Conflict detection in dual-process theory:

Are we good at detecting when we are biased in decision making?

by

Gordon Pennycook

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2011

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Gordon Pennycook

Abstract

In the domain of reasoning and decision making, some dual-process theorists have suggested that people are highly efficient at detecting conflicting outputs engendered by competing intuitive and analytic processes (De Neys & Glumicic, 2008; De Neys, Vartanian & Goel, 2008). For example, De Neys and Glumicic (2008) demonstrated that participants' reason longer about problems that are characterized by a conflict between a stereotypical personality description and a base-rate probability of group membership. Crucially, this increase occurred even when participants gave the nominally erroneous stereotypical response (i.e., "neglecting" the base-rate probability), indicating that their participants detected that there was a conflict and, as a result, engaged in slow, analytic processing to resolve it. However, this finding, and much of the additional support for the efficient conflict detection hypothesis, has come from base-rate neglect problems constructed with probabilities (e.g., 995 doctors and 5 nurses) that were much more extreme than typically used in studies of base-rate neglect. I varied the base-rate probabilities over five experiments and compared participants' response time for conflict problems with non-conflict problems. It was demonstrated that the integral increase in response time for stereotypical responses to conflict problems was fully mediated by extreme probabilities. I conclude that humans are not as efficient at detecting when they are engaging in biased reasoning as De Neys and colleagues have claimed.

Acknowledgments

I would like to thank Dr. Jonathan Fugelsang and Dr. Derek Koehler for their help and input with the project. I would also like to thank Dr. Derek Besner and Dr. Colin MacLeod for agreeing to review this manuscript. Finally, special thanks to Manar Shaker for her assistance in collecting data.

Table of Contents

List of Tables

List of Figures

Introduction

It has long been known that reasoning and decision making are often heavily influenced by systematic biases. These biases have been shown to produce behaviour that tends to deviate substantially from ostensibly normative performance on a variety of reasoning tasks (see Kahneman, Slovic, & Tverksy, 1982, for a review). For example, people violate utility theory, demonstrate overconfidence, display confirmation bias, fail to properly calibrate degrees of belief, allow prior belief to bias deductive reasoning, and, crucially for the current work, deviate from Bayes Theorem when calculating probabilities (for reviews, see Baron, 1994; 1998; Evans, 1989; Evans & Over, 1996; Kahneman, 2003; Kahneman et al., 1982; Shafir & Tverksy, 1995; Stanovich & West, 2000).

Such findings have sparked a long-running debate concerning the nature and quality of human rationality (e.g., Gigerenzer, 1996; Stanovich & West, 2000; Stein, 1996). Whereas some argue that these biases indicate that cognition is characterized by troubling irrationalities, others argue that reasoning deviates from normative standards for other reasons, thereby avoiding any attribution of irrationality to participants (Stanovich, 1999). For example, one possibility is that reasoning deviates from normative standards because of simple performance errors caused by sporadic and temporary information processing mishaps, such as lapses of attention or memory deactivation (Stanovich & West, 2000).

In a series of recent studies, De Neys and colleagues have demonstrated that participants are aware, albeit implicitly, when they are being biased (De Neys, Comheeke, & Osman, 2011; De Neys & Franssens, 2009; De Neys & Glumicic, 2008; De Neys, Vartanian, & Goel, 2008). The purpose of the current work is to assess De Neys and colleagues' evidence supporting the view that reasoning deviates from normative standards due to an additional temporary information processing mishap: namely, the failure to inhibit a prepotent response cued by diagnostic information. Specifically, I will show that participants only detect when they are biased under a fairly narrow set of conditions—that their awareness is much more limited than De Neys and colleagues have claimed. Additionally, De Neys and colleagues have drawn fairly broad conclusions from their research, implying a level of generalization for which I have not found empirical support.

Base-Rate Neglect

One of the most heavily studied of the aforementioned biases is the neglect of prior probability in lieu of diagnostic information during judgments of likelihood (see Barbey & Sloman, 2007, for a review). Kahneman and Tversky (1973) first presented evidence for this base-rate "neglect" in what would later be referred to as the 'Tom W.' problem. For the Tom W. problem, participants were divided into three groups. The base-rate group was given a list of graduate student areas of specialization and asked to rank them based on the probability of membership. In contrast, the similarity group was given a description of a university student consisting of stereotypical diagnostic information and asked to rank the same areas of specialization based on the similarity of the prototypical student in each area to the description. Finally, the prediction group was given the personality description and asked to rank the areas of specialization based on the likelihood that Tom W. was enrolled as a graduate student. In terms of Bayes Theorem, to produce the best answer, it is necessary to take into account both the diagnosticity of the personality description and the prior probability (i.e., the base-rate). However, Kahneman and Tverksy found that the mean rank for the prediction group closely matched the ranking for the similarity group (r = .97) and deviated greatly from the mean ranking for the base-rate group (r = -.67). Thus, the participants neglected the base-rate information and based their response instead almost entirely on the representativeness of the personality description.

The Tom W. problem was of particular importance because it demonstrated a strong bias in a type of judgment that is made frequently in everyday life. Kahneman and Tverksy (1973) went on to show that base-rate information is underweighted in surprising additional contexts. For example, in the "lawyer-engineer" problem, participants are explicitly presented the base-rate information:

> A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers.
>
> Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.
>
> The probability that Jack is one of the 30 engineers in the sample of 100 is ___%.

Kahneman and Tversky found that participants relied on the representativeness of the personality description despite the apparent accessibility of the explicitly presented base-rate information. While explicitly presenting the base-rate information may make the lawyer-engineer problem somewhat less ecologically valid than the Tom W. problem, it does demonstrate the truly systematic nature of the bias.

It should be noted, however, that there are conditions where participants will deviate from the representativeness heuristic and make better use of the base-rate information. For example, Evans, Handley, Over, and Perham (2002) found that, when the relative size of the base-rates was manipulated either within or between subjects, participants gave probability estimates that more closely reflected the base-rates for the larger (400 out of 1000) relative to the smaller (100 out of 1000) base-rates. This suggests that participants are at least somewhat sensitive to variations in prior probabilities over repeated problem presentation. In a similar vein, Gigerenzer, Hell, and Blank (1988) found that participants gave probability estimates that conformed closer to Bayes Theory when they performed and observed the random sampling themselves. Findings such as these suggest that base-rates are simply underweighted as opposed to wholly neglected. This, in turn, indicates that the use of base-rates is strongly influenced by the context in which they are presented.

Dual-Process Theory and Base-Rate Neglect

According to dual-process theories (for reviews, see Evans, 2008; Frankish & Evans, 2009; Stanovich, 2004), reasoning and decision making are based on two qualitatively different types of cognitive systems (Kahneman, 2003; Stanovich, 2004) or processes (Evans, 2009): System 1 (S1; also referred to as "Type 1") is defined by fast, frugal, unconscious, intuitive, and heuristic types of processes whereas System 2 (S2; also referred to as "Type 2") is slow, deliberative, conscious, and analytical. The key behavioural result of this distinction is that the output of S1 processing tends to dominate reasoning such that participants will generally have quite poor performance on problems that elicit an incorrect intuitive response (Evans, 2008). This performance deficit (as compared to normative standards) is usually explained by a lack of S2 override – i.e., participants were unable to engage in sufficient analytic processing to overcome their initial intuitive response (Evans, 2008; Frankish & Evans, 2009; Kahneman, 2003; Stanovich, 2004).

