Single or Dual Processing in Base-Rate Problems?

Using Signed Difference Analysis to Test the Distinction

Student Number [redacted]

6,999 words

This thesis is submitted in partial fulfillment of the Honours degree of Bachelor of Psychological Sciences (Honours)

Table of Contents

The Base-Rate Problem 9

Single and Dual-Process Theories of Base-Rate Insensitivity 11

Distinguishing Between Single and Dual-Process Accounts 12

Signal Detection Models 13

Testing Models of Reasoning Using Signed Difference Analysis 16

The Current Study 19

Method 21

Participants 21

Design and Analysis 22

Materials and Procedure 22

Base-Rate Task 23

Raven's Advanced Progressive Matrices: Short Form. 27

Results 31

Cognitive Ability Group Allocations 31

Analysis of Base-Rate Task Responses 32

Summary of General Base-Rate Task Effects 32

The Influence of Cognitive Ability on Base-Rate Task Performance 42

Signed Difference Analysis 45

Discussion 49

Summary of Findings 49

Discussion of Findings 49

Base-Rate Task Effects 49

The Influence of Cognitive Ability on Base-Rate Task Performance 50

Theoretical Implications 52

Further tests of the 1D model 52

Alignment with existing evidence 53

Practical Implications 54

Conclusion 55

References..56

Appendix A 64

Appendix B 66

Appendix C 68

List of Figures

1. 1D and 2D SDT models of responses in a base-rate task under different levels of cognitive ability 15
2. Example permitted and forbidden signed difference vectors 18
3. Stimulus for a base-rate trial 24
4. Stimuli presentation and duration 27
5. Raven's Advanced Progressive Matrices: Short Form Example Trial 29
6. Operation Span Main Task Example Trial 30
7. Mean Profession Endorsement Ratings (AR) 36
8. Mean Profession Endorsement Ratings (WMC) 37
9. Mean Profession Endorsement Ratings (GCA) 38
10. Interaction Between Base-Rate Consistency and Base-Rate Ratio by AR 39
11. Interaction Between Base-Rate Consistency and Base-Rate Ratio by WMC 40
12. Interaction Between Base-Rate Consistency and Base-Rate Ratio by GCA 41
13. Two-Way Interaction Between Base-Rate Consistency and AR 43
14. Two-Way Interaction Between Base-Rate Consistency and GCA 44
15. Mean Conclusion Profession Endorsement Rates based on Dichotomised Ratings (AR) 46
16. Mean Conclusion Profession Endorsement Rates based on Dichotomised Ratings (WMC) 47
17. Mean Conclusion Profession Endorsement Rates based on Dichotomised Ratings (GCA) 48

List of Tables

1. Kahneman and Tversky's (1973) Lawyer-Engineer Problem 10
2. Base-rate task instructions based on Bago and De Neys (2017) 23
3. Example Stimuli Demonstrating Factorial Manipulation of Base-Rate and Stereotype Consistency 26
4. Descriptive Statistics and Welch's Unpaired T-Test Results for Cognitive Ability Measures Across Overall, High and Low Ability Conditions 31
5. Factorial ANOVA Results for Conclusion Profession Endorsement Ratings (AR) 33
6. Factorial ANOVA Results for Conclusion Profession Endorsement Ratings (WMC) 34
7. Factorial ANOVA Results for Conclusion Profession Endorsement Ratings (GCA) 35

Abstract

When reasoning on classic base-rate problems, people's judgements often underweight base-rate information and rely heavily on stereotype information. These patterns of performance have been shown to vary according to cognitive ability. Single- and dual-process theories are two competing classes of theory that have been applied to explain the impact of cognitive ability on base-rate task performance. Dual-process theories propose that low and high cognitive ability individuals differentially rely on two distinct processes to evaluate base-rate problems: one that is more readily deployed and based more on stereotypes (Type 1 processing), and one that is less readily deployed and based more on base-rates (Type 2 processing). In contrast, single-process theories suggest that both low and high ability individuals draw on a common type of processing that integrates both base-rate and stereotype information. To help adjudicate between the competing theories, the current study instantiated them as formal Signal Detection models and rigorously tested their predictions against novel base-rate reasoning data using Signed Difference Analysis (SDA). In an online experiment, 101 participants completed a base-rate task that included several within-subjects' manipulations designed to probe dual-process assumptions (e.g., base-rate ratio), then completed two cognitive ability tasks (Raven's matrices and operation span). The results of the cognitive ability tasks were used to categorise participants into low or high ability for later analysis as a between-subjects factor. Consistent with dual-process predictions, high ability participants were more sensitive to base-rate information than low-ability participants. However, SDA revealed that the single-process model was also consistent with the data. This finding demonstrates the viability of a single-process theory of base-rate insensitivity and further challenges the value of standard dual-process views of the phenomenon.

Declaration

This thesis contains no material which has been accepted for the award of any other degree of diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

[redacted]

September 2023

Contribution Table

ROLE	ROLE DESCRIPTION	STUDENT	SUPERVISOR 1	SUPERVISOR 2
CONCEPTUALIZATION	Ideas; formulation or evolution of overarching research goals and aims.		X	
METHODOLOGY	Development or design of methodology; creation of models.	X	X	X
PROJECT ADMINISTRATION	Management and coordination responsibility for the research activity planning and execution.			
SUPERVISION	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.		X	X
RESOURCES	Provision of study materials, laboratory samples, instrumentation, computing resources, or other analysis tools.			
SOFTWARE	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code.			X
INVESTIGATION	Conducting research - specifically performing experiments, or data/evidence collection.			X
VALIDATION	Verification of the overall replication/reproducibility of results/experiments.			
DATA CURATION	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.			
FORMAL ANALYSIS	Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data.	X	X	X
VISUALIZATION	Visualization/data presentation of the results.	X		
WRITING – ORIGINAL DRAFT	Specifically writing the initial draft.	X		
WRITING – REVIEW & EDITING	Critical review, commentary or revision of original draft		X	

Single or Dual Processing in Base-Rate Problems?
Using Signed Difference Analysis to Test the Distinction

Picture yourself at law awards ceremony where most attendees are lawyers. Amid a formally dressed crowd, a man in a graphic t-shirt and jeans stands out. He is cracking jokes, talking about his amateur band and obsession with surfing. Based on the setting and your observations, how likely do you think it is that this 'mystery man' is a lawyer? His casual outfit and conversation might lead you to conclude that it is very unlikely, despite the knowledge that most attendees are lawyers.

This tendency to rely on generalisations about individual characteristics (i.e., stereotypes) instead of underlying frequencies (i.e., base-rates) when judging probability is known as base-rate insensitivity (Bar-Hillel, 1980). Insufficient attention to base-rates causes poor decision-making in a number of areas. In medical diagnostics, prioritising specific symptoms over the general prevalence of a disease can lead to misdiagnosis (Hamm, 1996). Base-rate insensitivity also hinders reasoning performance within law enforcement (Dahlman et al., 2016), insurance (Jaspersen & Aseervatham, 2017), and financial investment (Barberis & Thaler, 2002). Given the applied impact, addressing base-rate insensitivity has been a priority for cognitive science (Stengård et al., 2022). However, the phenomenon's causes remain ambiguous due to conflicting theoretical explanations.

Dual-process theory attributes base-rate insensitivity to two distinct reasoning processes: A more readily deployed "Type 1" intuitive process (i.e., typically characterised as fast, automatic and effortless) and a "Type 2" deliberative process (i.e., typically characterised as slow, conscious and effortful) that is less readily deployed but needed for considering probabilities (Evans & Stanovich, 2013). In contrast, single-process theories of reasoning do not posit a dichotomy between processes (Osman, 2004; Stephens et al., 2019).

Key evidence for a dual-process theory of base-rate insensitivity and other reasoning biases is that reliance on background knowledge such as stereotypes increases under low cognitive ability, when Type 2 processing is thought to be less readily applied (Evans & Stanovich, 2013). However, recent studies applying more conclusive formal modelling approaches to data from various reasoning tasks have found that such patterns of responding are also consistent with a single-process account (Stephens et al., 2018; Scott, 2021; Sikora-Przibilla, 2022). These findings challenge the need for a dual-process explanation of performance on reasoning tasks. However, formal models have not yet been used to clarify the number of processes underlying reasoning on base-rate tasks under different levels of cognitive ability. Accordingly, this thesis will do so, rigorously testing single and dual-process theories.

The Base-Rate Problem

Consider the problem in Table 1. This problem is the original version of a task used to investigate base-rate insensitivity called the "lawyer-engineer" problem (Kahneman & Tversky, 1973). Subsequent studies have tested many different versions of this problem, including alterations to the answer format (e.g., requiring forced-choice responses instead of a probability estimate), and whether the base-rate and stereotype information suggest conflicting solutions (Pennycook & Thompson, 2017). The base-rate information in the lawyer-engineer problem is the proportion of engineers as compared to lawyers in the sample. The stereotype information is the description of Jack's characteristics. In the original lawyer-engineer problem, the base-rate implies that Jack is more likely to be a lawyer, but the stereotype information is strongly consistent with him being an engineer. In general, people consistently display base-rate insensitivity when the base-rate and stereotype information point to conflicting solutions (Barbey & Sloman, 2007). This effect still occurs even when people are given a single word stereotype cue (e.g., "Person X is funny") instead of a more elaborate personality description (Bago & De Neys, 2017).

Despite these findings, people are not completely insensitive to base-rates. On the contrary, there is considerable evidence to suggest that base-rates influence people's judgements in a range of contexts, such as when they are presented in a natural frequency format (Pennycook & Thompson, 2016). People also tend to give significantly higher probability estimates (correlated with the base-rates) for extreme base-rate ratios (e.g., 995:5) compared with moderate (e.g., 700:300) and balanced ratios (e.g., 505:495), suggesting that they are giving some consideration to base-rates (Newman et al., 2017; Pennycook et al., 2015; Yang et al., 2023). However, given that base-rates are often underweighted, there is a need to understand the cognitive processes involved in thinking about base-rates. Single and dual-process theories offer competing explanations for why base-rate insensitivity occurs and how it might be addressed.

Table 1

Kahneman and Tversky's (1973) Lawyer-Engineer Problem

A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields [...] 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 _____%.

Single and Dual-Process Theories of Base-Rate Insensitivity

The dominant explanation for base-rate insensitivity is dual-process theory (Pennycook et al., 2016). Type 1 processing is autonomous, does not rely on working memory and tends to be fast, independent of cognitive ability and to produce non-normative or biased responses (Evans & Stanovich, 2013). Type 2 processing is deliberative, slow, requires substantial working memory resources, depends upon cognitive ability and tends to result in normative responses (Evans & Stanovich, 2013). The dual-process framework is ubiquitous in psychology, including dual-process theories of persuasion (Chaiken, 1980), personality (Epstein, 1998) and emotion (Smith & Neumann, 2005). Applied to base-rate problems, dual-process theories generally propose that Type 1 evaluations rely more on stereotype information and less on base-rates, while Type 2 evaluations rely more on base-rates and less on stereotypes (Stanovich & West, 2000). Stereotype information is thought to be processed relatively efficiently through mental shortcuts and is often consistent with preconceived beliefs (Kahneman, 2011). In contrast, base-rate information is thought to require effortful deliberation and often violates existing assumptions (Kahneman & Frederick, 2002).

Important evidence for distinct Type 1 and Type 2 processes comes from investigations of how working memory capacity and cognitive ability impact performance for a range of reasoning tasks (Osman, 2004). The first observation is that reliance on stereotypes or background knowledge increases and reliance on base-rates or logic decreases when cognitive capacity is constrained by the imposition of either a deadline or working memory load (De Neys, 2006; Evans & Curtis-Holmes, 2005; Newman et al., 2017). The second observation is that differences in participants' response patterns on reasoning tasks correlate reliably with measures of cognitive ability, such that individuals who provide normative answers show higher cognitive ability compared with those who give non-normative, intuitively based answers (De Neys et al., 2005; Stanovich & West, 2000; Toplak et al., 2011). These observations support the view that those with greater cognitive resources use more Type 2 processing, while those with lower resources use more Type 1.

Importantly, similar results have been found for the base-rate task. Two studies found a significant, positive relationship between scores on a composite measure of cognitive ability and degree of base-rate usage (Stanovich & West, 1998b, 1998c). Another investigation found that concurrent working memory load decreased base-rate but not stereotype responding, in line with dual-process predictions (Franssens & De Neys, 2009).

In contrast, single-process theories propose that one continuum of processing underlies performance on a range of reasoning tasks (De Neys, 2021). The single-process perspective does not deny that reasoning can at times appear intuitive and at other times deliberative. However, it interprets these differences as indicating two extremes of a common core process (De Neys, 2006). In regard to reasoning on base-rate problems, this account suggests that responses, including across levels of cognitive ability or working memory capacity, are based upon a unified evaluation of likelihood that integrates base-rate and stereotype information (Stephens et al., 2019).

Distinguishing Between Single and Dual-Process Accounts

Despite the theoretical importance, there is a lack of strong empirical evidence distinguishing dual- and single-process theories (Dunn & Kirsner, 1988; Stephens et al., 2018). Part of the issue is that dual-process models have traditionally been framed conceptually and advanced through a heavy reliance on the tenuous logic of functional dissociations (Newell & Dunn, 2008). Functional dissociations suggest that distinct processes can be inferred when the hypothesised functions of two latent variables show differential effects from an experimental manipulation (Machery, 2012); for example, when reliance on stereotype information increases and reliance on base-rate information decreases under constrained working memory capacity. A major limitation of this approach is that these dissociations can also be consistent with the operation of a single process (Stephens et al., 2018). The principle of parsimony suggests that the burden of proof falls more heavily on the more complex theory (Cassimatis et al., 2008). Therefore, to make a strong case for multiple processes, one must show that the evidence is *inconsistent with a single-process account,* as well as compatible with dual-process theory (Stephens et al., 2018). To test the accounts, it is thus necessary to go beyond verbal models and instantiate both theories as formal, quantitative models. Formal models use explicit mathematical formulations to define the relationships between variables and do not rely on observations of functional dissociation to make inferences about latent processes. This approach affords the necessary precision to fully test theories against data (Lewandowsky & Farrell, 2010).

