Exploring individual differences in deductive reasoning as a function of 'autistic'-like traits

Andrew J. B. Fugard

Doctor of Philosophy
University of Edinburgh
2009

Abstract

From a logical viewpoint, people must reason *to* as well as *from* interpretations in deductive reasoning tasks. There are two main interpretative stances (e.g., Stenning & van Lambalgen, 2004, 2005, 2008): *credulous*, the act of trying to infer the speaker's intended model; and *sceptical*, an adversarial strategy. A range of contextual factors influence interpretation, but there are also differences between individuals across situations. Taking an individual differences approach, this thesis focuses on reasoning in relation to milder variants of the autism spectrum condition (ASC) phenotype in a typically developing (TD) population. Earlier work on discourse processing in ASC using the 'suppression' task (van Lambalgen & Smid, 2004; Pijnacker et al., In press) shows that some aspects of reasoning to interpretations are different in the ASC population. Given that autistic traits involve impairment, e.g., in pragmatic language, and peaks of ability, e.g., in perceptual tasks, it was hypothesised that autistic traits would predict features of the inferences people in the TD population draw.

Data were collected from university students on a range of reasoning tasks making it possible to investigate the extent to which interpretation is consistent across task within individuals. Tasks chosen were: conditional reasoning using the 'suppression' task and Wason's selection task; one and two-premise Aristotelian quantifier reasoning; the Linda problem; and Raven's Advanced Progressive Matrices. Autistic traits were assessed using the Autism Spectrum Quotient (Baron-Cohen et al., 2001), used previously to study autistic traits in TD individuals, and the Broad Autism Phenotype Questionnaire (Hurley et al., 2007).

Autistic traits predicted patterns of inference in many of the tasks. The earlier suppression task result in ASC was replicated and extended in our TD population. Different dimensions of autistic trait related differentially to features of the inferences drawn. Some of the inferences drawn were recognisably related to the credulous versus sceptical distinction and correlated cross-task whilst others were seemingly related to more general top-down versus bottom-up processing preferences. These results provide further evidence of the existence of qualitative individual differences in deductive reasoning. They also show the importance of seeking cross-task correlates to move beyond studies of individual tasks and study reasoning to and from interpretations in the same individual.

Acknowledgements

First and foremost my family deserve special thanks for remaining calm and supporting me when I announced my intention to migrate to cognitive science and do a PhD. Thank you! I would like to thank my first supervisor Keith Stenning for patiently helping me make the transition from computer science to cognitive science. Keith's encouragement to embrace the data, become at one with the data, emerge myself in the data, paid off eventually. I've really enjoyed our chats about reasoning and autism, and found helpful the encouragement and counterexamples. Thanks to my second supervisor, Bob Logie, for discussions about cognitive psychology, and especially for pointing me in the direction of work relating cognitive styles to abilities; John Lee and Guy Politzer for an interesting viva discussion; Alan Bundy and Mateja Jamnik, whose work on the constructive ω-rule was the final push I needed to move towards cognitive science in the first place; Claes Strannegård, who gave me the first opportunity to learn about the psychology of reasoning, and importantly gave me a job in Gothenburg when I was in dire need of income just before I moved to Edinburgh—for this I am deeply grateful; John Raven, for letting me read his copy of his father's MSc thesis and working notes for the Progressive Matrices, and for interesting discussions on everything from psychometrics to politics; Mary Stewart at Heriot-Watt helped with participant recruitment and taught me loads about the autism spectrum; Pat Ferguson, who does a great job of running the Neuroinformatics machine; Karen Fleet, for being a fantastic librarian; Mike Allerhand's course on R helped me get a useful analysis machine rolling; Alan Smaill gave useful pointers on monadic predicate logic; Vladeta Jovović helped me get a sequence in the *On-Line Encyclopedia of Integer Sequences* (which I think is super cool!); and for earlier support and encouragement thanks to Ruth Foster; Jonathan Campbell; and Maurice Clint.

Conversations with many other people were a great help. I'd like to mention especially (randomised): Brendan McGonigle, George Malcolm, Michael Jin, Andrew Wawrzyniak, Kendra DeBusk, Anne Finucane, Thomas Bak, Alexandra Dima, Martha Whiteman, John Davey, Ian Deary, Alex Weiss, Lorenzo Vigentini, Tom Stafford, Autism Diva, Michelle Dawson, and Tim Bates. Thanks to fellow PsychStatsBANTERers for fun and enlightening times. Others helped me maintain some constantish level of sanity over the past few years, for which I am exceedingly thankful. Especially (randomised): Anisha Patel, Jude

Law, Rachel Hanson, Julia Davison, James Withers, Sam Halliday, Andrea Chai, Christos Gkogkas, Gareth Vile, Hicham Rifai, Julija Romanuka, Eve McCann, Susie Jackson, Kim Van Keymeulen, Gauthier Fleutot, Nine, Monika Hunter, Pete Gammie, Charlotte Blease, Nicky Saunders, Antimo Buonocore, Anna Farreras, Carole Lomas, Rita Poberschnigg, Adrianna Teriakidis, Lucy Dixon, Rob 'Dopamine' Kyle, Nelly Lakestani, Trevor Carpenter, Chryssa Retsa, Beth Thomson, and David Nimmo. The Philosophy Society folk also provided entertainment and education: thank you! Thanks to Enrico Mastropaolo, Annalisa Urbano, Mari Solicka, and Malin Nauwerck for being fantastic flatmates during the final scary months of the thesis—the distractions from work were very appreciated $\ddot\smile$

I am very grateful for financial support from an Engineering and Physical Sciences Research Council and Medical Research Council Studentship through the Neuroinformatics and Computational Neuroscience Doctoral Training Centre.

This thesis was typeset with LaTeX 2$_\varepsilon$, using a class written by Martin Reddy and Mary Ellen Foster, with a little help from WinEdt, JabRef, and LaTable. Data analysis was performed using R and the model enumeration code was written using Haskell. Nutritional support was provided by Cappadocia (as were Turkish lessons: merhaba!), Uncle T's, Golden House, Bonsai, the Mosque Kitchen, Palmyra, and Rudy's; and emergency coffee fixes were provided by the Coffee Box and Peter's Yard.

Greetings to my new friends in Salzburg, especially Gernot Kleiter and Niki Pfeifer. Thank you for all the help in getting up and running in Austria!

Finally, thanks very much to all the participants!

$\Gamma \vdash \heartsuit$

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Andrew J. B. Fugard*)

To my parents

Contents

Chapter 1

Introduction

This thesis is on the psychology of deductive reasoning, the process of drawing conclusions from a set of premises. Deduction is a central ability in human cognition. Falmagne and Gonsalves (1995, p. 526) capture the sense in which I consider the ability in this thesis:

> Deductive inference is an activity coordinated with other linguistic and non-linguistic activities and carried out in context, so the proper object of study is not a disembodied process but a person, though the focus is on one particular segment of that person's cognitive activity.

I don't intend to use the term 'reasoning' to refer only to, e.g., the activity of philosophers debating an interpretation of Kant or mathematicians proving theorems in pure mathematics. Reasoners may not necessarily be consciously aware of the premises from which they are drawing inferences and there is a non-trivial mapping from task to cognitive deductive apparatus. Deduction is involved in discourse comprehension, recognising perceived objects, reconstructing memories from partial information, making plans for the future, inferring others' mental states, engaging in argument, flirtation, and a range of other activities.

In psychology a relatively small number of deductive tasks have been considered, and with a narrow notion of what counts as deductive inference. Examples of tasks include especially tests of conditional reasoning (sentences containing 'if'), reasoning about quantified sentences like 'some A are B', and solving various logical puzzles like Sudoku. Initially the focus was on examining how rational people are. Demonstrations by Peter Wason (1966, 1968) of people's failure to reason with classical logic on an ostensibly simple task

> Below are four cards. Each card has a number on one of its sides and a letter on the other. You can see only the exposed face. There is a rule which applies only to the four cards: 'If there is a vowel on one side, then there is an even number on the other side.' Which cards (if any) must you turn in order to decide if the rule is true? Don't turn unnecessary cards.

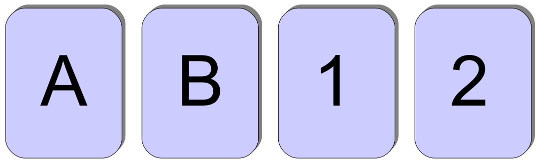

Figure 1.1: Wason's selection task. (If the moon is made of green cheese, then the answer is on p. 1024.)

(see Figure 1.1) are well known. More generally researchers investigate factors influencing the inferences people draw, for instance information ordering (e.g., Johnson-Laird & Steedman, 1978); how the task is interpreted (e.g., Geis & Zwicky, 1971; Politzer, 1986; Bott & Noveck, 2004); the degree to which attitudes are integrated with the presented information (e.g., Gorden, 1953). Each of these factors often result in invalid inferences according to the commonly chosen competence models, usually one particular interpretation of the task material in classical logic or probability. There is evidence, however, that many divergencies from the intended interpretation are reasonable and related to social and communicative processes.

Mathematically one way to model the relationship between premises and a conclusion is using a derivability relation, $\vdash$, where

$$\Gamma \vdash \varphi$$

means that conclusion φ follows from the sentences in the set of premises Γ using only the rules of the particular topic under discussion. The expression $\Gamma \vdash \varphi$ may itself be true or false. There are many possible definitions of $\vdash$: for example, what happens if Γ contains contradictory sentences? In general is it the case that $\Gamma \vdash \varphi \vee \neg\varphi$ for any Γ

and φ?[1] Can old conclusions be withdrawn in light of new evidence? Even tiny sets of premises entail many conclusions, generally infinitely many—how are conclusions chosen? Another approach involves models, which intuitively may be thought of as objects against which the truth of sentences may be judged. If φ is true in all models of the set of premises, Γ, we write $\Gamma \models \varphi$. It is also possible to restrict attention to a subset of models of the premises. A number of disciplines including philosophy, mathematical logic, and computer science study properties of various logical systems, produce algorithms for drawing inferences, finding models, and debate situations in which it is appropriate to use each system. (Chapter 3 has more on notions of logic.)

Psychologically, one might wonder how Γ and φ are represented. How cognitive processes select sentences as premises—recall it cannot be assumed they come naively and directly from the presented information. How inferences are drawn from those premises, or models selected. What causes mistakes and indeed what constitutes a mistake. To illustrate, consider the task of reading the following news article excerpt:[2]

> A man pulled over for driving slowly in the Australian outback has been charged with reckless driving—because police found him driving in reverse.

A man is introduced, we learn he was driving, in Australia, and reasons for his arrest. Now what follows from this description? One might feel that the driver must have been drunk, or suffering from mental distress, or that he decided one day that driving forward was passé. Later in the article we learn:

> The driver, 22, told police he chose to drive in reverse when his gears failed. [...] Police breathalysed the man, who was found not to be under the influence of alcohol.

Do people wait until the final sentence before drawing conclusions about what happened? The reader might wonder why the driver's age was introduced here or might infer something else about the driver based on this age.

Traditionally discussions of reasoning begin with examples of the dire consequences of poor reasoning. One such example is the problem of inferring what follows from these premises:

---

[1]The symbol $\vee$ is disjunction (or) and $\neg$ is negation, so '$\varphi \vee \neg\varphi$' is read 'φ or not φ'.

[2]From http://news.bbc.co.uk/1/hi/world/asia-pacific/5412814.stm (6 October 2006).

If the test is to continue, then the turbine must be rotating fast enough.
The turbine is not rotating fast enough.

Drawing the classically valid conclusion 'the test is not to continue' would have prevented the Chernobyl disaster in 1986, argues Johnson-Laird (1999, p. 110). But this is an example of a clinically sterilized inference. Presumably at the time, the situation was more difficult than an inference about a conditional written on a page, for instance some rules had been explicitly suspended for the purposes of the test, including the rule about turbine rotation (Medvedev, 1991). Interpretation matters in reasoning, i.e., viewed through the lens of logic from where do the premises come and what are the properties of the derivability relation chosen? To continue for a moment the dire-consequences theme, take the example of the disastrous atomic bombing of Hiroshima and Nagasaki in 1945. The USA, UK, and China issued a surrender demand, to which the Japanese premier at the time, Kantaro Suzuki, responded with the word *mokusatsu*, derived from the word for 'silence'. This utterance has at least two interpretations: one is something like 'no comment at the moment, but give us a while and we'll get back to you', the other, unfortunately broadcast by the news agencies, is 'not worthy of comment' (National Security Agency, 1968). An important lesson here is that the act of choosing the relevant premises to draw inferences from—and possibly seeking more information before committing oneself to serious consequences—is an important aspect of inference. Fortunately reasoning is not all about nuclear explosions.

There are two complementary approaches to the study of how people reason. A common choice is experimental intervention in an attempt to minimise individual differences. For instance discovering the instructions that enable the majority of children to perform counterfactual reasoning (Dias & Harris, 1988), improving mean sample performance in conditional reasoning (e.g., Cheng & Hollyoak, 1985; Moutier, Angeard, & Houdé, 2002), nudging people towards a classical interpretation of quantifiers (J. R. Schmidt & Thompson, 2008). This kind of research can be thought of addressing the question of how to increase the likelihood that an arbitrary person will understand the intended meaning of an utterance and what the experimenter wants in response.

Another approach is to investigate individual differences. Typically studies of individual difference involve exploring whether competence on one reasoning task (again according to a particular model) predicts competence on another. Stanovich and West (1998), in their article *Individual differences in rational thought*, sought correlations between measures of intelligence and reasoning performance on a range of tasks, as did Newstead, Handley, Harley, Wright, and Farrelly (2004) in their article *Individual differences in deductive reasoning*. One might be forgiven for believing that this is the only kind of possible investigation of individual difference.

A more recent trend in the study of reasoning is to explore qualitatively different responses beyond gradients of ability, e.g., what kinds of interpretation participants make on several tasks. This thesis explores these issues. To give a taste of what is to come, interesting questions which may be explored in this fashion include:

1. Is everyone trying to do the same thing, but some are succeeding better than others? Or are people adopting different goals?

2. Are the chosen competence models reasonable? Wason (1966, p. 146) briefly considered that an interpretation of the conditional as having three truth values might be a better model of what participants were using on his selection task, but immediately dismissed the idea.

3. How homogenous are people's inferences cross-task? For instance if a response in one task may be interpreted as reflecting sensitivity to issues of language pragmatics, is a similar kind of response drawn in other tasks?

4. What kinds of psychological processes relate to interpreting a task in a logic?

1.1 Outline of thesis

Participants were asked to solve a number of reasoning tasks, each of which was thought to have interpretable individual differences in terms, e.g., of social and communicative competence versus classical logical interpretation. To tap into individual differences in social and communicative ability, I investigated autistic-like traits in typically developing (TD) individuals. These traits are similar to but milder than the traits found in people with autism spectrum condition (ASC): impairment in social and communicative competence, restricted interests, repetitive behaviours, and also peaks in ability, especially in perceptual reasoning problems. It is important to stress that these traits may actually be positive traits to possess, especially in milder form. Following earlier work on discourse processing in ASC (see later), and given the nature of autistic traits, it was hypothesised that autistic traits would predict features of the inferences people draw.

Previous research on ASC and autistic-like traits (reviewed in Chapter 2, pp. 9 ff.) and on interpretation in reasoning (reviewed in Chapter 3, pp. 37 ff.) points towards relevant psychological phenomena and explanations thereof. We shall see we have several starting points for hypotheses to test relating autistic-traits and reasoning, and relating reasoning tasks themselves. However the theories are all still under-specified. As Stenning (2002, p. 143) writes:

> Testing hypotheses is fine so long as we have a well-developed body of theory and observation that give us criteria for evaluating which hypotheses are worth testing, how our tests might generalise, and how to operationalise the concepts that we need to express our hypothesis.
>
> . . . Our study is typical of where psychological investigations should begin. It is a fishing expedition. . . Of course like all explorers, we set out with a conceptual system—to have left it at home would have been both unhelpful and impossible, if only because, like all conceptual systems, some of ours was explicit, and some of it implicit.

The present thesis is an exploration, carried out in the fashion prescribed by Stenning.

Chapter 4 (pp. 73 ff.) introduces the tasks and our participants. Chapter 5 (pp. 87 ff.) investigates the psychometric properties of the self-report measures of autistic traits I used. Chapter 6 (pp. 119 ff.) investigates performance on Raven's Advanced Progressive Matrices, originally included in the data collection to estimate the general factor in intelligence, but later found to contain items which tap differentially into visual and verbal-analytic processes. Chapter 7 (pp. 139 ff.) investigates performance on the so-called suppression task, which people with ASC have previously been shown to interpret differently to TD individuals. Chapter 8 (pp. 165 ff.) investigates the 'immediate inference' task, a one-premise quantifier reasoning task, typically used to tap into language pragmatics. Chapter 9 (pp. 185 ff.) investigates the categorical syllogisms two-premise quantifier reasoning task which has previously been shown to involve a range of different factors, including executive function and language pragmatics. Chapter 10 (pp. 219 ff.) investigates cross-task correlates between the various tasks. Finally Chapter 11 (pp. 233

ff.) concludes the thesis with a general discussion of the findings and directions for the future.

Chapter 2

The Autism Spectrum Phenotype: Categories, Dimensions, and Theory

In later chapters we explore patterns of reasoning as a function of autistic traits in typically developing (TD) individuals. This chapter explains the autism spectrum phenotype, theories explaining differences between TD individuals and people with ASC, and how these may be used to gain a greater understanding of cognitive function in the general population, and vice versa.

2.1 The Clinical Phenotypes

Autism is a developmental condition characterised independently by Leo Kanner (1943), who described 'inborn autistic disturbances of affective contact', and Hans Asperger (1944/1991), who described an autistic psychopathy, later named Asperger syndrome by Wing (1981). The term 'autism', from Latin *autismos*, derived from the Greek αὐτός, meaning 'self', was originally used by Eugen Bleuler in the early 1900s to describe how people with schizophrenia lose contact with reality. The *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV-TR*; American Psychiatric Association, 2000) has separate definitions for conditions often referred to collectively as 'autism', including Autistic Disorder, which is closer to Kanner's definition, and Asperger's Disorder. Autistic Disorder is clinically defined by impairment in social interaction and communication, and restricted activities and interests. By definition, onset must be before 3 years. The main distinguishing feature of Asperger's Disorder is a lack of delay in early language development. A classification of High Functioning Autism (HFA), not officially in DSM, is often given if Full Scale IQ (from a Wechsler test) is at least 70 and the DSM criteria for Autism Disorder are met (see e.g., Ozonoff, South, & Miller, 2000, for an example of this usage). Appendix A (pp. 249 ff.) gives descriptions from *DSM-IV-TR* for Autistic Disorder and Asperger's Disorder. ICD-10 (World Health Organization, 2007) criteria are practically identical.

It is common to refer to any pervasive developmental disorder in DSM—Autistic Disorder, Rett's Disorder, Childhood Disintegrative Disorder, Asperger's Disorder, and Pervasive Developmental Disorder Not Otherwise Specified—as an Autism Spectrum Disorder (ASD; see Ozonoff, Goodlin-Jones, & Solomon, 2005) or Autism Spectrum Condition (ASC; e.g., Jolliffe & Baron-Cohen, 1999; Crespi & Badcock, 2008). In this thesis, I will use ASC to mean someone with a diagnosis of a pervasive developmental disorder, usually DSM Autistic Disorder or Asperger's Disorder.

Diagnosis often involves the use of structured diagnostic aids. The Autism Diagnostic Interview-Revised (ADI-R; Le Couteur, Lord, & Rutter, 2003) is a structured interview carried out by a trained interviewer with the main carer of the child. ADI-R contains over one hundred questions and can take up to three hours to complete. The Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2001) consists of several modules, chosen depending on the age of the child or adult to be diagnosed and level of linguistic ability. Each module consists of a series of activities involving social 'presses'—opportunities for the participant to display appropriate competencies—including observation during time which appears to be a break between sessions. Overall ADOS takes 30–45 minutes to complete.

To give an indication of the complexities involved in diagnosis, it's worth mentioning that a group chaired by Couteur (2003) recommended a multi-agency, multi-disciplinary assessment approach including: developmental and family history, e.g., using ADI-R; observations of the child to be diagnosed over several time periods and structured, e.g., using ADOS; a cognitive assessment; communication assessment; possible co-morbid mental-health problems assessed; needs and strengths of family members assessed; chromosome karyotype and fragile X DNA analysis; other assessments suggested include investigation of unusual sensory responses, motor planning, and difficulties with self-care.

Estimating prevalence is a thorny issue, often discussed with passion in the news (e.g., The Observer, 2007), and depends (unsurprisingly) on exact diagnostic criteria, methods used, and sample chosen. As a recent example, Baird et al. (2006) estimated prevalence in a sample of 57,000 children, aged 9–10, in South Thames, London UK. The authors searched for recorded diagnoses of ASC or statements of special educational needs to select participants. Of the original group, 1,700 met the criteria and were screened using a parent-report questionnaire; 1,035 returned the questionnaires and agreed to take part in the rest of the study. A two-way stratified random sample was used, resulting finally in 255 children receiving an in-depth assessment. Predicted prevalence in the general population ranged from a mean of 0.25% (lower 95% CI = 0.18%) for narrow childhood autism to 1.16% (upper 95% CI = 1.41%) including other ASC. The male to female ratio for all ASC together was around 3:1.

2.2 Cognitive Theories of Autism Spectrum Condition

This section outlines important cognitive theories of ASC. Importantly I want to allow that the term 'cognitive' includes processes supporting social and affective processes (see, e.g., LeDoux, 1995, p. 224); these are important in studies of autistic-like traits. Three main cognitive theories will be addressed here: impaired theory of mind, executive dysfunction, and weak central coherence, followed by a discussion of how they may overlap or depend on each other.

2.2.1 Impaired theory of mind

Premack and Woodruff (1978) defined theory of mind (ToM), as the ability to impute mental states (e.g., desires and beliefs) of oneself and others. Baron-Cohen, Leslie, and Frith (1985) performed the first experiment to investigate whether children with autism have intact ToM, using a paradigm developed by Wimmer and Perner (1983). Participants, children with ASC, Down's syndrome, and TD children, were shown two dolls, Sally and Anne. Sally placed a marble in her basket and then left. Unknown to Sally, Anne stole the marble and hid it in a box. Sally returned. The child was asked, ‘Where will Sally look for her marble?’ The experiment was repeated with a new hiding place, the experimenter’s pocket. There were three controls questions to check that the child could remember the names of the two dolls; knew where the marble really was; and that the child remembered where the marble was originally prior to Anne’s deception. All children in all three groups passed the control questions. Around 85% of both the TD and Down’s syndrome group passed the key ToM test, compared to only 20% of the ASC group. In passing, it’s worth highlighting that the authors argue (p. 42) that the questions were equivalent in terms of ‘psycholinguistic complexity’ but not ‘conceptual complexity’. There is a possibility that children with ASC actually interpret the crucial question differently.

Some research has examined mental state attribution in discourse. F. G. Happé (1994), for instance, tested people with ASC on a series of ‘strange stories’ which aimed to tap into ability to understand concepts such as lying, telling white lies, joking, pretence, figures of speech, sarcasm. An example of a ‘joke’ story:

> Today James is going to Claire’s house for the first time. He is going over for tea, and he is looking forward to seeing Claire’s dog, which she talks about all the time. James likes dogs very much. When James arrives at Claire’s house Claire runs to open the door, and her dog jumps up to greet James. Claire’s dog is huge, it’s almost as big as James! When James sees Claire’s huge dog he says, ‘Claire, you haven’t got a dog at all. You’ve got an elephant!’

Participants were then asked, if what, in this James, says is true, and why he would say this. Those with ASC tended to give fewer appropriate mental state explanations for why the characters said what they said, for instance nearly half of the ASC group (8 out of 18) made at least one error for a pretence story compared to 2 out of 26 TDs; half of the ASC group made an error on white lies versus four TDs.