In their review of base-rate neglect, Barbey and Sloman (2007) concluded that dual-process theory best explains the various forms of the phenomenon. Specifically, Barbey and

Sloman state that participants tend to neglect or underweight prior probabilities because of an overreliance on heuristics that fail to fully represent the structure of the problem. In terms of base-rate problems that contain diagnostic information in the form of stereotypes (such as the lawyer-engineer problem above), S1 processes are thought to engender an intuitive response based on the representativeness of the stereotypical personality description. Base-rate neglect is then caused by failure of S2 processing to override the prepotent response (Stanovich & West, 1998; 2000). This focus indicates that the mechanism that initiates S2 processing is critical in explaining base-rate neglect. One proposed mechanism for such initiation is conflict detection (De Neys & Glumicic, 2008).

In the lawyer-engineer problem presented above, the personality description suggests a different response (i.e., “engineer”) than does the base-rate information (i.e., “lawyer”). De Neys and Glumicic (2008) have claimed that this represents a conflict between S1 and S2: “Dual process theories generally state that the analytic system is monitoring the output of the heuristic system. When a conflict with analytic knowledge (e.g., sample size considerations) is detected, the analytic system will attempt to intervene and inhibit the prepotent heuristic response [i.e., based on the stereotype]” (p. 1250). This argument has important implications for the rationality debate because, if conflict detection can be shown to be efficient, the primary cause of systematic biases may be a failure to inhibit and override an intuitive response as opposed to a failure to recognize that an error is being made.

De Neys and colleagues have assembled an impressive roster of evidence supporting their claim that conflict monitoring is highly efficient, thereby suggesting that base-rate neglect, in particular, is caused primarily by a failure of inhibition and S2 override. The first and most basic piece of evidence came from response time data. De Neys and Glumicic (2008) showed that participants took longer to answer base-rate problems when the personality description and base-rates were incongruent (i.e., in conflict) than when they were congruent (i.e., non-conflicting). The crucial assumption here is that increased response time indicates higher likelihood of S2 processing. Thus, participants apparently detected the conflict and engaged in increased S2 processing to resolve it. This finding is striking for two reasons. First, participants succumbed to base-rate neglect – choosing the stereotypical response 78% of the time on incongruent trials – despite the apparent increase in S2 processing. Second, participants took longer on incongruent trials relative to congruent both when they selected the more analytic base-rate response *and* when they selected the stereotypical response. Thus, they showed evidence of conflict detection even when they gave the so-called "biased" or "incorrect" response. Indeed, as De Neys and Glumicic point out, the response time increase for incorrect (stereotypical) responses to incongruent problems is the "crucial question" because these participants, according to protocol analysis, are not readily utilizing the base-rate information in their explicit reasoning (p. 1252). Thus, an increase in response time for incorrect incongruent problems as compared to correct congruent problems is explained as a product of conflict detection because participants are basing their response primarily on the stereotypes in both cases.

The fundamental response time difference between incorrect incongruent and correct congruent problems has been bolstered by a host of additional measures, such as an increase in memory retrieval for base-rates found in incongruent relative to congruent problems (De Neys & Glumicic, 2008), an increased tendency to review the base-rate information for incongruent relative to congruent problems (De Neys & Glumicic, 2008), and a decrease in confidence for incongruent relative to congruent problems (De Neys, Comheeke, & Osman, 2011). De Neys, Vartanian, and Goel (2008) used fMRI to show that there was increased activation in the anterior cingulate cortex – which is often referred to as the conflict detection centre of the brain (Carter, Braver, Barch, Botvinick, Noll, & Cohen, 1998) – when participants responded to incongruent relative to congruent problems (regardless of whether the participant gave the stereotypical or the base-rate response). Additionally, they also showed an increase in right lateral prefrontal cortex activation when participants refrained from giving the stereotypical response to incongruent problems, suggesting that they were engaging in response inhibition. Together, these findings provide compelling evidence for highly efficient conflict detection coupled with an unreliable override or inhibition process.

There is, however, reason to question the generalizability of DeNeys and colleagues' findings beyond some particular characteristics of the set of base-rate problems used in each of the studies cited above. Consider the following problem form used in each of the above studies:

> In a study 1000 people were tested. Among the participants there were 995 nurses and 5 doctors. Paul is a randomly chosen participant of this study.
>
> Paul is 34 years old. He lives in a beautiful home in a posh suburb. He is well spoken and very interested in politics. He invests a lot of time in his career.
>
> What is most likely?
> a) Paul is a nurse
> b) Paul is a doctor

Here, as with the lawyer-engineer problem above, there is a conflict between the base-rate information and the personality description. However, there is a potentially critical difference between this problem and that used by Kahneman and Tversky (1973): The base-rates used by De Neys and colleagues are substantially more extreme (995/996/997 out of 1000). Indeed, base-rates are often left implicit in other studies (such as for the Tom W. problem) to present participants with a more ecologically valid type of base-rate problem. It is only De Neys and colleagues who have used such extreme values.

The use of extreme base-rates to demonstrate efficient conflict monitoring raises a couple of important questions. How obvious does a conflict need to be for it to be efficiently detected? Do extreme base-rates change the nature of the base-rate neglect phenomenon? If conflict monitoring is highly efficient, as De Neys and colleagues have claimed, participants should be able to detect a conflict between a stereotypical personality description and an implicit or moderate base-rate probability. Such base-rates are both more common in the literature (see Barbey & Sloman, 2007, for a review) and, arguably, more ecologically valid (Gigerenzer, 2007). Since base-rate neglect has been demonstrated over a very broad set of conditions (e.g., with moderate and implicit base-rates), De Neys and colleagues' implication that their demonstration of conflict detection for base-rate problems with extreme base-rates holds for base-rate neglect more generally is questionable.

The purpose of the current work is to address these questions. To do so, a novel set of base-rate problems was created so that the type of base-rate information could be manipulated over a series of experiments. I created and pretested a set of stimuli that utilizes information about University of Waterloo academic majors and that differs based on students' beliefs about relative size or base-rate (e.g., there are more students who major in general arts than in statistics) and characteristics of the prototypical student (e.g., general arts students are thought to be creative and statistics students are thought to be orderly). The base-rates were then either left implicit (Experiment 1), made explicit at a moderate level (e.g., 70%/30%; Experiment 2) or made explicit at an extreme level (e.g., 995/5; Experiment 3). Finally, I replicated my findings using De Neys and Glumicic's (2008) set of base-rate problems with extreme base-rates (e.g., 995/5; Experiment 4) and moderate base-rates (e.g., 70/30; Experiment 5). If conflict detection is highly efficient across a broad range of base-rate problems, response time for incorrect responses to incongruent problems should be higher than those for correct responses to congruent problems in each of the following studies. Alternatively, conflict detection in reasoning may be mediated by contextual factors; suggesting low efficiency. To maximize the ease of report Experiments 1 to 3 will be presented together and Experiments 4 and 5 will be presented together.