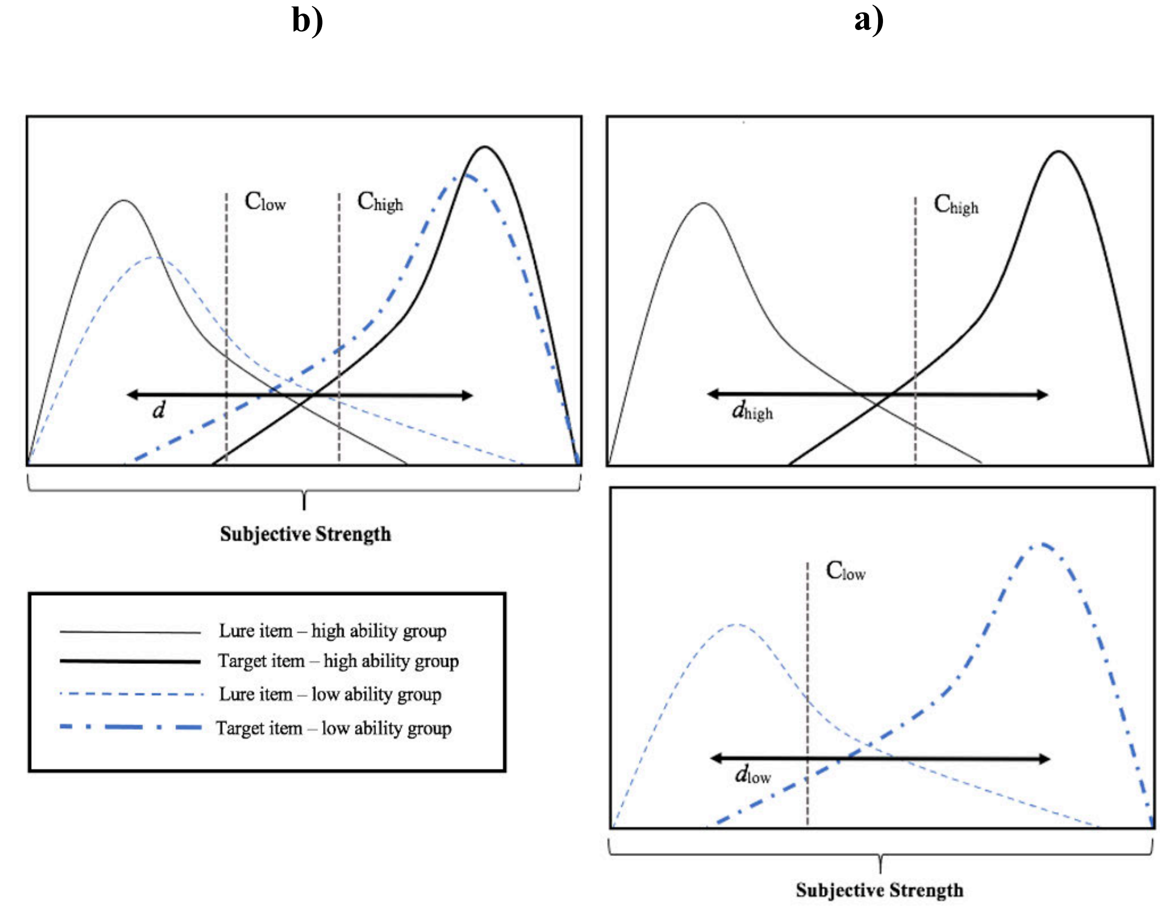
Signal Detection Models

Because of the problems associated with verbal models, several researchers have recently applied formal single- and dual-process models grounded in the Signal Detection Theory (SDT) framework to a variety of reasoning tasks (e.g., Rotello & Heit, 2009; Stephens et al., 2020). A key aspect of SDT is that it distinguishes between the information analysed in relation to a decision, and the criterion or decision threshold used by a decision-maker to evaluate this evidence against (Green & Swets, 1966). On the lawyer-engineer problem framed as a forced-choice judgment ("is Jack an engineer"?), this evidence could include a subjective assessment of conclusion strength based on stereotype as well as base-rate information, while the decision threshold would capture the degree of subjective strength required by an observer to endorse the conclusion of "engineer". Critically, in a single-process SDT model, the subjective strength associated with a base-rate problem is represented as a single point along a single dimension (1D) of subjective strength. In a dual-process SDT model, subjective strength is represented as a point along two separate and distinct dimensions (2D) of subjective strength. If reasoning on the base-rate task under different levels of cognitive ability or working memory differentially involves two distinct processes, then a 2D model will be required to account for the data. If reasoning involves just one process, a 1D model will be sufficient to account for the data (Stephens et al., 2018).

Within an SDT model, the subjective strength dimensions are each captured by the discriminability parameter, "*d*", which represents an observer's ability to distinguish between "targets" (conclusions that are consistent with base-rates) and "lures" (conclusions that are rejected by the base-rate). SDT models conceptualise targets and lures as two different distributions of subjective strength and *d* as the distance between them. Because 1D SDT models are unidimensional and 2D SDT models are two dimensional, they have one and two discriminability parameters, respectively. Figure 1 displays general 1D and 2D SDT models of judgements for low and high cognitive ability on the base-rate task, which also include separate decision thresholds for low and high ability groups. In applying the SDT models to differentiate between theories, I focus on testing for evidence against a single *d* parameter, which would reject the single-process model.

Figure 1

1D and 2D SDT models of responses in a base-rate task under different levels of cognitive ability



Note. General SDT models of reasoning performance on the base-rate task for high and low cognitive ability groups and lure and target distributions. Gaussian response distributions are not assumed, due to SDA requiring minimal assumptions. High ability refers to judgements made by groups with greater processing resources such as abstract reasoning and working memory capacity. Low ability refers to groups with lesser processing resources. Panel a) depicts a 1D single-process model with three parameters: two decision thresholds for low ability (C_{low}) and high ability (C_{high}) and one discriminability parameter (*d*). Panel b) depicts a 2D dual-process model with four parameters. It has the same decision thresholds as in the first model, however there are two separate discriminability parameters for the low and high ability groups (d_{low} and d_{high}). One of these dimensions might represent a strength assessment that is more reliant on Type 2 processing (i.e., based on deliberation when cognitive ability or working memory is high), while the other might represent an assessment that is more reliant on Type 1 processing (i.e., based on intuition when cognitive ability or working memory is low).

To this end, the current study will use Signed Difference Analysis (SDA) (Dunn & Anderson, 2018) to evaluate the core parameters of the models. The task to be modelled will involve judgements for target and lure base-rate problems under high versus low cognitive ability and working memory capacity. Other factors will include stereotype information and base-rate ratio.

Testing Models of Reasoning Using Signed Difference Analysis

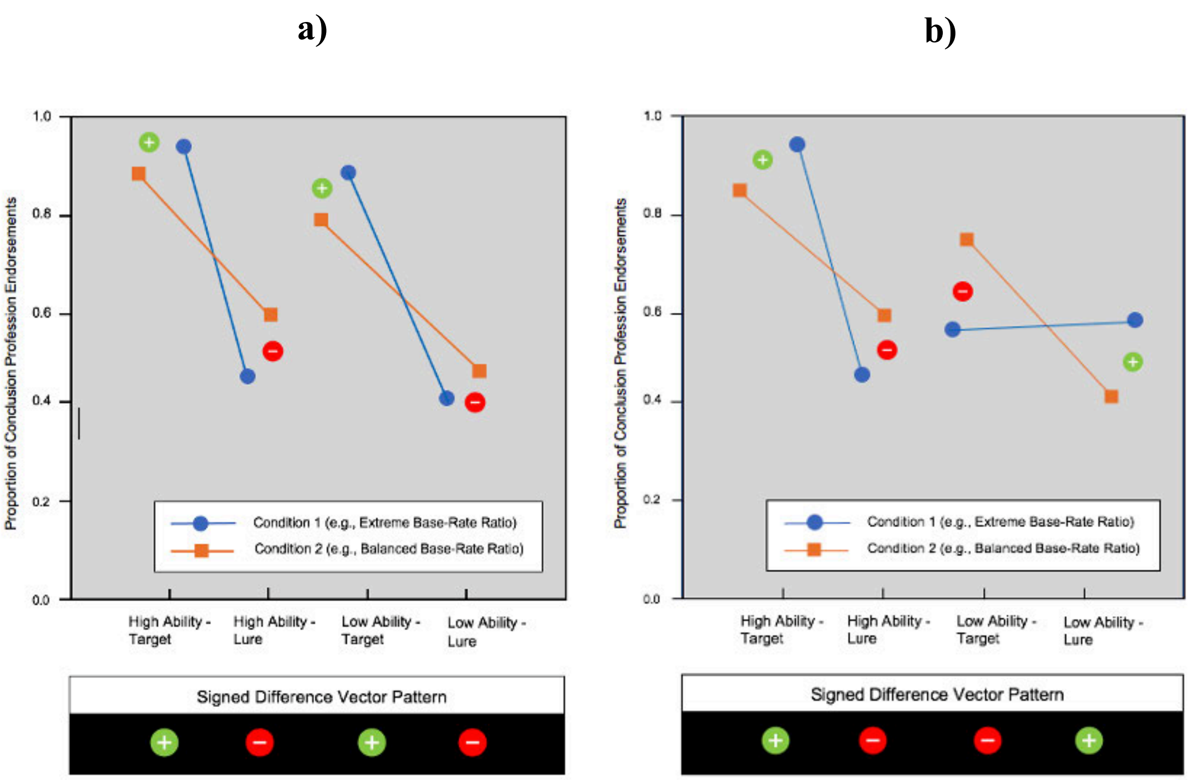
SDA tests models by searching for forbidden patterns of ordinal data that their core parameters cannot accommodate (see Dunn & Anderson, 2018). Under SDA, the only assumption about the relationship between model parameters and the dependent variable is monotonicity. Monotonicity requires that a positive or negative change in model parameters will result in a change in the same direction or no change in the dependent variable (Creelman, 2004). This assumption means that SDA can test the SDT models in their most general form, avoiding more auxiliary assumptions about the shape(s) of the response distributions. The ordinal patterns of data are contained within signed difference vectors, which list the direction (i.e., 'sign') of the differences between two conditions of an independent variable for each category of a dependent variable. The sign of the difference between conditions (+, -, or 0) indicates whether one condition is higher, lower or the same compared to a second condition. Permitted vectors are those that the model can generate given its set of parameters. Conversely, if a model's parameters cannot be adjusted to produce a vector, this vector is forbidden. When SDA is performed, the vectors derived from the observed data can be compared to a model's permitted and forbidden subsets. If an observed vector falls into the forbidden subset, it indicates that the model is inconsistent with the data.

SDA can be used to test the models described in Figure 1. The dependent variables in this application of SDA correspond to mean binary judgements made by participants within four categories defined by the variables of cognitive ability (i.e., high or low) and base-rate information type (i.e., target or lure): High Ability-Target, High Ability-Lure, Low Ability-Target and Low Ability-Lure. The binary judgements are captured as "conclusion profession endorsements"; the mean proportion of affirmative responses to a conclusion question such as "is individual X a lawyer?". The four dependent variable categories are displayed along the horizontal axes in Figure 2. The conditions for this analysis are levels of combinations of independent variables that should theoretically impact base-rate sensitivity, such as differences in the extremity of the base-rate ratio (Newman et al., 2017). The direction of the differences between these conditions for each of the four dependent variables constitute the signed difference vectors.

The permitted and forbidden subsets of the possible signed difference vectors for the general 1D and 2D SDT models have previously been partitioned (Stephens et al., 2018). Figure 2 provides examples of vectors that are permitted and forbidden by the models in Figure 1. Figure 2a depicts a vector pattern that is permitted by both the 1D and 2D models ±(+,-,+,-). In contrast, Figure 2b depicts a vector pattern that is permitted by the 2D model but forbidden by the 1D model: ±(+,-,-,+). The 1D model contains only a single subjective strength dimension, so *d* cannot move independently for the different ability groups. It forbids a pattern in which *d* for High Ability-Target and Low Ability-Target, plus High Ability-Lure and Low Ability-Lure changes in two different directions. The 2D model can accommodate this pattern because it includes separate *d* parameters for high and low ability groups.

Figure 2

Example permitted and forbidden signed difference vectors



Note. Each of the conditions could represent a combination of factors such as base-rate ratio (e.g., extreme [995:5] vs. balanced [505:495] base-rates) and consistent/inconsistent stereotype information. Panels: a) An example of a signed difference vector pattern ±(+, -, +, -) that is permitted by both the 1D and 2D models. In signal detection terms, this is consistent with a change in the discriminability (*d*) of target and lure items *in the same direction* across, for example, extreme and balanced base-rates for both high and low ability groups. Said differently, as the base-rate ratio becomes more extreme, both groups are more likely to endorse a profession when its base-rate is higher (target item) and less likely to endorse a profession when its base-rate is lower (lure item); b) an example of a signed difference vector pattern ±(+, -, -, +) that is forbidden by the 1D model but permitted by the 2D model. This vector corresponds to a pattern in which the high ability group shows improved target and lure discrimination across base-rate ratio conditions (consistent with a higher *d*), while the low ability group shows a simultaneous reduction in performance on these two item types (consistent with a lower *d*).

The Current Study

A key source of evidence for dual-process theories of base-rate insensitivity are ability-based dissociations in base-rate task performance (Stephens et al., 2019). Accordingly, this thesis will target cognitive ability and working memory with the study and model design. The initial aim of this thesis is to test for traditional dissociations in base-rate sensitivity between groups of high versus low cognitive ability. This test will be achieved by experimentally investigating how different configurations of stereotype information, base-rate information and base-rate ratios impact the judgements of both low and high cognitive ability groups. There will be three different versions of this test, with cognitive ability instantiated as either Working Memory Capacity (WMC), Abstract Reasoning (AR) or a General Cognitive Ability (GCA) composite. Multiple measures of cognitive ability were included in order to maximise opportunities for observing the forbidden pattern.

In light of previous research, this study expects to observe the following key results:

- Participants will rate the conclusion profession as more likely when the conclusion is in agreement with the base-rates (Pennycook et al., 2016).
- Participants will rate the conclusion profession as more likely when the conclusion agrees with the stereotype information (Pennycook et al., 2016).

- Participants' sensitivity to the base-rate will increase as a function of the base-rate ratio, such that trials with extreme base-rate ratios (995:5) will elicit the greatest base-rate sensitivity followed by moderate (700:300) and then balanced (510:490) ratios (Newman et al., 2017; Pennycook et al., 2015; Yang et al., 2023).
- High ability individuals will show greater sensitivity to the base-rate compared with low ability individuals (Stanovich & West, 1998, 2008; Toplak et al., 2011).
- Low ability individuals will show greater sensitivity to the stereotype information compared with high ability individuals (Stanovich & West, 1998).

The primary aim of this thesis is to then test the general 1D and 2D SDT models against any observed dissociations, using SDA (Dunn & Anderson, 2008). The goal is to determine whether the 1D SDT model can be disconfirmed through the observation of its forbidden ordinal pattern.

Method

This online quasi-experiment used a mixed-subjects factorial design to analyse the main and interaction effects of cognitive ability, base-rate consistency, stereotype consistency, and base-rate ratio on mean conclusion profession endorsements. Adult participants completed three tasks in a fixed-order: a base-rate task, a short form version of Raven's Advanced Progressive Matrices (RAPM-S) to measure abstract reasoning, and a shortened, computerised version of the Operation Span (OSPAN-S) task to measure working memory capacity. The results were analysed using three mixed ANOVAs based on the different cognitive ability measures and later, SDA. The study methodology was pre-registered prior to data collection and can be viewed here: https://osf.io/qhzfm/?view_only=cf64bf2c06d74888810c456f932ee5f6.

