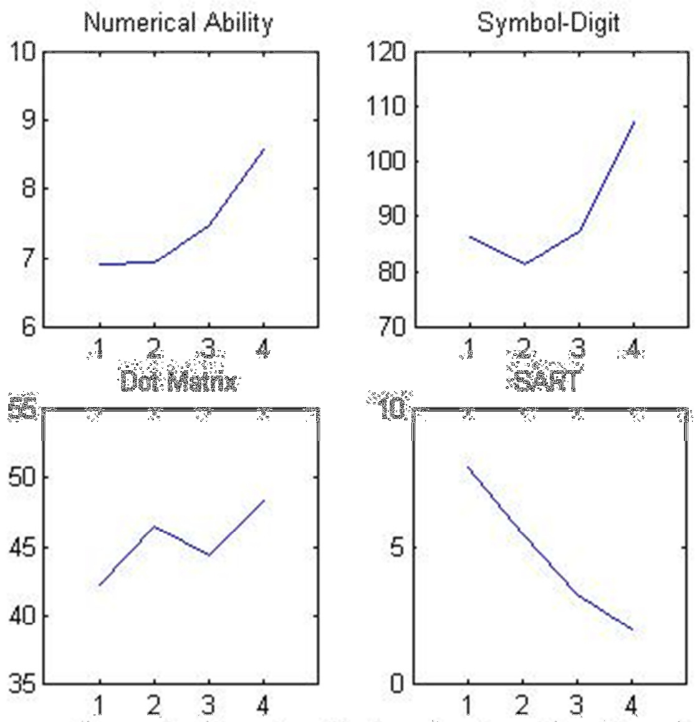
Figure Captions

Figure 1. Histogram of participants' estimated rates

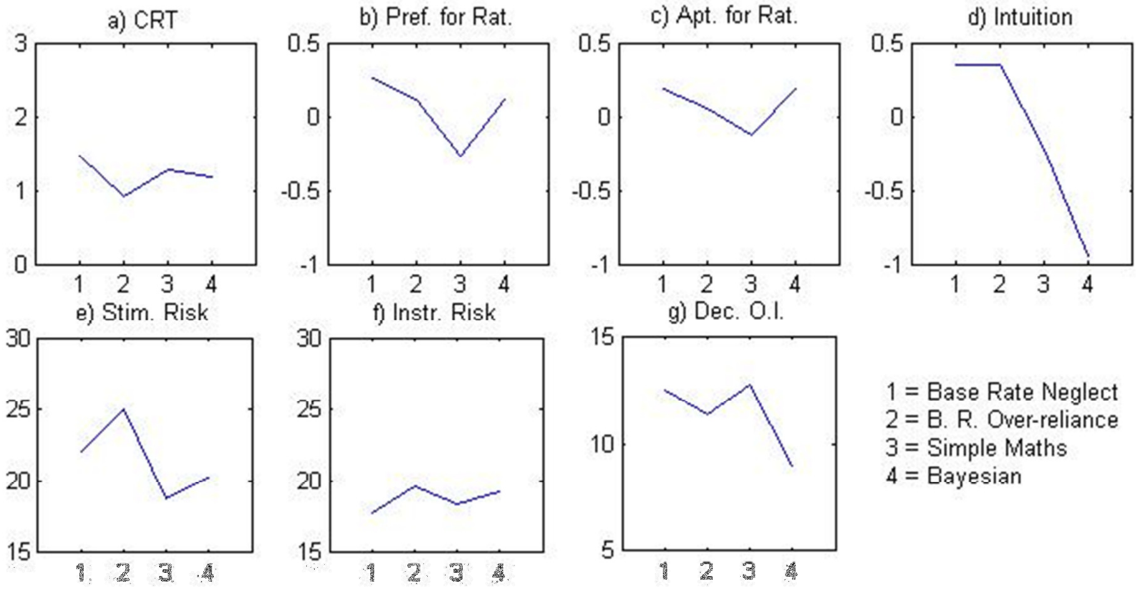
Figure 2. Mean cognitive ability scores by base rate response group.

Figure 3. Mean scores on decision style/personality variables by base rate response group.


Number of Participants
40
35
30
25
20
15
10
5
0
0
10
20
30
40
50
60
70
80
90
100
Estimated Probability of Blue Cab



Response Group: 1 = Base Rate Neglect; 2 = Base Rate Overreliance;
3 = Simple Maths; 4 = Bayesian


a) CRT
b) Pref. for Rat.
c) Apt. for Rat.
d) Intuition
e) Stim. Risk
f) Instr. Risk
g) Dec. O.I.
1 = Base Rate Neglect
2 = B. R. Over-reliance
3 = Simple Maths
4 = Bayesian