Experiments 1-3

Method

Experiment 1: Implicit Base-Rates. Fifty-two University of Waterloo students (13 male, 39 female; average age = 19.8 years) participated for credit in a psychology course. Participants completed 18 base-rate problems with implicit base-rates (see Appendix A for a full list). Here is an example problem:

> Person 'A' was selected at random from a group consisting of all UW students majoring in either GENERAL ARTS or STATISTICS.
>
> Person 'A' is orderly, organized, precise, practical and realistic.
>
> Is Person A's major more likely to be:
> a) GENERAL ARTS
> b) STATISTICS

Problems were presented on a computer monitor using *E-Prime v1.2* (Schneider, Eschman, & Zuccolotto, 2002). The personality descriptions for each problem contained individuating information (stereotypes) that always clearly favoured one group over the other. For congruent problems, the alternative favoured by the stereotype was consistent with the large base-rate. For incongruent problems, the alternative favoured by the stereotype was consistent with the small base-rate, thereby creating a conflict. The base-rate probability was incongruent with the stereotypes in the personality description (as above) for half of the problems, and congruent for the other half. Taking the above example, if the personality description were to be consistent with "general arts," it would be a congruent problem. Congruency was counterbalanced such that for each pair of majors (e.g., general arts and statistics), half of the participants saw individuating information pointing to the large major (i.e., congruent) and half saw individuating information pointing to the small major (i.e., incongruent).

The order that the majors were mentioned was also counterbalanced such that each major was presented first 50% of the time and second 50% of the time, both within and across participants.[1] Participants answered the problems by pressing ‘a’ or ‘b’ on the keyboard. Whether a major was presented as ‘a’ or ‘b’ corresponded with the order of presentation in the first sentence of the problem. This discrete choice paradigm was chosen to remain consistent with De Neys and Glumicic (2008). Problems were presented in full, one at a time. Response time was measured from the outset of problem presentation, and therefore included problem reading time.

Instructions were adapted from De Neys and Glumicic (2008) to fit the university major problems. Specifically, participants were told to suppose that every undergraduate student at the University of Waterloo had completed a questionnaire that identified a set of personality traits that best described him or her. Participants were then asked to decide to which of the two majors the student most likely belongs.

The problems were constructed using an online pretest (*N* = 130). For this, we selected a set of the largest and smallest majors on campus and asked participants (all of whom were University of Waterloo students) to rank them based on perceived size at the University of Waterloo. Participants were then asked to select the two personality traits that best described the prototypical student for each of the majors. We constructed the personality descriptions using the

[1] **Two majors were unable to be presented in such a way. Medieval studies and statistics were always presented to the participant either first twice or second twice. They were nevertheless presented first an equal number of times across participants.**

five personality traits selected most for each major. The pairings that were used were then selected on difference in both ranking and personality traits.

To validate the pairings, I used a couple of additional measures in Experiment 1 (each measure was completed following the presentation of the base-rate problems). First, for each of the 18 comparisons, participants were asked to consider both University of Waterloo majors as a group and, on that basis, to assign the proportion that they thought were in each major. The mean proportion across all comparisons was 70.2 for the actual large major and 29.8 for the actual small major. The individual item means for each comparison ranged from 62% / 38% (English / independent studies) to 82% / 18% (mechanical engineering / medieval studies).

Second, participants were asked to rate on a 7-point scale the similarity of the personality description to the prototypical major in each comparison. The scale was counterbalanced such that the major intended to be linked to the stereotypes (based on the pretest) was at the high end of the scale 50% of the time and at the low end of the scale 50% of the time. Responses for linked majors at the low end of the scale were then subtracted from 8 so that a high response always suggested that the stereotype was similar to the prototypical major in each comparison. The mean rating across all comparisons was 5.2 out of 7, which is greater than 4, $t(51) = 16.98$, $p < .001$. This indicates that the personality descriptions were deemed to provide diagnostic support in favour of the targeted major in each pair.

Participants were tested individually, with testing taking approximately 15 minutes.

Experiment 2: Moderate Base-Rates. Fifty-two University of Waterloo students (11 male, 41 female; average age = 19.3 years) participated for credit in a psychology course. Participants completed the same 18 base-rate problems used in Experiment 1, but this time with explicitly presented moderate base-rates (see Appendix B for a full list).

> Person ‘A’ was selected at random from a group consisting of all UW students majoring in either GENERAL ARTS or STATISTICS. In this group 67% are GENERAL ARTS majors and 33% are STATISTICS majors.
>
> Person ‘A’ is orderly, organized, precise, practical and realistic.
>
> Is Person A’s major more likely to be:
> a) GENERAL ARTS
> b) STATISTICS

The explicitly-presented base rates for the 18 problems in Experiment 2 were the mean proportions given by participants for each comparison in Experiment 1 (rounded up to the next whole number). Across the 18 problems, the mean base-rate was 70% for the large major and 30% for the small major, ranging from 62%/23% to 82%/18%. Unlike in Experiment 1, post-task ratings of base-rate and similarity were not elicited. Otherwise, problem set and procedure in Experiment 2 were identical to those in Experiment 1.

Experiment 3: Extreme Base-Rates. Thirty-two University of Waterloo students (9 male, 23 female; average age = 19.3 years) participated for credit in a psychology course. Participants completed the 18 base-rate problems from Experiments 1 and 2, this time with explicitly presented extreme base-rates (see Appendix C for a full list).

> Person ‘A’ was selected at random from a group consisting of 1000 UW students, with 995 majoring in GENERAL ARTS and 5 majoring in STATISTICS.
>
> Person ‘A’ is orderly, organized, precise, practical and realistic.
>
> Is Person A’s major more likely to be:
> a) GENERAL ARTS
> b) STATISTICS

The proportions used for the 18 problems in Experiment 3 were identical to those used by De Neys and Glumicic (2008). Following De Neys and Glumicic, three base-rate ratios were presented with equal frequency: 995/5, 996/4, 997/3. Otherwise, problem set and procedure were identical in Experiment 3 to those in Experiments 1 and 2.

Results

Analysis Strategy and Design. Following De Neys and Glumicic (2008), responses consistent with the base-rate information were considered correct. De Neys and Glumicic (2008) excluded from analysis response times for nominally incorrect responses to congruent problems (i.e., cases in which the participant selected the low base-rate, non-stereotypical major), leading to the exclusion of 3% of the responses for congruent problems. This was because such responses are not consistent with either base-rate or stereotypical information, so their basis cannot be inferred. As a result, response time was analysed using a repeated measures ANOVA with 3 levels [responses to congruent problems consistent with base-rate *and* stereotype (i.e., correct), responses to incongruent problems consistent with base-rate (i.e., correct), responses to incongruent problems consistent with stereotype (i.e., incorrect)]. In addition, differences between levels were assessed using planned paired-samples *t*-tests. To cut down on variance and skew, response times were converted to $\log^{10}$ prior to analysis. Only extreme outlying response times (falling 4 SDs or more from mean) were excluded from analysis. This represented 1.3% of the data.

Accuracy. Accuracy was analysed using a repeated measures ANOVA with 2 levels (congruent, incongruent). As is evident from Table 1, a substantial decrease in accuracy was observed for incongruent problems relative to congruent problems in each of the following experiments, all F's > 68.1, all p's < .001. This indicates a neglect of the base-rate information, the degree of which differs across the three experiments, suggesting sensitivity (between-subjects) both to the explicit presentation and to the change in extremity of base-rate probabilities.