Participants

This study was granted ethics approval by the School of Psychology Human Research Ethics Subcommittee at the University of Adelaide (ethics number 21/24). A total of 103 participants were recruited anonymously through the Prolific Academic participant pool and compensated according to the platform's fair payment principles (£6.00 per ~ 40 minutes). Three participants were removed because their accuracy on the arithmetic equation component of the OSPAN-S was below 85%, as per Conway et al. (2005). This left 101 participants for the final analysis. Participants were recruited from six countries: The United Kingdom (n = 67), Australia (n = 11), Canada (n = 10), the United States (n = 6), New Zealand (n = 5) and Ireland (n = 2). Participants could take part if they were fluent in English and over 18. Participants were between 21-77 years old (M = 41.6, SD = 12.3). 54 (53.5%) participants were male, and 47 (46.5%) participants were female.

Design and Analysis

The general design was a 2 (cognitive ability: low or high) x 2 (base-rate consistency: consistent or inconsistent with the conclusion profession) x 2 (stereotype consistency: consistent or inconsistent with the conclusion profession) x 3 (base-rate ratio: balanced, moderate or extreme) mixed-subjects factorial design. The cognitive ability factor varied between-subjects while all other factors were manipulated within-subjects.

Three mixed ANOVAs were conducted to test all main effects and interactions. The dependent variable for all analyses was participants' mean conclusion profession endorsement ratings. Cognitive ability was assessed using two different measures: WMC and AR. The data underwent three separate analyses, with cognitive ability instantiated as either WMC, AR or a GCA composite. GCA was scored by summing participants' standardised Z-scores for both the AR and WMC measures. Participants were split into high and low cognitive ability groups (scoring above vs. below the median, respectively). These binary groups were needed for the SDA model testing. SDA was run to evaluate the 1D SDT model against the data for the WMC, AR and GCA analyses.

Materials and Procedure

The experiment was built using the jsPsych JavaScript framework (de Leeuw et al., 2023) and accessed by participants through an internet browser. Participants first saw a landing page instructing them to use a device with a big screen, and to complete all experimental tasks within a single sitting. Participants were then shown the study's information sheet and gave their consent before proceeding. After this, participants were prompted to provide their gender identity and age in years. This step was followed by a basic overview of the three tasks included in the experiment and their presentation order.

Base-Rate Task

The next page provided instructions for the base-rate task, which was based on the design from Bago and De Neys (2017) (see Table 2). Three instruction check questions were presented to ensure participants had successfully grasped the key concepts. Participants who answered any of these questions incorrectly were automatically routed back to the start of the task instructions. Once participants answered all of the questions correctly, they commenced the three practice trials, which involved presentations of each of the three levels of base-rate ratio. These trials did not contain content that was used in the main experimental trials. Following the final practice trial, participants were told that clicking the "begin experimental trials" button would show them the first experimental trial.

Table 2

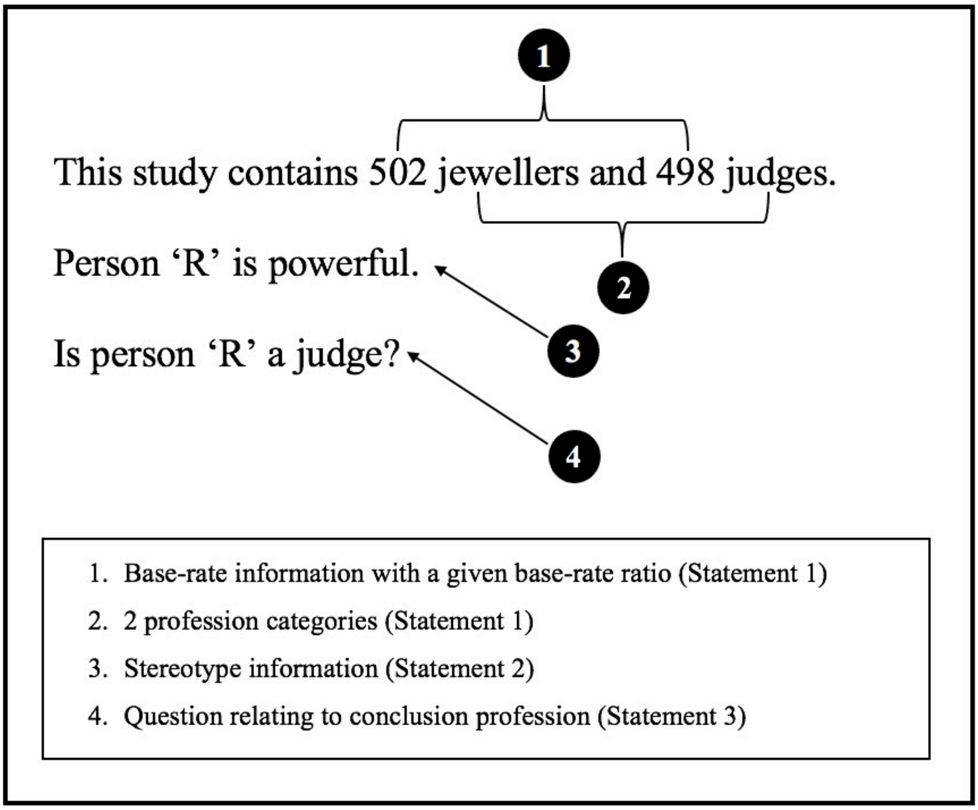
Base-rate task instructions based on Bago and De Neys (2017)

In a big research project, a large number of studies were carried out where a psychologist made short personality descriptions of the participants. You will see information about some of the studies. Note that for each study: 1. There were participants from two population groups (e.g. plumbers and models). 2. You'll see information about the composition of the population groups tested in the study in question (e.g., 995 plumbers and 5 models). 3. One participant was drawn at random from the sample of two population groups. You'll see one personality trait of this randomly chosen participant (e.g. "practical"). You'll be asked to indicate the chance that the participant belongs to one of the two population groups (e.g. "Is the person a plumber?"). Rate the chance on a 6-point scale from 0 (Definitely Not) to 6 (Definitely Yes).

The base-rate task consisted of 48 trials per participant. The stimulus for each trial, as with the practice trials, included three statements consisting of the same elements: Statement 1: two profession categories and base-rate information with a given base-rate ratio (balanced: 510:490, moderate: 700:300, or extreme: 995:5); Statement 2: stereotype information; and Statement 3: a question about one of the profession categories (henceforth, the 'conclusion profession') (see Figure 3). The stereotype information associated with the profession categories was drawn from Sikora-Przibilla (2022) who piloted them previously. To introduce a small amount of variation within the base-rate ratios, each of the three levels were allowed to vary randomly by a count of up to three points. For example, the moderate ratio could range from 749:251 to 752:248. This variation was fixed across participants.

Figure 3

Stimulus for a base-rate trial



Each participant saw 4 trials of each cell of the 3x2x2 stimulus design, forming 48 trials. See Table 3 for example stimuli illustrating how base-rate and stereotype consistency were manipulated factorically. The 48 pairings of conclusion profession and stereotype content were randomly allocated to each cell per participant. Position order was counterbalanced for the profession categories in Statement 1, and the trials were shown in a random order for each participant. Each statement in the stimulus became sequentially visible to participants at intervals of 2000ms (Figure 4). Along with the conclusion question, six response buttons appeared beneath the problem without a response deadline and remained visible until participants made their selection. Participants could make their judgements by clicking on one of the buttons, labelled: “Definitely Not”, “Probably Not”, “Possibly Not”, “Possibly Yes”, “Probably Yes” or “Definitely Yes”. Participants pressed a button to begin the next trial.

Table 3

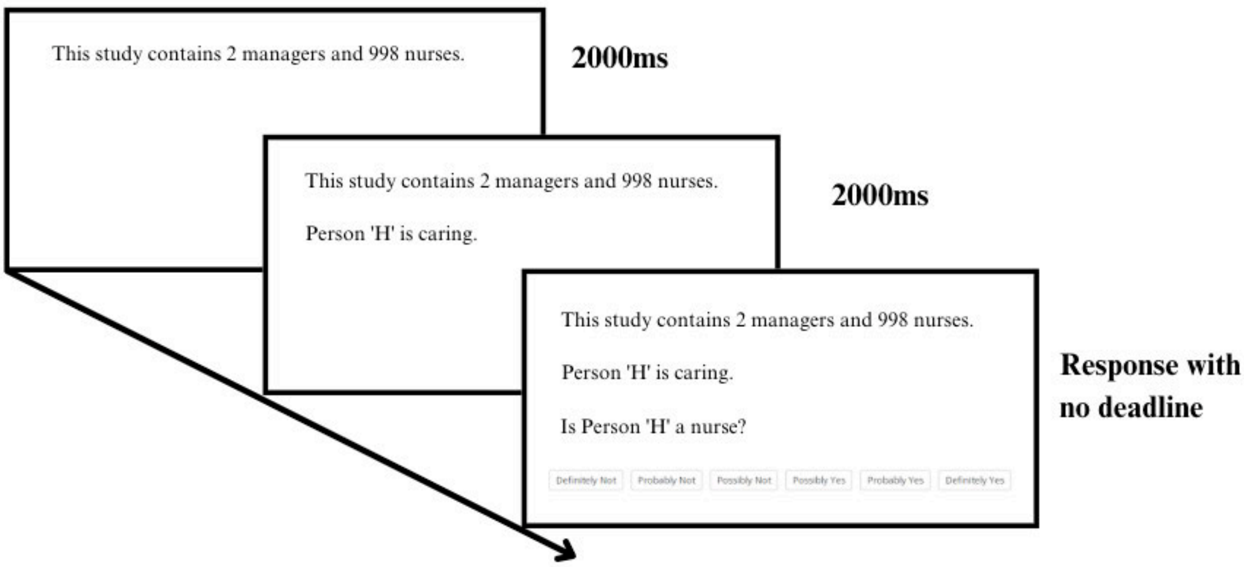
Example Stimuli Demonstrating Factorial Manipulation of Base-Rate and Stereotype Consistency

	Base-Rate Consistent	Base-Rate Inconsistent
Stereotype Consistent	This study contains 995 politicians and 5 librarians. Person 'Y' is opinionated. Is person 'Y' a politician?	This study contains 5 politicians and 995 librarians. Person 'Y' is opinionated. Is person 'Y' a politician?
Stereotype Inconsistent	This study contains 5 politicians and 995 librarians. Person ‘Y’ is opinionated. Is person ‘Y’ a librarian?	This study contains 995 politicians and 5 librarians. Person ‘Y’ is opinionated. Is person ‘Y’ a librarian?

Note. Conclusion professions that are both stereotype and base-rate consistent or both stereotype and base-rate inconsistent constitute a *non-conflict* trial. In contrast, conclusion professions that are stereotype consistent and base-rate inconsistent or stereotype inconsistent and base-rate consistent constitute a *conflict* trial.

Figure 4

Stimuli presentation and duration



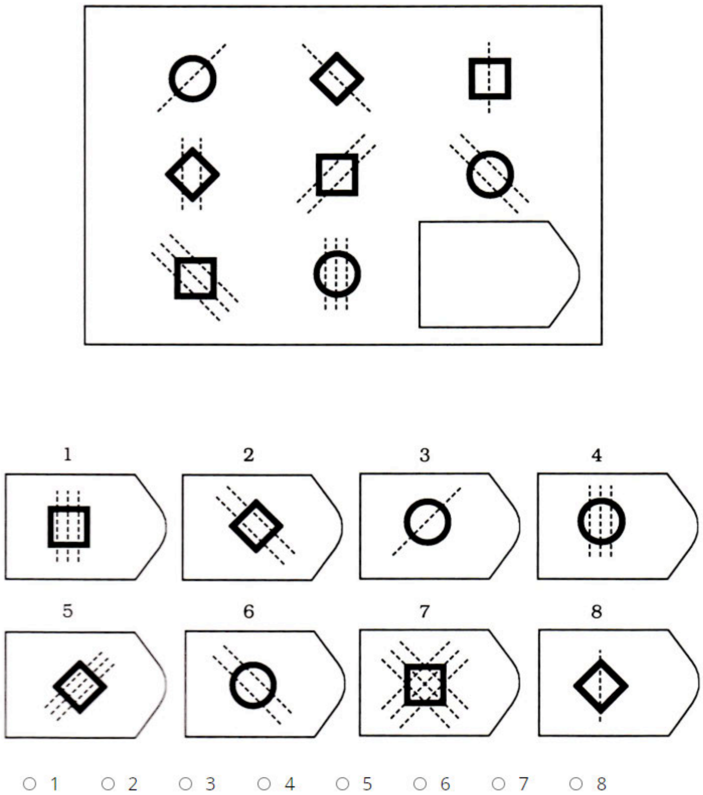
Note. This figure shows the timeline of a single experimental trial. Participants started at the screen in the top left panel and ended at the bottom right. ms = milliseconds.

Raven's Advanced Progressive Matrices: Short Form. Immediately after completing the base-rate task, participants completed the Raven's Advanced Progressive Matrices: Short Form (Arthur & Day, 1994), designed to measure AR (see Figure 5). The measure was a web-adapted version of standard procedures (Raven & Raven, 2003) (see Appendix A for the adapted instructions). Similar to the base-rate task, participants were required to correctly answer 3 instruction check questions before commencing the problems. The 12 matrix problems were presented in the standard progressive order with no ability to return to previous questions, and there was no time limit placed on responding. The RAPM-S is scored by summing the total number of problems correctly solved. As such, each participant received an accuracy score from 0 to 12.

Operation Span Task. Finally, participants completed a shortened version of the operation span task (OSPAN-S) (see Figure 6) for measuring working memory capacity (Luthra & Todd, 2019). Prior to the experimental trials, participants completed practice trials: 3 for the consonant memorisation component of the task, 5 for the arithmetic equation component and 2 that involved both components presented together. These final two practice trials involved viewing a series of arithmetic equations and consonants in repeated turns, as in the main task. The arithmetic equations served as a secondary task to engage participants' working memory. Participants were asked to identify whether each equation was correct or incorrect and remember the consonants. After the series was complete, participants serially recalled the list of presented consonants. The series size ranged from 3-7, and there were three administrations for each series size (i.e., 75 total equation-consonant pairs). Scoring followed the partial credit scoring procedure outlined in Oswald et al. (2015), in which participants were rewarded for both absolute and partial recall. Task instructions were identical to those in the experimental code (Luthra & Todd, 2019). After completing the OSPAN-S, participants were informed that the experiment had finished.