Baron-Cohen (1997) develops a theory of a mindreading system which enables performance on ToM tasks. This consists of four components. In the theory, the intentionality detector (ID) interprets motion, e.g., approach and avoidance, in terms of underlying goals that may have initiated the movements. The Eye-Direction Detector (EDD) detects eye(-like object)s; infers their direction; and interprets gaze as seeing. The Shared-Attention Mechanism (SAM) represents relationships between an external agent, self, and a third object (which may also be an agent). An example given is that of recognising that you and I both see an object. Finally, the Theory-of-Mind Mechanism (ToMM) represents mental states, such as pretending, believing, imagining, and allows inferences to be drawn about them. The theory posits that the ID and EDD function correctly in people with ASC, but the SAM and ToMM are impaired.

2.2.2 Executive dysfunction

Gilbert and Burgess (2008) define executive functions as 'high-level cognitive processes that facilitate new ways of behaving, and optimise one's approach to unfamiliar circumstances.' Much work on executive function is derived from research on the effect of lesions to various parts of the cortex to cognitive function. Of particular importance was the work of Aleksandr Romanovich Luria (1962/1966), *Higher Cortical Functions in Man*, which carefully detailed the impact on function of lesions to the temporal, occipital, parietal, sensorimotor, and frontal cortex. Functions affected by damage to the frontal lobes in particular are often labelled executive functions and now include: action initiation, action monitoring, impulse control, inhibition, planning, set shifting, and working memory. It's clear that there is no singular, homogenous, executive function, but rather a set of heterogenous executive functions.

Executive functions appear in a range of theories. One of the most well known, originating from Norman and Shallice (1980) (see also Shallice, 1982), is a model of deliberate attention, inspired by production systems used, e.g., in symbolic artificial intelligence. The model consists of the following elements:

1. perceptual processes;
2. an effector system;
3. action or thought schemas, each of which controls an 'overlearned' action or skill;
4. a contention scheduler;
5. and the Supervisory Attention System (SAS).

Schemas are activated by perceptual processes, or (recursively) by other schemas, and many can run in parallel. Both the contention scheduler and SAS select the schemas to run, though the latter is involved only in the selection of ill-learned or novel schema sequences, or where planning is required. Baddeley (1986) incorporated this model into his model of working memory to specify the operations of the 'central executive' which is a central component.

A series of correlational studies have investigated the degree to which various tasks which purport to test executive functions relate to each other. Correlations between different tests of one construct would be expected as a necessary but not sufficient condition for the existence of that construct. Duncan, Johnson, Swales, and Freer (1997) discovered relatively low correlations between the various tasks conventionally used to test executive function, with R^2's ranging from .03 to .12. Larger correlations were found by jumping out to other supposedly non-executive tasks, for instance object naming and verbal learning ($R^2 = .30$). Miyake et al. (2000) investigated the relationships between three hypothesized executive functions: shifting back and forth between different tasks (e.g., between local and global processing), updating and monitoring working memory (e.g., remembering letters), and inhibition of prepotent responses (e.g., the Stroop task[1]), using a latent variable approach (see, e.g., Bollen, 2002; Borsboom, Mellenbergh, & Van Heerden, 2003). Correlations between the three operationalisations of each hypothesized function were low, with R^2's ranging from .02 to .10, so performance on one supposed test of a function is not particularly informative of performance on another, though correlations between tasks defined to measure a construct were higher than those cross-construct (e.g., between a shifting task and an updating task). The final latent variable model selected by the authors consisted of three latent variables, each pair of which was allowed to correlate. Correlations between the latent variable scores ranged from .4 to .6 and were statistically significantly different to 1, from which the authors argue that no 'pair of the three latent variable factors is in fact the same construct.'

Executive functions have been studied extensively in ASC. One of the most studied tests is the Wisconsin Card Sorting Test (WCST), which requires participants to sort cards according to a rule they must discover. Participants receive feedback (correct or incorrect) on their accuracy. As the test progresses, the rule changes without warning, for instance to begin with the correct rule for sorting cards might be according to their colours, then soon

[1]Stroop requires participants to read a series of words naming colours, e.g., 'RED', 'GREEN'. In congruent conditions, the colour of the ink is the same as the word; in incongruent conditions they are different. The task is always to read aloud the colour of the ink. Participants typically show slower reaction times for the incongruent condition.

after according to shape.[2] Impairment on the task in people with ASC has been found in a number of studies (e.g., Rumsey, 1985; Ozonoff, Pennington, & Rogers, 1991; B. R. Lopez, Lincoln, Ozonoff, & Lai, 2005). B. R. Lopez et al. (2005) found that controlling for verbal ability removed the effects, though they argue that this is a result of insufficient statistical power. Hughes, Russell, and Robbins (1994) investigated in more detail the processes required to solve WCST-like tasks and discovered that performance on extra-dimensional shift (transfer of learning to new exemplars and shifting attention to different features) was particularly impaired. Planning, measured using the Tower of London (e.g., Hughes et al., 1994) or Tower of Hanoi (e.g., Ozonoff, Pennington, & Rogers, 1991) tasks, both of which involve moving discs between pegs whilst following rules, is also often found to be impaired in those with ASC versus TD individuals.

The discussion by Hughes et al. (1994) based on the set-shifting and planning deficits interpreted the results with respect to Norman and Shallice's theory, outlined above. They argue that the results would be found with an impairment to ('functional absence of') the SAS, requiring the contention scheduler to step in, triggering schemas from the environment. They use this to explain deficits in ASC such as distractibility and adherence to routine. This might sound paradoxical at first, but the idea is that one becomes more dependent on one's environment: how that changes has more of an impact on cognition in the absence of the SAS.

A series of studies have shown that although lesions to the rostral prefrontal cortex (rPFC) do not affect performance on, e.g., WCST, Tower of London, working memory tasks, or Stroop, they do affect multitasking and performance on ill-structured tasks whose processes of solution is not cued by the environment (P. W. Burgess, Simons, Dumontheil, & Gilbert, 2005). Extensive evidence of ability and impairment from the lesion studies, and functional imaging studies of neurologically intact individuals, led Burgess and colleagues to formulate the gateway hypothesis of rPFC function (recently exposited in P. W. Burgess, Gilbert, & Dumontheil, 2007; P. W. Burgess, Dumontheil, & Gilbert, 2007). This posits the following basic distinction between cognitive processes: *Stimulus-oriented (SO)*, relating to current sensory input and requiring well-learned processes to respond; and *Stimulus-independent (SI)*, e.g., task irrelevant processes including 'mind-

[2] See Dehaene and Changeux (1991) for a detailed analysis of performance on WCST.

wandering' and goal-directed behaviour where all the information required to proceed is not available in the stimulus. At any time, the authors argue, individuals typically engage in a combination of these types of processes, but certain tasks require more of one than the other. The rPFC acts as a gateway between these two different types of process.

Hill and Bird (2006) argue that in their tests of executive function in ASC, the greatest difficulty appeared to come from engaging and disengaging actions to achieve an overall goal. Following these authors, Gilbert, Bird, Brindley, Frith, and Burgess (2008) tested their model of SI/SO switching in ASC with functional magnetic resonance imaging (fMRI). The task they used consists of two conditions cued by different colours. In the SO condition, participants were shown uppercase letters, one at a time, in alphabetical order, and had to classify them according to whether they consisted only of straight lines versus containing any curves (Arial typeface was used). In the SI condition, the letters were displayed in random order and participants had to do the classification continuing for themselves the alphabetical order beginning on the first letter displayed. There was a (non-statistically significant) trend for those with ASC to be slower to respond than TD individuals on all trials. Comparing the blood-oxygen-level dependent (BOLD) signal (indicative of level of metabolic activity; see Heeger & Ress, 2002) during the SO condition with the SI condition, the ASC group showed more activation than the TD group in medial rPFC (and other areas including amygdala and cerebellum), and TD individuals showed more activation in bilateral occipital cortex than the ASC group. Why the ASC group shows more activation is not made clear, but the authors argue is somehow indicative of dysfunction. The occipital cortex activation increase in TD individuals might suggest, the authors argue, that they 'were able to modulate activity in early visual cortex according to the attentional demands of the task to a greater degree than the ASD group.'

2.2.3 Weak central coherence (or enhanced perceptual functioning?)

One of the well replicated observations in people with ASC is that they do well on tasks like Block Design (Figure 2.1) and embedded figures type tests (Figure 2.2), relative to other tasks such as metaphor comprehension. (See Chapter 6 for more detail.) Both tasks require the segmentation of a complex figure, ignoring global features. For block design the segmentation area is imposed by the dimensions and shapes on the individual blocks;

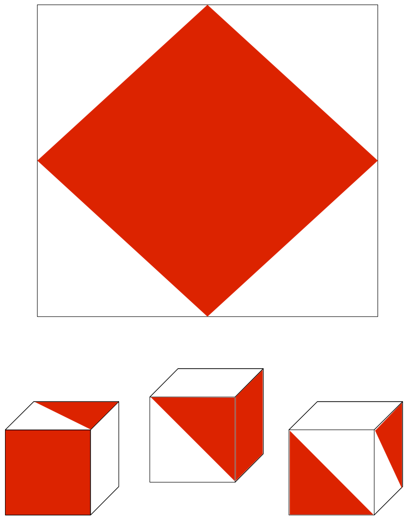

Figure 2.1: A sketch to convey the basic idea of the block design. Participants are provided with a number of blocks and are asked to arranged them to reconstruct a complex figure.

for embedded figures an external target shape must be used to try to extract a matching feature from a more complex collection of lines. Performance on embedded figures tests is often associated with the field dependent and independent (FD/FI) cognitive styles (Witkin, 1950; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962). Those who are FD perform poorer on embedded figures than those who are FI.

An introduction of the idea of cognitive styles may be useful as the term has been used in the autism literature (e.g., F. Happé & Frith, 2006). Cognitive styles are often conceptualised as bipolar dimensions, position at either end of which indicates a strong tendency to process information in a particular way but where there is no normative polarity. More of a cognitive ability is always preferred whereas a style's adaptive value depends on the situation (see e.g., Messick, 1976). This is quite a controversial concept. Kozhevnikov (2007) notes that many researchers do believe in the existence of styles but that styles' effects are 'often overwhelmed by other factors'. Can cognitive styles just be viewed as specific abilities once variation in 'overall ability', e.g., the general factor in

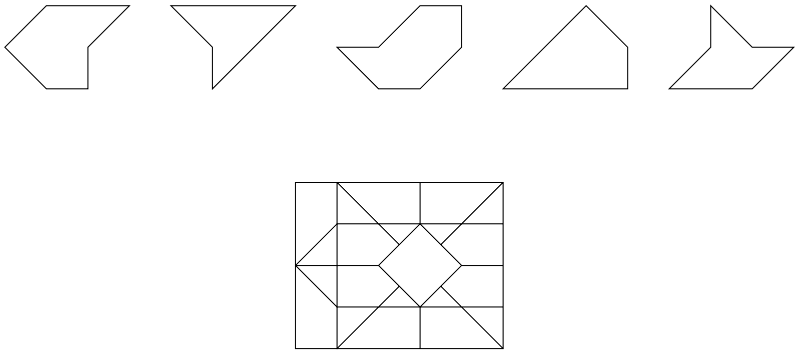

Figure 2.2: Example of a hidden figures test item. The task is to select which outline figure from the top row fits into the more complex figure below.

intelligence (g; Spearman, 1904), has been partitioned out? What is one to make of the correlations between measures of FI and tests which load on g (e.g., McKenna, 1984)? How do styles relate to personality? These are questions to be addressed elsewhere. The interested reader may enjoy Zhang and Sternberg (2006) for a recent defense and model of what they call intellectual styles.

U. Frith and Snowling (1983) investigated what has been argued to be a related contextual effect using homographs: words which are spelt the same but have different meanings depending on context. They chose homographs which had different pronunciations depending on the meaning, for instance 'he had a pink *bow*' compared to 'he made a deep *bow*', thus making it possible test the effect of context in a reading-aloud test. The study consisted of three groups: ASC, dyslexic, and TD. All three groups differed in accuracy of pronunciation (with respect to the context), with the ASC group showing the worst performance, followed by the TD group and then—interestingly!—the dyslexic group score was highest.

U. Frith (1989) produced the concept of central coherence, drawing on field (in)dependence research, to explain ASC superiority on block design and embedded figures. She writes (p. 96):

> One way of describing field independence is the ability to disregard context. Field independent individuals are also good at Block Design. Moreover, this

> ability goes together with social field independence! From this point of view it was perhaps not surprising that our autistic children, who are supremely socially independent, indeed aloof, did extremely well on this test.

It may not be obvious that block design and embedded figures require similar processes, though the suggestion that they do was not first made by U. Frith (1989). Witkin (1950) argued, for instance, that the Kohs Test, a predecessor of Block Design, lay somewhere between embedded figures and tests typically used to test IQ. He also reported correlational evidence of overlap between the tasks ($r \approx .5$, discovered by Thurstone in the early 1940's), arguing for the importance of perceptual factors in the Kohs Test, further noting (p. 14) that 'perceptual tendencies shown by people in the embedded-figures test are important in determining how easy or difficult they find the Kohs Test.'

U. Frith (1989) argues that TD individuals combine large amounts of information when they try to make sense of the world. They have a strong tendency to take context into account. This ability, she argued, is impaired in people with an ASC, whose 'high-level central cohesive force' is weak. Their 'more peripheral input and output processes' are intact, however. Rephrasing, she argues that people with ASC have the ability not to take context into consideration, thus helping to explain their success on tasks like embedded figures. Typically developing children have to be taught this ability. Clearly the notion of central coherence is not here characterised in any detail.

Is weak central coherence identical to field independence? Frith writes (p. 98):

> A weak central cohesive force, weak relative to lower level cohesive forces, would *simulate* 'field independence' and all that it entails for performance on embedded figures. It would entail thought detachment and social detachment, but this would not be the same as detachment in an older normal child. In the normal child detachment is the sophisticated end-product of education, a sign of control over the high-level central force towards cohesion. In the case of Autism I propose that such control is lacking and that this results in an incoherent world of fragmented experience. [Emphasis mine.]

In a more recent description of WCC (F. Happé & Frith, 2006, p. 16) the distinction is summarised as follows:

> [. . .] while FI people are conceptualised as succeeding on EFT because of their ability to see, but resist, the gestalt, people with weak coherence are postulated to be good at this test precisely because they do not spontaneously attend to the gestalt, instead seeing the figure first in terms of its parts.

Recently F. Happé and Booth (2008) argue that WCC may be separable into two components: integrative weakness and local enhancement. They leave as an open question how best to investigate this. In contrast to the weak central coherence theory, the enhanced perceptual function (EPF; the most recent version of which was exposited by Mottron, Dawson, Soulières, Hubert, & Burack, 2006) theory posits that perceptual advantages in ASC are due to the possession of more developed low-level perceptual processes. They argue that people with ASC are biased towards local processing of information, as opposed to the hierarchical nature of processing more common in TD individuals. Tasks which require simple processing (as defined by the hierarchical organisation of neural processing) are performed better by those with ASC. They argue

> ...the characteristics that differentiate autistic from non-autistic perception plausibly correspond to an overall superior functioning, involvement, and autonomy of posterior regions of the perceptual visual cortex (for the hierarchical, antero-posterior axis) and of the central part of this cortex (for the dorsoventral, specialization axis).

Interestingly Mottron et al. (2006) argue that atypical behaviour shown by people with ASC when perceiving objects, for instance staring at an object with pupils in the corner of the eyes, is due to a strategy for regulating perceptual input.

2.2.4 Cross-border cooperation

This section addresses the question of to what extent the theories mentioned above may be integrated with each other. For instance, are two supposedly different theories describing the same mechanisms using only different language? This may be the case if the same phenomena are accounted for by two theories and if the theories are indistinguishable empirically. Is there evidence of correlation between operationalisations of supposedly different theories? The latter would be insufficient to prove a common causal mechanism, but is helpful as a contributing step towards integration.

One of the earliest attempts at integration was made between theory of mind and executive function. Ozonoff, Pennington, and Rogers (1991) discuss a series of possibilities for explaining the co-occurrence of impairments on tests of ToM and executive functions: one deficit may be primary and cause the other; correlations may be due to neurological proximity; some third variable could be causing impairments both to ToM and executive function tasks; or the deficits could be independent. For instance a deficit in inhibition might cause impairment on the ToM task, considering that the planning and information update aspects of the task might be involved in both executive function and the theory of mind task impairment. Hughes et al. (1994) argue that two alternatives are plausible: the executive function deficits seen in ASC underlie the difficulty in acquiring a ToM or the frontal lobes cause both executive and theory of mind problems.

Ozonoff, Rogers, and Pennington (1991) (reanalysing data reported by Ozonoff, Pennington, & Rogers, 1991) note a difference in performance by people from two sub-groups of ASC. People with High Functioning Autism (HFA) and Asperger Syndrome (AS) are impaired with respect to TD people on Tower of Hanoi, WCST, and emotion perception whereas only those with HFA were impaired on theory of mind. Also those with AS tended to do better than those with HFA on a verbal memory task. Interestingly there was a trend for the AS group to do better on an executive function composite score than the HFA group, perhaps evidence in favour of an executive account of theory of mind?

Zelazo, Jacques, Burack, and Frye (2002) performed a study using a range of tasks, seeking to investigate if there were correlations with ToM. A WCST-type task correlated strongly with ToM in mildly impaired (verbal IQ > 40) participants ($r \approx .8$). They also used a ramp task with two input and two output holes. Marbles were rolled down the ramps; where they appeared, in a hole parallel to or diagonal to the input hole, depended on whether a light was switched on. There were two configurations, told to the participants by conditional rules: 'when the light is on, the marbles roll across' or 'when the light is on, the marbles roll straight' combined with a diagrammatic depiction. Input holes were randomised and participants had to respond with from which hole the marbles would appear. Performance on this task also correlated with ToM ($r \approx .7$). The WCST-esq task and this ramp task also correlated ($r \approx .5$). There were no correlations between ToM and any other tasks in the more severely impaired participants.

Theory of mind has also been related correlationally to tests of weak central coherence. Burnette et al. (2005) found a correlation between a homographs test, taken to test WCC, and a ToM task ($r \approx .6$) in a sample of children aged around 11 with HFA, and this remained marginally significant when partialling out Performance IQ and Verbal IQ (both dropping to $r \approx .4$). There was no correlation found between ToM and an averaged composite of block design, embedded Figures, and a pattern construction task. Jarrold, Butler, Cottington, and Jimenez (2000) found ToM performance in children with ASC was related to embedded figures only when verbal IQ was partialled out ($r \approx .6$).

Booth, Charlton, Hughes, and Happé (2003) argue based on an empirical result that WCC is independent of executive function. However they examined only planning, and tested both WCC and planning using a drawing task. Given a heterogenous view of executive function, many possibilities for connections remain. The distinction of top-down versus bottom-up processing seems highly relevant, and is more general in cognition. What exactly is the distinction? Bottom-up processes tend to seen as stimulus driven, especially by more basic features of the stimulus, whereas top-down processes are seen as directed more by goals and knowledge (e.g., Hahn, Ross, & Stein, 2006). Integrative processes, required for central coherence, are often characterised as top-down, and are thought to depend more on feedback connections in the brain (C. Frith, 2004). The distinction between stimulus-oriented (SO) versus stimulus-independent (SI) thought described above is a related concept. Those with ASC are perhaps more biased towards SO thought.

Stenning and van Lambalgen (2007) use a logical approach to specifying executive function, focusing on goal maintenance, planning, and inhibition, and show how this can be used to model a ToM task. Their central argument is that if one wants a logical characterisation of planning, classical logic is not the way to do it because of the frame problem: the impossibility of taking into consideration all eventualities. Instead they propose the use of logic with closed world reasoning in which every proposition for which one has no reason to believe in its truth is assumed to be false. Central to this is to have general rules which may be applied assuming no exceptions; by default, and given no evidence to the contrary, it is always assumed there are no exceptions.[3] Their model gives them a framework for characterising aspects of performance on a false belief task: the Maxi task where a child and a doll (Maxi) see where chocolate has been hidden in a box, but then known to the child but unknown to Maxi who has temporarily left the room, the chocolate is moved to the drawer. The child has to say where Maxi will look when she returns.

The analysis exposes how a child attempting the task has to keep in mind the goal to say

[3]This description is a preview of more detailed explanations to come in Chapter 7.

where Maxi will look, not where she should look if she wants to acquire the chocolate. The prepotent response, barring exceptions, would be to report where the chocolate actually is, so this must be inhibited. There are also processes related to drawing inferences about beliefs from the perception of, e.g., what Maxi did and did not see (this aspect seems similar to the theory by Baron-Cohen (1997) discussed above).

2.3 Reasoning in Autism Spectrum Condition

The overall aim of this chapter is to guide the reader towards research on the broader phenotype, including that in TD individuals, and to argue that investigating links between autistic traits and reasoning in TD individuals is likely to be successful. Let's pause for a moment and consider the research that has been done on reasoning in ASC and how to date unexplored areas of reasoning may explain and be informed by research on ASC. Already in this chapter we have seen hints of reasoning processes as typically defined: for instance theory of mind as drawing inferences about mental states, planning (often considered part of reasoning; see Rips & Conrad, 1989) in executive functions, perceptual processing (shown to be important, e.g., by Prado, Kaliuzhna, Cheylus, & Noveck, 2008, for reasoning about conditionals). Gilhooly, Logie, Wetherick, and Wynn (1993), using the working memory theoretical framework devised by Baddeley (1986), investigated reasoning in syllogistic inference (e.g., from *all A are B* and *all B are C*, what follows?). They found that a concurrent task of random number generation, hypothesised to require parts of the central executive, tended to force participants towards guessing on syllogistic inference. Given the performance by people with ASC on tests of various executive functions, one might expect differences in reasoning performance here too. The study of deductive reasoning has been relatively neglected in the ASC literature. Narrow conceptualisations of deductive reasoning have also tended to isolate the study of reasoning from these areas just encountered.

Scott and Baron-Cohen (1996) were the first I'm aware of to test reasoning in non-ToM tasks in ASC. They tested transitive and analogical reasoning in children aged around 12. ASC and TD children performed similarly, though it does appear there is a trend to a difference on analogical reasoning: all 17 of the control group passed at least 3 trials, compared to 14 of the ASC group; 14 children in the control group passed at least 5 trials, compared to 10 of the ASC group. These differences were not statistically significant, however. Scott, Baron-Cohen, and Leslie (1999) investigated counterfactual reasoning in children with ASC, children with learning disabilities, and TD children. For instance children were told that the experimenter had a story where all cows go 'Quack' and Freda is a cow; the task was to decide if Freda says 'Quack'. Interestingly, children with ASC performed better at the tasks than TD children and children with learning disabilities. However if the children played a set of pretense games before the task, the ASC group's performance was impaired, but the pretence improved the other groups' mean scores. This suggests that both groups chose, and required, different strategies to solve the problems.

Leevers and Harris (2000) note that this experiment had a possible confound: the correct response was always 'yes'. They attempted to replicate in children with ASC and TD controls who had a mean verbal mental age of around 5 (mean chronological age in the TD group was around 4 versus 11 in the ASC group). Children with ASC tended to give fewer 'no' responses and were more likely to restate premises to justify their inferences. Also children with ASC tended to be less consistent in their choice of classical logic or empirical responses.

Noveck, Guelminger, Georgieff, and Labruyere (2007) investigated how TD adults, adults with ASC and TD children (mean age around 5) interpreted sentences of the form 'Every horse did not jump over the fence'. Such sentences have two common interpretations: one where not every horse jumped and the other where no horses jumped. Stories were presented to the children about three people or things, with accompanying pictures. The task was to decide if the presented sentence was true or false of the picture. For one of the conditions, both the 'not every' and the 'no' interpretations were false (for instance every horse did jump over the fence in the picture). For the second condition, for instance where two out of three horses jumped, the response could distinguish the interpretation: false would imply the 'no' interpretation and true would imply the 'not every' interpretation. In TD adults they found a preference (mean $\approx 90\%$) for the 'not every' interpretation. In the ASC and TD children groups no strong preference was shown (40% and 45%, respectively). The authors argue that adults use pragmatic processes which are unavailable to both children and people with ASC.