Table 1. *Mean accuracy as a function of problem type for Experiments 1 to 5 (standard errors are in parentheses).*

	E1	E2	E3	E4	E5
Congruent	.84 (.02)	.90 (.02)	.95 (.02)	.91 (.02)	.95 (.01)
Incongruent	.21 (.02)	.39 (.04)	.59 (.06)	.24 (.05)	.26 (.05)
Neutral				.68 (.05)	.69 (.04)

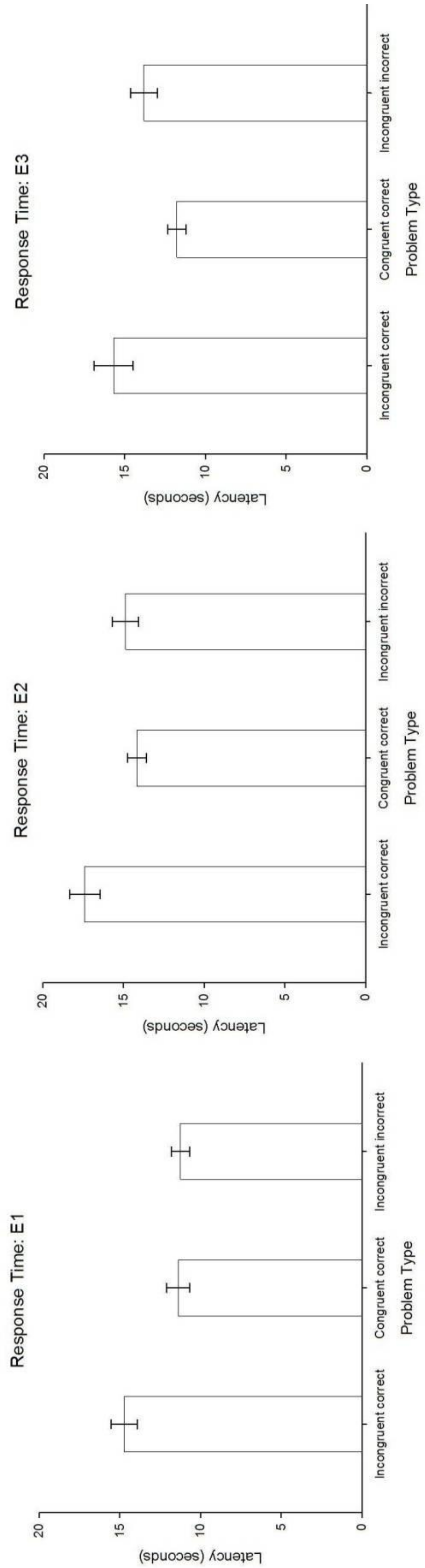
Note: E1 = implicit base-rates; E2 = moderate base-rates; E3 = extreme base-rates; E4 = extreme base-rates; E5 = moderate base-rates.

Experiment 1: Implicit Base-Rates. First, it should be noted that the within-subject design of the response time analysis dictated that data from any participant who scored 100% (N = 1) or 0% (N = 6) on the incongruent problems be excluded from analysis because they did not contribute an observation to each of the cells in the design.

Response time differed as a function of congruency, $F(1.5, 63.9) = 19.03$, $MSE = .020$, $p < .001$ (reported using Greenhouse-Geisser correction for violation of sphericity). However, as Figure 1a shows, this effect was caused by responses consistent with the base-rate information (i.e., incongruent correct) and not by an increase in response time for stereotypical (i.e., incorrect) responses to incongruent problems, in contrast to the crucial finding from De Neys and Glumicic (2008). Indeed, a *t*-test revealed no significant difference between correct congruent and incorrect incongruent response times, $t(50) = 0.25$, SE = .016, $p = .80$. This finding indicates that participants apparently did not detect the conflict between the implicit base-rate information and the stereotypical personality description, at least as measured by response time.

The increase in response time for responses consistent with the base-rate information is unsurprising as the participants in such cases are either actively considering the more analytic base-rate information or just are not very convinced by the particular stereotype. Both cases would likely require additional System 2 processing compared to relying on a salient stereotype to respond.

Figure 1. *Mean response time as a function of problem type for Experiments 1-3.*



Note: E1 = implicit base-rates; E2 = moderate base-rates; E3 = extreme base-rates.

Experiment 2: Moderate Base-Rates. Participants who scored 100% (N = 4) or 0% (N = 5) on the incongruent problems were excluded from analysis.

As with E1, response time differed as a function of congruency, $F(2, 84) = 10.24$, $MSE = .009$, $p < .001$. Again, as Figure 1b shows, this effect was caused by an increase in response time for correct (base-rate) responses to incongruent problems. The crucial difference in response time between correct congruent and incorrect incongruent did not approach significance, $t(47) = 0.43$, SE = .019, $p = .667$. This suggests that participants did not reliably detect the conflict between base-rate and stereotypical information in incongruent problems, even when the base-rates were explicitly presented to them, albeit in a moderate (e.g., 70/30) form. Importantly, this finding cannot be attributed to a total neglect of base-rates because participants actually had higher accuracy for incongruent problems (39%) than did De Neys and Glumicic's (2008) participants (19-22%). This result may be partially attributable to the use of less diagnostic personality descriptions (i.e., our stereotypes were likely less salient) in our problems compared to those used originally by De Neys and colleagues. The possible implications of this are discussed below.

Experiment 3: Extreme Base-Rates. Participants who scored 100% (N = 9) or 0% (N = 2) on the incongruent problems were excluded from analysis.

Again, response time differed as a function of congruency, $F(1.3, 26.1) = 6.01$, $MSE = .017$, $p = .015$ (reported using Greenhouse-Geisser correction for violation of sphericity). However, as Figure 1c shows, this effect was caused by an increase in response time for both correct and incorrect responses to incongruent relative to congruent problems. Participants took significantly longer to answer incongruent problems correctly (as in E1 and E2), $t(29) = 3.59$, SE = .024, $p = .001$, *and* incorrectly, $t(22) = 2.46$, SE = .021, $p = .022$. Thus, the De Neys and

Glumicic (2008) finding replicated once the extreme base-rates (e.g., 995/5) that they used were incorporated into the present set of base-rate problems. It therefore appears that the so-called "highly efficient" conflict detection that has been reported by De Neys and colleagues over a series of studies (De Neys, Comheeke, & Osman, 2011; De Neys & Glumicic, 2008; De Neys, Vartanian, & Goel, 2008) is dependent upon the extremity of the base-rates that they used.

Discussion

The findings of Experiments 1 to 3 suggest that the efficiency of conflict monitoring in base-rate problems has been overstated by De Neys and colleagues. Not only was it necessary to explicitly present the base-rates to participants for them to readily detect a conflict with diagnostic information, the base-rates needed to be presented in an extreme form. Previous demonstrations of base-rate neglect have shown that it occurs in contexts involving implicit base rates and in contexts involving explicitly-presented moderate base rates (for example, Kahneman & Tversky, 1973). The present results are compatible with the idea that, in such cases, the conflict between individuating information and the relevant base rate is not reliably detected (much less overridden). Evidence for such conflict detection seems to be limited to cases in which extreme base rates are explicitly presented to participants. It is notable that, when such extreme base rates were presented in Experiment 3, participants typically (though not always) chose the high base-rate option even when it was incongruent with the individuating information. Thus, Experiments 1-3 collectively provide little evidence for a clear dissociation between conflict detection and intervention by System 2: For the most part, it seems, participants presented with incongruent problems either failed to detect the conflict with base rate information and consequently chose the low-base-rate option, or successfully detected the conflict and chose the high-base-rate option despite its conflict with the individuating information.