Figure 5

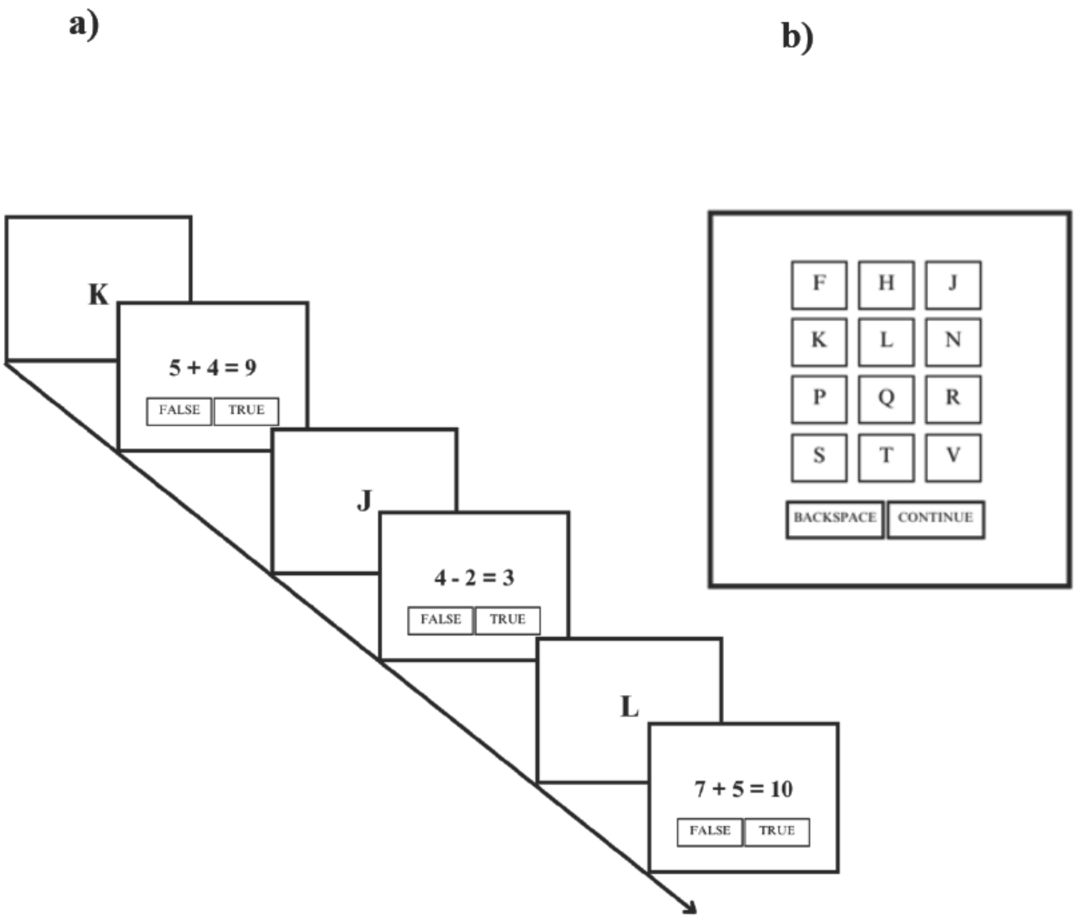
Raven's Advanced Progressive Matrices: Short Form Example Trial



Note. Each trial contained a pattern inside a rectangular frame with a piece missing from the bottom right corner. Participants were required to select the piece from the panel of pieces below the pattern (numbered 1-8) that completed the pattern correctly.

Figure 6

Operation Span Main Task Example Trial



Note. Panel a): Example main trial for the operation span task showing consonants and arithmetic equations in repeated turns. The series progressed from the top left to the bottom right. Panel b): the panel of buttons participants used to input their consonant selections at the end of the trial.

Results

Cognitive Ability Group Allocations

To examine how cognitive ability impacted the other experiment factors, participants were first split at the median into high versus low groups for each of the three ability measures. Because the AR measure captured discrete scores from 0-12, participants who scored 6 or below were assigned to the low group and those who scored 7 or above were assigned to the high group. This approach seemed suitable because 6 was both the mid-point of the AR scale and the median score of the current sample. The descriptive statistics for the ability measures and corresponding Welch's unpaired t-test results for the group allocations can be viewed in Table 4. The t-test results showed significant differences in mean ability between the high and low groups across all three measures. These results, together with the precedent set by a prior study (Quayle & Ball, 2000) that used a similar procedure on a reasoning task and found cognitive ability effects, made the current approach to group allocation seem justified.

Table 4

Descriptive Statistics and Welch's Unpaired T-Test Results for Cognitive Ability Measures Across Overall, High and Low Ability Conditions

Variable	Condition	Median	Mean	SD	N	t	df	p
WMC	Overall	0.79	0.74	0.19	101	11.07	92	<.001
	High	0.89	0.89	0.07	47			
	Low	0.66	0.61	0.17	54			
AR	Overall	6	6.20	2.93	101	14.30	92	<.001
	High	8	8.87	1.71	45			
	Low	4	4.05	1.65	56			
GCA	Overall	0.13	0.00	1.60	101	12.07	92	<.001
	High	1.10	1.24	0.86	50			
	Low	-0.79	-1.21	1.16	51			

Note. WMC = Working Memory Capacity, AR = Abstract Reasoning, GCA = General Cognitive Ability.

Analysis of Base-Rate Task Responses

Traditional dissociation evidence used to support dual-process theory has often been based upon interactions identified through conventional statistical tests, such as Analysis of Variance (ANOVA). Thus, three mixed factorial ANOVAs were conducted to determine whether this evidence was replicated in the current data (mean endorsement ratings). Several preliminary checks confirmed that the ANOVA assumptions were met.

Summary of General Base-Rate Task Effects

The full ANOVA results for the three separate analyses can be found in Tables 5, 6 and 7. However, only findings of the greatest theoretical relevance will be described in detail here. The mean endorsement ratings for the 24 conditions within each analysis are presented in Figures 7, 8 and 9. At a general level, I replicated many of the standard effects observed in base-rate tasks. All three analyses showed large and significant main effects for base-rate and stereotype consistency. As predicted, participants gave higher endorsement ratings for base-rate-consistent conclusion professions (M = 3.25, SD = 1.05), compared with those that were base-rate-inconsistent (M = 2.42, SD = 1.12). Similarly, endorsement ratings were higher for stereotype-consistent conclusion professions (M = 3.48, SD = 0.96) compared with those that were stereotype-inconsistent (M = 2.20, SD = 0.98). Therefore, overall, participants' endorsement ratings were sensitive to both base-rate and stereotype information. As expected, participants displayed more sensitivity to base-rate information as the base-rate ratio increased in extremity. This effect of base-rate ratio is shown by changes in the difference between participants' endorsement ratings across the base-rate consistency conditions as a function of the base-rate ratio (also see Figures 10, 11 and 12). There was a larger difference between base-rate-consistent and base-rate-inconsistent conditions for the extreme ratio, a smaller difference for the moderate ratio, and an even smaller difference for the balanced ratio condition. This result further indicates that participants' judgements were impacted by base-rates.

Table 5

Factorial ANOVA Results for Conclusion Profession Endorsement Ratings (AR)

Effect	DFn	DFd	*F*	*p*	*ges*
AR	1.00	99.00	5.785	0.018*	0.009
ST	1.00	99.00	229.720	0.000*	0.375
BR	1.00	99.00	178.122	0.000*	0.213
Ratio	2.00	198.0	8.364	0.000*	0.006
AR x ST	1.00	99.00	1.803	0.182	0.005
AR x BR	1.00	99.00	7.373	0.008*	0.011
AR x Ratio	2.00	198.0	0.352	0.703	0.000
ST x BR	1.00	99.00	1.972	0.163	0.001
ST x Ratio	2.00	198.0	1.101	0.334	0.001
BR x Ratio	1.54	152.1	65.087	0.000*	0.099
AR x ST x BR	1.00	99.00	1.972	0.163	0.001
AR x ST x Ratio	2.00	198.0	1.410	0.246	0.001
AR x BR x Ratio	1.54	152.1	7.195	0.003*	0.012
ST x BR x Ratio	2.00	198.0	4.005	0.020*	0.003
AR x ST x BR x Ratio	2.00	198.00	1.097	0.336	0.001

Note. AR = Abstract reasoning, ST = Stereotype consistency, BR = Base-rate consistency, Ratio = Base-rate ratio. Dfn = "degrees of freedom in the numerator", DFd = "degrees of freedom in the denominator", *ges* = "generalised eta squared".

Table 6

Factorial ANOVA Results for Conclusion Profession Endorsement Ratings (WMC)

Effect	DFn	DFd	*F*	*p*	*ges*
WMC	1.00	99.00	2.798	0.098	0.005
ST	1.00	99.00	231.562	0.000*	0.374
BR	1.00	99.00	162.917	0.000*	0.205
Ratio	2.00	198.00	8.335	0.000*	0.006
WMC x ST	1.00	99.00	0.758	0.386	0.002
WMC x BR	1.00	99.00	0.959	0.330	0.002
WMC x Ratio	2.00	198.00	2.187	0.115	0.002
ST x BR	1.00	99.00	2.283	0.134	0.001
ST x Ratio	2.00	198.00	1.068	0.346	0.001
BR x Ratio	1.51	149.79	60.263	0.000*	0.094
WMC x ST x BR	1.00	99.00	0.313	0.577	0.000
WMC x ST x Ratio	2.00	198.00	0.658	0.519	0.000
WMC x BR x Ratio	1.51	149.79	2.781	0.080	0.005
ST x BR x Ratio	2.00	198.00	3.394	0.036*	0.002
WMC x ST x BR x Ratio	2.00	198.00	1.323	0.269	0.001

Note. WMC = Working memory capacity, ST = Stereotype consistency, BR = Base-rate consistency, Ratio = Base-rate ratio. Dfn = “degrees of freedom in the numerator”, DFd = “degrees of freedom in the denominator”, *ges* = “generalised eta squared”.

Table 7

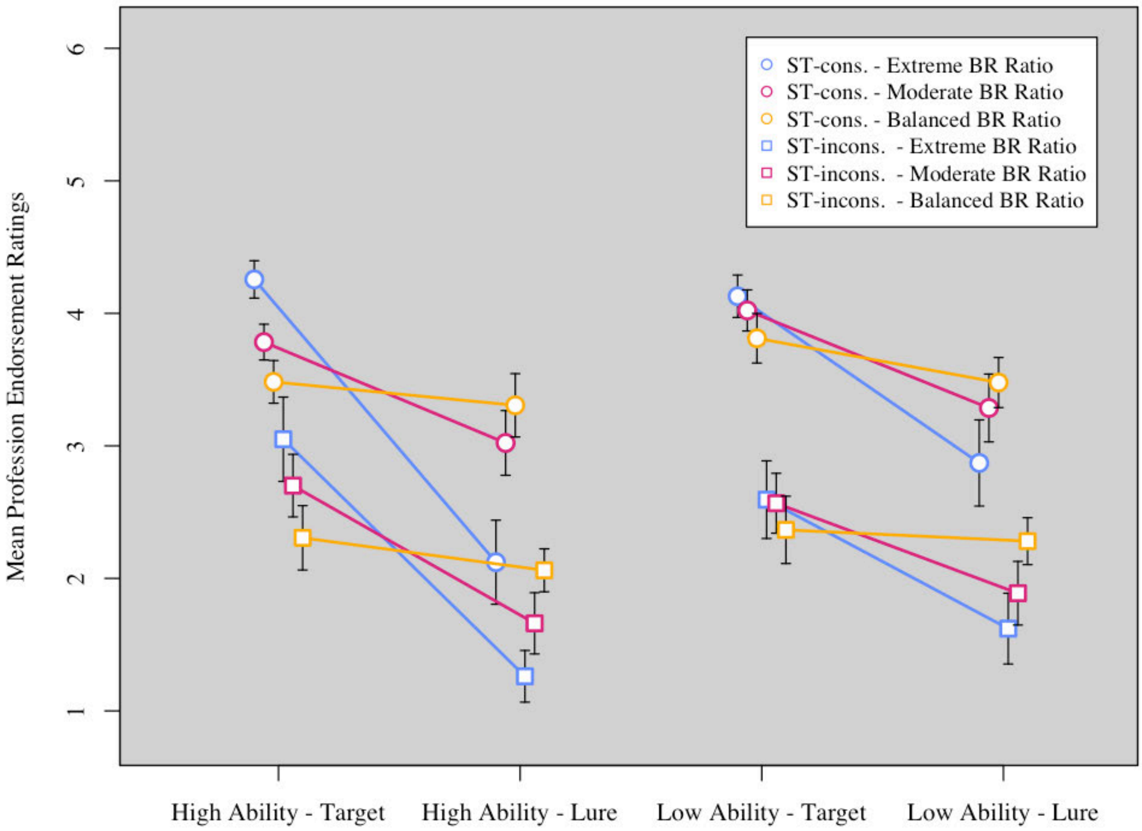
Factorial ANOVA Results for Conclusion Profession Endorsement Ratings (GCA)

Effect	DFn	DFd	*F*	*p*	*ges*
GCA	1.00	99.00	3.848	0.053	0.006
ST	1.00	99.00	237.372	0.000*	0.381
BR	1.00	99.00	170.320	0.000*	0.207
Ratio	2.00	198.00	8.922	0.000*	0.007
GCA x ST	1.00	99.00	2.169	0.144	0.006
GCA x BR	1.00	99.00	5.742	0.018*	0.009
GCA x Ratio	2.00	198.00	3.293	0.039*	0.002
ST x BR	1.00	99.00	2.397	0.125	0.001
ST x Ratio	2.00	198.00	0.999	0.370	0.001
BR x Ratio	1.53	151.32	60.736	0.000*	0.094
GCA x ST x BR	1.00	99.00	0.173	0.678	0.000
GCA x ST x Ratio	2.00	198.00	1.926	0.148	0.001
GCA x BR x Ratio	1.53	151.32	5.564	0.009*	0.009
ST x BR x Ratio	2.00	198.00	3.659	0.028*	0.003
GCA x ST x BR x Ratio	2.00	198.00	0.852	0.428	0.001

Note. GCA = General cognitive ability, ST = Stereotype consistency, BR = Base-rate consistency, Ratio = Base-rate ratio. Dfn = "degrees of freedom in the numerator", DFd = "degrees of freedom in the denominator", *ges* = "generalised eta squared".