Pijnacker et al. (In press) found differences in the ability of adults to deal with exceptions in conditional reasoning. When reasoning about conditionals like 'If she has an essay to write, she'll be in the library', people with ASC were less likely retract inferences when cued about possible exceptions, e.g., that the library may be closed, than were typically developing individuals. They were however sensitive to alternatives for a conclusion being true. This result will be discussed in further detail in Chapter 7.

2.4 How Broad and Continuous is the Autism Spectrum?

2.4.1 Distinctions within clinical diagnoses

There are two approaches to understanding clinical conditions: the categorical or diagnostic approach, which aims to divide people into distinct types of healthy and disordered; and the dimensional approach, which characterises disorder as one end of a spectrum merging with the general population (see, e.g., Eysenck & Eysenck, 1976; Beglinger & Smith, 2001, for discussion). The two approaches are not incompatible. A dimensional system may easily be translated into a categorical system by choosing appropriate thresholds, e.g., such that position below the threshold indicates no requirement for psychiatric care. Eysenck and Eysenck (1976, p. 1) argue that psychiatrists 'seem to adhere to the categorical system, but they also often talk about intermediate types [. . .] as if recognising the existence of continuity.' There are hints of this attitude in a recent handbook of psychiatry (Semple, Smyth, Burns, Darjee, & McIntosh, 2005, p. 6):

> There is a non-absolute, value judgement involved in the diagnosis of mental disorder—e.g. wheeze and dyspnoea are abnormal and a sign of disease, but some degree of anxiety at times is a common experience and the point at which it is pathological is debatable.

Do the differences between categorical and dimensional systems say anything deeper than this? Baron-Cohen, Wheelwright, Skinner, Martin, and Clubley (2001) argue that 'There is [. . .] an assumption, still under debate, that autism and AS [Asperger syndrome] lie on a continuum of social-communication disability, with AS as the bridge between autism and normality.' Wing (1981), for instance, noted that the defining features of Asperger syndrome, e.g., impairment in social interaction and the reading of non-verbal cues, pursuit of special interests, are found to varying degrees in the general population, though she adds, 'Even though Asperger syndrome does appear to merge into the normal continuum, there are many cases in whom the problems are so marked that the suggestion of a distinct pathology seems a more plausible explanation than a variant of normality.' It's not clear how 'distinct pathology' would be inconsistent with an extreme position on a continuous dimension. In later work she argues (Wing, 1988) that ASC is best viewed as a continuum, though that this 'represents a concept of considerable complexity, rather than simply a straight line from severe to mild' (p. 92). She suggests categorising ASC based on level of intelligence and social and communicative ability, as these are the best predictors of the support required (note how the use to which the categorisation is put informs its structure), or, as a compromise, having a category for autism as defined by Kanner (1943), another category for the remaining developmental conditions, and then dimension both according to intelligence.

Miller and Ozonoff (2000) give evidence that *DSM-IV-TR* Asperger's Disorder is higher-IQ (recall, this is usually Full Scale IQ assessed using a Weschler test, which is not necessarily g; see Colom, Abad, García, & Juan-Espinosa, 2002) HFA. Ozonoff, Rogers, and Pennington (1991) argue that there is another distinction between (again *DSM-IV-TR*) Asperger Disorder and HFA: the HFA group scored best on a spatial task, which was one of tasks with worst performance in the Asperger Disorder group; and the Asperger Disorder scored best on first-order theory of mind, one of the worst in the HFA group, though the authors consider that the differences may be due to differences in language ability which defines the diagnostic categories.

A series of cluster analysis studies have been performed to examine co-occurrence of traits in ASC along severity continuums. For instance Spiker, Lotspeich, Dimiceli, Myers, and Risch (2002) performed a cluster analysis of a series of measures on children who had received a diagnosis of an ASC. The variables were subscores from ADI-R, whether the child had functional three-word phrase speech at time of test administration, and non-verbal IQ (obtained depending on the available data, e.g., from a Wechsler test performance IQ or Stanford-Binet scale). Taking two to five clusters always showed the same pattern of monotonically increasing impairment for the variables, except for three out of ten of the ADI-R variables. Taking the three-cluster solution, cluster membership mostly explained non-verbal IQ ($R^2 = .80$) and verbal status ($R^2 = .20$). For the ADI-R variables, cluster membership mostly explained social reciprocity ($R^2 = .10$) and peer relationships

($R^2 = .08$).

2.4.2 The broader phenotype

Kanner (1943, p. 250) observed that among the parents of the children with ASC he studied,

> ...there are very few really warmhearted fathers and mothers. For the most part, the parents, grandparents, and collaterals are persons strongly preoccupied with abstractions of a scientific, literary, or artistic nature, and limited in genuine interest in people.

Kanner wondered to what extent the familial environment was a causal factor in ASC, but noted that

> The children's aloneness from the beginning of life makes it's difficult to attribute the whole picture exclusively to the type of early parental relations...

He appeared to favour a genetic explanation, describing the subjects of his case studies as 'examples of inborn autistic disturbances of affective contact.' Empirical evidence of the clinical phenotype eventually emerged to support this view. The first twin study of ASC by Folstein and Rutter (1977) investigated concordance in 11 monozygotic (MZ) and 10 dizygotic (DZ) twins, in which at least one of the pair had a diagnosis of autism. They found that 36% of the MZ pairs showed concordance for autism compared to 0% of the DZ pairs. Relaxing the diagnostic criteria, 82% of the MZ showed concordance for some form of cognitive disorder (including autism), compared to 10% of the DZ pairs. A series of studies have replicated this finding, e.g., Bailey et al. (1995) who found 60% concordance for MZ twins versus 0% for DZ, and 92% MZ concordance for a broader spectrum of traits versus 10% of DZ pairs. Note especially the high MZ concordance for autistic-like traits.

The notion of traits in people without a diagnosis of ASC which are qualitatively similar to but milder than the traits seen in ASC has received attention since the 1990s. This has come to be referred to as the broader autism phenotype (BAP; see e.g., Piven, 2001). For instance Piven, Palmer, Jacobi, Childress, and Arndt (1997) found that, compared to children with Down syndrome, the parents, grandparents, aunts and uncles of children with ASC tended to have more social impairment and show more stereotyped behaviour. Aunts and uncles of children with ASC were also more likely to show communication deficits, than were aunts and uncles of children with Down syndrome.

Since Gottesman and Shields (1973) introduced to psychiatry the distinction by John and Lewis (1966) of exophenotypes (the external phenotype) versus endophenotypes (closer, in some sense, to the genetic mechanisms), there has been a slow trend to move away from the diagnostic criteria. Flint and Munafò (2007, p. 165) give examples labelled endophenotypes used to study a range of conditions (e.g., ADHD, schizophrenia, and autism) including patterns of eye movements, olfactory sensitivity, performance on tests of executive functions, and age at first word. The hope is that the genetic mechanisms of such endophenotypes would be easier to unravel than more complex and vague clinical phenotypes. To be considered an endophenotype, the measure must be shown to be heritable, present in people with the condition under study, but also present in the absence of a clinical diagnosis.

Work is beginning to appear on cognitive endophenotypes in ASC. For executive function, Hughes, Leboyer, and Bouvard (1997) tested parents of children with ASC, parents of learning disabled children, and adults from unaffected families on an attentional flexibility task similar to the Wisconsin Card Sorting Task; a planning task similar to the Towers of Hanoi; and tests of spatial memory. The flexibility task required categorisation of shapes with feedback. After answering correctly six times out of 50 trials, the task changed, otherwise the task ended. At the 'extradimensional shift' stage (recall Hughes et al., 1994, outlined above), where participants had to attend to a new dimension, only half of the parents in the ASC group successfully completed the task, as compared to nearly all parents in the learning disability group and controls. For the planning task, problems requiring four or five moves distinguished the groups. The autism-parents group provided fewer perfect solutions than the controls (50% versus 65%). The ASC parents group also made more errors on a combined search and spatial working memory task than controls (33 versus 17). There was no differences between ASC parents and controls on spatial sequence span.

Wong, Maybery, Bishop, Maley, and Hallmayer (2006) tested parents and siblings of children with ASC on planning (Towers of London); set-shifting (same task as explained before); a test of inhibition created by the authors; and a generativity task: five meaningless line drawings to which participants were asked, 'What could this be?', and had to respond with as many answers as possible. Parents of children with ASC showed poorer performance than controls on generativity (mean of 31 versus 36 responses), as did the siblings of children with ASC versus controls (33 versus 37). Fathers of children with ASC showed problems with set-shifting to a dimension which was previously irrelevant. The authors found no significant differences between groups on planning nor on inhibition.

G. L. Schmidt et al. (2008) examined non-word repetition in parents of children with ASC, as part of a battery of other tasks. The numbers of syllables in the words increases from one to seven, stopping when there were three consecutive errors in repetition. ASC parents performed worse than controls on this task on two (around 60% correct versus around 95%) and three syllable words (around 90% correct versus around 70%; no other lengths were analysed as too few participants progressed further than three).

2.4.3 The broadest phenotype

Previously the broader autism phenotype has been used to refer specifically to traits of the relatives of people with a diagnosis of ASC. Baron-Cohen, Wheelwright, Skinner, et al. (2001) derived the first self-report measure of the broader phenotype, the Autism Spectrum Quotient (AQ), by comparing clinical groups with controls and refining items and threshold scores until the groups could be distinguished. They showed that the traits appeared as a continuous dimension in TD individuals—importantly not only in the relatives of those with ASC, and as yet there is no evidence that those with high AQ will go on to have children with ASC. Science students showed more autistic traits than humanities students, fuelling the speculation (see James, 2003) that autistic traits may be associated with abilities required for science and also (implicitly) that the absence of autistic traits may be associated with the skills required for humanities subjects. Interestingly they observed that participants in the control group who had a score indicating it was likely they had an ASC reported that within the university, the traits they possessed were not considered odd and the skills they had were often valued. There is evidence that AQ may be used to screen for Asperger syndrome and High Functioning Autism (Woodbury-Smith, Robinson, Wheelwright, & Baron-Cohen, 2005), for instance using a threshold score of 26 (out of 50) gave $\sim$95% sensitivity, 50% specificity, and correctly classified $\sim$80%.

The broader phenotype maps partially across to the five factor model of personality, a much studied descriptive theory of personality consisting of dimensions of openness, conscientiousness, extroversion, agreeableness, and neuroticism (see e.g., McCrae & John, 1992). Austin (2005) examined the personality correlates of AQ and found a positive correlation with neuroticism ($r = .2$), negative with extraversion ($r = -.4$), and negative with agreeableness ($r = -.4$). In a similar study performed in Japan using a Japanese translation of AQ, AQ-J (Wakabayashi, Tojo, Baron-Cohen, & Wheelwright, 2004) comparable correlations were found between AQ, and neuroticism and extraversion, however no association was found with agreeableness. Also there was an additional negative correlation with conscientiousness ($r \approx -.3$).

Kunihira, Senju, Dairoku, Wakabayashi, and Hasegawa (2006) investigated correlations between AQ and a range of self report measures in a Japanese sample. From the Temperament and Character Inventory (TCI), AQ was negatively correlated with novelty seeking and reward dependence (both r's $\approx -.3$) and positively correlated with harm avoidance ($r \approx .6$); no correlation was found with the persistence subscale. AQ was also positively correlated with the Self-Rating Depression Scale ($r \approx .4$), the State-Trait Anxiety Inventory ($r \approx .5$); and with experience of being bullied ($r \approx .2$) and, to a lesser extent, having bullied others ($r \approx .1$).

Ronald et al. (2006) investigated the heritability of autistic traits in a general population in children aged around 8 years, using the parent-completed Childhood Asperger Syndrome Test (CAST). They used a twin-study, with ~3400 pairs, MZ, DZ (both male and female, ~550–650 in each group) and ~1000 opposite sex DZ. The best model indicated broad heritability of 86% for males (78% from additive genetic influences and 8% from nonadditive genetic influences) and only additive factors of 81% for females. There was no significant effect of the shared environment and nonshared environment effects were small: 14% for males and 19% for females. They found moderate correlations between the three triads of impairment, as measured by separating CAST into three subscales according to DSM-IV criteria: $r \approx .3$ between social impairments (SIs) and communicative impairments (CIs), ~.2 between SIs and restricted, repetitive behaviours and interests (RRBIs), and ~.4 between CIs and RRBIs. There was no correlation found between the genetic influences of SIs and RRBIs, a modest correlation between SIs and CIs ($r_g \approx .4$ for males and .3 for females), and high between CIs and RRBIs ($r_g \approx .9$ for males and females). This is evidence for heterogeneity in the genetics of autism.

The traits measured by AQ have also been shown to be heritable. Hoekstra, Bartels, Verweij, and Boomsma (2007) examined heritability in a twin family sample from the Netherlands. The sample consisted of approximately equal numbers of MZ and DZ male, female, and mixed sex twin pairs, all of whom were aged around 18 years ($n = 370$), and their siblings ($n = 94$). The authors fitted a series of structural equation models, the most parsimonious of which indicated that additive genetic influences explained 57% of the variance and nonshared environmental effects explained 43% of the variance in AQ. The shared environment had no significant effect in the model and so was not included. Neither was there an effect of sex. Bishop et al. (2004) found that the parents of children with ASCs have higher scores in two of the five subscales of AQ, social skills (2.2 vs. 1.8 for mothers and 3.8 vs. 2.3 for fathers) and communication (2.1 vs. 1.8 for mothers and 3.4 vs. 2.4 for fathers), than do the parents of TD children. (These subscores were found to correlate, $r \approx .66$.) No differences were found between the groups for measures of attention switching, imagination, or attention to detail.

One of the important defining features of autistic-traits is impaired social functioning. Jobe and White (2007) examined variation in friendships and romantic relationships as a function of AQ. They found that those with a higher AQ reported more loneliness ($r \approx .5$; measured using the UCLA loneliness scale). AQ was negatively correlated with the duration of best friendship ($r \approx -.2$), but interestingly was positively correlated with duration of romantic relationships ($r = .34$). The authors interpret this latter result as reflecting preference for sameness, however why the result is not the same for friendship is not made clear.

Two other self-report questionnaires commonly associated with autistic-traits are the systemizing quotient (SQ; Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003) and empathising quotient (EQ; Baron-Cohen & Wheelwright, 2004) scales. Wheelwright et al. (2006) found AQ and EQ were moderately negatively correlated ($r = -.5$) and AQ and SQ-R (a revised version) were weakly positively correlated ($r \approx .3$). They also regressed AQ on the SQ-R and EQ, finding the linear relationship $AQ = 0.089 SQ\text{-}R - 0.25 EQ + \alpha$ where $\alpha = 21.6$ for males and $\alpha = 22.7$ for females.[4] Comparing those with and without ASC, the ASC group scored higher for SQ-R (mean of 77 vs. 56) and the TD

[4] The R^2 is not reported.

group scored higher for EQ (44 vs. 19). Wakabayashi et al. (2007) also compared SQ-R and EQ in people with ASC and TD controls in a Japanese sample. People with ASC scored lower on EQ than TD individuals (mean of 25 vs. 34), and higher on SQ-R than TD individuals (mean of 32 vs. 24).

Nettle (2007) studied the relationship with the five factor model. Regressing EQ on the five factors explained around 60% of the variance; only the estimates for agreeableness (standardised $\beta \approx .70$) and extraversion (standardised $\beta \approx .10$) were significantly different to zero. The relationships found with SQ were less striking, explaining around a quarter of the variance. The largest contributors were openness (standardised $\beta \approx .35$) and conscientiousness (standardised $\beta \approx .30$).

In the first study relating AQ and cognitive function (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), participants were given the Reading the Mind in the Eyes task, which consists of pictures of the eyes of faces displaying various emotions. The task is to identify the emotion expressed from a list of four adjectives. There were three groups tested: those with an ASC, IQ matched controls, and students. AQ correlated negatively with performance at the task ($r \approx -.5$ for all groups combined). There was also a negative correlation in all subgroups, including $r \approx -.3$ in the TD group (Sally Wheelwright, personal communication, February 25, 2008). For students, social skills and communication subscores were also negatively correlated (both $r \approx -.25$) with performance at the task.

The correlation between performance in this emotion recognition task and AQ has been replicated in another study (Lombardo, Barnes, Wheelwright, & Baron-Cohen, 2007, though these data are not reported in the text). Across all participants the correlation was $r \approx -.5$; in the TD group $r \approx -.5$ and in the ASC group $r \approx -.4$ (Michael V. Lombardo, personal communication, February 23, 2008). This has been replicated in an Austrian sample ($r \approx -.2$; Voracek & Dressler, 2006) but there was a failure to replicate in Japan (Kunihira et al., 2006). Losh and Piven (2007) argue based on a study in the parents of children with ASC who show impaired performance on the task, that the task is a marker of genetically meaningful endophenotype.

Golan, Baron-Cohen, Hill, and Rutherford (2007) tested an auditory analogue of the task, the Reading the Mind in the Voice task in TD individuals and people with ASC. This consists of spoken sentences such as, 'There's... uh,... there is something I want to ask you' (this is 'nervous') and 'Katherine, perhaps you'd come to help' (this is 'hopeful').[5] They found it correlated with the eyes task ($r \approx .4$). Both the eyes and the voice task were negatively correlated with AQ ($r \approx -.4$).[6]

Bayliss, Pellegrino, and Tipper (2005) analysed the relationships between cueing of attention and AQ, as part of a study into sex differences. Participants were shown faces, looking either towards the participant, left, or right. After either 100, 300, or 700 ms, a letter appeared to the left or to the right of the face; participants were asked to identify the letter as quickly as possible. They were told to ignore the direction the face was looking as this did not predict the target location. AQ was negatively correlated with response time when the letter came 700 ms after eyes cueing away from the target letter ($r \approx -.4$), so more autistic traits implies less of an effect of eye cueing.

Bayliss and Tipper (2005) ran a similar experiment used two cues: either a face gazing a particular direction or arrows pointing. To either side of the cue there were also faces, in one condition scrambled so they no longer resembled faces, in the other intact. The target was a red patch, which was superimposed over the faces. The task this time was just to respond when a target appeared, which it did in half the trials. AQ predicted the effect of the target type, with low AQ faster than high AQ participants at locating the patch on an intact face, and high AQ faster than lower AQ at locating the patch when on the scrambled face. Crucially, in this experiment the cue type had no effect, just the target. A second experiment used power tools instead of faces and the same patterns of results emerged: the target type affected behaviour. Overall, then, the high AQ group showed more cueing towards scrambled objects than the low AQ group, and the low AQ group showed more cueing towards coherent objects than the high AQ group.

Self-referential memory effects have been studied using AQ in both TD individuals and people with ASC (Lombardo et al., 2007). Participants were provided a list of trait adjectives in four different conditions. Depending on condition, they responded with the degree to which each adjective was descriptive of self, descriptive of a close friend, or descriptive of a 'dissimilar close other' (Harry Potter; participants were excluded if they had managed successfully to avoid all of the books and films). Following a distraction task, participants performed a recognition task on the words they had earlier assessed.

[5] The tone is (hopefully) important. The author read 'Katherine, perhaps you'd come to help' as 'bossy'.
[6] Performance within the TD group was not given (and not obtainable by the author).

People in the ASC group showed worse recognition memory for items judged in the self and friend condition than did the TD group. Pooling the two groups, higher AQ was associated with worse memory of words judged in the self condition ($r \approx -.3$).

Lombardo et al. (2007) also explored the relationship with performance on a self-focus sentence completion task in which participants are asked to complete sentences beginning, e.g., 'I think. . . ', 'If I had my way. . . '. The ASC group tended to use slightly fewer first-person pronouns in their completions than did the TD group (9% versus 11% of the pronouns used). Intriguingly, in TD group there was a positive correlation between AQ and the number of first-person pronouns ($r = .5$) but a negative correlations in the ASC group ($r \approx -.4$), hinting at a non-linear relationship.

Bach and Tipper (2007) investigated the relationships between observed and produced actions and their influence on personality judgements. Participants were shown a series of videos and had to identify as quickly as possible which of two actors was in each video. There were two actions performed by actors in the video: kicking a ball or sitting at a desk and pressing a key on keyboard. One actor was identified with a foot-key, the other with a keyboard press, and the mapping to actors was counterbalanced. Although there was an overall effect that the action perceived increased the speed of response for the congruent input method (e.g., foot-press for an actor kicking a ball), this was not moderated by AQ. At the end of the experiment, participants were also asked to judge to what extent the two actors were sporty and to what extent they were academic. The actor identified with a keyboard press was more likely to be considered academic than the actor identified with the foot-pedal; similarly the actor identified with a foot-key was more likely to be considered sporty than the actor identified with the keyboard. Interestingly, AQ did moderate this effect: higher AQ individuals were more likely to be affected by the type of response ($r \approx .5$).

Rawlings and Locarnini (2008) examined different kinds of creativity in artists and scientists as a function of AQ, using the Kent-Rosanoff Word Association Test.[7] Responses were classified into three disjoint sets: modal (non-opposite) response; idiosyncratic (given by fewer than 3% of participants, and not an opposite); and opposite. There was a positive correlation between AQ and the number of opposites produced ($r \approx .3$), but none of the

[7]The authors also studied relationships with psychosis proneness: see Section 2.4.4, pp. 35 ff., for a discussion of overlap between psychosis and autism.

other scores. This result is not given much comment by the authors, but it is interesting tendency when viewed through the lens of reasoning: perhaps (speculatively) AQ predicts a tendency to seek counterexamples.

Although not investigating the broader phenotype, the study by Hill and Bird (2006) is interesting as they also included AQ in the dataset in a study of people with Asperger syndrome (AS). Differences between groups were found in a range of subtasks in the Behavioural Assessment of the Dysexecutive Syndrome (BADS) battery, Hayling sentence completion, and the trail-making test. Within the AS group, higher AQ predicted fewer tasks completed in the six elements task ($r \approx -.6$), which consists of six tasks (two dictation, two arithmetic, and two picture-naming tasks) performed concurrently but each individually, and lower Hayling overall score ($r \approx -.5$).[8]

2.4.4 How specific are autistic traits?

It has been mentioned that AQ may be used to screen for ASC. It is important to note that AQ measures traits which have been shown not to be specific to ASC. In a Dutch study (Ketelaars et al., 2008), the composite score for a Dutch translation of AQ (AQ-D) could not distinguish individuals with a mild ASC from those in general outpatient clinical who had a psychiatric condition but not an ASC. However one of the five subscores, communication, did show a difference between groups (the ASC group had a mean of 4.5, versus 2.8 in the outpatient group). In a Japanese study, Kurita and Koyama (2006) showed that AQ-J tends to reflect non-psychotic anxiety or depressive state in addition to autistic traits. Hambrook, Tchanturia, Schmidt, Russell, and Treasure (2008) found that patients with anorexia nervosa had higher AQ scores than healthy controls (mean of 23 versus 15). Hurst, Nelson-Gray, Mitchell, and Kwapil (2007) found a range of correlations between AQ and a measure of psychosis proneness, the Schizotypal Personality Questionnaire (SPQ), in around 600 college students. The correlation between total scores was around $r = .5$, with the largest correlation between subscores was between a measure of interpersonal ability in SPD and social skills in AQ ($r = .5$). A study by Rawlings and Locarnini (2008) found a positive correlation (again around $r = .5$) between Introvertive Anhedonia in OLIFE, another measure of psychosis proneness, and AQ.

[8]No data are reported for the TD group and AQ, nor were any differences in scores on the subtests of AQ reported.