Before discussing the implications of this finding, it is first necessary to discuss a possible limitation of the previous experiments. Specifically, the personality descriptions used in Experiments 1-3 may have been less stereotypical than those used by De Neys and Glumicic (2008) and therefore may have conflicted less strongly with the base rate information in incongruent problems. Evidence for this comes from the increase in accuracy for incongruent problems in Experiment 2 and Experiment 3 relative to that found by De Neys and Glumicic. In particular, when De Neys and Glumicic's extreme base-rates were used (Experiment 3), participants answered according to the base-rates 59% of the time. This is substantially higher than the 19-22% accuracy for incongruent problems reported by De Neys and Glumicic. Increased base-rate usage is precisely what would be expected if the saliency of stereotypes were to be decreased. It is possible that participants failed to detect the conflict in Experiment 1 and Experiment 2 because the stereotypes used therein were not strong enough to cause a conflict (although this runs counter to our pretest). Of course, if that were the case, it would still call into question the true efficiency of conflict detection. Regardless, this possible explanation needs to be addressed. In the following two experiments, therefore, the above findings were replicated using the problem set used by De Neys and Glumicic.

Experiments 4 and 5

Method

Experiment 4: Extreme Base-Rates. Thirty-two University of Waterloo students (14 male, 18 female; average age = 20.5 years) participated for credit in a psychology course. Participants completed the 18 base-rate problems used by De Neys and Glumicic (2008) with extreme base-rates (i.e., 995/5, 996/4 or 997/3). Of these 18 problems, 6 were congruent, 6 were incongruent, and 6 were neutral. The neutral problems were constructed by De Neys and Glumicic with personality descriptions that did not contain any stereotypes. All other methods and procedures were identical to the previous experiments.

Experiment 5: Moderate Base-Rates. Forty University of Waterloo students (23 male, 17 female; average age = 19.9 years) participated for credit in a psychology course. Participants completed the same 18 base-rate problems from Experiment 4, this time with moderate base-rates. Parallel to Experiment 4, there were three base-rate ratios used an equal number of times: 70/30, 71/29, 72/28. All other methods and procedures were identical to the previous experiments.

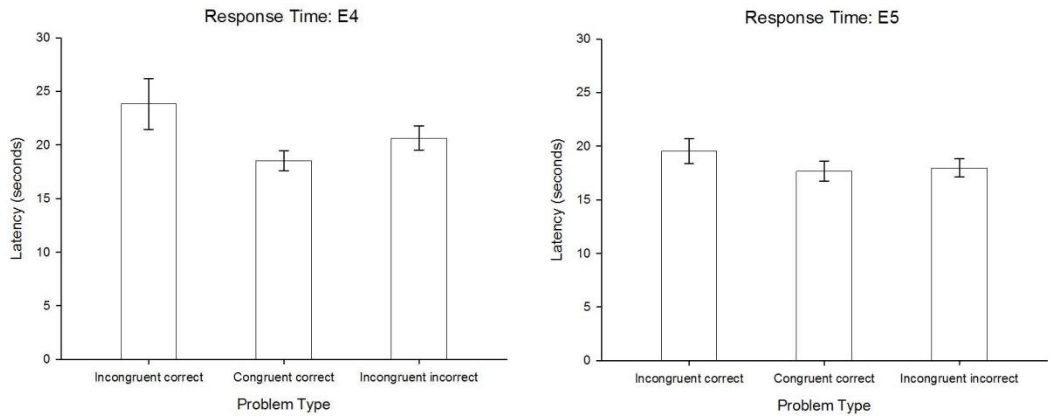
Results

Analysis Strategy and Design. In the interest of continuity, response times for neutral problems were excluded from analysis. Thus, as before, response time was analysed using a repeated measures ANOVA with 3 levels (responses to congruent problems consistent with base-rate *and* stereotype, responses to incongruent problems consistent with base-rate, responses to incongruent problems consistent with stereotype). As before, response times were converted to $\log^{10}$ prior to analysis and only extreme outliers (4 SD's) were excluded from analysis. This represented 0.8% of the data. As with Experiments 1-3, participants who scored 100% (N = 2) or 0% (N = 10) on the incongruent problems were excluded from analysis of response times.

Accuracy. Accuracy was analysed using a repeated measures ANOVA with 3 levels (congruent, incongruent, neutral). As with Experiments 1-3, a decrease in accuracy for incongruent problems was observed in both Experiments 4 and 5, all F's > 105.3, all p's < .001. Accuracy for Experiments 4 and 5 (see Table 1) closely paralleled that reported by De Neys and Glumicic (2008; 19-22% incongruent, 97% congruent, and 80% neutral).

Experiment 4: Extreme Base-Rates. Response time differed across congruency, $F(2, 38) = 4.04$, $MSE = .015$, $p = .041$ (reported using Greenhouse-Geisser correction for violation of sphericity). Crucially, as Figure 2a shows, this effect was caused by an increase in response time for both correct and incorrect responses to incongruent problems relative to congruent. Replicating Experiment 3 and De Neys and Glumicic (2008), participants took significantly longer to answer incongruent problems correctly, $t(21) = 2.36$, SE = .035, $p = .028$, *and* incorrectly, $t(29) = 2.48$, SE = .015, $p = .019$. Thus, participants were able to detect the conflict using De Neys and Glumicic's problem set with extreme base-rates.

Figure 2. *Mean response time as a function of problem type for Experiments 4 and 5.*



Note: E4 = extreme base-rates; E5 = moderate base-rates.

Experiment 5: Moderate Base-Rates. Unlike the previous experiments, response time only marginally differed across congruency, $F(2, 34) = 2.76$, $MSE = .004$, $p = .078$. There was, however, a significant increase in response time for correct responses to incongruent problems relative to congruent, $t(21) = 2.42$, SE = .018, $p = .025$. Thus, this difference is simply more modest than in the previous four experiments. At any rate, as Figure 2b shows, the crucial difference in response time between incorrect responses to incongruent problems and correct responses to congruent problems did not approach significance, $t(35) = 0.67$, SE = .013, $p = .50$. Critically, this demonstrates that the finding that conflict detection is limited to the case of explicitly-presented extreme base-rates cannot be attributed to the possibly weaker stereotypes used in Experiments 1-3.