Figure 7

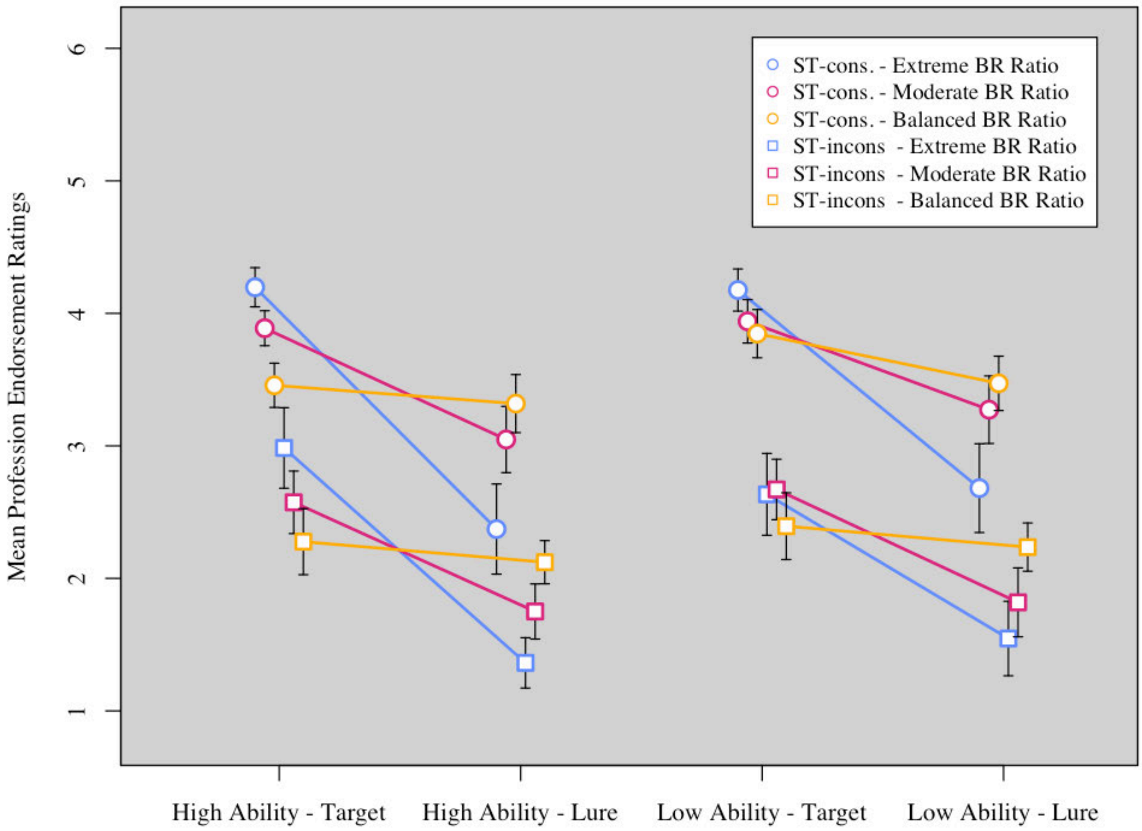
Mean Profession Endorsement Ratings (AR)



Note. ST = Stereotype consistency (cons. = consistent; incons. = inconsistent), Target = Base-rate is consistent with the conclusion profession, Lure = Base-rate is inconsistent with the conclusion profession and BR = base-rate. Error bars indicate 95% confidence intervals.

Figure 8

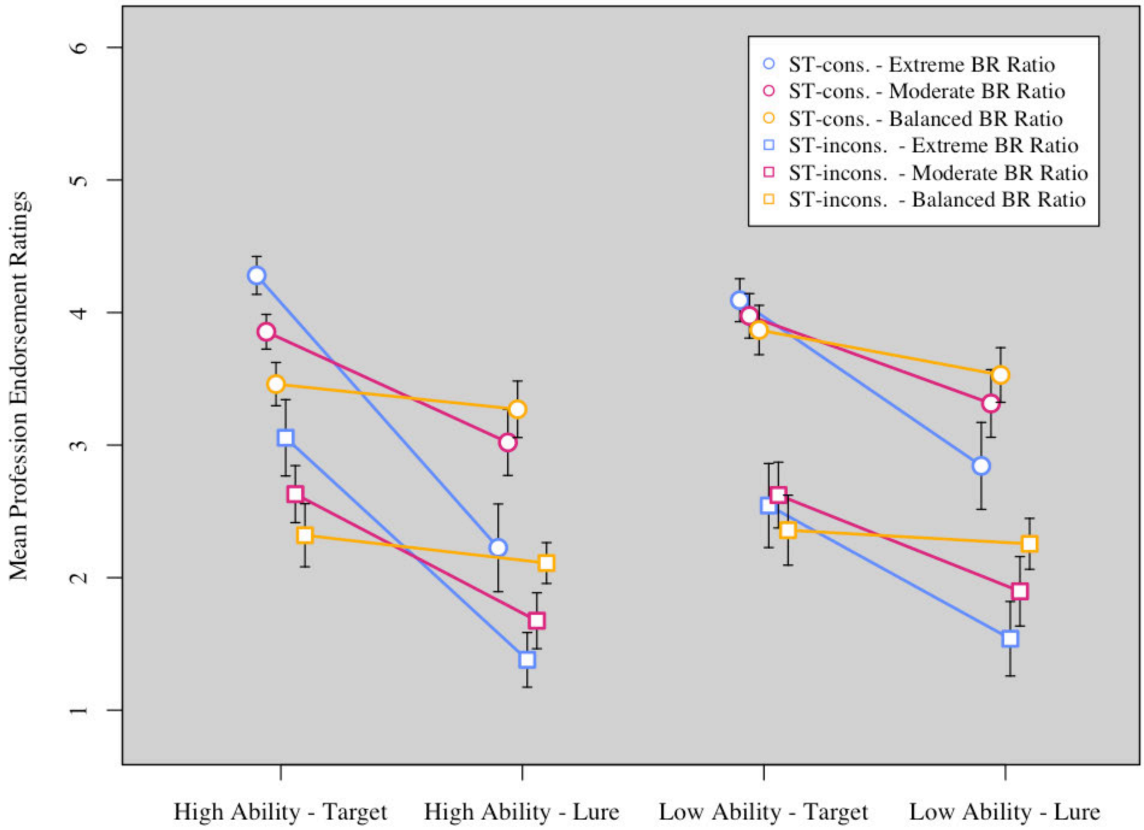
Mean Profession Endorsement Ratings (WMC)



Note. ST = Stereotype consistency (cons. = consistent; incons. = inconsistent), Target = Base-rate is consistent with the conclusion profession, Lure = Base-rate is inconsistent with the conclusion profession and BR = base-rate. Error bars indicate 95% confidence intervals.

Figure 9

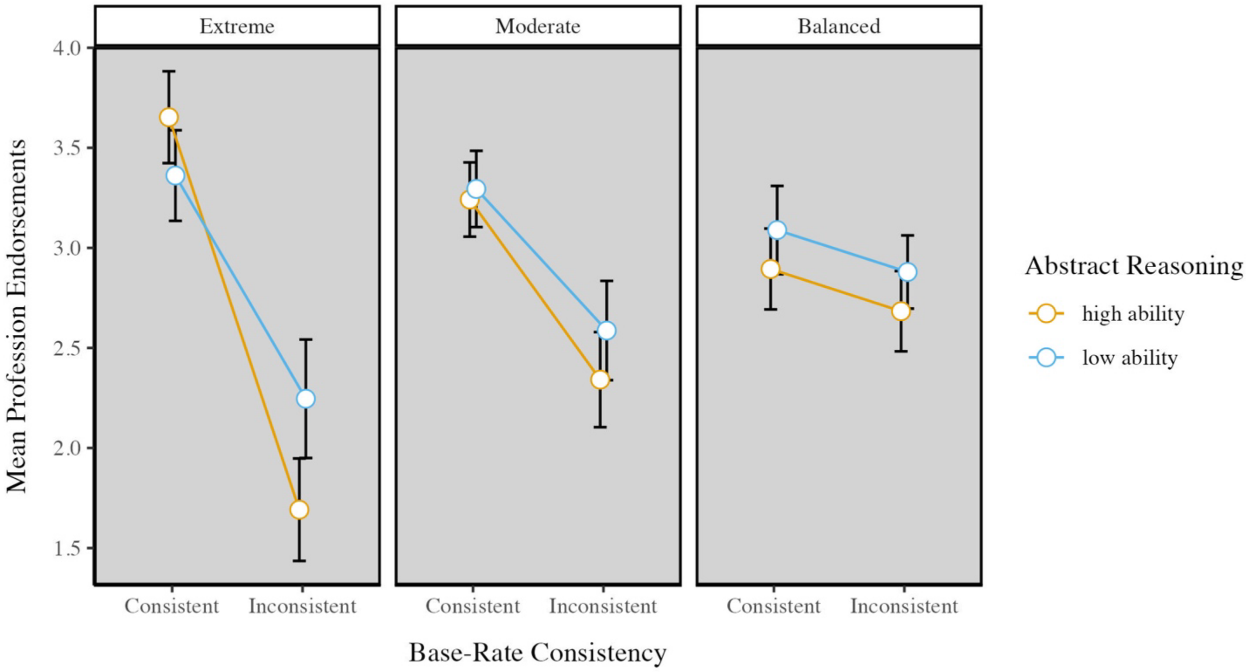
Mean Profession Endorsement Ratings (GCA)



Note. ST = Stereotype consistency (cons. = consistent; incons. = inconsistent), Target = Base-rate is consistent with the conclusion profession, Lure = Base-rate is inconsistent with the conclusion profession and BR = base-rate. Error bars indicate 95% confidence intervals.

Figure 10

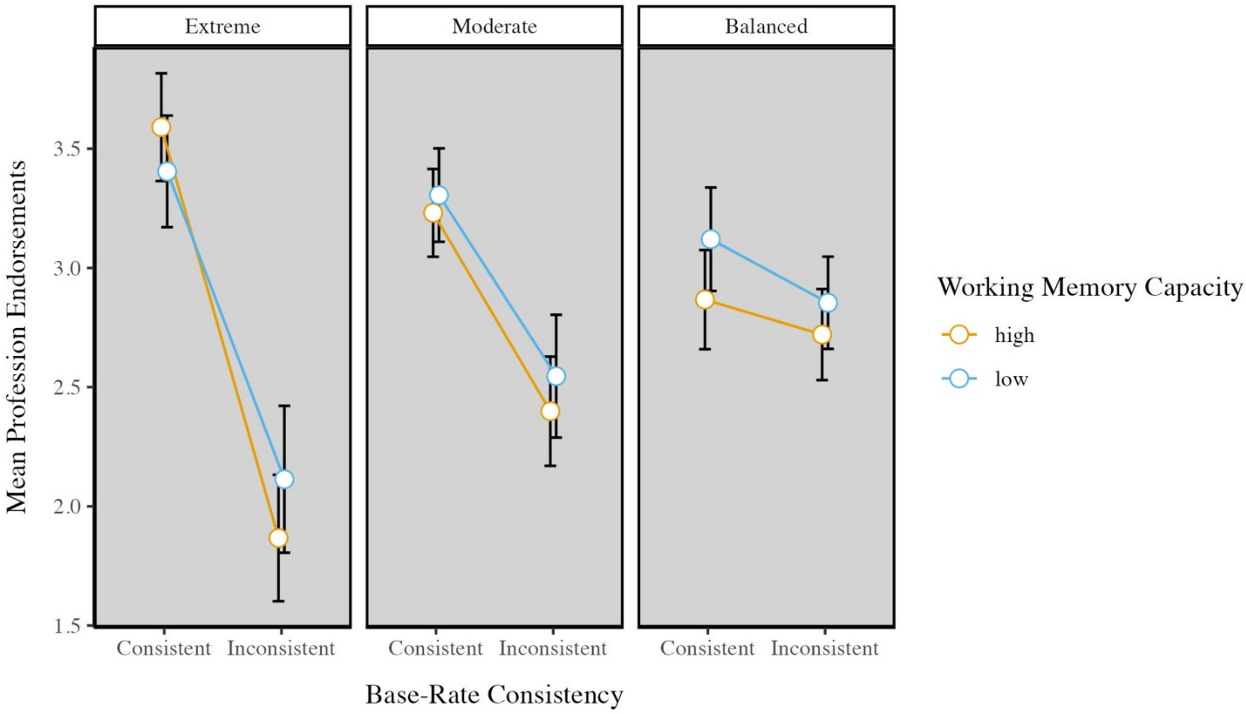
Interaction Between Base-Rate Consistency and Base-Rate Ratio by AR



Note. Extreme = extreme base-rate ratio, Moderate = moderate base-rate ratio, Balanced = balanced base-rate ratio, Consistent = base-rate is consistent with the conclusion profession (i.e., Target), Inconsistent = base-rate is inconsistent with the conclusion profession (i.e., Lure), High Ability = high abstract reasoning ability, Low Ability = low abstract reasoning ability. Error bars represent 95% confidence intervals.

Figure 11

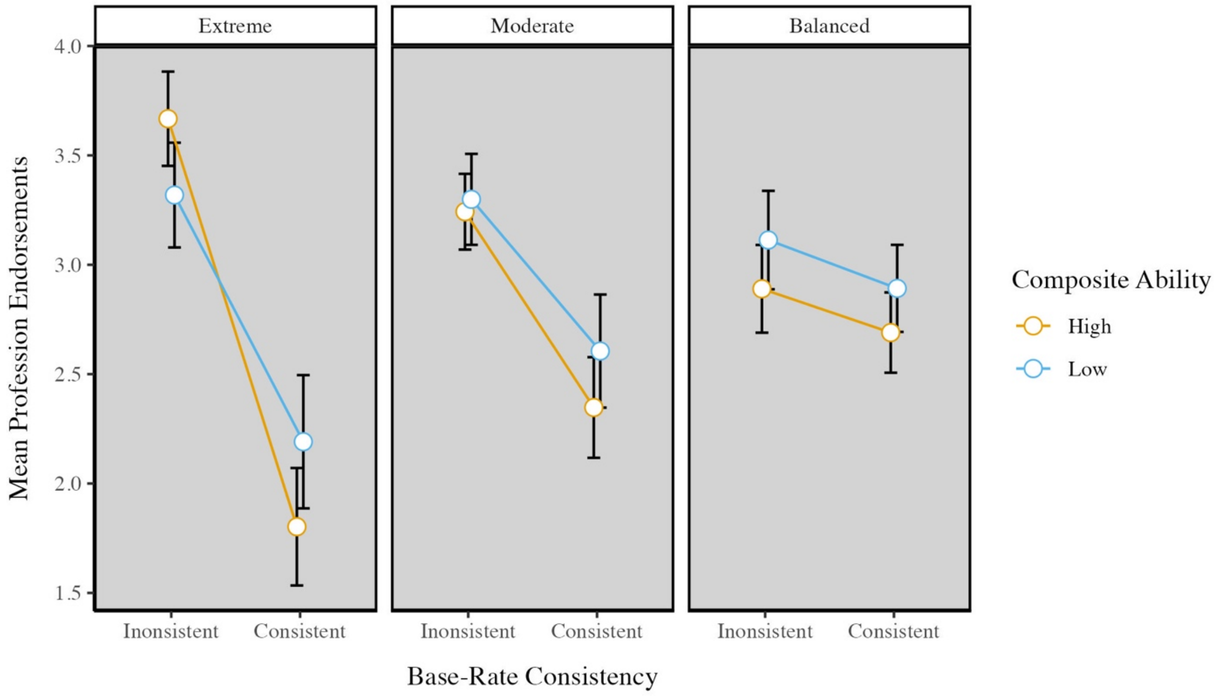
Interaction Between Base-Rate Consistency and Base-Rate Ratio by WMC



Note. Extreme = extreme base-rate ratio, Moderate = moderate base-rate ratio, Balanced = balanced base-rate ratio, Consistent = base-rate is consistent with the conclusion profession, Inconsistent = base-rate is inconsistent with the conclusion profession, High Ability = high working memory capacity, Low Ability = low working memory capacity. Error bars represent 95% confidence intervals.