How should these results be interpreted? There is evidence that the discrete boundaries imposed by DSM and ICD-10 are not always respected by the conditions described. For instance Gillberg (1983) wondered whether a ‘common biochemical disturbance’ may cause ASC in young males and anorexia nervosa in young girls, and recently evidence has been found of weak central coherence in anorexia (C. Lopez et al., 2008; Southgate, Tchanturia, & Treasure, In press). Depression is often comorbid with autism (e.g., Ghaziuddin, Ghaziuddin, & Greden, 2002). *DSM-IV-TR* warns clinicians not to confuse Schizophrenia and Asperger Disorder as aspects of the conditions are quite similar. Given this overlap between conditions, it is unavoidable for measures of ‘autistic’ traits to detect traits associated with conditions other than ASC. Driven by results like these, researchers have argued for a transdiagnostic or complaint oriented approach to psychiatric and developmental conditions (e.g., Harvey, Watkins, Mansell, & Shafran, 2004; Bentall, 2006). There seems to be no shortage of continuums, overlapping and distinct, within and between typical and atypical development and clinical and non-clinical conditions of existence. I will strive to discuss the specific traits concerned where possible, e.g., pragmatic language, social interaction, and perseveration, rather than only ‘autistic’ traits.

2.5 Summary

In this chapter I gave an overview of autism spectrum conditions and how various theoretical frameworks have been developed to explain differences in cognitive processes. I have argued that it is meaningful to explore the autistic trait continuums in typically developing individuals, in particular that self-reported autistic-like traits have been shown to covary with milder variants of cognitive features seen in people with ASC, and that autistic traits seem likely to relate to deductive reasoning processes.

Chapter 3

Using Mathematical Theories of Reason to Study People Reasoning

> ... in the same way as optics is the geometry of light rays, logic is the mathematics of reasoning systems. No one would maintain that the technicalities of optics are irrelevant to how the visual system infers, say, form from motion. Likewise, no consideration of actual human reasoning processes can proceed without careful attention to mathematical constraints on these processes.
>
> Stenning and van Lambalgen (2005, p. 921)

It is common in the psychology of reasoning to begin with theories from logic and probability to produce stimuli for participants to reason about and psychological theories of how they are doing the reasoning. This section outlines three main sources of theory: proof theory and model theory of classical logic, and Bayesian probability theory, together with examples of how the theories have been used to inform psychological theory.[1] Often these theories are used only to characterise the most common inferences that groups of people draw. The chapter concludes with a review of empirical evidence that there are important individual differences in reasoning which go beyond issues of accuracy, and a summary of attempts to characterise these differences which I draw on in later chapters in analyses of my own data.

[1] Probability can be given a proof and model theoretical treatment; this issue is sidestepped here for simplicity.

3.1 Interpretational Knots and Derivational Machinery

An analysis in formal logic of natural language first requires an *interpretation* into the language of the logic. In one of the classic books on logic, Quine (1966, p. 40) summarises the process as follows:

> The premisses and conclusions may treat of any topics and are couched, to begin with, in ordinary language rather than in the technical ideography of modern logic. It is as an aid to establishing implications that we proceed to mutilate and distort the statements, introducing schematic letters in order to bring out the relevant skeletal structures, and translating varied words into a few fixed symbols such as '$\supset$' and '$\vee$' in order to gain a manageable economy of structural elements. *The task of thus suitably paraphrasing a statement and isolating the relevant structure is just as essential to the application of logic as is the test of proof of implication for which that preliminary task prepares the way.* [Emphasis mine.]

This is well known, though not always made obvious in the psychology literature. The process described by Quine was for classical logic, but generalises beyond to other logics. The choice of logic is itself a stage in the interpretation process and has a direct bearing on the degree of mutilation and distortion of the original discourse required. If one is interested in tense, then it makes sense to choose a logic which has as primitives operators for describing time (e.g., van Lambalgen & Hamm, 2005). Modelling of action is possible using Floyd-Hoare logic (Hoare, 1969) or its successors. Once the choice of logic has been made and the various parameters set, then derivation is possible.

Interpretation is nicely illustrated by novice students of logic. When provided with an English sentence of the form 'if P then Q', e.g., 'If John leans out the window any further, he'll fall', and the classical logical conditional $P \Rightarrow Q$, read as 'if P then Q', they often formalise the English as the more complicated expression, $(P \Rightarrow Q) \wedge (\neg P \Rightarrow \neg Q)$ (e.g., Geis & Zwicky, 1971). Clearly given this interpretation, the inferences they draw will be different to those drawn if they had interpreted the conditional only as $P \Rightarrow Q$. We will see other examples later from the reasoning literature, but it is worth noting in passing that issues of interpretation are important outside the logic class. For instance Justice Stephen G. Breyer, commenting on a court ruling, notes some factors influencing interpretation (Barnes, 2008):[2]

[2]Found thanks to *Leiter Reports* (http://leiterreports.typepad.com/blog/2008/01/supreme-court-i.html).

> When I call out to my wife, 'There isn't any butter,' I do not mean, 'There isn't any butter in town.' The context makes clear to her that I am talking about the contents of our refrigerator.
>
> That is to say, it is context, not a dictionary, that sets the boundaries of time, place and circumstances within which words such as 'any' will apply.

These issues of interpretation were most famously taken up by Henle (1962). She lists sources of apparent error in reasoning tasks, all of which relate to how the material to be reasoned about was interpreted, for instance: disagreeing with premises provided and refusing to draw inferences from them; restating a premise with changed meaning; ignoring a premise and drawing inferences from those remaining; adding a premise that is taken to be commonsensical. Henle (1962) concludes that fallacies are possible in reasoning, but quotes J. S. Mill who says it is 'scarcely ever possible decidedly to affirm that any argument involves a bad syllogism.' Further she argues that supposed errors may be due to the reasoner's understanding of the materials. Interestingly she notes that if people really could not reason logically, it would be impossible for them to understand each other.

Clark (1969, p. 403) argues that deductive reasoning cannot be studied in isolation from linguistic processes:

> ... the principal difficulties inherent in many reasoning problems are not due to cognitive processes specific to these problems, but to the very language in which the problems are stated. Linguistic processes like these arise in every situation in which a problem is stated in linguistic terms.

He gives the example of three-term series problems, e.g., 'If John is better than Dick, and Pete is worse than Dick, then who is best?' The exact wording of the statement affects the inferences drawn, for instance the following is harder: 'If Dick is worse than John, and Dick is better than Pete, then who is best?' Smedslund (1970) argues that researchers of reasoning have two choices: assume the reasoner is logical and test comprehension, or assume the reasoner has interpreted the premises as intended and test logicality. Interestingly it seems he might allow his notion of comprehension to be affected by limitations to working memory. He notes that failure to reason logically is often modelled computationally (implicitly, using a model equivalent to a Turing machine). As is well known, the state of a Turing machine is easily captured by a logical description (e.g., Tärnlund, 1977, demonstrates that Turing comptuable functions can be implemented using logic programming). Thus, he claims, the models explain the performance (including that constrained by working memory) 'as a problem of faulty understanding, because they attempt to find alternative premises on the basis of which the performance may be simulated.' This line of argument is not without difficulties. As others have argued for analogous problems (e.g., Oaksford & Chater, 2007, pp. 48–50), the description of the state of the Turing machine is at the wrong level of analysis to characterise comprehension, and allowing anything that is implemented in a Turing-equivalent computational theory to be considered a set of logical premises says very little.

A distinction often made in the reasoning literature is that of competence versus performance. Chomsky (1965, p. 4) introduced these terms in the context of language use:

> We thus make a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations). [...] A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on. The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that have been mastered by the speaker-hearer and that he puts to use in actual performance.

Two related ideas are introduced here: the problem of how a theoretician determines the intended output of some operation, in spite of errors, and how people learn from others and detect errors in what others are trying to do. People try to behave in a way which is consistent with their own notion of competent behaviour, but their actual performance falls short because of various constraints, perhaps working memory limitations, time constraints. As children develop they see other people trying and failing to be consistent with their competence model, and from this try to infer the competence model. People sometimes can report and reflect upon notions of competence. Education presumably can change competence models and encourage performance to be more compatible with a person's competence model (where that person could be an observer, e.g., a theoretician). Someone who has taken a course on formal proof might be expected (and expect) to perform differently on a reasoning task in an exam setting than someone who has not. An important issue in studying reasoning is to try to infer what it is that people are trying to do or, at some more mechanistic level of analysis, what perhaps competing goals cognitive processes may have. (Section 11.2 has more on these sorts of knots.)

In the remainder of this section we will examine three main sources of inspiration for psychological models of reason. A sketch will be given of psychological uses of the theories, in particular how the natural language and context of a situation are encoded and how inferences are drawn. The syllogism, a fragment of monadic predicate logic, was chosen as the running example as this fragment of logic has been the subject of extensive empirical and theoretical psychological investigation over the years.

3.1.1 Model theory

In logic, a *model*, also known as a *structure*, is a mathematical object with respect to which the truth of a theory, and of sentences within a theory, may be judged. If a proposition P is true in a model $\mathcal{M}$ ('$\mathcal{M}$ models P') then we write $\mathcal{M} \models P$. If P is true in all models, then we write $\models P$. Often the notation is abused slightly and sentences (rather than a model) are written on the lefthand side of the turnstyle: $P_1, P_2, \ldots, P_n \models Q$ expresses that in all models for which $P_1, P_2, \ldots, P_n$ are true, Q is also true. We can now characterise in terms of models when a syllogistic inference is classically valid. Let's take the example

$$\begin{array}{c} \text{All } A \text{ are } B \\ \text{All } B \text{ are } C \\ \hline \text{All } A \text{ are } C \end{array}$$

This argument is valid if and only if in all models where *all* A *are* B and *all* B *are* C are both true, *all* A *are* C is also true. Another way to write this is

$$\text{All } A \text{ are } B, \text{All } B \text{ are } C \models \text{All } A \text{ are } C$$

More generally,

$$\begin{array}{c} Q_1AB \\ Q_2BC \\ \hline Q_3AC \end{array}$$

is valid when

$$Q_1AB, Q_2BC \models Q_3AC$$

and analogously for the other arrangements of premises and conclusion. The details of this definition will become clearer in the next sections as we examine particular kinds of model.

3.1.1.1 First-order models

A first-order model $\mathcal{M}$ consists of:

1. A domain (or universe), $\mathsf{dom}(\mathcal{M})$. This is a set of constants, e.g., the set of natural numbers, $\{0, 1, 2, 3, \ldots\}$.

2. For each individual, i, in the language, an object $i^{\mathcal{M}}$ in the domain.

3. For each function symbol, f, in the language, a mapping, $f^{\mathcal{M}}$, between elements in the domain.[3]

4. For each predicate, P, in the language, a function, $P^{\mathcal{M}}$, mapping constants (which may be the result of function application) to truth values. Classically, there are only two truth values, true and false.

Typically, the following mapping is used from the syllogistic language into the language of classical logic:

All A are B	$\rightsquigarrow$	$\forall x.\ A(x) \Rightarrow B(x)$
No A are B	$\rightsquigarrow$	$\forall x.\ A(x) \Rightarrow \neg B(x)$
Some A are B	$\rightsquigarrow$	$\exists x.\ A(x) \wedge B(x)$
Some A are not B	$\rightsquigarrow$	$\exists x.\ A(x) \wedge \neg B(x)$

The choice of domain is often not considered and it is assumed that inferences are about any possible domain.

Much is left outside this particular choice of interpretation—by design. We lose the distinction between subject and predicate which matters in the natural language (e.g., Westerståhl, 1989). The morphology of the *no* quantifier in the English is very different to the classical logic expression: the negation in the former is pushed to the head of the sentence; in the latter, it rests close to the B term. The context in which the sentence is introduced is ignored with this interpretation; for instance not allowing that some of the time, some (but not all) people interpret *some* as meaning some but not all, $[\exists x.\ A(x) \wedge B(x)] \ \wedge\ [\exists x.\ A(x) \wedge \neg B(x)]$. However the notion of classical validity is more easily seen and separated from linguistic packaging.

[3] Note that syllogisms don't contain function symbols.

An interpretation is a mapping from sentences through models to truth values. If ϕ is a sentence and $\mathcal{M}$ a model, then $[\![\phi]\!]_{\mathcal{M}}$ is the interpretation of ϕ in $\mathcal{M}$.[4] See Figure 3.1 for a definition. Essentially it says things like '$A \wedge B$ is true when A is true and B is true' and '$\forall x.\ P(x)$ is true when $P(x)$ is true for every x', so it characterises a correspondence with the mathematical vernacular.[5] The general approach of defining models in this way has its critics, for instance Girard (1999) writes:

> In fact the notion of *truth à la Tarski* avoids complete triviality by the use of the magical expression 'meta': we presuppose the existence of a meta-world, in which logical operations already make sense; the world of discourse can therefore be interpreted in the meta-world, typically the truth of A becomes 'meta-A', and we can in turn explain 'meta-A' by 'meta-meta-A' [...] We are facing a transcendental explanation of logic *'The rules of logic have been given to us by Tarski, which in turn got them from Mr. Metatarski'*, something like *'Physical particles act this way because they must obey the laws of physics'*.

This is in fact the idea: the formalism is mapped to a language where operations do already make sense, similar to how one can write an interpreter for one programming language in another where primitives already exist for selection, composition, and recursion. The important point is that 'truth à la Tarski' is not the final word on what these logical expressions 'really' mean, but it's an important mechanism for understanding logics, as I will now outline.

Consider the sorts of operations which may be performed on models. Checking the truth of an expression in a model is checking individual cases. For instance take discrete countable models, each of which has 1 individual, call it i. For syllogisms, there are 8 distinct ways the predicates A, B, and C can be defined for i. Each row of Table 3.1 defines a possible interpretation of A, B, and C. So for instance in the model represented by the second row, call it $\mathcal{M}$, the following holds:

$$\begin{aligned}
[\![A(i)]\!]_{\mathcal{M}} &= \textsf{true}\\
[\![B(i)]\!]_{\mathcal{M}} &= \textsf{true}\\
[\![C(i)]\!]_{\mathcal{M}} &= \textsf{false}
\end{aligned}$$

[4] The formalisation of this idea is usually attributed to Tarski and Vaught (1956).

[5] Other variants make the definition in terms of arithmetic by representing true as 1 and false as 0, so for instance $P \wedge Q = min(P, Q)$, $P \vee Q = max(P, Q)$, and so on. In this case, arithmetic becomes the metalanguage.

$$
\begin{aligned}
[\![\top]\!]_{\mathcal{M}} &= \mathsf{true} \\
[\![\bot]\!]_{\mathcal{M}} &= \mathsf{false} \\
[\![P(t_1,\ldots,t_n)]\!]_{\mathcal{M}} &= P^{\mathcal{M}}(t_1^{\mathcal{M}},\ldots,t_n^{\mathcal{M}}) \\
[\![\phi \land \psi]\!]_{\mathcal{M}} &= \text{true iff } [\![\phi]\!]_{\mathcal{M}} \text{ and } [\![\psi]\!]_{\mathcal{M}} \text{ both give true} \\
[\![\phi \lor \psi]\!]_{\mathcal{M}} &= \text{true iff at least one of } [\![\phi]\!]_{\mathcal{M}} \text{ and } [\![\psi]\!]_{\mathcal{M}} \text{ gives true} \\
[\![\phi \Rightarrow \psi]\!]_{\mathcal{M}} &= \mathsf{false} \text{ iff } [\![\phi]\!]_{\mathcal{M}} = \mathsf{true} \text{ and } [\![\psi]\!]_{\mathcal{M}} = \mathsf{false} \\
[\![\neg\phi]\!]_{\mathcal{M}} &= \text{true iff } [\![\phi]\!]_{\mathcal{M}} = \mathsf{false}\text{, otherwise true} \\
[\![\forall x.\ P(x)]\!]_{\mathcal{M}} &= \text{true iff } \mathsf{false} \notin \{P^{\mathcal{M}}(y) \mid y \in \mathsf{dom}(\mathcal{M})\} \\
[\![\exists x.\ P(x)]\!]_{\mathcal{M}} &= \text{true iff } \mathsf{true} \in \{P^{\mathcal{M}}(y) \mid y \in \mathsf{dom}(\mathcal{M})\}
\end{aligned}
$$

Figure 3.1: Classical interpretation for first-order logic

$[\![A(i)]\!]_{\mathcal{M}}$	$[\![B(i)]\!]_{\mathcal{M}}$	$[\![C(i)]\!]_{\mathcal{M}}$
true	true	true
true	true	false
true	false	true
true	false	false
false	true	true
false	true	false
false	false	true
false	false	false

Table 3.1: A table enumerating each possible specification of $A(i)$, $B(i)$, and $C(i)$ for a given individual, i.

Individuals	Possible models
1	8
2	64
3	512
4	4096
5	32768
...	...
n	2^{3n}

Table 3.2: The number of models to check with n individuals and 3 classes

As the number of individuals increases, so too does the number of models. In general for models containing n individuals and 3 classes, there will be 2^{3n} models (see Table 3.2). Also note that infinitely many non-countable (and infinite) domains exist. Naïvely following this definition of first-order models to check if a syllogistic argument is valid requires checking infinitely many models, some of which will themselves be uncountably infinite.

Syllogisms are a fragment of monadic logic, for which there is a theorem that allows us to restrict attention to finitely many finite first-order models (see Boolos & Jeffrey, 1980, Ch. 25).

Theorem 3.1 *If P is a monadic formula which is satisfiable, then P is true in some interpretation whose domain contains at most $2^k \cdot r$ members, k being the number of predicates and r being the number of variables in P.*

From this, it is possible to derive an algorithm for checking if a syllogistic inference is valid, since it suffices to check if its negation is satisfiable—if not, then the syllogism is valid. Syllogistic inference can be written as

$$Q_1x.\ A(x)B(x) \wedge Q_2x.\ B(x)C(x) \ \Rightarrow\ Q_3x.\ A(x)C(x)$$

where the Q_i's are the quantifiers. To check if this is true in all models means to check if

$$\neg[Q_1x.\ A(x)B(x) \wedge Q_2x.\ B(x)C(x) \ \Rightarrow\ Q_3x.\ A(x)C(x)]$$

is false in all models with at most $2^3 \cdot 3 = 24$ individuals. That's a lot of models. A lower bound can be found by implementing the combinatorial generation and determining the largest model which is needed to filter out the valid from the invalid syllogisms (i.e., when we already know which syllogisms are valid and invalid). It turns out that it suffices to restrict attention to countable models with 3 individuals since for such models only the conclusions of valid syllogistic inferences are indeed true. However this is still a costly checking algorithm since at worst there are $512 \times 8 = 4096$ models to check for one pair of premises. (See Appendix D for an explanation of the enumeration.)

3.1.1.2 Herbrand models

Herbrand models are models in the propositionalisation of the quantified expressions. For syllogisms, the first step to propositionalisation is performing the following interpretation:

All A are B	$\rightsquigarrow$	$A(x) \Rightarrow B(x)$	, where x is free
No A are B	$\rightsquigarrow$	$A(x) \Rightarrow \neg B(x)$	, where x is free
Some A are B	$\rightsquigarrow$	$A(c) \wedge B(c)$	, where c is a new constant
Some A are not B	$\rightsquigarrow$	$A(c) \wedge \neg B(c)$	, where c is a new constant

Given this interpretation, the *some* and *some... not* statements are now sentential propositions, i.e., without quantifiers and without free variables. Each constant introduced stands for any individual in any model. The two cases of universal quantifier, *all* and *no*, are slightly more complicated. Note that the free variable allows us to view universal quantification as a *family* of implications, one implication for each possible x. After propositionalising the existentials, we know what these x's can be—namely, any one of the constants introduced, so an implication must be introduced for each constant. It is also necessary to deal with arbitrary x's, for instance to allow us to prove the transitivity[6] of *all*. For syllogisms, this may be achieved by transforming $A(x) \Rightarrow B(x)$, where x is free, into $A(a) \Rightarrow B(a)$, where a is another constant, but considered (at the meta-level) to be a special kind of arbitrary constant. To ensure soundness, it's crucial to ensure that a non-arbitrary constant cannot be transformed into an arbitrary constant. A conjunction $P_1(c_1) \wedge P_2(c_2) \wedge P_3(c_3) \wedge \ldots \wedge P_n(c_n)$ may be interpreted as the existential $\exists x.\ P_1(x) \wedge P_2(x) \wedge P_3(x) \wedge \ldots \wedge P_n(x)$ iff $c_i = c_j$ for all i, j.

Some examples should make the interpretation clearer:

[6] A relation, R, is transitive if $xRy \wedge yRz \equiv xRz$.

Example 3.1 *The following premise pair has no valid syllogistic conclusion:*

$$\begin{array}{lll} \textit{Some A are B} & \rightsquigarrow & A(c_1) \wedge B(c_1) \\ \textit{Some B are C} & \rightsquigarrow & B(c_2) \wedge C(c_2) \end{array}$$

We can derive $A(c_1) \wedge C(c_2)$*, but* $c_1 \neq c_2$*.*

Example 3.2 *These universals do have a valid conclusion,* $A(a) \Rightarrow C(a)$*, i.e., all A are C:*

$$\begin{array}{lll} \textit{All A are B} & \rightsquigarrow & A(a) \Rightarrow B(a) \\ \textit{All B are C} & \rightsquigarrow & B(a) \Rightarrow C(a) \end{array}$$

Example 3.3 *This mixed existential and universal pair also has two valid conclusions, some A are C (*$A(c_1) \wedge C(c_1)$*) and some C are A (*$C(c_1) \wedge A(c_1)$*):*

$$\begin{array}{lll} \textit{Some A are B} & \rightsquigarrow & A(c_1) \wedge B(c_1) \\ \textit{All B are C} & \rightsquigarrow & B(a) \Rightarrow C(a) \\ & & B(c_1) \Rightarrow C(c_1) \end{array}$$

Example 3.4 *Another mixed existential and universal, this time with no valid conclusion:*

$$\begin{array}{lll} \textit{All A are B} & \rightsquigarrow & A(a) \Rightarrow B(a) \\ & & A(c_1) \Rightarrow B(c_1) \\ \textit{Some B are C} & \rightsquigarrow & B(c_1) \wedge C(c_1) \end{array}$$

Example 3.5 *This example assumes that A, B, and C each hold for at least one individual, thus allowing us to conclude* some A are C *(*$A(c_1) \wedge C(c_1)$*),* some C are A *(*$C(c_1) \wedge A(c_1)$*), as well as the universal* all A are C *as described above.*

$$\begin{array}{lll} \textit{Some A} & \rightsquigarrow & A(c_1) \\ \textit{Some B} & \rightsquigarrow & B(c_2) \\ \textit{Some C} & \rightsquigarrow & C(c_3) \\ \textit{All A are B} & \rightsquigarrow & A(a) \Rightarrow B(a) \\ & & A(c_1) \Rightarrow B(c_1) \\ & & A(c_2) \Rightarrow B(c_2) \\ & & A(c_3) \Rightarrow B(c_3) \\ \textit{All B are C} & \rightsquigarrow & B(a) \Rightarrow C(a) \\ & & B(c_1) \Rightarrow C(c_1) \\ & & B(c_2) \Rightarrow C(c_2) \\ & & B(c_3) \Rightarrow C(c_3) \end{array}$$

The models of this propositionalised world are called *Herbrand models*. The domain consists of one object for each of the propositional variables. For instance, for Example 3.5 above, that would be

$$\{A(a),\ B(a),\ C(a),\ A(c_1),\ A(c_2),\ A(c_3),\\ B(c_1),\ B(c_2),\ B(c_3), C(c_1),\ C(c_2),\ C(c_3)\}.$$

It is important to emphasise that each of these objects cannot be decomposed further. Each is a propositional variable which, classically, is either true or false.

Checking if, say, some A are C follows from the premises can be decided naïvely by building a truth table with $2^{12} = 4096$ rows. Each row in the table represents a model. The additional sentences needed to represent this conclusion are $A(c_1) \wedge C(c_1)$, $A(c_2) \wedge C(c_2)$, and $A(c_3) \wedge C(c_3)$. The provable representation of *some A are C* is $A(c_1) \wedge C(c_1)$. However the number can be drastically reduced by clever choice of models, e.g., taking advantage of properties of implication.