General Discussion

> "Interestingly, past studies pointing to the pervasive impact of Heuristics and biases (e.g., Tversky & Kahneman, 1974) have progressively deemphasized the importance of normative standards in human thinking.... One could say that the present work helps the pendulum swing back in the other direction.... At least in case of the classic base rate neglect phenomenon, heuristic thinking seems to be always accompanied by successful analytic monitoring." (De Neys & Glumicic, 2008; *Cognition*; p. 1280)

> "Hence, this behavioral study is consistent with the present imaging findings in indicating that successful conflict detection is omnipresent, regardless of whether participants answer problems correctly or incorrectly.... At a more theoretical level, the evidence for successful conflict detection helps to sketch a less bleak picture of human rationality. Our findings indicate that people's thinking is more normative than the infamous failure to solve classic decision making tasks suggests. If people did not know or care about the implications of sample-size considerations, for example, they would not detect conflicts between their intuitive responses and base rates." (De Neys, Vartanian, & Goel, 2008; *Psychological Science*; p. 488-489)

The above quotes clearly represent the position taken by De Neys and colleagues based on evidence garnered from base-rate problems with extreme prior probabilities. The present results indicate, however, that conflict detection during base-rate neglect is limited to the case of explicit presentation of extreme base-rates. Specifically, base-rate probabilities that are left implicit (Experiment 1) or are moderate rather than extreme (e.g., 70/30; Experiments 2 and 5) do not elicit increased response time when they are neglected. This crucial increase using extreme base-rates was demonstrated here both for the De Neys and Glumicic (2008) problem set (Experiment 4) and for a novel problem set (Experiment 3). Thus, it appears that De Neys and colleagues' description of conflict (or "analytic") monitoring as always present and highly efficient is overstated in that it implies a level of generalizability over different types of base-rate problems that is not empirically supported. In fact, their pattern appears to occur only under very limited circumstances.

It should be noted that the present results do not directly contradict any of the primary findings from De Neys and colleagues. De Neys and Glumicic (2008) established the existence and efficiency of conflict detection under conditions where there is a salient conflict. De Neys, Vartanian. and Goel (2008) advanced this story by differentiating between the brain region activated for conflict detection (anterior cingulate cortex) and that activated for intuitive response inhibition (right lateral prefrontal cortex). In addition, De Neys' work with Samuel Franssens (De Neys & Franssens, 2009; Franssens & De Neys, 2009) has demonstrated the effortless nature of conflict detection and the powerful effects of response inhibition. De Neys and Franssens were also able to demonstrate conflict detection with syllogisms. Finally, De Neys, Comheeke, and Osman (2011) demonstrated decreased confidence for conflict problems, again suggesting that participants reliably detected the salient conflict.

Although the current work does not contradict De Neys and colleagues' primary findings, it does call into question their scope and especially their interpretation. That is, the current research suggests that a conflict between System 1 and System 2 outputs must reach a high level of saliency for conflict detection to be successful. What this indicates, then, is that conflict detection is not as broadly efficient as has been claimed. This is an important point because reasoning conflicts in everyday life surely are less obvious than even the moderate base-rate problems used here. Indeed, in terms of everyday probabilistic reasoning, base-rate problems where the prior probabilities are not explicitly presented (Experiment 1) are likely the most ecologically valid (Gigerenzer, 2007). Such problems showed no evidence of conflict detection whatsoever.

The current work also has implications for the rationality debate. Whereas De Neys and colleagues claim that the ability to detect a conflict between intuition and base-rate information suggests that reasoners "care" about "sample size considerations" (i.e., probabilistic information), we show that reasoners will only "care" if the probabilistic information is made very obvious. In line with the pioneering work by Kahneman, Slovic, and Tversky (1982), this does not paint an optimistic picture of human rationality. If people require such specific and arguably unusual circumstances to implicitly detect conflict between intuitive and analytic outputs, it stands to reason that people simply are not very good at detecting when they are biased. This is consistent with the idea that cognition is characterized by troubling irrationalities and also runs counter to the idea that reasoning only deviates from normative standards due to a temporary information processing mishap (i.e., the failure to inhibit a prepotent response cued by diagnostic information). Instead, the information processing mishap seems to have been that conflict detection is a highly efficient process.

References

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioural and Brain Sciences, 30*, 241-256.

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences, 17*, 1–42.

Baron, J. (1998). *Judgment misguided: Intuition and error in public decision making.* City: Oxford University Press.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D. & Cohen, J. D. (1998). Anterior cingulated cortex, error detection, and online monitoring performance. *Science, 280*, 747-749.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*, 1284-1299.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113*, 45–61.

De Neys W., Comheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE, 6*, e15954.

De Neys, W., Vartanian, O., & Goel, V. (2008). When our brains detect that we are biased. *Psychological Science, 19*, 483-489.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences.* City: Lawrence Erlbaum Associates.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278.

Evans, J. St. B. T. (2009). How many dual process theories do we need: One, two, or many? In J. Evans and K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 33-54). City: Oxford University Press.

Evans, J. St. B. T., Handley, S. J., Over, D.E., & Perham, N. (2002). Background beliefs in Bayesian inference. *Memory & Cognition*, *30*, 179-190.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning. City:* Psychology Press.

Frankish, K., & Evans, J. St. B. T. (2009). The duality of mind: An historical perspective. In J. Evans and K. Frankish (Eds.). *In two minds: Dual processes and beyond* (pp. 1-32). Oxford: Oxford University Press.

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking and Reasoning*, *15*, 105-128.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103,* 592–96.

Gigerenzer, G. (2007). *Gut feelings*. London, UK: Penguin.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance.* 14. 513-525.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697-720.

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

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases. City:* Cambridge University Press.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide.* Pittsburgh, PA: Psychology Software Tools.

Shafir, E., & Tversky, A. (1995). Decision making. In E. E. Smith & D. N. Osherson (Eds.), *Thinking, vol. 3* (pp. 77-100). City: MIT Press.

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Stanovich, K. E. (2004). *The Robot's rebellion: Finding meaning in the age of Darwin*. Chicago: The University of Chicago Press.

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. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161–88.

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

Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science.* City: Oxford University Press.

Appendix A

Experiment 1

Person ‘A’ was selected at random from a group consisting of all UW students majoring in either General Arts or Statistics.

Person ‘A’ is orderly, organized, precise, practical and realistic.

Is Person A’s major more likely to be:
a) GENERAL ARTS
b) STATISTICS

Person ‘B’ was selected at random from a group consisting of all UW students majoring in either French or Computer Science.

Person ‘B’ is shy, quiet, precise, practical and organized.

Is Person B’s major more likely to be:
a) FRENCH
b) COMPUTER SCIENCE

Person ‘C’ was selected at random from a group consisting of all UW students majoring in either History or Psychology.

Person ‘C’ is intuitive, helpful, self-confident, friendly and idealistic.

Is Person C’s major more likely to be:
a) HISTORY
b) PSYCHOLOGY

Person ‘D’ was selected at random from a group consisting of all UW students majoring in either Mechanical Engineering or Atmospheric and Planetary Sciences.

Person ‘D’ is precise, practical, competitive, orderly and arrogant.

Is Person D’s major more likely to be:
a) Mechanical Engineering
b) Atmospheric and Planetary Sciences

Person ‘E’ was selected at random from a group consisting of all UW students majoring in either Electrical Engineering or Independent Studies.

Person 'E' is independent, unconventional, self confident, adaptive and shy.

Is Person E's major more likely to be:
a) Electrical Engineering
b) Independent Studies

Person 'F' was selected at random from a group consisting of all UW students majoring in either Drama or Accounting.