Figure 12

Interaction Between Base-Rate Consistency and Base-Rate Ratio by GCA



Note. Extreme = extreme base-rate ratio, Moderate = moderate base-rate ratio, Balanced = balanced base-rate ratio, Consistent = base-rate is consistent with the conclusion profession, Inconsistent = base-rate is inconsistent with the conclusion profession, High Ability = high general cognitive ability, Low Ability = low general cognitive ability. Error bars represent 95% confidence intervals.

The Influence of Cognitive Ability on Base-Rate Task Performance

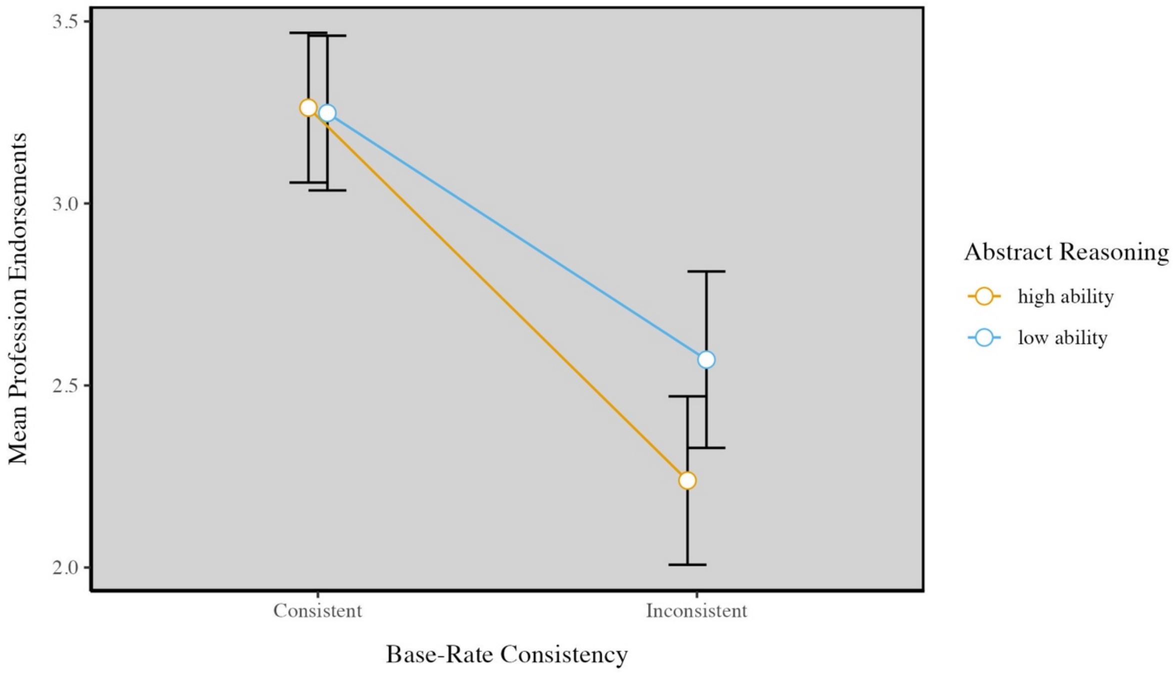
There was a main effect of AR; however, the main effects of WMC and GCA were not significant (Tables 5-7). Those in the low AR group gave higher endorsement ratings overall (M = 2.91, SD = 1.18) than those in the high AR group (M = 2.75, SD = 1.14). This difference is consistent with a response bias shift towards more lenient responding for low AR participants.

More importantly, as expected, the observed base-rate task effects were moderated by interactions with the cognitive ability factors. However, these interactions were significant only for the AR and GCA analyses. Thus, the following discussion will not apply to the WMC analysis. Most notably, base-rate consistency interacted with cognitive ability in the expected direction. Participants with high cognitive ability showed a larger difference in endorsement ratings across the base-rate consistency conditions, compared to those with low ability (see Figures 13 and 14). This result suggests that participants displayed differential sensitivity to base-rates depending on their cognitive resources: those with more resources showed greater sensitivity to base-rates compared to those with less. Dissociations such as this are often used to support dual-process theories. This two-way interaction was complicated by a significant three-way interaction between base-rate consistency, base-rate ratio and cognitive ability (see Figures 10 and 12). Follow-up analyses revealed significant simple two-way interactions between cognitive ability and base-rate consistency for the extreme but not for the moderate or balanced base-rate ratio conditions (see Appendix B). Statistical significance was accepted at the $p<0.017$ level to correct for multiple comparisons. Therefore, the effect of cognitive ability on participants' sensitivity to base-rates depended upon the presentation of an extreme base-rate ratio.

Against expectations, cognitive ability showed no interactions with stereotype consistency for any analyses. In other words, stereotype information informed the judgements of high and low ability participants to an equal degree. This result conflicts with a possible prediction of dual-process theory: differences in cognitive resources did not lead to differential use of stereotype information.

Figure 13

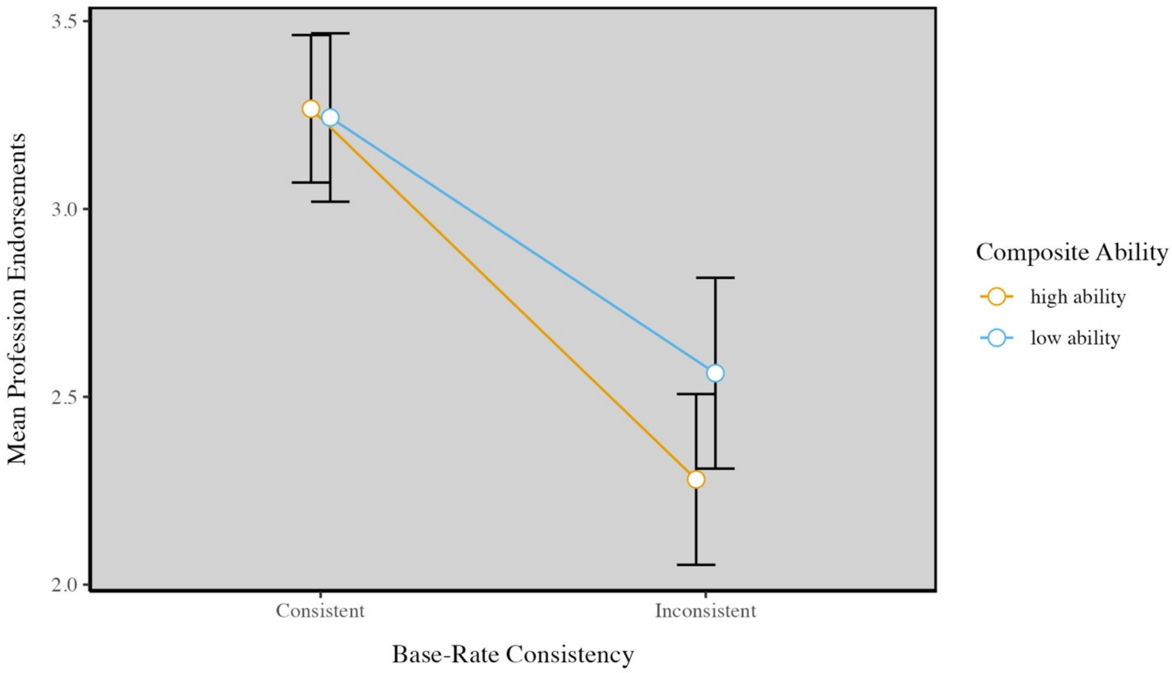
Two-Way Interaction Between Base-Rate Consistency and AR



Note. Consistent = base-rate is consistent with the conclusion profession, Inconsistent = base-rate is inconsistent with the conclusion profession, High Ability = high Abstract Reasoning ability, Low Ability = low Abstract Reasoning ability. Error bars represent 95% confidence intervals.

Figure 14

Two-Way Interaction Between Base-Rate Consistency and GCA



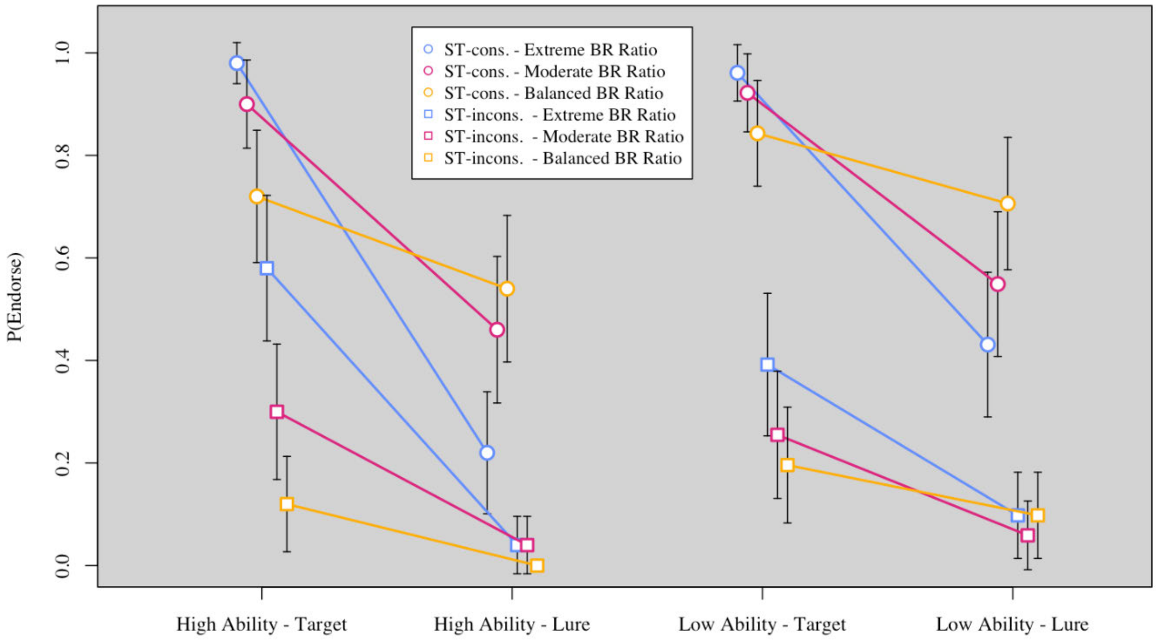
Note. Consistent = base-rate is consistent with the conclusion profession, Inconsistent = base-rate is inconsistent with the conclusion profession, High Ability = high composite cognitive ability, Low Ability = low composite cognitive ability. Error bars represent 95% confidence intervals.

Signed Difference Analysis

Both the AR and GCA ANOVAs identified a larger influence of base-rate information on endorsement ratings under high versus low cognitive ability. This interaction is consistent with the multiple latent processes typically proposed by dual-process theories and the 2D SDT model. However, it is not necessarily inconsistent with a single latent process. Thus, SDA was used to more conclusively test the most general version of the 1D SDT model against the obtained dissociation data. For testing the 1D and 2D signal detection models using SDA, binary judgements are required. Therefore, the 6-point Likert scale used to capture participants' endorsement ratings was dichotomised such that ratings above 3 indicated conclusion endorsements and ratings below 3 indicated rejections. Figures 13, 14 and 15 depict the resulting mean endorsement rates for the AR, WMC and GCA analyses, respectively. Visual inspection of these plots shows that the key effects observed for the mean ratings in Figures 5, 6 and 7 were left largely unchanged. However, most importantly, the dichotomised ratings in Figures 13 to 15 do not display the vector pattern (+, -, -, +) that is forbidden by the 1D model and so this model can fit the data perfectly ($p = 1$ from conjoint monotonic regression tests; see Kalish et al., 2016). In the current data, the high ability group never showed higher target and lure discrimination across two conditions, while the low ability group simultaneously showed lower discrimination on these same items (or vice versa).

Figure 15

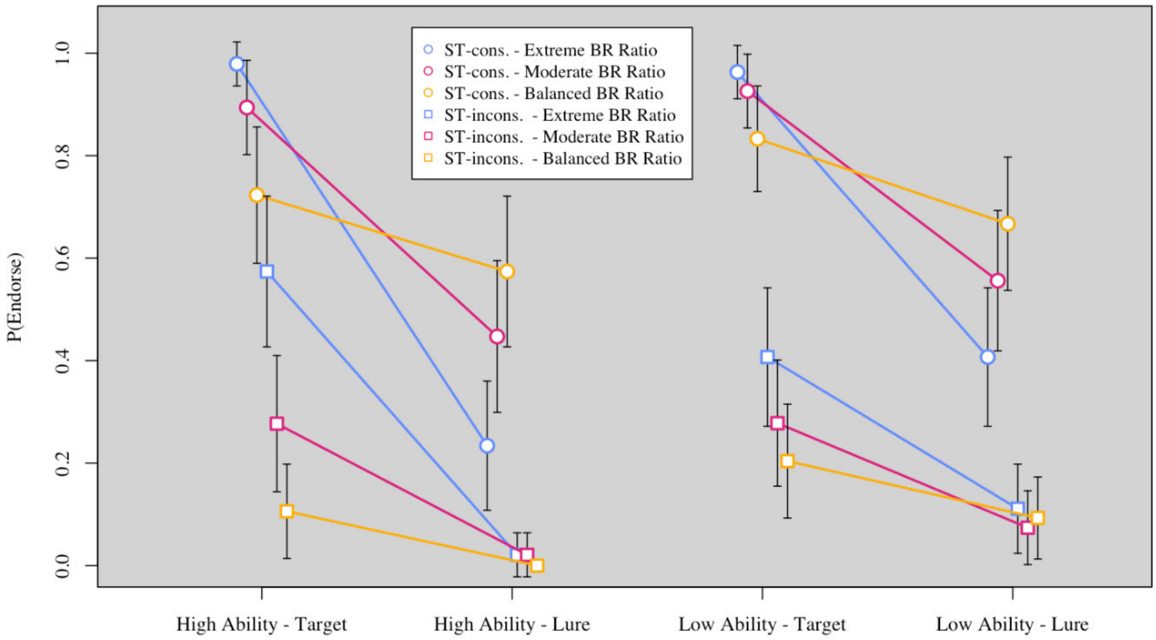
Mean Conclusion Profession Endorsement Rates based on Dichotomised Ratings (AR)



Note. ST = Stereotype consistency (cons. = consistent; incons. = inconsistent), Target = Base-rate is consistent with the conclusion profession, Lure = Base-rate is inconsistent with the conclusion profession and BR = base-rate. Error bars indicate 95% confidence intervals.