3.1.1.3 Johnson-Laird and colleagues' mental models theory

The most well known attempt to model syllogistic reasoning performance using first-order model construction is by Johnson-Laird and colleagues. The key insight of this approach is that using just finite (first-order) models of individuals, and—of key importance—a suitable heuristic for model construction, can give a proof procedure which, in terms of inferences drawn, is equivalent to more explicitly inference based approaches. Inferences made during performance can also be modelled; differences to classical logic are explained as being due to an unsound model selection algorithm or due to working memory limitations. The theory has undergone two main revisions, e.g., for the first see Johnson-Laird (1983, Ch. 5). I shall describe a more recent version (see Johnson-Laird & Byrne, 1991, pp. 118–124).

First a model is constructed for the initial premise of the problem:

$$\text{All } A \text{ are } B \quad \rightsquigarrow \quad \begin{array}{ll} [A] & B \\ [A] & B \\ \multicolumn{2}{c}{\ldots} \end{array} \qquad \text{No } A \text{ are } B \quad \rightsquigarrow \quad \begin{array}{ll} [A] & \\ [A] & \\ & [B] \\ & [B] \\ \multicolumn{2}{c}{\ldots} \end{array}$$

$$\text{Some } A \text{ are } B \quad \rightsquigarrow \quad \begin{array}{ll} A & B \\ A & B \\ \multicolumn{2}{c}{\ldots} \end{array} \qquad \text{Some } A \text{ are not } B \quad \rightsquigarrow \quad \begin{array}{ll} A & \\ A & \\ A & [B] \\ A & [B] \\ \multicolumn{2}{c}{\ldots} \end{array}$$

Each row represents an individual. The presence of square brackets around a property name indicates that the property has been fully represented. For instance, the initial model of All A are B has exactly 2 individuals with the property A and at least 2 with the property B. They write, ‘The number of individuals remains arbitrary, but, for simplicity, is likely to be small’ (Johnson-Laird & Byrne, 1991, p. 119), which I take to mean that the 2 could be any small constant.

The second stage of construction is to add information from the second premise. For instance from All A are B and All B are C ‘the following sort of initial model’ is constructed (Johnson-Laird & Byrne, 1991, p. 121):

$$\begin{array}{lll} [[A] & B] & C \\ [[A] & B] & C \\ \multicolumn{3}{c}{\ldots} \end{array}$$

From this one model, the conclusion that All A are C is easily drawn, which it happens is true in all first-order models. In the computer simulation[7] of this version of the theory, All C are A is also true in this model. This conclusion is classically invalid since it doesn’t hold in all models, so according to the implementation to correctly withhold this conclusion requires examining more than one model.

Consider the example All B are A and No B are C. The initial model constructed is (Johnson-Laird & Byrne, 1991, p. 122):

$$\begin{array}{lll} [A & [B]] & \\ [A & [B]] & \\ & & [C] \\ & & [C] \\ \multicolumn{3}{c}{\ldots} \end{array}$$

[7] Thanks to Phil Johnson-Laird for providing this.

In this model, No C are A is true. However this conclusion is not true in all models, for instance

$$\begin{array}{lll} [A & [B]] & \\ [A & [B]] & \\ A & & [C] \\ & & [C] \\ & \ldots & \end{array}$$

So again this would be considered a multiple model problem.

Recall again from Section 3.1.1.1 that there are many models which could be chosen—in the worst case, infinitely many. Hence the under specification of the theory ensures that any data can characterised, given appropriate selection and manipulation of the models chosen. For instance, problems currently designated as being multiple model could occasionally be single model given a fortuitous choice. Stenning and Oberlander (1995) show that a model using Euler Circles which does not need to make appeal to members of classes, but rather the classes themselves, is indistinguishable empirically from mental models theory. This is an example of the importance of equivalence classes of models, rather than individual models themselves.

3.1.2 Proof theory

3.1.2.1 Natural Deduction Calculi

One way of formalising proof is using natural deduction.[8] Proofs in natural deduction can be viewed as trees where each point on the tree is a sentence $P_1, P_2, \ldots, P_n \vdash Q$. Intuitively the P's are what is 'known' (or assumed, or imagined to be true) at that point of the proof; Q is the claim which, if the proof is valid, follows from $P_1, P_2, \ldots, P_n$. At the root of the tree is the statement which has been proved. A proof is complete when all of its branches end with a judgement of the form $P_1, P_2, \ldots, P_n \vdash P_i$, where i is one of $1, 2, \ldots, n$. To illustrate the idea see the proof below of $P \wedge Q \Rightarrow Q \wedge P$.

$$\dfrac{\dfrac{\dfrac{\dfrac{}{P \wedge Q \vdash P \wedge Q}\,\text{ID}}{P \wedge Q \vdash Q}\,\wedge\text{-elim} \qquad \dfrac{\dfrac{}{P \wedge Q \vdash P \wedge Q}\,\text{ID}}{P \wedge Q \vdash P}\,\wedge\text{-elim}}{P \wedge Q \vdash Q \wedge P}\,\wedge\text{-intro}}{\vdash P \wedge Q \Rightarrow Q \wedge P}\,\Rightarrow\text{-intro}$$

[8] Originally due to Gentzen (1935).

$$\frac{}{P_1, P_2, \ldots, P_n \vdash P_i}\ \mathsf{ID}\text{, where } 1 \leq i \leq n$$

$$\frac{\Gamma \vdash P \qquad \Gamma \vdash P \Rightarrow Q}{\Gamma \vdash Q}\ \mathsf{MP} \qquad \frac{\Gamma \vdash P \Rightarrow Q \qquad \Gamma \vdash \neg Q}{\Gamma \vdash \neg P}\ \mathsf{MT}$$

$$\frac{\Gamma \vdash P \wedge Q}{\Gamma \vdash P}\ \wedge\text{-}\mathsf{elim} \qquad \frac{\Gamma \vdash P \wedge Q}{\Gamma \vdash Q}\ \wedge\text{-}\mathsf{elim} \qquad \frac{\Gamma \vdash P \qquad \Gamma \vdash Q}{\Gamma \vdash P \wedge Q}\ \wedge\text{-}\mathsf{intro}$$

$$\frac{\Gamma \vdash \neg\neg P}{\Gamma \vdash P}\ \neg\neg\text{-}\mathsf{elim} \qquad \frac{\Gamma \vdash P}{\Gamma \vdash \neg\neg P}\ \neg\neg\text{-}\mathsf{intro}$$

$$\frac{\Gamma, P \vdash Q}{\Gamma \vdash P \Rightarrow Q}\ \Rightarrow\text{-}\mathsf{intro} \qquad \frac{\Gamma \vdash P(x)}{\Gamma \vdash P(c)}\ \mathsf{FV\text{-}inst}$$

Figure 3.2: Natural deduction rules required for syllogisms. Note how the only rule which modifies the assumptions is ⇒-intro. The rule FV-inst instantiates a free variable x with any constant c. All rules except ¬¬-elim are constructively valid.

The trees are conventionally written with the root placed at the bottom as shown. Reading from bottom to top, the story is somewhat like the following:

> To prove $P \wedge Q \Rightarrow Q \wedge P$ we assume $P \wedge Q$ and prove $Q \wedge P$. To prove $Q \wedge P$ using this assumption we prove, separately Q and P. To prove Q we prove $P \wedge Q$. To prove $P \wedge Q$ we extract $P \wedge Q$ from the list of premises. Similarly, to prove P we prove $P \wedge Q$. To prove $P \wedge Q$ we extract $P \wedge Q$ from the list of premises.

Hopefully this mechanical story makes explicit the recursive nature of the proofs. To find the justification for a point in the tree we just check if the conclusion is a member of the premises at that point. If not, then the justification depends on the branches, which we can check by traversing up the tree.

To solve syllogisms in natural deduction calculi, the following interpretation is used.

All A are B	$\rightsquigarrow$	$A(x) \Rightarrow B(x)$	, where x is free
No A are B	$\rightsquigarrow$	$A(x) \Rightarrow \neg B(x)$	, where x is free
Some A are B	$\rightsquigarrow$	$A(c) \wedge B(c)$	, where c is a new constant
Some A are not B	$\rightsquigarrow$	$A(c) \wedge \neg B(c)$	, where c is a new constant

See Figure 3.2 for the rules required. We do not explicitly use quantifiers here; constants take the role of an existentially quantified variable, free variables take the role of a universally quantified variable.

The following example requires all of the rules except $\Rightarrow$-intro. First the sentences are translated into predicate logic as follows:

$$
\begin{array}{lcll}
\text{No } A \text{ are } B & \rightsquigarrow & A(x) \Rightarrow \neg B(x) & \text{, where } x \text{ is free} \\
\text{Some } B \text{ are } C & \rightsquigarrow & B(c) \wedge C(c) & \text{, where } c \text{ is a new constant}
\end{array}
$$

Let's prove $A(x) \Rightarrow \neg B(x), B(c) \wedge C(c) \vdash C(c) \wedge \neg A(c)$. Since none of the proof steps manipulates the premises, we shan't bother making them explicit.

$$
\begin{prooftree}
\AxiomC{}
\RightLabel{ID}
\UnaryInfC{$B(c) \wedge C(c)$}
\RightLabel{$\wedge$-elim}
\UnaryInfC{$C(c)$}
\AxiomC{}
\RightLabel{ID}
\UnaryInfC{$B(c) \wedge C(c)$}
\RightLabel{$\wedge$-elim}
\UnaryInfC{$B(c)$}
\RightLabel{$\neg\neg$-intro}
\UnaryInfC{$\neg\neg B(c)$}
\AxiomC{}
\RightLabel{ID}
\UnaryInfC{$A(x) \Rightarrow \neg B(x)$}
\RightLabel{FV-inst}
\UnaryInfC{$A(c) \Rightarrow \neg B(c)$}
\RightLabel{MT}
\BinaryInfC{$\neg A(c)$}
\RightLabel{$\wedge$-intro}
\BinaryInfC{$C(c) \wedge \neg A(c)$}
\end{prooftree}
$$

Note that $A(x) \Rightarrow \neg B(x), B(c) \wedge C(c) \vdash \neg A(c) \wedge C(c)$ can also easily be proved by flipping over the tree above, but doesn't have a translation back into the syllogistic language.

3.1.2.2 Mental logic theory

Rips (1994) gives a well known book-length treatment of a mental logic approach to reasoning, based closely on natural deduction. The theory assumes that when people reason they attempt—even though they may be unsuccessful—to construct a mental proof in working memory. The process is constrained by working memory capacity, availability of the required rules, and also unsound rules may be present which allow the derivation of statements that are incompatible with classical logic.

The theory covers a wide range of statements expressible in classical logic. Categorical syllogisms are encoded as described above:

$$
\begin{array}{lcl}
\text{All } A \text{ are } B & \rightsquigarrow & \forall x.\ A(x) \Rightarrow B(x) \\
\text{No } A \text{ are } B & \rightsquigarrow & \forall x.\ A(x) \Rightarrow \neg B(x) \\
\text{Some } A \text{ are } B & \rightsquigarrow & \exists x.\ A(x) \wedge B(x) \\
\text{Some } A \text{ are not } B & \rightsquigarrow & \exists x.\ A(x) \wedge \neg B(x)
\end{array}
$$

The theory consists of a set of rules which will be familiar to those who use natural deduction. See Figure 3.3 for how $\forall$ introduction (p. 52) is represented. A search mechanism, similar to that used in automatic theorem provers, takes a set of premises and using forward and backwards search, attempts to rewrites them to a conclusion. From the automatic theorem proving community, Sieg and Byrnes (1998) developed the first

1. If $P(a)$ holds in a given domain,
2. and does not occur as a subscript in $P(a)$,
3. and a was not produced by FOR SOME elimination,
4. and a does not occur in any suppositions that hold in the domain,
5. and a does not occur within the scope of (FOR ALL v) or (FOR SOME v) in $P(a)$,
6. and $P(v)$ is the result of replacing all occurrences of a in $P(a)$ by v,
7. then (FOR ALL v) $P(v)$ can be added to the domain.

Figure 3.3: The rule for $\forall$ introduction (Rips, 1994).

systematic proof procedure for natural deduction, though this has not yet been applied to the study of psychological phenomena.

3.1.2.3 Source-founding model

The source-founding model (Stenning & Yule, 1997) is an algorithm specifying how features of the syllogistic stimulus drive the specification of a critical individual at a representational level of abstraction more general than graphical or sentential representations. The existence of such an individual implies that the syllogism has a valid conclusion. Here I use Gentzen natural deduction to explain its operation. A critical individual is a constant which, using the premises, can be related to both A and C via B. The model makes a prediction about conclusion term-order. The key detail is in dealing with existential premises: if there are two, then it's known there is no valid conclusion; if there is one, however, then it is chosen as the 'source premise' and its two terms become the beginning of the critical individual description. For instance consider the premises *all B are A* and *some B are C*. The second premise gives an individual, i, such that $B(i) \wedge C(i)$. The first premise $B(x) \Rightarrow A(x)$ takes $B(i)$ to $A(i)$, so giving a complete individual description of $B(i) \wedge C(i) \wedge A(i)$, and thus the conclusion some C are A. This strategy of integrating existential and universal premises has a long history going back to Aristotle's proof by ecthesis (proof by exposition).[9]

A proper formal reconstruction of the algorithm would alter the authors' treatment of universal quantifiers. They say: 'We think of the no-empty-sets axiom as providing a choice of starting points in the process of building up a description of a single critical individual' (Stenning & Yule, 1997, p. 120). This confusion between constants (introduced from existential quantifiers) and arbitrary variables (introduced from universal quantifiers) has the possibility to introduce a contradiction. However the error is easily rectified if universal quantifiers and implication are dealt with using the machinery of natural deduction's ⇒-intro and ∀-elim rules:

$$\frac{\Gamma, P \vdash Q}{\Gamma \vdash P \Rightarrow Q} \Rightarrow\text{-intro} \qquad \frac{\Gamma \vdash \forall x.\ P(x)}{\Gamma \vdash P(a)} \forall\text{-elim}$$

Here the suggestion is that there ought to be two kinds of critical individual: constants and arbitrary constants. Constants come initially from an existential quantifier, e.g., from the premise *some A are B* there is an individual, i, such that $A(i)$ and $B(i)$. Arbitrary constants come from an assumption, which comes from a universally quantified statement. Take the proof of

$$\frac{\begin{array}{c}\text{All } A \text{ are } B\\ \text{All } B \text{ are } C\end{array}}{\text{All } A \text{ are } C}$$

The argument goes: *suppose there were an arbitrary constant* a such that $A(a)$, then $B(a)$ by the first premise and $C(a)$ by the second premise. The critical individual a represents any constant, possibly *no* constant, if we do not assume existential presupposition. The individual description is $A(a) \wedge B(a) \wedge C(a)$. However there is now not such an obvious mapping to Euler Circles, as the authors would desire.

3.1.2.4 Type Theoretical

Another approach closely related to natural deduction above is type theoretical, here described in Martin-Löf's framework (e.g., Martin-Löf, 1984). The Curry-Howard correspondence shows that proofs may be viewed as terms (programs, functions) and propositions may be viewed as types (see Girard, Taylor, & Lafont, 1989/2003). A proof of $A \wedge B \Rightarrow B$

[9] See Politzer and Mercier (In press) for a discussion of the history.

corresponds to a function of type $A \times B \rightarrow B$, i.e., a function which extracts the second element from a tuple. A proof of $A \Rightarrow A$ corresponds to a function of type $A \rightarrow A$, for instance the identity function.

In type theory, Σ corresponds to $\exists$ and Π corresponds to $\forall$. The description by Ranta (1994, Ch. 3) of formalisations for quantifiers in English suggests the following interpretation for syllogisms:

All A are B	$\rightsquigarrow$	$\Pi x \in A.\ B(x)$
No A are B	$\rightsquigarrow$	$\Pi x \in A.\ B(x) \rightarrow \bot$
Some A are B	$\rightsquigarrow$	$\Sigma x \in A.\ B(x)$
Some A are not B	$\rightsquigarrow$	$\Sigma x \in A.\ B(x) \rightarrow \bot$

where $A \in \star$, $B \in A \rightarrow \star$, and $\star$ is the class of all types.[10] This formalisation is appealing since it retains the asymmetry of the original surface form which is especially important for existential quantification ($\Sigma x \in A.\ B(x)$ versus $\exists x.\ A(x) \wedge B(x)$). (The treatment of negation is not perfect but could easily be improved by defining $\neg\phi$ as an abbreviation of $\phi \rightarrow \bot$ or adding rules for a primitive for negation. For now we shall make do with this interpretation.)

Example 3.6 *The following premise pair has no valid syllogistic conclusion:*

Some A are B	$\rightsquigarrow$	$\Sigma x \in A.\ B(x)$
Some B are C	$\rightsquigarrow$	$\Sigma x \in B.\ C(x)$

It is instructive to see why. The two premises give proof objects $\langle a \in A, p_1 \in B(a)\rangle$ and $\langle b \in B, p_2 \in C(b)\rangle$, i.e. we have an A which we know is a B and we have a B which we know is a C. Unfortunately we don't have a proof that $a = b$ so the end terms cannot be connected.

Example 3.7 *These universals have the valid conclusion, a proof object of type $\Pi x \in A.\ C(x)$, i.e. All A are C:*

All A are B	$\rightsquigarrow$	$\Pi x \in A.\ B(x)$
All B are C	$\rightsquigarrow$	$\Pi x \in B.\ C(x)$

The premises give two proof objects $\lambda x.\ b(x)$ and $\lambda x.\ c(x)$. We would like the object $\lambda x.\ c(x) \in \Pi x \in A.\ C(x)$. Take an arbitrary $a \in A$. This fed into $\lambda x.\ b(x)$ gives a

[10] I am using slightly different notation to Ranta (1994).

proof object, $b(a) \in B(a)$. Now take the same a, which we now know proves $B(a)$ and give to $\lambda x.\ c(x)$. This gives a proof $a \in A \vdash c(a) \in C(a)$ which can be used to introduce $\vdash \lambda x.\ c(x) \in \Pi x \in A.\ C(x)$.

The logic differs from classical logic above in that a proof of a disjunction, $\varphi \vee \psi$ is a proof of one of the disjuncts (classically the disjunction may be provable even if neither disjunct is, e.g., $\varphi \vee \neg\varphi$) and a proof of an existential statement must include a constant together with a proof that the conjecture holds for this constant (classically existential statements may be proved without this so-called 'witness' constant). This is arguably closer to the notions of disjunction and existential quantification held by people. To the best of my knowledge, this sort of calculus has not yet been used to model psychological phenomena.

3.1.2.5 Monotonicity Calculus

Hodges (2005) argues that *On the Purity of the Art of Logic* (*De Puritate Artis Logicae Tractatus Longior*), written by Walter Burley in the late 1320s, gives a treatment of monotonicity which is almost a calculus. Geurts and Slik (2005) suggest that the idea can be traced back 'in part to medieval times and in part even to Aristotle'. In this section I follow Geurts (2003), with my own notational preferences. The monotonicity profiles for syllogistic sentences are as follows:

All $A{\downarrow}$ are $B{\uparrow}$	No $A{\downarrow}$ are $B{\downarrow}$
Some $A{\uparrow}$ are $B{\uparrow}$	Some $A{\uparrow}$ are not $B{\downarrow}$

This specifies what kind of object may replace the marked term. Examples help here:

(3.1) Some men↑ [like electro]↑.

(3.2) Some people↑ [like electro]↑.

(3.3) Some people↑ [like music]↑.

Sentence (3.1) implies (3.2) since 'men' has been replaced with 'people'. The ↑ arrow indicates that the term it marks may be replaced by a superset, or something entailed by that term. Sentence (3.1) also implies (3.3) and (3.2) implies (3.3) for the same reason.

(3.4) All people↓ are people↑.

(3.5) All musicians$\downarrow$ are people$\uparrow$.

(3.6) All musicians$\downarrow$ are mammals$\uparrow$.

Sentence (3.4) is tautological. It implies (3.5) since if someone is a musician then he is a person, or taking a set interpretation, the set of musicians is a subset of, hence smaller than, the set of people. Sentence (3.5) implies (3.6) if someone is a person then she is a mammal; the set of people is included in the set of mammals.

More generally, if a term P is upward entailing (or monotone increasing), then it may be replaced by a term Q which is entailed by P, or, taking a set interpretation, is a superset of P. If a term P is downward entailing (or monotone decreasing), then it may be replaced by a term Q which entails P, or, taking a set interpretation, is a subset of P.

The same interpretation as before is used, this time with the monotonicity markers added:

$$\begin{array}{lcl}
\text{All } A\downarrow \text{ are } B\uparrow & \rightsquigarrow & A(x)\downarrow \;\Rightarrow\; B(x)\uparrow \\
\text{No } A\downarrow \text{ are } B\downarrow & \rightsquigarrow & A(x)\downarrow \;\Rightarrow \neg(B(x)\downarrow)\uparrow \\
\text{Some } A\uparrow \text{ are } B\uparrow & \rightsquigarrow & A(c)\uparrow \;\wedge\; B(c)\uparrow \\
\text{Some } A\uparrow \text{ are not } B\downarrow & \rightsquigarrow & A(c)\uparrow \;\wedge\; \neg(B(c)\downarrow)\uparrow
\end{array}$$

The rules required for syllogistic reasoning, expressed in a natural deduction notation, are in Figure 3.4.

$$\frac{P \wedge Q}{Q \wedge P}\,\wedge\text{-symm} \qquad \frac{P \Rightarrow \neg Q}{Q \Rightarrow \neg P}\,\text{no-symm}$$

$$\frac{P(x) \Rightarrow Q(x) \qquad P(y)\uparrow \diamond R(y)}{Q(y)\uparrow \diamond R(y)} \quad \frac{P(x) \Rightarrow Q(x) \qquad R(y) \diamond P(y)\uparrow}{R(y) \diamond Q(y)\uparrow}\,\text{Mon-}\uparrow$$

$$\frac{Q(x) \Rightarrow P(x) \qquad P(y)\downarrow \diamond R(y)}{Q(y)\downarrow \diamond R(y)} \quad \frac{Q(x) \Rightarrow P(x) \qquad R(y) \diamond P(y)\downarrow}{R(y) \diamond Q(y)\downarrow}\,\text{Mon-}\downarrow$$

Figure 3.4: Rules for symmetry and monotonicity (derived from Geurts, 2003), where $\diamond$ is either $\Rightarrow$ or $\wedge$, x is free, and y is free or a constant.

Here is a proof of $A(x) \Rightarrow \neg B(x), B(c) \wedge C(c) \vdash C(c) \wedge \neg A(c)$.

$$\dfrac{\dfrac{\dfrac{}{A(x)\downarrow \;\Rightarrow\; \neg(B(x)\downarrow)\uparrow}\,\text{ID}}{B(x)\downarrow \;\Rightarrow\; \neg(A(x)\downarrow)\uparrow}\,\text{no-symm} \qquad \dfrac{\dfrac{}{B(c)\uparrow \;\wedge\; C(c)\uparrow}\,\text{ID}}{C(c)\uparrow \;\wedge\; B(c)\uparrow}\,\wedge\text{-symm}}{C(c)\uparrow \;\wedge\; \neg(A(c)\downarrow)\uparrow}\,\text{Mon-}\uparrow$$

This logic gives a better handling of the morphology of *no*, however it does not deal with contextual information, and as yet the process model appears ad-hoc and underspecified (see Newstead, 2003).

3.1.3 Probabilistic theories

Syllogistic inference has been modelled using probability theory (which may be viewed as a logic; see Paris, 1995). According to the Probability Heuristics Model (Chater & Oaksford, 1999; Oaksford & Chater, 2007), the following mapping is made from the language of the syllogism into the language of probability.