Person 'F' is competitive, ambitious, orderly, organized and practical.

Is Person F's major more likely to be:
a) Drama
b) Accounting

Person 'G' was selected at random from a group consisting of all UW students majoring in either Music or Biomedical Sciences.

Person 'G' is ambitious, competitive, precise, helpful and practical.

Is Person G's major more likely to be:
a) Music
b) Biomedical Sciences

Person 'H' was selected at random from a group consisting of all UW students majoring in either English or Women's Studies.

Person 'H' is self-confident, independent, unconventional, energetic and loud.

Is Person H's major more likely to be:
a) English
b) Women's Studies

Person 'I' was selected at random from a group consisting of all UW students majoring in either Earth Sciences or Medieval Studies.

Person 'I' is adaptive, intuitive, realistic, ambitious and organized.

Is Person I's major more likely to be:
a) Earth Sciences
b) Medieval Studies

Person ‘J’ was selected at random from a group consisting of all UW students majoring in either Atmospheric and Planetary Sciences or General Arts.

Person ‘J’ is curious, unconventional, intuitive, independent and quiet.

Is Person J’s major more likely to be:
a) Atmospheric and Planetary Sciences
b) General Arts

Person ‘K’ was selected at random from a group consisting of all UW students majoring in either Computer Science or Music.

Person ‘K’ is creative, energetic, friendly, colourful and extraverted.

Is Person K’s major more likely to be:
a) Computer Science
b) Music

Person ‘L’ was selected at random from a group consisting of all UW students majoring in either Psychology or Statistics.

Person ‘L’ is orderly, organized, precise, practical and realistic.

Is Person L’s major more likely to be:
a) Psychology
b) Statistics

Person ‘M’ was selected at random from a group consisting of all UW students majoring in either Medieval Studies or Mechanical Engineering.

Person ‘M’ is unconventional, independent, quiet, conservative and unpredictable.

Is Person M’s major more likely to be:
a) Medieval Studies
b) Mechanical Engineering

Person ‘N’ was selected at random from a group consisting of all UW students majoring in either Women’s Studies or Electrical Engineering.

Person ‘N’ is competitive, precise, serious, ambitious and organized.

Is Person N’s major more likely to be:
a) Women’s Studies

b) Electrical Engineering

Person ‘O’ was selected at random from a group consisting of all UW students majoring in either Accounting or French.

Person ‘O’ is friendly, cheerful, colourful, curious and adaptive.

Is Person O’s major more likely to be:
a) Accounting
b) French

Person ‘P’ was selected at random from a group consisting of all UW students majoring in either Biomedical Sciences or History.

Person ‘P’ is conservative, quiet, organized, orderly and shy.

Is Person P’s major more likely to be:
a) Biomedical Sciences
b) History

Person ‘Q’ was selected at random from a group consisting of all UW students majoring in either Independent Studies or English.

Person ‘Q’ is creative, organized, colouful, punctual and helpful.

Is Person Q’s major more likely to be:
a) Independent Studies
b) English

Person ‘R’ was selected at random from a group consisting of all UW students majoring in either Earth Sciences or Drama.

Person ‘R’ is energetic, extraverted, loud, creative and colourful.

Is Person R’s major more likely to be:
a) Earth Sciences
b) Drama

Appendix B

Experiment 2

Person ‘A’ was selected at random from a group consisting of all UW students majoring in either General Arts or Statistics. In this group, 70% are General Arts majors and 30% are Statistics majors.

Person ‘A’ is orderly, organized, precise, practical and realistic.

Is Person A’s major more likely to be:
a) GENERAL ARTS
b) STATISTICS

Person ‘B’ was selected at random from a group consisting of all UW students majoring in either French or Computer Science. In this group, 29% are French majors and 71% are Computer Science majors.

Person ‘B’ is shy, quiet, precise, practical and organized.

Is Person B’s major more likely to be:
a) FRENCH
b) COMPUTER SCIENCE

Person ‘C’ was selected at random from a group consisting of all UW students majoring in either History or Psychology. In this group, 28% are History majors and 72% are Psychology majors.

Person ‘C’ is intuitive, helpful, self-confident, friendly and idealistic.

Is Person C’s major more likely to be:
a) HISTORY
b) PSYCHOLOGY

Person ‘D’ was selected at random from a group consisting of all UW students majoring in either Mechanical Engineering or Atmospheric and Planetary Sciences. In this group, 70% are Mechanical Engineering majors and 30% are Atmospheric and Planetary Sciences majors.

Person ‘D’ is precise, practical, competitive, orderly and arrogant.

Is Person D’s major more likely to be:
a) Mechanical Engineering
b) Atmospheric and Planetary Sciences

Person ‘E’ was selected at random from a group consisting of all UW students majoring in either Electrical Engineering or Independent Studies. In this group, 71% are Electrical Engineering majors and 29% are Independent Studies majors.

Person 'E' is independent, unconventional, self confident, adaptive and shy.

Is Person E’s major more likely to be:
a) Electrical Engineering
b) Independent Studies

Person ‘F’ was selected at random from a group consisting of all UW students majoring in either Drama or Accounting. In this group, 28% are Drama majors and 72% are Accounting majors.

Person ‘F’ is competitive, ambitious, orderly, organized and practical.

Is Person F’s major more likely to be:
a) Drama
b) Accounting

Person ‘G’ was selected at random from a group consisting of all UW students majoring in either Music or Biomedical Sciences. In this group, 30% are Music majors and 70% are Biomedical Sciences majors.

Person ‘G’ is ambitious, competitive, precise, helpful and practical.

Is Person G’s major more likely to be:
a) Music
b) Biomedical Sciences

Person ‘H’ was selected at random from a group consisting of all UW students majoring in either English or Women’s Studies. In this group, 71% are English majors and 29% are Women’s Studies majors.

Person ‘H’ is self-confident, independent, unconventional, energetic and loud.

Is Person H’s major more likely to be:
a) English
b) Women’s Studies

Person ‘I’ was selected at random from a group consisting of all UW students majoring in either Earth Sciences or Medieval Studies. In this group, 72% are Earth Sciences majors and 28% are Medieval Studies majors.

Person ‘I’ is adaptive, intuitive, realistic, ambitious and organized.

Is Person I’s major more likely to be:
a) Earth Sciences
b) Medieval Studies

Person ‘J’ was selected at random from a group consisting of all UW students majoring in either Atmospheric and Planetary Sciences or General Arts. In this group, 30% are Atmospheric and Planetary Sciences majors and 70% are General Arts majors.

Person ‘J’ is curious, unconventional, intuitive, independent and quiet.

Is Person J’s major more likely to be:
a) Atmospheric and Planetary Sciences
b) General Arts

Person ‘K’ was selected at random from a group consisting of all UW students majoring in either Computer Science or Music. In this group, 71% are Computer Science majors and 29% are Music majors.

Person ‘K’ is creative, energetic, friendly, colourful and extraverted.

Is Person K’s major more likely to be:
a) Computer Science
b) Music

Person ‘L’ was selected at random from a group consisting of all UW students majoring in either Psychology or Statistics. In this group, 72% are Psychology majors and 28% are Statistics majors.

Person ‘L’ is orderly, organized, precise, practical and realistic.