Figure 16

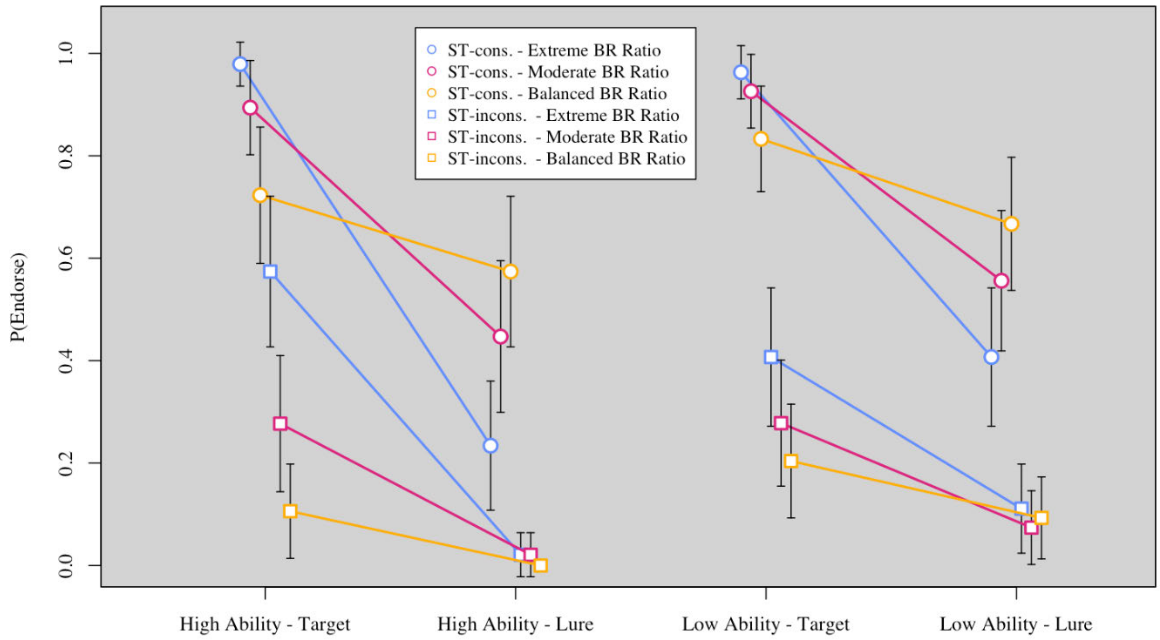
Mean Conclusion Profession Endorsement Rates based on Dichotomised Ratings (WMC)



Note. ST = Stereotype consistency (cons. = consistent; incons. = inconsistent), Target = Base-rate is consistent with the conclusion profession, Lure = Base-rate is inconsistent with the conclusion profession and BR = base-rate. Error bars indicate 95% confidence intervals.

Figure 17

Mean Conclusion Profession Endorsement Rates based on Dichotomised Ratings (GCA)



Note. ST = Stereotype consistency (cons. = consistent; incons. = inconsistent), Target = Base-rate is consistent with the conclusion profession, Lure = Base-rate is inconsistent with the conclusion profession and BR = base-rate. Error bars indicate 95% confidence intervals.

Discussion

Summary of Findings

This thesis aimed to gain insight into the processes underlying reasoning via the comparison of single- and dual-process theories of performance on the base-rate task. ANOVA was conducted to test for dissociations based on cognitive ability in base-rate task effects. Subsequently, SDA was performed to assess the single-process account's compatibility with the data. The study replicated many of the standard task effects. Participants endorsed conclusions more when they were consistent with either base-rates or stereotypes, indicating sensitivity to both cues. Participants displayed greater base-rate sensitivity in the extreme base-rate ratio condition compared to the moderate and balanced conditions. WMC showed no interactions with base-rate task effects. In line with dual-process theory, high AR and GCA individuals displayed greater base-rate sensitivity under an extreme base-rate ratio. In contrast to dual-process theory predictions, cognitive ability did not interact with stereotype consistency across any analyses. Critically, while some traditional dissociation evidence for dual-process theory was observed, the SDA showed that the 1D SDT model could not be rejected.

Discussion of Findings

Base-Rate Task Effects

Participants' were sensitive to both base-rate and stereotype information, consistent with previous findings (Macchi, 2005; Pennycook & Thompson, 2012). Also consistent with previous findings (e.g., Newman et al., 2017; Pennycook et al., 2015; Yang et al., 2023), base-rate sensitivity was impacted by base-rate ratio. These results add further evidence that people do not universally neglect base-rates and can use them effectively under a range of conditions (Birnbaum, 2004; Koehler, 1996). In this study, exposure to many problems (i.e.,

48) with varying base-rate ratios likely heightened participants' base-rate sensitivity (Fischhoff & Bar-Hillel, 1984).

The Influence of Cognitive Ability on Base-Rate Task Performance

The most notable effect observed was that high AR and GCA participants were more sensitive to base-rate information than low AR and GCA participants under an extreme base-rate ratio. Similar correlations between cognitive ability and logical responding have been reported on a wide range of reasoning tasks, including causal base-rate problems (Stanovich & West, 1998, 2008; Toplak et al., 2011). Ability-based dissociations like this constitute traditional evidence for multiple processes and can be explained by dual-process theory (Evans & Stanovich, 2013). The greater cognitive resources of high ability participants better facilitated the Type 2 (analytic) processing necessary to use base-rate information and to override the default, stereotypical response cued by their Type 1 (intuitive) processing. However, based on the testing of formal 1D and 2D signal detection models, a single-process account cannot be ruled out. The 1D model describes participants' judgements using three parameters: two decision thresholds ($C_{low\ and}$ C_{high}) and one discriminability parameter (*d*). Thus, instead of two dimensions of subjective strength, it situates differences in task performance between ability groups with the three parameters. For example, according to this account, the observed dissociation could be attributed to high ability participants setting a more conservative decision threshold (Stephens et al., 2019). This finding strengthens and extends related work that has shown that the 1D model can account for a large range of reasoning data, previously thought to support dual-processes (Stephens et al., 2018, 2020).

High and low ability participants made similar use of stereotypes in their judgements. Therefore, this study did not observe a double dissociation (i.e., opposing effects of a variable theorised to distinguish multiple processes on logical and intuitive responses), which has been reported for other reasoning tasks and is typically considered strong evidence of dual-processes (Evans & Stanovich, 2013). For example, studies of belief bias have reported a decrease in logical accuracy *and* an increase in belief-based responding under time-pressure (Evans & Curtis-Holmes, 2005). If accurate, this finding poses a potential challenge for dual-process theories of base-rate task performance. Alternatively, perhaps participants in this study lacked the motivation to override stereotypical responses. For instance, high ability individuals have been shown to suppress belief biases only when explicitly instructed to answer logically (Evans, 2010). To investigate this possibility, future studies could include an instruction manipulation to direct participants away from or towards base-rate versus stereotype responding (see Handley et al., 2011).

The lack of effects for WMC is surprising given the many studies that have implicated working memory in performance on reasoning tasks (De Neys, 2006; Evans & Curtis-Holmes, 2005). One interpretation is that WMC actually plays a limited role in base-rate sensitivity. Thus, the effects seen for AR could reflect its closer relationship to the reasoning skills required for processing probability information. While RAPM-S more directly assesses reasoning ability via the use of logic in new situations, OSPAN-S measures the capacity to hold and manipulate information, which is just one component of reasoning (Unsworth & Engle, 2005). If accurate, this interpretation challenges dual-process perspectives that consider working memory essential to base-rate use (Evans & Stanovich, 2013). Alternatively, a potential ceiling effect caused by the OPSPAN-S measure could have limited the detection of group differences. Indeed, WMC scores were skewed towards the higher end of the scale (see histogram in Appendix C). Although initial evaluations of the OPSPAN-S have been promising and operation span tasks are generally psychometrically sound (see

Conway et al., 2005), further research may be required to more conclusively establish its reliability and validity (Oswald et al., 2015). Future work on base-rate sensitivity could involve assessing WMC using a more extensively evaluated measure. However, the OPSPAN-S uses a standard and rigorous procedure involving several rounds of practice questions and feedback. Given the online nature of the study, it is also possible that some participants used writing aids to remember the consonants. If true, this would have amplified a potential ceiling effect. Ensuring compliance is a common challenge for online studies (Kees et al., 2017) and future research may benefit from a lab-based approach.

Theoretical Implications

Further tests of the 1D model

Previous research has shown that a 1D SDT model can explain a large range of reasoning data from logic tasks (Stephens et al., 2018, 2020), in which inductive and deductive judgements distinguish Type 1 and Type 2. The primary contribution of the current project was extending the success of the 1D model to a different task, the base-rate task, targeting a key distinguishing feature of Type 1 and Type 2 – cognitive ability. The success of a single-process model on another popular reasoning task dominated by the dual-process perspective underscores and strengthens the value of using formal models to precisely define and test the predictions of competing accounts. Prior to this study, only one other investigation had formally compared 1D and 2D models on the base-rate task (Sikora-Przibilla, 2022). One notable advantage of SDT models is that they can be applied to provide a clearer understanding of reasoning performance than traditionally used rates of overall accuracy, which obscure the relative contributions of discriminability and bias. Moving forward, it is recommended that researchers give more consideration to the impact of bias and discriminability. The formal models described here will be a useful tool for this task.

This study targeted cognitive ability and working memory to try to capture the Type 1/2 distinction. However, perhaps other experimental designs would reveal the need for a model with two discriminability parameters. A direction for future research is thus to determine whether alternative manipulations theorised to distinguish Type 1 and 2 produce similar results. A prior study has already used time-pressure to distinguish Type 1 and Type 2 on the base-rate task, and its findings regarding the 1D model are consistent with those reported here (Sikora-Przibilla, 2022). Future studies could consider combining the current approach and that of Sikora-Przibilla (2022) by investigating both cognitive ability and time-pressure factors in a single design. Researchers could also determine whether the 1D model generalises to other cognitive ability measures commonly used by dual-process theorists, like the Scholastic Aptitude Test (SAT) (Stanovich & West, 1998). More broadly, the described models should continue to be extended to other popular tasks in the reasoning field, such as the ratio-bias task (Thompson, 2021) and the transitive reasoning task (Ameel et al., 2007; Scott, 2021).

Alignment with existing evidence

The retention of a 1D model suggests that a dual-process theory of base-rate task performance may be unwarranted. This finding converges with recent evidence in the wider literature that has challenged a clear distinction between the operation of Type 1 and Type 2 processes (Handley et al., 2011; Hayes et al., 2023). For example, base-rates may be accessible to Type 1 processing (Pennycook et al., 2014). In response, contemporary theorising has focused on accommodating these findings within more elaborate dual-process frameworks that blur the classic distinction between Type 1 and Type 2. For example, the logical intuitions model of De Neys (2012) proposes that people may access Type 2 processing, but only after a conflict is detected between Type 1 *intuitive* and Type 1 *logical* processing (also see Handley & Trippas, 2015). An alternative response is to embrace a single-process account, removing the problematic distinction altogether. The results of the current study support this second solution and add to the growing view that single-process theories have been dismissed prematurely (Loftus et al., 2004; Newell et al., 2010; Newell & Dunn, 2008). It is thus recommended that future research shift away from the differentiation of ever more complex dual-process explanations and towards the investigation of single-process accounts. Further research is needed to understand the unidimensional mechanism for evaluating subjective strength. Possibilities for this mechanism include weighing-and-adding evidence (Juslin et al., 2009) or Bayesian belief revision (Oaksford & Chater, 1994).

Practical Implications

Base-rate sensitivity is crucial to decision-making performance in many practical contexts, including petroleum exploration (Milkov, 2017) and human resources (Whyte & Sue-Chan, 2002). Although people do not completely ignore base-rates, they often underweight them (Pennycook et al., 2016). This observation supports the value of applied interventions to promote more accurate probability-based judgements. Currently, many such interventions emphasise strategies that shift people from Type 1 to Type 2 processing (Borodin, 2016; Lambe et al., 2016; Tsalatsanis et al., 2015). If, as the current study suggests, the Type 1/2 distinction is unwarranted, then we must re-evaluate the dual-process approach to intervention design and explanation. Without a convenient qualitative distinction that applies broadly, interventions will likely proceed on a more task-specific basis. Examined from the SDT perspective, designing helpful interventions may become more about: 1) enhancing people’s ability to discriminate and/or 2) ensuring that response thresholds are set optimally. For example, if an individual has a systematic tendency to answer leniently, an intervention may highlight the costs of this bias.

Conclusion

This study provides further insight into the nature of human judgement and decision-making in scenarios such as that presented at the beginning of this thesis. The results showed that people's conclusions about whether or not the 'mystery man' is a lawyer are likely to be impacted by factors such as his hobbies and personality traits (stereotype information), the relative frequency of lawyers at the awards ceremony (base-rate information), and the interaction between these factors and peoples' available cognitive resources (cognitive ability). The main contribution of this study was clarifying the underlying structure of base-rate task performance under low and high cognitive ability, which is widely believed to reflect multiple processes. While traditional dissociations in support of dual-processes were observed, a targeted test of competing single- and dual-process accounts showed that the data were compatible with a single process. This finding supports and extends related investigations of reasoning with formal models (e.g., Stephens et al., 2018), which have similarly confirmed the viability of a single-process account. Based on considerations of parsimony, this study challenges the value of retaining the dual-process distinction in explanations of base-rate task performance. If the dual-process distinction is unwarranted, then the many applied interventions and theories that are grounded in it will need to be revaluated—perhaps instead drawing on a single-process account.

References

Allen, T. J., Sherman, J. W., Conrey, F. R., & Stroessner, S. J. (2009). Stereotype strength and attentional bias: Preference for confirming versus disconfirming information depends on processing capacity. Journal of Experimental Social Psychology, 45(5), 1081–1087. https://doi.org/10.1016/j.jesp.2009.06.002

Ameel, E., Verschueren, N., & Schaeken, W. (2007). The relevance of selecting what's relevant: A dual process approach to transitive reasoning with spatial relations. Thinking & Reasoning, 13, 164–187. https://doi.org/10.1080/13546780600780671

Arthur, W., & Day, D. V. (1994). Development of a Short form for the Raven Advanced Progressive Matrices Test. Educational and Psychological Measurement, 54(2), 394–403. https://doi.org/10.1177/0013164494054002013

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. Cognition, 158, 90–109. https://doi.org/10.1016/j.cognition.2016.10.014

Barberis, N., & Thaler, R. H. (2002). A Survey of Behavioral Finance (SSRN Scholarly Paper 327880). https://doi.org/10.2139/ssrn.327880

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, *30*(3), 241–254. https://doi.org/10.1017/S0140525X07001653

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. Acta Psychologica, *44*(3), 211–233. https://doi.org/10.1016/0001-6918(80)90046-3

Birnbaum, M. H. (2004). Base rates in Bayesian inference. In *Cognitive Illusions*. Psychology Press.