All A are B	$\rightsquigarrow$	$\mathsf{Pr}(B\|A) = 1$
No A are B	$\rightsquigarrow$	$\mathsf{Pr}(B\|A) = 0$
Some A are B	$\rightsquigarrow$	$(\mathsf{Pr}(B\|A) > 0) \wedge (\exists x.\ A(x) \wedge B(x))$
Some A are not B	$\rightsquigarrow$	$(\mathsf{Pr}(B\|A) > 0) \wedge (\exists x.\ A(x) \wedge \neg B(x))$

(Their analysis covers non-Aristotelian quantifiers, which I omit here for simplicity.) Most of the work here is done using conditional probabilities, but note how the existential statements (*some* and *some... not*) include a conjunction with the standard classical logical formalisation (e.g., see Section 3.1.1.1). The model includes a variety of heuristics which model the modal inferences drawn in the task. Key is the authors' analysis of informativeness (Chater & Oaksford, 1999, Appendix A), inversely related to a statement's probability, which provides the following ordering on the quantifiers:

$$\textit{all} \succ \textit{some} \succ \textit{no} \succ \textit{some}\ldots\textit{not}$$

This depends on the assumption that it's extremely likely that for a given A and B, *some A are not B* is true. The premise with the most informative quantifier is named the *max-premise* and that with the least informative is the *min-premise*. Then there are three *generative heuristics* for drawing inferences (Oaksford & Chater, 2007, p. 217).

Definition 3.1 (min-heuristic) *Find the premise with the lease informative premise and use this quantifier for the conclusion.*

Definition 3.2 (p-entailment) *The conclusion quantifier next favoured is the* ***p-entailment*** *of the conclusion predicted by the min-heuristic. With Aristotelian quantifiers, these are: all $\Rightarrow$ some, no $\Rightarrow$ some... not, some $\Rightarrow$ some... not, and some... not $\Rightarrow$ some.*

Definition 3.3 (attachment) *If only one of the premises has a subject that could be a possible subject in a conclusion, e.g., an A or a C, then choose that subject.*

Finally there are two *test heuristics* (Oaksford & Chater, 2007, p. 219) that according to the theory some participants apply.

Definition 3.4 (max-heuristic) *Confidence in a conclusion generated by the generative heuristics is proportional to the informativeness of the max-premise.*

Definition 3.5 (some... not-heuristic) *Avoid drawing some... not conclusions.*

An example (from Oaksford & Chater, 2007, pp. 217–218) will help explain this:

Example 3.8 *Take the sentences:*

all B are A
some C are B

The min-premise is some C are B, so by the min-heuristic the conclusion should have the some quantifier. Term order is decided by the attachment heuristic which says to take C are the subject, thus the conclusion is some C are A.

In terms of explaining individual differences, Oaksford and Chater (2007, p. 251) argue these could arise from differences in the heuristics, for instance weighting given to the max-heuristic, the 'sophistication' of the attachment-heuristic. Assumptions about the size of the domain may also affect the informativeness ordering.

3.2 Grey, my dear friend, is all theory...

Readers of a linear persuasion have now been treated with a second dose of theories. What should we do with these mathematical theories of reason? As with the theories of the autism spectrum gradient surveyed in Chapter 2, we ought to discuss just how distinct the theories are, for instance to what extent they may be distinguished empirically.

One issue relates to the distinction between syntax (often associated with proof theory) and semantics (often associated with model theory). On this, J. P. Burgess (2008) writes:

> ...there can hardly be any question that what 'semantics' conveyed and conveys to the mind of the general reader is a theory of meaning, which Tarski's theory most emphatically was not. By calling his theory 'semantics,' Tarski opened the door to endless misunderstandings on this point. There has been significant damage to logic arising from such misunderstandings, from confusion of model theory or 'semantics' improperly so-called with meaning theory or 'semantics' properly so-called.

One need not look far to see an example of this in the psychology of reasoning. Take Johnson-Laird, Byrne, and Schaeken (1992), for instance, an arbitrary (but small) set of individuals from the mental models school, who argue (p. 419) that:

> the underlying deductive machinery depends not on syntactic processes that use formal rules but on semantic procedures that manipulate mental models. This theory is in part inspired by the model-theoretic approach to logic.

This is misleading. For instance the theory may be embedded in a rules system using disjunctive normal forms (Evans, Over, & Handley, 2005), i.e., formulae of the form

$$
\begin{array}{ll}
 & (P_{1,1} \wedge P_{1,2} \wedge \ldots \wedge P_{1,n_1}) \\
\vee & (P_{2,1} \wedge P_{2,2} \wedge \ldots \wedge P_{2,n_2}) \\
\ldots & \\
\vee & (P_{m,1} \wedge P_{m,2} \wedge \ldots \wedge P_{m,n_m})
\end{array}
$$

where $P_{i,j}$, is either an atom or the negation of an atom. Proofs, using formal rules, may be used to describe semantics, e.g., as shown by Heyting semantics (see, e.g., Girard et al., 1989/2003, pp. 4–7). Stenning and Oberlander (1995) show that mental models are equivalent to a graphical system of logic based on Euler Circles.

The Probability Heuristics Model (PHM) defines an ordering of informativeness which was derived from a Bayesian analysis. This ordering may also be obtained by the enumeration of first-order models outlined in Appendix D (pp. 259 ff.). Ignore syllogisms which are classically valid, and those whose conclusions are false in all models of the premises, and focus on those for which the conclusion is true in at least one but not all models. These syllogisms show the ordering where those with most true models (least informative) have the conclusion quantifier *some* or *some...not*, and those with the fewest true models (most informative) have the conclusion quantifier *all* or *no*. Also some features of the model

$(A$	$\Rightarrow$	$B)$	$\vee$	$(B$	$\Rightarrow$	$A)$
T	*T*	T	**T**	T	*T*	T
T	*F*	F	**T**	F	*T*	T
F	*T*	T	**T**	T	*F*	F
F	*T*	F	**T**	F	*T*	F

Figure 3.5: A proof by truth-table of $\models (A \Rightarrow B) \vee (B \Rightarrow A)$. The columns in italics show the truth values for the two implications in the four models and the column in bold shows the truth values for the overall disjunction.

appear close to a mental logic approach, e.g., *p-entailment*, with rules like *all* $\Rightarrow$ *some*. Attachment gives behaviour very similar to that resulting from the source-founding model.

Classical logic often receives criticism from psychology of reasoning theorists, but as Makinson (2005, p. 13) reassures us, writing in a book about nonmonotonic logic, 'there is nothing wrong with classical logic.' Indeed it is used to reason about other non-classical logics. Here for the sake of completeness, I give an example of the sort of bashing classical logic receives. Take the sentence:

> If it's raining, I bring an umbrella or if I bring an umbrella, it's raining.

Suppose the conditional 'if A then B' is interpreted as the material conditional $A \Rightarrow B$, then the sentence above might be expected to be given the interpretation $(A \Rightarrow B) \vee (B \Rightarrow A)$, a tautology (see Figure 3.5), however few would say this of the English statement above. See the chaos that ensues if we allow classical logic near psychology, the arguments often go. Sometimes somebody somewhere might actually believe the above interpretation, especially if provided with pen and paper. (And we see later in this thesis some signs of interpretations which appear consistent with classical logic.) Perhaps another mapping may be used from the natural language to classical logic (recall the remark by Quine at the beginning of this chapter). Education complicates proceedings as people can be trained to reason according to classical logic, as exam results and the existence of classical logicians testify. Classical logic is used by some people in some situations: it should not be hastily abandoned in its entirety because of an overly naive conceptualisation of interpretation.

Still, it seems that classical logic does a poor job of characterising how many people typically reason, so something must be wrong—I do not want to deny that! Translating the syllogism into classical logic directly doesn't, for instance, reveal the effect that order of presentation affects the inferences people draw (see later chapters). The dependence between sentences as information is added to the set of premises is not naturally captured by classical logic. Importantly, as Makinson (2005) shows, classical logic may be used as a good starting point for developing logics which do have the desired properties. Key is to consider the details of classical logic, what properties it has. Already we have seen an example of a minor modification to classical logic: removal of the assumption that $\varphi \vee \neg\varphi$ is true, regardless of what sentence replaces φ. It is also possible to modify the logic so that adding new premises can cause old conclusions to withdrawn. The notion of classical logical dependence and independence, i.e., whether or not a conclusion necessarily follows from a set of premises, turns out to be important when interpreting the data on so-called immediate inference.

If the material conditional $A \Rightarrow B$ is not the best interpretation of 'if A, then B', what is? Although Wason (1966) assumed the answer is the material conditional he did briefly consider the possibility (p. 146) that participants interpreted the conditional as having three truth values: true, false, and irrelevant. A number of researchers have argued that the correct mapping is to the conditional probability $\mathsf{Pr}(B|A)$ (Over, Hadjichristidis, Evans, Handley, & Sloman, 2007; Pfeifer & Kleiter, In press). As I will argue in the next section, searching for one true interpretation of 'if A, then B' or any other fragment of natural language is arguably the wrong quest as the evidence indicates that the interpretation depends on the context and traits of people doing the interpreting.

3.3 Embracing qualitative individual differences

> I'm a reasonable man
> Get off, get off, get off my case
> I'm a reasonable man
> Get off my case, get off my case
>
> From *Packt Like Sardines in a Crushd Tin Box* by Radiohead

I might be wrong
I might be wrong

From *I Might Be Wrong* by Radiohead

When studying individual differences in reasoning typically the focus is on how good people are at a task according to one particular competence model, and correlations with performance on other tasks, distressingly often without any detailed characterisation of the processes that may be enabling the performance. A relatively more recent trend is to accept that people are attempting to and (sometimes) succeeding at doing different things (cf. the competence versus performance distinction above) and to try to model this. The challenge is how to make sense of different strategies with respect to the various cognitive theories derived from the logical literature. In the present section I shall highlight the different conceptual and statistical models that researchers have used to comprehend the empirical data.

A series of studies have examined the relationships between external representation use and the inferences people draw. Such approaches to studying cognition have a long history stretching back (at least) to Luquet (1927/2001) who studied his children's drawings to infer properties of cognitive representations. A little over a decade ago, Ford (1995) performed a heroic study of syllogistic reasoning (the article describing which weighs in at 71 pages) in twenty participants. They were asked to solve the problems twice: once when speaking aloud, and they could use scribbles if they so desired; a second time with scribbles emphasising the explanation of what they were doing. The results were striking. Around half the participants drew diagrams resembling Euler Circles (see Stapleton, Masthoff, Flower, Fish, & Southern, 2007, for a thorough mathematical treatment); the other half attempted a form of algebraic rewriting. Ford found that classical correctness could be used for some problems to predict whether someone had used an algebraic or graphical method of solution; also that the term-order of responses could, for some problems, reveal strategy used. Interestingly it appeared that no participants tried to represent (at least in their scribbles) individuals of the classes to be reasoned about, as might be predicted by the mental models account, instead focusing on the classes themselves.

Bacon, Handley, and Newstead (2003) attempted to replicate Ford's study in around fifty participants. They again found people tended to use different strategies, and that there were differences in the inferences people drew for some problems. In terms of classical accuracy, the direction and magnitude of differences between graphical and algebraic reasoners were not always consistent between Ford and Bacon et al's studies. For instance Bacon et al's algebraic reasoners were more likely to use a naive strategy of substitution than were Ford's algebraic reasoners. Perhaps a reason for inconsistencies is overall level of ability: Ford (1995) used students at Stanford University whereas Bacon et al. (2003) used students at the University of Plymouth, which has weaker selection criteria.[11] This suggests that the messy details considered in both these works need to be considered to do justice to the different cognitive processes driving the use of external representations. In another study using the same methodology and Plymouth students, Bacon, Handley, and McDonald (2007) investigated whether people with dyslexia would be more likely to use a graphical than an algebraic strategy. As predicted, they were: around 65% of the group which used a graphical strategy had dyslexia; only around 27% of the algebraic group had dyslexia.

There is extensive research indicating that linguistic pragmatic factors influence how people reason with quantifiers. Take the following simple 'immediate inference' problem: whether *some A are not B* follows from *some A are B*. Classically this conclusion is true in some situations but false when *all A are B*. Most listeners assume that speakers would say *all A are B* if they believed this were true and many reasoners in the lab also draw this inference (e.g., Begg & Harris, 1982; Newstead & Griggs, 1983; Politzer, Henst, Luche, & Noveck, 2006). This is an instance of assumed cooperativeness. In the language of Grice (1975), *some A are not B* is *implicated*. See Figure 3.6 for a list of Gricean conversational maxims, thought by Grice to be implicit standards of conversation understood by people.

Given that immediate inference is one-premise reasoning about Aristotelian quantifiers and categorical syllogisms require reasoning about two-premises containing Aristotelian quantifiers, one might expect evidence of homogeneity in interpretations, specifically in terms of Gricean implicature. Newstead (1995) argued that although there is plenty of Gricean interpretation in immediate inference, there was little sign in syllogisms. By 'Gricean' he meant specifically interpreting 'some' as 'some but not all'. Stenning and Cox (2006) took a broader view, classifying responses to immediate inference according to deviations from classical logic, and inferences in syllogisms according to how the conclusion

[11] Recall the comments made in Section 2.2.3 about cognitive style and its relations to ability.

1. Quantity
    (a) Make your contribution as informative as is required (for the current purposes of the exchange).
    (b) Do not make your contribution more informative than is required.
2. Quality: Try to make your contribution one that is true.
    (a) Do not say what you believe to be false.
    (b) Do not say that for which you lack adequate evidence.
3. Relation: be relevant.
4. Manner: be perspicuous.
    (a) Avoid obscurity of expression.
    (b) Avoid ambiguity.
    (c) Be brief (avoid unnecessary prolixity).
    (d) Be orderly.

Figure 3.6: Paul Grice's conversational maxims.

term-order could be predicted by the arrangement of terms in the premises (i.e., the figure), quantifiers present, and also the degree to which the premise ordering affected the resulting term-order. These two characterisations were connected using logistic regression, with the key predictors being interaction terms modelling how immediate inference classifications moderate the modal response term-order to syllogisms. For now it suffices to say that interactions were found, providing evidence of the existence of homogenous interpretative processes, some recognisably Gricean and some not. (See Section 10.1.2, pp. 220 ff., for the full details.)

Individual differences in conditional reasoning have also been studied. Suppose that an English sentence 'if A, then B' is translated into the classical logic material conditional, $A \Rightarrow B$, then the rules shown in Figure 3.7 apply. A long series of research has shown that people tend not to be consistent with classical logic (given the interpretation above), for instance they often fail to draw MT but do draw DA and AC.

Of particular importance for the study of reasoning traits is to what degree responses to particular items are due to state effects as opposed to longer-term trait effects. Bonnefon,

$$\frac{P \qquad P \Rightarrow Q}{Q}\ \textsf{MP (classically valid)} \qquad \frac{P \Rightarrow Q \qquad \neg Q}{\neg P}\ \textsf{MT (classically valid)}$$

$$\frac{\neg P \qquad P \Rightarrow Q}{Q}\ \textsf{DA (classically invalid)} \qquad \frac{P \Rightarrow Q \qquad Q}{\neg P}\ \textsf{AC (classically invalid)}$$

Figure 3.7: Valid inferences (modus ponens [MP] and modus tollens [MT]) and invalid inferences (denial of the antecedent [DA] and affirmation of the consequent [AC]) in classical propositional logic

Vautier, and Eid (2007) investigated this for MT problems. They used six conditionals including 'If there is a low pressure system, then it will rain' and 'If a company makes a big profit, then the price of their shares will go up'. Participants solved each problem twice, separated by around three weeks. The authors modelled classical logic accuracy using a multistate-multitrate latent-variable model, which made it possible to partition state- and trait-variance. They found the test-retest reliabilities of the variables ranged from around .6 to .9. The lowest was for 'If there is a low pressure system, then it will rain'—perhaps the interpretation depends on the present weather!—and the highest for 'If someone has broken an item in the store, then they must pay for it'. The consistences were all high, all but one over .9. They argue that this shows the items were tapping into traits and not exposing state effects.

The authors also argue that there is evidence that more than one trait is being measured as the trait scores show a range of correlations, at their lowest, .27. The authors do not elaborate further on this comment. Examining the correlations provided in the article, it appears they hint at the detection of different kinds of interpretation. For instance, the lowest correlation was between responses for the following two items:

(3.7) If a company makes a big profit, then the price of their shares will go up.

(3.8) If a restaurant sells liquor, then it must have a liquor license.

Sentence (3.7) is rather law like, descriptive, and sentence (3.8) is deontic. It is known that people tend to be more consistent with classical logic in responses to deontic MT (e.g., Schaeken & d'Ydewalle, 1996). I used exploratory factor analysis (with varimax

	Factor 1	*Factor 2*
If the content of the bottle is poisonous, then it must be labeled 'poison'	**.84**	.22
If someone has broken an item in the store, then they must pay for it	**.69**	.17
If a restaurant sells liquor, then it must have a liquor license	**.66**	.17
If a patient has malaria, he makes a quick recovery	.53	.27
If a company makes a big profit, then the price of their shares will go up	.16	**.99**
If there is a low pressure system, then it will rain	.36	**.51**

Table 3.3: Loadings from an exploratory factor analysis (with varimax rotation) of the correlation matrix provided by Bonnefon, Vautier, and Eid (2007, p. 1411).

rotation) to summarise the correlations. See Table 3.3 for the loadings. Factor 1 appears to represent deontic conditionals (if one ignores the blemish of the malaria sentence) and factor 2 descriptive.

Building on work by Rumain, Connell, and Braine (1983), Rijmen and Boeck (2003) modelled interpretations in the inferences of high school students. They hypothesised that the following three types of reasoner exist:

1. *Unsophisticated reasoners.* People who draw MT because they have accepted the invited inferences, i.e., interpreting 'If you mow the lawn, you'll receive five dollars' as the biconditional (if you don't mow the lawn then you won't).

2. *Intermediate reasoners.* Do not draw AC and DA but fail to draw MT 'because they don't master the reductio strategy', i.e., the strategy of assuming the opposite of the proposition to be proved and showing that doing so leads to a contradiction.

3. *Sophisticated reasoners.* Resist invited inferences and draw MT.

Using a combination of truth tables and MP, DA, MT, AC integrated with disjunction, conjunction, and disjunctive modus ponens (given $P_1 \vee P_2$ and $\neg P_1$ then P_2; $P_1 \vee P_2$ and $\neg P_2$ then P_1), they used latent class analysis in an attempt to discover the various groups. Their final model consisted of three classes: first was the unsophisticated and second the intermediate reasoners. The second class had higher overall classical propositional reasoning ability, but had more difficulty with MT. The third class did not correspond to their hypothesised 'sophisticated reasoners' and was too small to say much more about.

Bonnefon, Eid, Vautier, and Jmel (2007) building on this work set out to classify kinds of reasoner using a combination of item-response theory and latent class analysis, according to the answers given to a series of conditional logic problems testing MT, DA, and AC. Although the statistical model was setup to predict classical correctness, the responses were interpreted with respect to four kinds of response. They argue (p. 2):

> Reasoners cannot simply be ordered on an ability continuum, but they have to be qualitatively compared with respect to their response process, that is, with respect to the reasoning subsystem that underlies their answers.

A brief foray into dual process theories of reasoning is now required (see, e.g, Evans, 2008). In these theories, System 1 is seen as fast, operating outside of conscious awareness, feeling effortless, and independent of working memory. System 2 is seen as slower, operating with conscious awareness of its operation, feeling more effortful, and requiring working memory. These two systems have often been used to organise interpretations of reasoning performance. The authors divide up reasoning systems in the following manner:

1. *System 1 Pragmatic.* The argument here is that the conditional is interpreted as $P \Rightarrow Q$ and invites $\neg P \Rightarrow \neg Q$ and $Q \Rightarrow P$. This encourages AC and DA, but not MT: perfect classically incorrectness.

2. *System 1 Semantic.* These reasoners' responses will be depend more on background knowledge. Empirically, the authors argue, this will be shown as less-frequent drawing of AC and DA, depending on background knowledge accessed, than the System 1 Pragmatic group, but more frequent drawing of MT.

3. *System 2 Inhibitory.* This group will be able to inhibit background knowledge about pragmatic and semantic aspects of the sentences to be reasoned about, so, the authors argue, will tend to draw fewer DA and AC inferences. However they also argue that not all reasoners will be sophisticated enough to draw MT which requires 'an abstract strategy for reductio ad absurdum'.

4. *System 2 Generative.* These are 'the most sophisticated reasoners' and will get all problems correct with respect to classical logic.

The classifications are difficult to comprehend. Presumably pragmatic inference cannot take place in a semantic vacuum. Along with Rijmen and Boeck (2003), Bonnefon, Eid, et al. (2007) argue that MT requires reductio ad absurdum (RAA), which they claim is a difficult rule.[12] The authors do not point out that although MT may be proved using RAA it also has a proof which does not require this rule. Write $\neg Q$ as $Q \Rightarrow \bot$, where $\bot$ is the always false proposition. The goal is to prove $\neg P$, represented as $P \Rightarrow \bot$. This is proved by assuming P; then Q follows by modus ponens from $P \Rightarrow Q$, and $\bot$ follows by modus ponens from $Q \Rightarrow \bot$. Discharging the assumption allows us to conclude $P \Rightarrow \bot$, i.e., $\neg P$. There are other ways of doing this. The characterisation of MT being more difficult for 'Intermediate' or 'System 1 Pragmatic' reasoners than DA and AC depends on how many inference rules would be required to get to MT from the invited inferences. Clearly MT would still be derivable.

The authors argue that their data support three classes of reasoner: a mix of pragmatic reasoners and generative reasoners; semantic reasoners; and inhibitory reasoners. Again there are hints of effects depending on the different conditional types. The conditional 'If the content of the bottle is poisonous, it must be labelled "poison"' produces more MT-like responses than the other conditionals, so clearly how a sentence is interpreted influences the ease with which MT may be drawn: this seems to make the availability of RAA irrelevant.

Some researchers have attempted to connect qualitative individual differences in reasoning to processes often thought to involve deductive reasoning. Klaczynski and Daniel (2005) investigated individual differences in conditional reasoning (again, MP, MT, DA, AC) as a function of the PMA Verbal Meaning test (which requires participants to select

[12] RAA is a classical but not constructive inference rule (see e.g., Dalen, 1997, pp. 157–159).

which of four words means the same as the target) and a self-report measure of decontextualised thinking. They selected conditionals which were casual and deontic, and which have weak and strong alternatives (i.e., there are few or many alternative antecedents which imply the consequent; more on this in Chapter 7, pp. 139 ff.). As well as responding with 'yes', 'no', or 'can't be certain' to questions testing each of the four inferences, participants were asked to provide justifications for their choices.

Here I focus on results related to individual differences. Low and high scorers on the verbal meaning test were more likely to draw MT than were medium scorers—echoing results reported above. Low and medium verbal scorers were more likely to draw AC for weak alternative than strong alternative problems. MP and MT were positively correlated with greater decontextualised thinking on all problems. MT with weak alternatives was positively correlated with verbal ability. Decontextualised thinking and verbal ability predicted more AC and DA inferences for weak alternatives than strong alternatives. In summary, a range of individual differences were found in the details of the inferences that people draw.

Evans, Handley, Neilens, and Over (2007) investigated individual differences in conditional reasoning tasks in 120 students, using the Alice Heim 4 (AH4) test to estimate average performance in a range of cognitive abilities. Splitting ability into high and low, there was no difference in the proportion of participants drawing MP or MT, but there was a difference for DA and AC with these inferences drawn less frequently in the high than the low group (DA: 34% versus 50%; AC: 64% versus 78%). People who interpreted the conditional as $\mathsf{Pr}(B|A)$ rather than $\mathsf{Pr}(A \wedge B)$ tended to have slightly higher scores on AH4 (mean of 98 versus 92; note that the score does not have a mean of 100 and *SD* of 15 as common in IQ tests). Also those who interpreted the conditional as $\mathsf{Pr}(B|A)$ were less likely to draw AC, but there was no difference for the other inference types.