Is Person L’s major more likely to be:
a) Psychology
b) Statistics

Person ‘M’ was selected at random from a group consisting of all UW students majoring in either Medieval Studies or Mechanical Engineering. In this group, 30% are Medieval Studies majors and 70% are Mechanical Engineering majors.

Person ‘M’ is unconventional, independent, quiet, conservative and unpredictable.

Is Person M's major more likely to be:
a) Medieval Studies
b) Mechanical Engineering

Person 'N' was selected at random from a group consisting of all UW students majoring in either Women's Studies or Electrical Engineering. In this group, 29% are Women's Studies majors and 71% are Electrical Engineering majors.

Person 'N' is competitive, precise, serious, ambitious and organized.

Is Person N's major more likely to be:
a) Women's Studies
b) Electrical Engineering

Person 'O' was selected at random from a group consisting of all UW students majoring in either Accounting or French. In this group, 72% are Accounting majors and 28% are French majors.

Person 'O' is friendly, cheerful, colourful, curious and adaptive.

Is Person O's major more likely to be:
a) Accounting
b) French

Person 'P' was selected at random from a group consisting of all UW students majoring in either Biomedical Sciences or History. In this group, 70% are Biomedical Sciences majors and 30% are History majors.

Person 'P' is conservative, quiet, organized, orderly and shy.

Is Person P's major more likely to be:
a) Biomedical Sciences
b) History

Person 'Q' was selected at random from a group consisting of all UW students majoring in either Independent Studies or English. In this group, 29% are Independent Studies majors and 71% are English majors.

Person 'Q' is creative, organized, colouful, punctual and helpful.

Is Person Q's major more likely to be:
a) Independent Studies
b) English

Person ‘R’ was selected at random from a group consisting of all UW students majoring in either Earth Sciences or Drama. In this group, 72% are Earth Sciences majors and 28% are Drama majors.

Person ‘R’ is energetic, extraverted, loud, creative and colourful.

Is Person R’s major more likely to be:
a) Earth Sciences
b) Drama

Appendix C

Experiment 3

Person ‘A’ was selected at random from a group consisting of 1000 UW students, with 995 majoring in General Arts and 5 majoring in Statistics.

Person ‘A’ is orderly, organized, precise, practical and realistic.

Is Person A’s major more likely to be:
a) GENERAL ARTS
b) STATISTICS

Person ‘B’ was selected at random from a group consisting of 1000 UW students, with 4 majoring in French and 996 majoring in Computer Science.

Person ‘B’ is shy, quiet, precise, practical and organized.

Is Person B’s major more likely to be:
a) FRENCH
b) COMPUTER SCIENCE

Person ‘C’ was selected at random from a group consisting of 1000 UW students, with 3 majoring in History and 997 majoring in Psychology.

Person ‘C’ is intuitive, helpful, self-confident, friendly and idealistic.

Is Person C’s major more likely to be:
a) HISTORY
b) PSYCHOLOGY

Person ‘D’ was selected at random from a group consisting of 1000 UW students, with 995 majoring in Mechanical Engineering and 5 majoring in Atmospheric and Planetary Sciences.

Person ‘D’ is precise, practical, competitive, orderly and arrogant.

Is Person D’s major more likely to be:
a) Mechanical Engineering
b) Atmospheric and Planetary Sciences

Person ‘E’ was selected at random from a group consisting of 1000 UW students, with 996 majoring in Electrical Engineering and 4 majoring in Independent Studies.

Person 'E' is independent, unconventional, self confident, adaptive and shy.

Is Person E's major more likely to be:
a) Electrical Engineering
b) Independent Studies
Person 'F' was selected at random from a group consisting of 1000 UW students, with 3 majoring in Drama and 997 majoring in Accounting.

Person 'F' is competitive, ambitious, orderly, organized and practical.

Is Person F's major more likely to be:
a) Drama
b) Accounting

Person 'G' was selected at random from a group consisting of 1000 UW students, with 5 majoring in Music and 995 majoring in Biomedical Sciences.

Person 'G' is ambitious, competitive, precise, helpful and practical.

Is Person G's major more likely to be:
a) Music
b) Biomedical Sciences

Person 'H' was selected at random from a group consisting of 1000 UW students, with 996 majoring in English and 4 majoring in Women's Studies.

Person 'H' is self-confident, independent, unconventional, energetic and loud.

Is Person H's major more likely to be:
a) English
b) Women's Studies

Person 'I' was selected at random from a group consisting of 1000 UW students, with 997 majoring in Earth Sciences and 3 majoring in Medieval Studies.

Person 'I' is adaptive, intuitive, realistic, ambitious and organized.

Is Person I's major more likely to be:
a) Earth Sciences
b) Medieval Studies

Person ‘J’ was selected at random from a group consisting of 1000 UW students, with 5 majoring in Atmospheric and Planetary Sciences and 995 majoring in General Arts.

Person ‘J’ is curious, unconventional, intuitive, independent and quiet.

Is Person J’s major more likely to be:
a) Atmospheric and Planetary Sciences
b) General Arts

Person ‘K’ was selected at random from a group consisting of 1000 UW students, with 996 majoring in Computer Science and 4 majoring in Music.

Person ‘K’ is creative, energetic, friendly, colourful and extraverted.

Is Person K’s major more likely to be:
a) Computer Science
b) Music

Person ‘L’ was selected at random from a group consisting of 1000 UW students, with 997 majoring in Psychology and 3 majoring in Statistics.

Person ‘L’ is orderly, organized, precise, practical and realistic.

Is Person L’s major more likely to be:
a) Psychology
b) Statistics

Person ‘M’ was selected at random from a group consisting of 1000 UW students, with 5 majoring in Medieval Studies and 995 majoring in Mechanical Engineering.

Person ‘M’ is unconventional, independent, quiet, conservative and unpredictable.

Is Person M’s major more likely to be:
a) Medieval Studies
b) Mechanical Engineering

Person ‘N’ was selected at random from a group consisting of 1000 UW students, with 4 majoring in Women’s Studies and 996 majoring in Electrical Engineering.

Person ‘N’ is competitive, precise, serious, ambitious and organized.

Is Person N’s major more likely to be:

a) Women's Studies
b) Electrical Engineering

Person 'O' was selected at random from a group consisting of 1000 UW students, with 997 majoring in Accounting and 3 majoring in French.

Person 'O' is friendly, cheerful, colourful, curious and adaptive.

Is Person O's major more likely to be:
a) Accounting
b) French

Person 'P' was selected at random from a group consisting of 1000 UW students, with 995 majoring in Biomedical Sciences and 5 majoring in History.

Person 'P' is conservative, quiet, organized, orderly and shy.

Is Person P's major more likely to be:
a) Biomedical Sciences
b) History

Person 'Q' was selected at random from a group consisting of 1000 UW students, with 4 majoring in Independent Studies and 996 majoring in English.

Person 'Q' is creative, organized, colouful, punctual and helpful.

Is Person Q's major more likely to be:
a) Independent Studies
b) English

Person 'R' was selected at random from a group consisting of 1000 UW students, with 997 majoring in Earth Sciences and 3 majoring in Drama.

Person 'R' is energetic, extraverted, loud, creative and colourful.

Is Person R's major more likely to be:
a) Earth Sciences
b) Drama