Borodin, A. (2016). *The Need for an Application of Dual-Process Theory to Mathematics Education*. https://www.repository.cam.ac.uk/handle/1810/294053

Cassimatis, N. L., Bello, P., & Langley, P. (2008). Ability, Breadth, and Parsimony in Computational Models of Higher-Order Cognition. *Cognitive Science*, *32*(8), 1304–1322. https://doi.org/10.1080/03640210802455175

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Creelman, N. A. M., C. Douglas. (2004). *Detection Theory: A User's Guide* (2nd ed.). Psychology Press. https://doi.org/10.4324/9781410611147

Dahlman, C., Zenker, F., & Sarwar, F. (2016). Miss rate neglect in legal evidence. *Law, Probability and Risk*, *15*(4), 239–250. https://doi.org/10.1093/lpr/mgw007

De Neys, W. (2006). Dual Processing in Reasoning Two Systems but One Reasoner. *Psychological Science*, *17*, 428–433. https://doi.org/10.1111/j.1467-9280.2006.01723.x

De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science*, *7*(1), 28–38. https://doi.org/10.1177/1745691611429354

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, *11*(4), 349–381. https://doi.org/10.1080/13546780442000222

Dunn, J. C., & Anderson, L. (2018). Signed difference analysis: Testing for structure under monotonicity. *Journal of Mathematical Psychology*, *85*, 36–54. https://doi.org/10.1016/j.jmp.2018.07.002

Epstein, S. (1998). Cognitive-experiential self-theory: A dual-process personality theory with implications for diagnosis and psychotherapy. In *Empirical perspectives on the* *psychoanalytic unconscious* (pp. 99–140). American Psychological Association. https://doi.org/10.1037/10256-004

Evans, J. S. B. T. (2010). Intuition and Reasoning: A Dual-Process Perspective. *Psychological Inquiry*, *21*(4), 313–326.

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*(4), 382–389. https://doi.org/10.1080/13546780542000005

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Fischhoff, B., & Bar-Hillel, M. (1984). Diagnosticity and the base-rate effect. *Memory & Cognition*, *12*(4), 402–410. https://doi.org/10.3758/BF03198301

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning*, *15*(2), 105–128. https://doi.org/10.1080/13546780802711185

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (pp. xi, 455). John Wiley.

Hamm, R. M. (1996). Physicians neglect base rates, and it matters. *Behavioral and Brain Sciences*, *19*(1), 25–26. https://doi.org/10.1017/S0140525X00041261

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 28–43. https://doi.org/10.1037/a0021098

Handley, S. J., & Trippas, D. (2015). Dual processes and the interplay between knowledge and structure: A new parallel processing model. *Psychology of Learning and Motivation - Advances in Research and Theory*, *62*, 33–58. https://doi.org/10.1016/bs.plm.2014.09.002

Hayes, B. K., Stephens, R. G., & Dunn, J. C. (2023). Unifying theories of reasoning and decision making. *Behavioral and Brain Sciences*, *46*, e126. https://doi.org/10.1017/S0140525X22002916

Jaspersen, J. G., & Aseervatham, V. (2017). The Influence of Affect on Heuristic Thinking in Insurance Demand: The Influence of Affect on Heuristic Thinking in Insurance Demand. *Journal of Risk and Insurance*, *84*(1), 239–266. https://doi.org/10.1111/jori.12088

Juslin, P., Nilsson, H., & Winman, A. (2009). Probability Theory, Not the Very Guide of Life. *Psychological Review*, *116*, 856–874. https://doi.org/10.1037/a0016979

Kahneman, D. (2011). *Thinking, fast and slow* (p. 499). Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251. https://doi.org/10.1037/h0034747

Kalish, M. L., Dunn, J. C., Burdakov, O. P., & Sysoev, O. (2016). A statistical test of the equality of latent orders. *Journal of Mathematical Psychology*, *70*, 1–11. https://doi.org/10.1016/j.jmp.2015.10.004

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141–155. https://doi.org/10.1080/00913367.2016.1269304

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*(1), 1–17. https://doi.org/10.1017/S0140525X00041157

Lambe, K. A., O'Reilly, G., Kelly, B. D., & Curristan, S. (2016). Dual-process cognitive interventions to enhance diagnostic reasoning: A systematic review. *BMJ Quality & Safety*, *25*(10), 808–820. https://doi.org/10.1136/bmjqs-2015-004417

Leeuw, J. R. de, Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an Open-Source Collaborative Ecosystem of Behavioral Experiments. *Journal of Open Source Software*, *8*(85), 5351. https://doi.org/10.21105/joss.05351

Lewandowsky, S., & Farrell, S. (2011). *Computational Modeling in Cognition: Principles and Practice*. https://doi.org/10.4135/9781483349428

Lin, Y., Osman, M., & Ashcroft, R. (2017). Nudge: Concept, Effectiveness, and Ethics. *Basic and Applied Social Psychology*, *39*, 1–14. https://doi.org/10.1080/01973533.2017.1356304

Loftus, G. R., Oberg, M. A., & Dillon, A. M. (2004). Linear theory, dimensional theory, and the face-inversion effect. *Psychological Review*, *111*(4), 835–863. https://doi.org/10.1037/0033-295X.111.4.835

Luthra, M., & Todd, P. (2019). Role of Working Memory on Strategy Use in the Probability Learning Task. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), Proceedings of the 41th Annual Meeting of the Cognitive Science Society, CogSci 2019: Creativity + Cognition + Computation (pp. 721-727). Cognitive Science Society. https://mindmodeling.org/cogsci2019/papers/0141/index.html

Macchi, G. P., Laura. (2005). The representation of the task: The case of the Lawyer–Engineer problem in probability judgement. In *The Shape of Reason*. Psychology Press.

Machery, E. (2012). Dissociations in Neuropsychology and Cognitive Neuroscience. *Philosophy of Science*, *79*(4), 490–518. https://doi.org/10.1086/668002

Milkov, A. (2017). Integrate instead of ignoring: Base rate neglect as a common fallacy of petroleum explorers. *AAPG Bulletin*, *101*, 1905–1916. https://doi.org/10.1306/0327171622817003

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*, *12*(8), 285–290. https://doi.org/10.1016/j.tics.2008.04.009

Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition*, *38*(5), 563–581. https://doi.org/10.3758/MC.38.5.563

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1154–1170. https://doi.org/10.1037/xlm0000372

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608–631. https://doi.org/10.1037/0033-295X.101.4.608

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, *11*(6), 988–1010. https://doi.org/10.3758/BF03196730

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, *47*(4), 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72. https://doi.org/10.1016/j.cogpsych.2015.05.001

Pennycook, G., & Thompson, V. A. (2012). Reasoning with base rates is routine, relatively effortless, and context dependent. *Psychonomic Bulletin & Review*, *19*(3), 528–534. https://doi.org/10.3758/s13423-012-0249-3

Pennycook, G., & Thompson, V. A. (2017). Base-rate neglect. In *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory, 2nd ed* (pp. 44–61). Routledge/Taylor & Francis Group.

Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 544–554. https://doi.org/10.1037/a0034887

Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *53A*(4), 1202–1223. https://doi.org/10.1080/02724980050156362

Raven, J., & Raven, J. (2003). Raven Progressive Matrices. In *Handbook of nonverbal assessment* (pp. 223–237). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-0153-4_11

Rotello, C. M., & Heit, E. (2009). Modeling the effects of argument length and validity on inductive and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1317–1330. https://doi.org/10.1037/a0016648

2(2–4), 48–71. https://doi.org/10.1007/s13162-012-0030-9

Scott, N. (2021). *Investigating single-process and dual-process theories of transitive reasoning: Applying Signal Detection Theory and Signed Difference Analysis.* [Honours thesis, The University of Adelaide]. https://digital.library.adelaide.edu.au/dspace/handle/2440/133947

Sikora-Przibilla, A. (2022). *Base-rates, fast and slow: testing single-process and dual-process models of reasoning.* [Unpublished honours thesis]. The University of Adelaide.

Smith, E. R., & Neumann, R. (2005). Emotion Processes Considered from the Perspective of Dual-Process Models. In *Emotion and consciousness* (pp. 287–311). The Guilford Press.

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

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*(2), 161–188. https://doi.org/10.1037/0096-3445.127.2.161

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665. https://doi.org/10.1017/S0140525X00003435

Stengård, E., Juslin, P., Hahn, U., & van den Berg, R. (2022). On the generality and cognitive basis of base-rate neglect. *Cognition*, *226*, 105160. https://doi.org/10.1016/j.cognition.2022.105160

Stephens, R. G., Dunn, J. C., & Hayes, B. K. (2018). Are there two processes in reasoning? The dimensionality of inductive and deductive inferences. *Psychological Review*, *125*(2), 218–244. https://doi.org/10.1037/rev0000088

Stephens, R. G., Dunn, J. C., Hayes, B. K., & Kalish, M. L. (2020). A test of two processes: The effect of training on deductive and inductive reasoning. *Cognition*, *199*, 104223. https://doi.org/10.1016/j.cognition.2020.104223

Stephens, R. G., Matzke, D., & Hayes, B. K. (2019). Disappearing dissociations in experimental psychology: Using state-trace analysis to test for multiple processes. *Journal of Mathematical Psychology*, *90*, 3–22. https://doi.org/10.1016/j.jmp.2018.11.003

Sunstein, C. R. (2003). Terrorism and Probability Neglect. *Journal of Risk and Uncertainty*, *26*(2/3), 121–136.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness* (pp. x, 293). Yale University Press.

Thompson, V. A. (2021). Eye-tracking IQ: Cognitive capacity and strategy use on a ratio-bias task. *Cognition*, *208*, 104523. https://doi.org/10.1016/j.cognition.2020.104523

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289. https://doi.org/10.3758/s13421-011-0104-1

Tsalatsanis, A., Hozo, I., Kumar, A., & Djulbegovic, B. (2015). Dual Processing Model for Medical Decision-Making: An Extension to Diagnostic Testing. *PLoS ONE*, *10*(8), e0134800. https://doi.org/10.1371/journal.pone.0134800

Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory & Cognition*, *33*(2), 213–220. https://doi.org/10.3758/BF03195310

Whyte, G., & Sue-Chan, C. (2002). The Neglect of Base Rate Data by Human Resources Managers in Employee Selection. *Canadian Journal of Administrative Sciences / Revue Canadienne Des Sciences de l'Administration*, *19*(1), 1–10. https://doi.org/10.1111/j.1936-4490.2002.tb00665.x

Yang, J., Hu, Z., Nie, D., & Zhu, D. (2023). Conflict detection and base-rate extremity. *Acta Psychologica*, *237*, 103960. https://doi.org/10.1016/j.actpsy.2023.103960

Appendix A

Adapted Raven's Advanced Progressive Matrices Instructions

TASK 2 INSTRUCTIONS

Please read the following instructions carefully. You will need to answer some questions about them correctly before you can begin the task.

This task consists of 12 picture puzzles, presented one at a time. For each puzzle, there is a piece missing in the lower right corner. Your task is to select the best option that completes the puzzle from several different options.

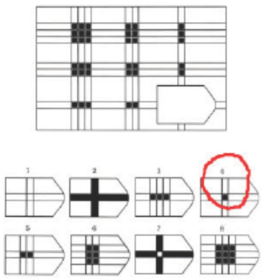
Use your mouse or trackpad to select the answer you believe is best (there are 8 options numbered 1-8 that can be selected by clicking the hollow circle on the left of the number). Once you have selected an answer, click "Next Question" to move to the next question. You cannot miss test items and you cannot go back to a previous question once you have clicked "Next Question". It is important to make sure that you are certain of your choice before submitting.

Click the button below to see the next page of instructions.

Next >

Look at the below example test item labelled "Example Figure". It contains a pattern inside a rectangular frame with a piece missing in the bottom right corner. Each of the pieces in the panel below the rectangular frame (numbered 1-8) are the right shape to fit the space but they do not all complete the pattern correctly.

Example Figure



Numbers 1, 2, 3, 5, 6, 7 and 8 are wrong in this example. Their shape fits into the blank space but they do not contain the right pattern.

- Number 1 has the correct pattern except the inner square has not been filled in with black.
- Number 7 is incorrect because the inner square is not filled in with black and the parallel lines are filled in with black.
- Number 4 (identified with a red circle in the example figure) is the right pattern in this example because it completes the missing pattern correctly.

In other words, it has the right number of inner squares (1), the right number of parallel lines (4), the inner square has been filled in with black and the parallel lines are not filled in with black.

Click the button below to see the next page of instructions.

Next >

For all of the 12 test items, there will be a similar pattern inside a rectangular frame with a piece missing in the bottom right corner. For each puzzle, your task is to decide which of the 1-8 numbered alternatives in the panel below completes the pattern. They are simple at the beginning and get progressively harder as you go on. All the questions are structured the same way and there are no trick questions. If you pay attention to the way the easier test items work, you will find the later test items less difficult. Please work at your own speed, without interruption, from the beginning to the end of the scale. See how many you can get right. If you are unsure of your answer, please make an educated guess than to leave it blank. You can have as much time as you like.

We encourage you to do your best on this assessment. Remember, some items are quite challenging, so don't worry if you find some of the puzzles difficult.

Click the button below to see the task instruction questions.

Next >

Appendix B

Simple Two-Way Interactions

Abstract Reasoning Simple-Two Way Interactions

Ratio	Effect	DFn	DFd	*F*	*p*	*ges*
Extreme	AR	1	99	2.82	.096	.008
	BR	1	99	148.00	<.001*	.520
	AR x BR	1	99	11.20	.001*	.075
Moderate	AR	1	99	3.05	.084	.017
	BR	1	99	113.00	<.001*	.334
	AR x BR	1	99	1.61	.207	.007
Balanced	AR	1	99	6.53	.012*	.037
	BR	1	99	10.80	.001*	.043
	AR x BR	1	99	0.00	.992	.000

Note. Extreme = extreme base-rate ratio, moderate = moderate base-rate ratio, balanced = balanced base-rate ratio, AR = abstract reasoning ability, BR = base-rate consistency, Ratio = base-rate ratio. To correct for multiple comparisons, statistical significance was accepted at the p<0.017 level.

General Cognitive Ability Simple-Two Way Interactions

Ratio	Effect	DFn	DFd	*F*	*p*	*ges*
Extreme	AR	1	99	0.064	.801	.000
	BR	1	99	138	<.001*	.502
	AR x BR	1	99	8.40	.005*	.058
Moderate	AR	1	99	3.44	.067	.019
	BR	1	99	111	<.001*	.332
	AR x BR	1	99	1.79	.184	.008
Balanced	AR	1	99	7.90	.006*	.045
	BR	1	99	11.0	.001*	.044
	AR x BR	1	99	0.026	.872	.000

Note. Extreme = extreme base-rate ratio, moderate = moderate base-rate ratio, balanced = balanced base-rate ratio, AR = abstract reasoning ability, BR = base-rate consistency, Ratio = base-rate ratio. To correct for multiple comparisons, statistical significance was accepted at the $p<0.017$ level.

Appendix C

Histogram of Overall Participant WMC Scores