Slugoski and Wilson (1998) investigated the relationships between communicative competence and performance on a series of reasoning tasks. Communicative competence was assessed using the 'method of reconstruction', which consists of giving participants a series of randomised cards containing fragments of conversation; the task is to order the cards in such a way as to reconstruct the conversation. Reasoning tasks were tests of: 'confirmation bias', Wason's selection task; a primacy effect task which consisted of a description of a person, the beginning of which had positive adjectives, the end of which had negative adjectives; a consensus-effect task; a 'conjunction fallacy' task, the Bill and Linda problems; a test of the dilution effect which determines how non-diagnostic but information about stereotypes affects inferences drawn about a person described; and finally a test of the degree to which participants used base-rate information. Greater pragmatic competence as determined by the card sort was associated with more confirmation and consensus bias, and less sensitivity to base rate information. Poorer pragmatic competence was associated with more of an effect of primacy, conjunction, and dilution. Sometimes biases are associated with better communicative competence, sometimes poorer.

3.4 Summary

This chapter introduced some details of logics that have been used to inspire theories of reasoning, using the syllogism as a case study for the various approaches. An important message I wanted to convey is that although different theories may look superficially different, they may also be empirically indistinguishable. Logics provide ideas for properties of inferences; they ought not to be treated as exam solution sheets. It is necessary to choose an appropriate level of abstraction and to focus on claims that may be tested. The well known issue of interpretation when applying logical models has been explained and evidence of qualitative individual differences in interpretation, some related to social and pragmatic competence, others related to more general cognitive abilities, has been reviewed. The majority of the individual differences work has focused on deviations from classical competence, as opposed to more positive characterisations of alternative logical processes, and assumed a single derivational theory with variation in interpretational processes. Clearly work on the characterisations of interpretation is still in its early stages. In this thesis I aim to contribute to the work.

Chapter 4

Study Design and Participants

> design, *n.* (1b) A scheme formed to the detriment of another.
>
> OED, 2nd ed., 1989

This thesis aims to investigate individual differences in reasoning *to* and *from* interpretations (Stenning & van Lambalgen, 2004, 2005, 2008). There are two main questions addressed:

1. How do qualitative individual differences in the inferences people draw in deductive reasoning tasks relate to autistic traits?

2. To what extent are individual differences in inference due to individual differences in interpretation?

Self-report measures of autistic traits take the role of reconnecting deductive reasoning to activities that matter to people outside the lab, for instance interacting with others, understanding language, making plans. The extent to which any differences found are interpretational will be investigated by examining consistency across task within individuals. Note how there is a symbiotic relationship between these questions, which is most easily seen for the social and communicative aspects of autistic traits. For instance one might expect participants who report being particularly adept at communicating to have different cognitive goals than those who don't, perhaps being more consistent with Grice's maxims.

A guiding principle in the explorations to come is the distinction between *credulous* and *sceptical* interpretation (see e.g., Stenning & van Lambalgen, 2005, p. 922). Credulous interpretation is the cooperative act of trying to make sense of the speaker's intended model, using shared context. Sceptical interpretation is an adversarial strategy, for instance that of finding models of the premises which are counter models of the conclusion, or drawing conclusions which minimise the possibility that a countermodel will be found.[1]

This chapter gives an overview of the design of the data collection stage of the study, describes our participants, and outlines some of the statistical machinery to come later.

4.1 Selecting the tasks

I aimed to cover four main task types in the data collection. Firstly some measure of autistic traits was required. Secondly, reasoning tasks which had previously been found to pick up differences between typically developing individuals and those with ASC were required. These were easy to choose as when I began there was one: van Lambalgen and Smid (2004) found differences for the suppression task, a test of conditional reasoning. Thirdly, I also required deductive reasoning tasks which showed differences in responses interpretable as being due to differences in interpretation. Finally, since it appears a range of abilities may be related to reasoning, a task known to load on g, the general factor in intelligence, was required. This section introduces the tasks.

4.1.1 Self-report measures of autistic traits

The Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, et al., 2001) was chosen as it is the most studied measure of autistic-like traits in typically developing populations (see Section 2.4.3). Another questionnaire, the Broad Autism Phenotype Questionnaire (BAPQ; Hurley, Losh, Parlier, Reznick, & Piven, 2007), was also selected as this was developed independently and it was thought it might be interesting to compare results.

There are 50 statements in AQ, divided into five theoretically defined dimensions. Participants are asked to indicate the degree to which they agree or disagree with each

[1] The terms 'credulous' and 'sceptical' have specific connotations in the field of non-monotonic logic, but I mean them more generally here.

statement on a four-point scale: 'definitely agree', 'slightly agree', 'slightly disagree', or 'definitely disagree'. The items were originally given a binary score by Baron-Cohen, Wheelwright, Skinner, et al. (2001), ignoring the degree of agreement or disagreement, and then summed; later studies (e.g., Austin, 2005; Hoekstra et al., 2007; Ring, Woodbury-Smith, Watson, Wheelwright, & Baron-Cohen, 2008) include all four levels in the sum-score. Example questions in each of these dimensions are as follows:

- *Social skill*, e.g., 'I prefer to do things with others rather than on my own' and 'I find social situations easy'.
- *Attention switching*, e.g., 'I prefer to do things the same way over and over again' and 'I frequently get so strongly absorbed in one thing that I lose sight of other things'.
- *Attention to detail*, e.g., 'I often notice small sounds when others do not' and 'I am fascinated by dates'.
- *Communication*, e.g., 'Other people frequently tell me that what I've said is impolite, even though I think it is polite' and 'I enjoy social chit-chat'.
- *Imagination*, e.g., 'If I try to imagine something, I find it very easy to create a picture in my mind' and 'I find making up stories easy'.

Hurley et al. (2007) recently developed the Broad Autism Phenotype Questionnaire (BAPQ) to study the genetics of ASC, e.g., in unaffected relatives of children with ASC, drawing on the authors' experience with assessment interviews. Their questionnaire consists of 36 statements, divided into three dimensions: aloof personality ('a lack of interest or enjoyment of social interaction'), rigid personality ('little interest in change or difficulty adjusting to change'), and pragmatic language problems ('deficits in the social aspects of language, resulting in difficulties communicating effectively or in holding a fluid, reciprocal conversation'). Participants are asked to respond with how often they consider each statement is true of them, with responses on a six-point scale: 'very often', 'often', 'somewhat often', 'occasionally', 'rarely', or 'very rarely'. Questions in each dimensions include:

- *Aloofness*, e.g., 'I like being around other people' and 'People find it easy to approach me'.
- *Pragmatic language ability*, e.g., 'It's hard for me to avoid getting sidetracked in conversation' and 'I am "in-tune" with the other person during conversation'.
- *Rigidity*, e.g., 'I am comfortable with unexpected changes in plans' and 'People have to talk me into trying something new'.

For both AQ and BAPQ, a higher score indicates the presence of more autistic-like traits. Chapter 5 (pp. 87 ff.) studies the properties of responses to these questionnaires in detail. See Section B.1 (pp. 253 ff.) for a complete list of all AQ items and Section B.2 (pp. 255 ff.) for a complete list of BAPQ items.

4.1.2 Reasoning

Reasoning tasks were implemented using the E-prime package (Schneider, Eschmann, & Zuccolotto, 2002). Testing took place over two sessions, the first consisting just of syllogisms and the second the remainder of the tasks. At the beginning of both sessions, participants were presented with the text:

> The questions that follow can be interpreted in different ways, and there is more than one correct answer. We are interested in how you interpret the tasks and solve them. We are very grateful for your time. Without the contribution of people like you doing these tasks we cannot make further progress in analysing thinking styles.

4.1.2.1 Categorical Syllogisms

Syllogisms are arguments containing two premises and one conclusion, each of which has a single quantifier, subject, and predicate. For instance:

> All bus drivers are Austrians
> Some cat owners are not Austrians, therefore
> Some cat owners are not bus drivers

Here 'bus drivers' and 'cat owners' are the subjects, and 'Austrian' is the predicate in both premises. Various tasks are used in the psychology literature; we present the premises and ask participants what follows. A total of 64 syllogisms are generated by varying the quantifiers present (*all*, *no*, *some*, *some... not*) and the order of the terms in both premises, given by the following schema:

$$\begin{array}{cccc} Q_1\mathbf{A} \text{ are } B & Q_1\mathbf{A} \text{ are } B & Q_1 B \text{ are } \mathbf{A} & Q_1 B \text{ are } \mathbf{A} \\ Q_2 B \text{ are } \mathbf{C} & Q_2\mathbf{C} \text{ are } B & Q_2 B \text{ are } \mathbf{C} & Q_2\mathbf{C} \text{ are } B \end{array}$$

The term B in this schema is called the middle-term; A and C are the end-terms. The arrangement of terms is called the *figure*. These are often numbered, but I resist as the numbers have changed over time (see, e.g., Bobzien, 2000) and psychologists continue to adhere to different conventions.[2] For ease of comprehension we name them by the term orders, first premise first: *ABBC*, *ABCB*, *BABC*, and *BACB*. The example above has the figure *ABCB*. I often also refer to the figure as premise term-order. *Columnar* figures are those there the middle terms are both either on the left or both on the right: *ABCB* and *BABC*. *Diagonal* figures are not columnar: *ABBC* and *BACB*.

The first computer-based session contained the 64 syllogisms with forms obtained by varying quantifiers and term-orders in both premises. The order each form was displayed was chosen randomly. Before answering the syllogisms, participants were told they would see 64 pairs of premises and that they were to imagine that each gave a true description of the people at a party. Their task was to decide what else they could infer about the people at the party choosing one option from the list. The subject and predicates were filled randomly from a list (see Appendix C) with the constraints that all the initial letters differed and that term A was a profession, term B a nationality, and term C a hobby. Each item was displayed full screen, for instance for one of the problems:

3/64

Some Hungarians are stockbrokers
All Hungarians are Christians

What follows?

1 – All stockbrokers are Christians

[2] Intriguingly, Michele Abrusci has shown that a proof-net representation, a geometric method of representing proofs in an attempt to reduce syntactic bureaucracy, of syllogisms reveals the original Aristotelian notion of figure (see Girard, 2003, pp. 141–142). The number of crossings in the graph corresponds to the figure number. Girard frequently makes teasing references to an unpublished technical report by Abrusci in his writings, however sadly I have been unable to obtain the original work.

2 – All Christians are stockbrokers
3 – Some stockbrokers are Christians
4 – Some Christians are stockbroker
5 – No stockbrokers are Christians
6 – No Christians are stockbrokers
7 – Some stockbrokers are not Christians
8 – Some Christians are not stockbrokers
0 – None of the above

4.1.2.2 Immediate Inference

The immediate inference task consists conjectures of the form $Q_1(A, B) \Rightarrow Q_2(A, B)$ where the Q's are the same quantifiers as used in syllogisms (*all*, *no*, *some*, *some...not*) and the term order in the conclusion can be either AB or BA. The task, as originally intended, is to decide whether the conjecture is true, false, or could be either true or false. An example item:

Assume
Some A are B
is true

So
Some A are not B
is

1 – True
2 – False
0 – Could be True or False

4.1.2.3 The Linda Problem

Based on the problem given by Tversky and Kahneman (1982), in this task participants are given descriptions of two people and are asked to rank the statements about them, most probable first. The descriptions given were:

1. Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.

A. Linda is a bank teller
B. Linda is active in the feminist movement
C. Linda is a bank teller and is active in the feminist movement

2. Bill is 34 years old. He is intelligent, very punctual, but unimaginative, rather compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.
A. Bill plays in a rock band for a hobby
B. Bill is an accountant
C. Bill is an accountant and plays in a rock band for a hobby

(See Section G.1 for the results of performance on this task.)

4.1.2.4 The letter and numeral selection task

Based on the original task by Wason (Wason, 1966, 1968). The following instruction was given:

> Below there are four cards. Each has a letter on one side and a number on the other side. As you can see, some have the letter side showing, some have the number side showing.

Below this were four cards, showing 'A', 'K', '4', and '7'. Participants were told:

> I claim that if there is a vowel on one side then there is an even number on the other side. Type the cards you'd need to turn to work out if my claim is correct.

4.1.2.5 The suppression task

The suppression task (Byrne, 1989) attempts (from the point of view of the analysis by Stenning & van Lambalgen, 2005) to induce a non-classical interpretation of a sequence of conditionals and an atom. There are three main conditions in the task, originally delivered between-participant but here within. The simple condition consists of two premises, a conditional, *if* P *then* Q, and an atom, P or Q, positive or negated. The additional condition adds another conditional, *if* R *then* Q, where the R is a condition which could possibly block Q, even if P were true. The alternative conditional of the same surface form except here the R is another possible reason for making Q true.

The following instructions were given:

> In each of the tasks that follow, you will be given some sentences about a girl and a library. Imagine you have overheard them in different conversations.
>
> Note, they are quite similar so pay close attention! Your task is to decide what follows.

The suppression task was originally performed by Byrne (1989) as a between-participants design. Lechler (2004) performed a qualitative exploration of how people reason with the material. Here as a compromise we provided participants with a large text window into which responses were typed. We used one of the content types produced by Byrne (1989). The following twelve forms (Simple, Additional, Alternative) $\times$ (MP, MT, AC, DA) were presented in random order.

1. If she has an essay to finish then she will study late in the library
   She has an essay to finish [Simple, P, MP]
2. If she has an essay to finish then she will study late in the library
   She doesn't have an essay to finish [Simple, $\neg P$, DA]
3. If she has an essay to finish then she will study late in the library
   She will study late in the library [Simple, Q, AC]
4. If she has an essay to finish then she will study late in the library
   She will not study late in the library [Simple, $\neg Q$, MT]
5. If she has an essay to finish then she will study late in the library
   If she has some textbooks to read then she will study late in the library
   She has an essay to finish [Alternative, P, MP]
6. If she has an essay to finish then she will study late in the library
   If she has some textbooks to read then she will study late in the library
   She doesn't have an essay to finish [Alternative, $\neg P$, DA]
7. If she has an essay to finish then she will study late in the library
   If she has some textbooks to read then she will study late in the library
   She will study late in the library [Alternative, Q, AC]
8. If she has an essay to finish then she will study late in the library
   If she has some textbooks to read then she will study late in the library
   She will not study late in the library [Alternative, $\neg Q$, MT]

9. If she has an essay to finish then she will study late in the library
   If the library stays open then she will study late in the library
   She has an essay to finish [Additional, P, MP]

10. If she has an essay to finish then she will study late in the library
    If the library stays open then she will study late in the library
    She doesn't have an essay to finish [Additional, $\neg P$, DA]

11. If she has an essay to finish then she will study late in the library
    If the library stays open then she will study late in the library
    She will study late in the library [Additional, Q, AC]

12. If she has an essay to finish then she will study late in the library
    If the library stays open then she will study late in the library
    She will not study late in the library [Additional, $\neg Q$, MT]

4.1.2.6 Raven's Advanced Progressive Matrices

As we were constrained by the time available with each participant, a 20 minute computer-based version of set II of Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 1998) was used. Hamel and Schmittmann (2006) argue that the resulting score for the standard Raven's test after 20 minutes is an adequate predictor of the full test score ($r \approx 0.7$ between the full score and the score after 20 minutes). See Figure 4.1 for an item capturing the general idea of those used in the task.

The instructions were as follows:

> In the tasks that follow you will be presented with a visual pattern like [one displayed on the screen, item 1 of the set].
>
> You have to decide which of the 8 items completes the pattern by pressing the appropriate number on the keypad. The correct answer for this example is 5. The items get progressively harder. Please speak to the experimenter when they're too hard to solve—please do not guess.

The participant was not told that the test was being timed. The test terminated after either (a) 20 minutes had elapsed and the participant answered the current item or said it was too difficult; or (b) before 20 minutes had elapsed if the participant requested that the session finish.

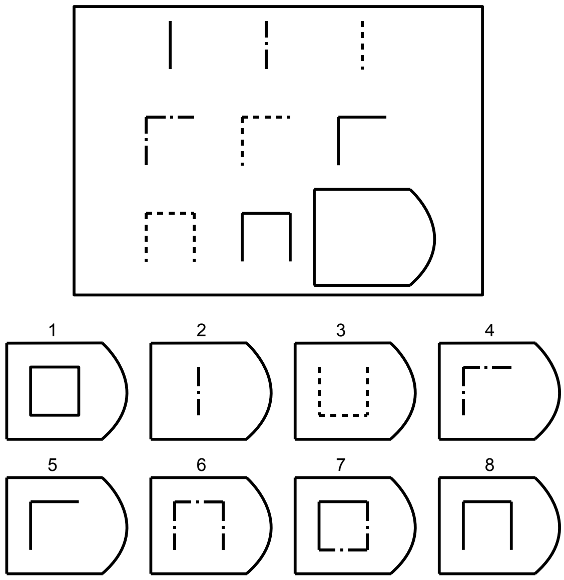

Figure 4.1: An example item analogous to those found in Raven's Progressive Matrices (redrawn from an example provided by John Raven)

4.2 Participants

The study was advertised by email to all students at the Edinburgh campus of Heriot-Watt University, except those studying computer science, mathematics and physics, who would be likely to have received an education in classical logic. All participants gave informed consent and ethical approval was obtained from the Ethics Committee of Heriot-Watt University. Questionnaires were collected and completed by participants at their leisure, after which they were returned in a sealed envelope to a school office.[3] Participants who returned questionnaires were asked to attend two computer-based sessions, each lasting at most one hour and separated by at least half-an-hour. The first computer-based session consisted of 64 categorical syllogism problems. The second session consisted of immediate inference, suppression task, selection task, the Linda problem, and the short version of

[3]This pedantry included as method of questionnaire delivery may affect degree of self-disclosure (see Bowling, 2005).

(Set II of) Raven's Advanced Progressive Matrices (RAPM). Task order was randomised, except for RAPM which came last as it is more obviously a measure of ability, and we were concerned this might have affected interpretative processes on any reasoning tasks that followed. Participants were tested in groups in a computer lab. On completion of the second computer-based session, participants were reimbursed £10 for their time.

105 participants completed the questionnaires, of whom 67 were female and 38 male (mean age = 23, *SD* = 5.1, range 17–49). For females, the mean age was 22 (*SD* = 4, range 18–38) and for males, the mean age was 24 (*SD* = 6.7, range 17–49; one male didn't record his age). Of these, 90 attended the first computer-based session, of whom 61 were female and 29 were male (mean age = 23, *SD* = 5, range 17–49). For females, the mean age was 22 (*SD* = 3.5, range 18–31) and for males, the mean age was 24 (*SD* = 7.2, range 17–49). Of these, 84 attended the second computer-based session, of whom 55 were female and 29 were male (mean age = 23, *SD* = 5.1, range 17–49). For females, the mean age was 22 (*SD* = 3.4, range 18–31) and for males, the mean age was 24 (*SD* = 7.2, range 17–49). Four participants failed to complete the syllogisms in one hour, but as the stimulus presentation was randomised, their partial responses are included in the analyses.

4.3 A brief note on model fitting and selection

In this thesis I use a variety of inferential statistical methods, especially various forms of regression, e.g., general linear mixed effects models (GLMMs), also known as multilevel models, and proportional odds logistic regression. The basic idea for each is the same: first the model structure is specified relating the variables, with some parameters left freely varying, e.g., $y = \beta_0 + \beta_1 x$, where a linear relationship is postulated to exist between y and x, and the intercept β_0 and slope β_1 are to be estimated. For linear regression, ordinary least squares may be used to estimate the parameters. Generally, no analytic solution exists and so an iterative method of maximum likelihood estimation (MLE) is used which maximises the likelihood of parameters given the observed data. Figure 4.2 gives a graphical description of three attempts to fit a univariate Gaussian distribution to simulated data with a mean of 100 and *SD* of 15. A probability density function is a way to represent a probability distribution, with larger densities where a value is particularly

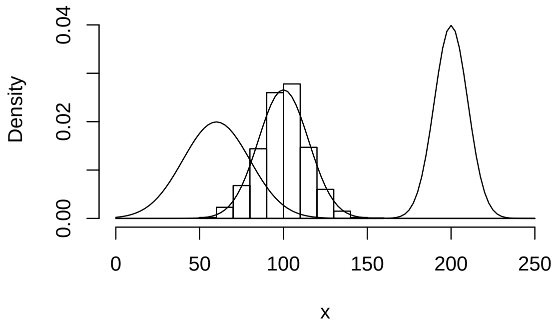

Figure 4.2: An example of three attempts to fit a univariate Gaussian distribution to a dataset. The histogram shows the distribution of the actual data and the curves show the density plots of the estimates.

likely. Take the far left parameter estimates ($M = 60$, $SD = 20$). The density at the value of 100 given this estimate is 0.0027. For the far right attempt ($M = 200$, $SD = 10$), the density at 100 is 7.7×10^{-24}. For the correct attempt in between, the density is 0.027.

The likelihood of a model's parameters given all the data points (the joint distribution) needs to be considered, which can be expressed as a product of densities (assuming independence of the points). Computationally, however, products of density estimates are troublesome as the numbers become very small, so the log is frequently taken to allow the likelihood to be expressed as an easier to manage sum (recall that $\log \prod x = \sum \log x$). For the simulated example shown in Figure 4.2, and the correct model, the largest density was 0.027 and the smallest 1.3×10^{-5}: the product quickly approaches zero. The logged densities, on the other hand, ranged from -11.2 to -3.6 and the sum was -4098. Log is monotonically increasing ($x > y \Rightarrow \log x > \log y$), so the maximum occurs at the same point. Reassuringly, the log-likelihoods for the other two incorrect estimates were smaller than -4098: -7598 and -26458.

Mercifully, statistical programs (e.g., R) estimate the maximum likelihood for various classes of model without requiring any special programming effort. The main challenge for the user of statistical methods is the choice of appropriate structure: what distribution do the variables have; which predictors are likely to be important for a given problem and how do they relate to each other and to the outcome variable to be predicted. The approach taken is to fit a collection of models which are then put into competition with each other and compared with respect to how comprehensible they are and how well they describe the data.

There are three main methods for assessing goodness of fit for models estimated using MLE. The most common is the log-likelihood ratio (LLR) for nested models:

$$LLR = -2(\ell_1 - \ell_0) \tag{4.1}$$

where ℓ_0 and ℓ_1 denote the maximised log-likelihood of the two models being compared, M_0 and M_1, and the models are identical except M_0 is missing the parameters one wants to test (i.e., M_0 is nested in M_1). (Why is this difference called a ratio? Remember logs again: $\log(x/y) = \log x - \log y$.) The LLR is asymptotically $\chi^2(n)$ distributed, where n, the degrees of freedom, is the difference in the number of parameters between the two models. Computing the p-value is the LLR-test.

Secondly, derived from the maximum log-likelihood are also Akaike's information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978) fit indices, defined as follows:

$$AIC = -2\ell + 2k \tag{4.2}$$

$$BIC = -2\ell + k \log n \tag{4.3}$$

where k is the number of parameters and n the number of observations. These indices are compared for models, with a lower value preferred. Both AIC and BIC penalise for the number of parameters in the model in an attempt to prevent overfitting. BIC is stricter, tending to favour models with fewer parameters (Schwarz, 1978, p. 463), so I will not report BIC if the AIC is unfavorable.

The final test discussed here is of the null hypothesis $\hat{\beta} = 0$, for a parameter estimate $\hat{\beta}$. This is done using a Wald test, the statistic for which is calculated as $\hat{\beta}/SE(\hat{\beta})$ which, depending on the exact nature of the model, typically has a Gaussian distribution (mean $= 0$, $SD = 1$, i.e., a z-score) or a t-distribution.[4] One problem with Wald tests is that for

[4] Sometimes one also finds the χ^2 distribution mentioned in relation to Wald tests. This is since if

logit models (which I use) they have been shown to lose power as the effect sizes increase, tending beyond a threshold to approach 0. Hauck and Donner (1977) illustrate this using simulation data and a reanalysis of data on the presence of the *T. vaginalis* organism in a sample of females. All the predictors were significant at the 0.05 level using LLR-tests. Using the Wald test, however, two were far from significance ($p = 0.7$ and $p = 0.9$). Any instances of a significant LLR but not significant Wald will be flagged in this thesis.

A variety of techniques are used for model selection, importantly guided by substantial theory as well as statistical fit summaries. Sometimes M_0 is the null-model, i.e., a model consisting only of an intercept, or the simplest known model which previous research has shown to be important. Another approach is to begin with the most complex model which can estimated, set this to M_1, and explore which parameters may be removed without significantly affecting the LLR, AIC, or BIC. The process of model selection is necessarily qualitative, with a tension between explanatory power, parsimony, and occasionally parameters for very complex models cannot be reliably estimated at all.

I will introduce more of the required details of the various models used where their use first occurs. For general background and practical words of wisdom, I found books by Agresti (2002), Venables and Ripley (2002), Faraway (2006), and Gelman and Hill (2007) especially helpful, as was the mailing list[5] for the R special interest group on Mixed Effect Models.

$x \sim N(0,1)$, then $x^2 \sim \chi^2(1)$ (see Leemis & McQueston, 2008).

[5] See https://stat.ethz.ch/mailman/listinfo/r-sig-mixed-models.

Chapter 5

Self-report Measures of Autistic Traits

In this chapter we examine the two self-report measures of autistic-like traits in the sample, the Autism-Spectrum Quotient (AQ) and the Broad Autism Phenotype Questionnaire (BAPQ). The focus is their psychometric properties, which might be tedious going for some readers, though it is interesting taking seriously participant self-report and how questionnaire responses relate to each other. See later chapters for details of association with patterns of reasoning.

5.1 Descriptives

5.1.1 AQ

Recall that AQ consists of 50 items, and answers are recorded on a four-point scale: 'definitely agree', 'slightly agree', 'slightly disagree', or 'definitely disagree'. AQ has been scored in two ways in the literature:

1. The original binary scoring (0,1) where *degree* of agreement or disagreement is ignored (Baron-Cohen, Wheelwright, Skinner, et al., 2001).

2. A scoring which uses all four levels, thus allowing finer-grained distinctions across the dimensions. Some workers score 1 to 4 (e.g., Austin, 2005; Hoekstra et al., 2007)

	Num of items	Min	Max	Mean	SD	α	n
AQ (0,1 item coding)	50	2	38	17.5	6.8	0.80	99
Social skill	10	0	10	2.6	2.4	0.76	103
Attention switching	10	0	10	4.4	2.2	0.59	103
Attention to detail	10	1	10	5.4	2.1	0.55	104
Communication	10	0	9	2.7	2.0	0.58	103
Imagination	10	0	8	2.4	1.7	0.44	102
AQ (0 to 3 item coding)	50	29	100	58.7	14.8	0.85	99
Social skill	10	1	28	10.1	5.1	0.81	103
Attention switching	10	4	25	13.7	4.3	0.67	103
Attention to detail	10	9	30	15.6	4.3	0.60	104
Communication	10	0	22	10.1	4.5	0.68	103
Imagination	10	0	19	9.5	3.8	0.54	102
AQ (1 to 4 item coding)	50	79	150	108.7	14.8	0.85	99
Social skill	10	11	38	20.1	5.1	0.81	103
Attention switching	10	14	35	23.7	4.3	0.67	103
Attention to detail	10	19	40	25.6	4.3	0.60	104
Communication	10	10	32	20.1	4.5	0.68	103
Imagination	10	10	29	19.5	3.8	0.54	102

Table 5.1: Table of descriptive statistics for AQ and sumscores, computed in all three ways currently found in the literature

and some 0 to 3 (e.g., Ring et al., 2008).

See Table 5.1 for descriptives for the total composite score and subscores for each of these calculation methods.

Cronbach's α's were all very similar to that found in the literature. For instance for the total AQ scored in the four-level fashion, $\alpha = 0.85$; Austin (2005) found $\alpha = 0.82$.

	Num of items	Min	Max	Mean	SD	α	n
BAPQ	36	1.4	4.8	2.8	0.6	0.92	97
Aloof	12	1.3	5.2	2.8	0.9	0.90	101
Pragmatic language	12	1.3	4.9	2.8	0.7	0.80	101
Rigid	12	1.3	4.9	2.9	0.7	0.84	102

Table 5.2: Table of descriptive statistics for BAPQ

Using the four-point scoring slightly improves the α's, as can be seen from Table 5.1.[1]

ASCs are more common among males than females. For instance Baird et al. (2006) report a ratio of around 3:1, so one might expect a reflection of this in the spread of autistic traits amongst TD individuals. Indeed, Baron-Cohen, Wheelwright, Skinner, et al. (2001) found a statistically significant difference in scores. However in the present study, no sex differences were found for any of the scores (all t-test $p > 0.3$). For the four-level coding, the mean difference ranged from 0.15–0.7 and for the binary coding, 0.08–0.6. This was surprising. However the difference found by Baron-Cohen, Wheelwright, Skinner, et al. (2001), though statistically significant, was only a mean 2 point difference (male mean of 18.5 versus a female mean of 16.5). More recently, Hurst, Mitchell, Kimbrel, Kwapil, and Nelson-Gray (2007) failed to find any difference in means in a sample of just over 1000 college students.

5.1.2 BAPQ

See Table 5.2. The internal consistency for BAPQ and its subscales were good, e.g., for total score $\alpha = 0.92$. Hurley et al. (2007) found $\alpha = 0.95$ across all items and similar values for the other subscales. The α's are high but not too high, which would indicate that the same question is just being asked in different ways as opposed to a deeper construct being measured (see, e.g., Streiner, 2003, p. 102).

[1]The maximum correlation between two variables x and y, with Cronbach's α's of α_x and α_y, is $\sqrt{\alpha_x} \cdot \sqrt{\alpha_y}$. This can be useful for interpreting what counts as a 'strong' correlation.

5.2 Bivariate correlations

Table 5.3 shows Pearson's product moment correlations between AQ, BAPQ, and their subscale scores. Immediately noticeable is that everything correlates with everything, except the Attention to Details subscore of AQ which correlates only with its superset AQ.

	AQ	(soc)	(att swit)	(att det)	(comm)	(imag)	BAPQ	(aloof)	(prag lang)
AQ									
Social skill	.84**								
Attention switching	.69**	.48**							
Attention to detail	.34**	.04	−.02						
Communication	.83**	.74**	.48**	.08					
Imagination	.70**	.51**	.41**	.07	.46**				
BAPQ	.85**	.83**	.68**	.05	.75**	.54**			
Aloof	.74**	.86**	.42**	.05	.69**	.43**	.85**		
Pragmatic language	.68**	.65**	.53**	.00	.74**	.36**	.83**	.59**	
Rigid	.66**	.51**	.75**	.07	.42**	.53**	.79**	.45**	.51**

Table 5.3: Pearson's correlations between AQ, BAPQ, and their subscale scores. ** $p < 0.01$; listwise elimination of missing values ($n = 94$).

5.3 Factor structure

The problem this section addresses is as follows: AQ may be viewed as a collection of 50 variables, BAPQ may be viewed as a collection of 36 variables; can latent variable modelling point to a way of reducing the number of variables, whilst retaining maximum interesting (with respect to our goals) variability?

In passing we should note that there is still controversy on the difference between factor analysis and principle components analysis and when one should be preferred over another (Velicer & Jackson, 1990; Loehlin, 1990; Borsboom, 2006). Lawley and Maxwell (1962) summarise the mathematical difference concisely. Suppose we have p observed variables $x_1, x_2, \ldots, x_p$. For principle components analysis, the problem is to estimate uncorrelated principle component scores, z, and weights, w, such that

$$x_i = \sum_{r=1}^{p} w_{ir} z_r$$

Note that there are as many components as there were original observed variables. Factor analysis, on the other hand, requires a specification, k, of the number of components (called factors here) to be extracted:

$$x_i = \sum_{r=1}^{k} w_{ir} z_r + \epsilon_i$$

The weights, w, are often named loadings. Note the addition of the residual term ϵ, representing the variance in x which is not explained by the k factors. It is often the case that using only the first k components of a principle components analysis is equivalent to fitting an exploratory factor analysis with the same k. See Loehlin (1990, p. 31) for a counterexample showing this is not always the case.

5.3.1 AQ

The first factor analytic study of AQ was done by Austin (2005) who used principle components analysis with an oblique rotation (allowing the factors be correlated). A scree plot suggested three factors, which Austin (2005) named Social Skills, Details/Patterns, and Communication/Mind Reading. These are not subsets of the original theoretically defined subscales; some of the items in the factors come from more than one of the original

	Num of items	Min	Max	Mean	SD	α	n
Social skills	12	13	45	23.9	6.8	0.86	104
Details/patterns	8	13	32	20.4	3.7	0.58	104
Communication/mindreading	6	6	18	12.0	3.0	0.66	104

Table 5.4: Descriptives for unit-weighted factor scores.

subscales. Table 5.4 gives descriptives for the present study of unit-weighted scores for the items which Austin (2005) found to be loaded > 0.4 on the factor.

The Social Skills and Communication/Mindreading scores were positively correlated ($r = 0.3$, $p < 0.001$), similarly to that found by Austin (2005) ($r = 0.2$). The Cronbach α's found by Austin (2005) were: Social Skills, 0.85; Details/Patterns, 0.70; and Communication/Mindreading, 0.66; the details/patterns value was lower in the present sample, the others are similar. These α's are not so much better than the original however α depends on the number of items and the solution is of interest as the number of subscales suggested is three, versus five in the original, and 26 items in total, versus 50 in the original.

Hurst, Mitchell, et al. (2007) attempted to replicate this analysis in a larger sample and found loadings 'somewhat consistent' with the solution provided by Austin (2005). Applying a loading threshold of 0.4: 3 (out of 12) items were dropped from the Social Skills factor, 2 (out of 8) from the Details/Patterns factor, and 3 (out of 6) from the Communication/Mindreading factor. The others remained in the factors described by Austin (2005). Hurst, Mitchell, et al. (2007) argue in favour of either a two or three factor solution based on examination of the scree plot.

Hoekstra, Bartels, Cath, and Boomsma (2008) examined the factor structure using confirmatory factor analysis (CFA) in two Dutch population, students and a general population, using the Dutch translation of AQ. The translation was performed by an official translator, then back translated to English by a second translator, and compared with the English version in consultation with Simon Baron-Cohen before finalisation. Sentences included, (11) 'Ik vind sociale situaties gemakkelijk' and (24) 'Ik zou liever naar het theater gaan dan naar een museum'. They compared three models by CFA: (a) five factors derived from the original five theoretically defined subscales, (b) two factors, one of which (labelled Social Interaction) contained four of the five original subscales (excluding Attention to Detail) and the second consisted only of Attention to Detail, and (c) a one factor model, to test if AQ may be considered a unitary structure. Examining various fit statistics, the unitary model fared worst for both populations. For the students, the two factor model gave the best fit; for the general population the two factor model and the five factor model were only distinguishable by the principle of parsimony, i.e., the two factor model won as it has fewer factors.

5.3.1.1 Confirmatory Factor Analysis

An attempt was made to test these models on our data using CFA with the sem package (Fox, 2007) in R. First we fitted the model due to Austin (2005), allowing for a correlation between the Communication and Social/Mindreading latent variables. Now the misery begins. According to χ^2, the model is not a good fit [$\chi^2(298) = 548$, $p < 0.001$]; a $SRMR < 0.1$ can be interpreted as showing adequate fit, however for this model, $SRMR = 0.11$; a $RMSEA \leq 0.05$ is taken as good, however for this model, $RMSEA = 0.09$. Two items had non-significant loadings: 'I like to plan any activities that I participate in carefully' (43) and 'Other people frequently tell me that what I've said is impolite, even though I think it is polite' (7).

Next we fit the model due to Hoekstra et al. (2008). Several attempts to fit their hierarchical model and simplifications thereof failed to converge on a solution. A model with their two-latent variables, but no hierarchical structure, did give similar fit statistics to that for the discovered structure by Austin (2005) [$\chi^2(1174) = 2042$, $p < 0.001$, $RMSEA = 0.09$, $SRMR = 0.11$], although we did not find a correlation between the two latent variables ($p = 0.7$). Paths for several items were not significant (for Social Interaction, 2, 4, 16, 25, 43, 7, 18, 3, 21, 23; for Attention to Detail, 5, 28, 30).

5.3.1.2 Exploratory Factor Analysis

The confirmatory factor analysis fits were not as good as hoped, so next an exploratory analysis was attempted. Firstly a parallel analysis (Horn, 1965) was performed to estimate how many factors to include. The paran program by Dinno (2008) was used with the modification suggested by Glorfeld (1995) of taking the 99th centile of the distribution of eigenvalues resulting from PCA of the randomly-generated data, rather than Horn's original suggestion of taking the mean. Figure 5.1 shows a plot of the unadjusted, adjusted, and random eigenvalues. The analysis suggested retaining four factors (i.e., there were four components with adjusted eigenvalues > 1).

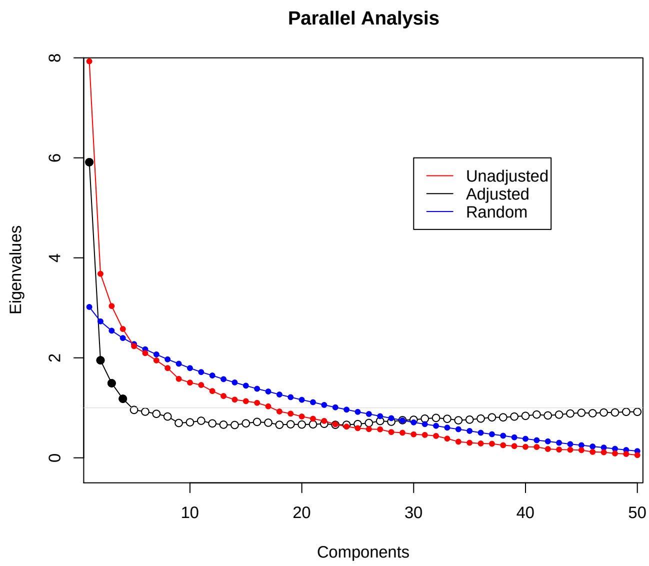

Figure 5.1: Parallel analysis of AQ: unadjusted, adjusted, and random eigenvalues. Suggested components to retain are marked with solid points.

Maximum likelihood factor analysis was then used to extract four factors. There was evidence that 4 factors were not sufficient ($\chi^2(1031) = 1279$, $p < 0.001$); only when a model with 10 factors was fitted was there evidence of adequate fit ($\chi^2(770) = 800$, $p = 0.2$).[2] Regardless, we chose 4, listening to the parallel analysis, and staying closer to the previously identified numbers of factors. Reiterating, there is likely to be other useful information in a deeper factor structure but for now we want to examine the data at the crudest level of analysis. See Table 5.5 for the loadings found, grouped according to whether items loaded ≥ 0.3 on a factor.

[2]Venables and Ripley (2002, p. 323) note that examples in the literature tend not to fit well or don't come with fit statistics, so we are in good company!

Table 5.5: Loadings for a maximum likelihood factor analysis with promax rotation on our AQ data, for items loading ≥ 0.3 on a factor (reported to one decimal place). Original subscale placement for each item is given surrounded by square brackets (SS: Social Skill; AS: Attention Switching; AD: Attention to Detail; C: Communication; and I: Imagination). Items loading ≥ 0.3 on more than one factor are listed at the end.

	1	2	3	4
Factor 1				
1. [SS] I prefer to do things with others rather than on my own	**0.3**	−0.2		
11. [SS] I find social situations easy	**0.8**		−0.2	**0.3**
15. [SS] I find myself drawn more strongly to people than to things	**0.6**	−0.1		
17. [C] I enjoy social chit-chat	**0.7**	−0.1		
27. [C] I find it easy to 'read between the lines' when someone is talking to me	**0.4**	0.2		−0.1
38. [C] I am good at social chit-chat	**0.9**		−0.1	
44. [SS] I enjoy social occasions.	**0.7**	−0.2	0.2	
47. [SS] I enjoy meeting new people	**0.8**			0.2
Factor 2				
20. [I] When I'm reading a story, I find it difficult to work out the characters' intentions		**0.6**	0.2	
31. [C] I know how to tell if someone listening to me is getting bored	0.1	**0.4**	0.2	
35. [C] I am often the last to understand the point of a joke		**0.5**		0.2
37. [AS] If there is an interruption, I can switch back to what I was doing very quickly		**0.5**	0.2	0.1
45. [SS] I find it difficult to work out people's intentions	0.2	**0.5**	0.1	0.2

(Continued on next page)

Table 5.5—continued from previous page

	1	2	3	4
Factor 3				
9. [AD] I am fascinated by dates		−0.3	**0.4**	
16. [AS] I tend to have very strong interests which I get upset about if I can't pursue	−0.2	0.1	**0.4**	0.1
25. [AS] It does not upset me if my daily routine is disturbed	−0.1		**0.7**	
34. [AS] I enjoy doing things spontaneously	0.2		**0.5**	
40. [I] When I was young, I used to enjoy playing games involving pretending with other children	0.1		**0.5**	−0.5
41. [I] I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.)	0.2	−0.2	**0.4**	
43. [AS] I like to plan any activities I participate in carefully	−0.3	0.2	**0.7**	0.1
Factor 4				
4. [AS] I frequently get so strongly absorbed in one thing that I lose sight of other things		0.2		**0.4**
5. [AD] I often notice small sounds when others do not	0.2	−0.3	−0.1	**0.4**
7. [C] Other people frequently tell me that what I've said is impolite, even though I think it is polite			0.1	**0.4**

(Continued on next page)

Table 5.5—continued from previous page

	1	2	3	4
Cross-factor				
13. [SS] I would rather go to a library than a party	**0.5**	−0.3	**0.4**	
18. [C] When I talk, it isn't always easy for others to get a word in edgeways	−0.3		**0.4**	**0.3**
22. [SS] I find it hard to make new friends	**0.7**	0.2	−0.2	**0.4**
26. [C] I frequently find that I don't know how to keep a conversation going	**0.7**	0.2	−0.2	**0.3**
32. [AS] I find it easy to do more than one thing at once		**0.4**	0.3	
33. [C] When I talk on the phone, I'm not sure when it's my turn to speak	**0.3**	0.2		**0.6**
36. [SS] I find it easy to work out what someone is thinking or feeling just by looking at their face	**0.3**	**0.4**	**0.3**	
39. [C] People often tell me that I keep going on and on about the same thing		0.2	**0.3**	**0.4**
46. [AS] New situations make me anxious	**0.3**	**0.3**		**0.5**
48. [SS] I am a good diplomat	**0.3**	**0.3**		

There was some overlap in the loadings; let us focus first on items which load mostly on one factor. Factor 1 appears to capture attitudes to conversation and socialising; factor 2 appears mostly about comprehension; factor 3 captures perseverance, inflexibility. Factor 4 is the most difficult to interpret, capturing a range of issues: language pragmatics, social anxiety, detail focus; this factor appears to be capturing correlations between linguistic-hints of theoretically interesting constructs.

We tried again with 3 factors, for easier comparison with Austin (2005), and with the hope that the factors will be more readily interpretable. (See Table 5.6 for the loadings.) These factors are more easily comprehensible: factor 1 is related to socialising and conversation (mostly the original Social Skills and Communication items); factor 2 captures perseverance (items from Attention Switching and some items from Communication); and factor 3 captures competence with numbers (all Attention to Detail items). Items which loaded ≥ 0.3 cross-factor are also included for reference. Now let us compare with the items found by Austin (2005). The social skills factor is a good match to our first factor (we excluded only three items). The Details/Patterns factor is a mixture of our 2nd and 3rd factors. Finally only one item (39) appears in our analysis from Austin's Communication/Mindreading factor, loading on our 2nd factor.

Table 5.6: A second attempt at factor analysis for AQ. Loadings for a maximum likelihood factor analysis with promax rotation for items loading ≥ 0.3 on a factor (reported to one decimal place). Original subscale placement for each item is given surrounded by square brackets (SS: Social Skill; AS: Attention Switching; AD: Attention to Detail; C: Communication; and I: Imagination). Items loading ≥ 0.3 on more than one factor are listed at the end.

	1	2	3
Factor 1			
1. [SS] I prefer to do things with others rather than on my own	**0.3**		0.2
8. [I] When I'm reading a story, I can easily imagine what the characters might look like	**0.3**	−0.2	−0.1
11. [SS] I find social situations easy	**0.8**	−0.2	
15. [SS] I find myself drawn more strongly to people than to things	**0.6**	−0.2	0.2
17. [C] I enjoy social chit-chat	**0.7**	−0.1	0.2
22. [SS] I find it hard to make new friends	**0.7**	−0.1	−0.1
26. [C] I frequently find that I don't know how to keep a conversation going	**0.7**	−0.2	−0.1
27. [C] I find it easy to 'read between the lines' when someone is talking to me	**0.4**		−0.1
36. [SS] I find it easy to work out what someone is thinking or feeling just by looking at their face	**0.3**	0.2	−0.3
38. [C] I am good at social chit-chat	**1.0**	−0.2	
44. [SS] I enjoy social occasions	**0.7**		0.3
47. [SS] I enjoy meeting new people	**0.8**		
48. [SS] I am a good diplomat	**0.3**		−0.3
50. [I] I find it very easy to play games with children that involve pretending	**0.3**		

(Continued on next page)

Table 5.6—continued from previous page

	1	2	3
Factor 2			
16. [AS] I tend to have very strong interests which I get upset about if I can't pursue	−0.2	**0.5**	
18. [C] When I talk, it isn't always easy for others to get a word in edgeways	−0.3	**0.5**	
25. [AS] It does not upset me if my daily routine is disturbed	−0.2	**0.8**	0.2
34. [AS] I enjoy doing things spontaneously	0.2	**0.5**	0.2
39. [C] People often tell me that I keep going on and on about the same thing		**0.4**	−0.1
43. [AS] I like to plan any activities I participate in carefully	−0.4	**0.8**	
Factor 3			
6. [AD] I usually notice car number plates or similar strings of information		0.2	**0.4**
19. [AD] I am fascinated by numbers	0.2		**0.4**
29. [AD] I am not very good at remembering phone numbers			**0.4**
49. [AD] I am not very good at remembering people's date of birth	0.1		**0.3**
Cross-factor			
9. [AD] I am fascinated by dates		**0.3**	**0.3**
13. [SS] I would rather go to a library than a party	**0.5**	**0.3**	**0.4**
23. [AD] I notice patterns in things all the time		0.3	**0.3**
41. [I] I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.)	0.2	**0.4**	**0.3**

5.3.2 BAPQ

5.3.2.1 Confirmatory Factor Analysis

No work has been done to investigate the factor structure of BAPQ, so we attempt to fit the structure as characterised by the original subscales (Hurley et al., 2007), allowing the latent variables to correlate as the subscores do. This wasn't a great fit to the data

[$\chi^2(592) = 1128$, $p < 0.001$, $RMSEA = 0.097$, $SRMR = 0.12$], but marginally better than a one-factor model [$\chi^2(594) = 1393$, $p < 0.001$, $RMSEA = 0.12$, $SRMR = 0.13$].

All paths were significant in both models, except for item 21 in the three factor model ('I can tell when someone is not interested in what I am saying') and again item 21 and 17 ('I have been told that I talk too much about certain topics') in the one factor model.

5.3.2.2 Exploratory Factor Analysis

We repeat the parallel analysis performed for AQ. Figure 5.2 shows a plot of the unadjusted, adjusted, and random eigenvalues. The analysis suggested retaining three factors, the same as the originally specified number of subscales. (There again was evidence that 3 is not sufficient [$\chi^2(525) = 767$, $p < 0.001$]. The first model showing evidence of good fit has 9 factors [$\chi^2(342) = 373$, $p = 0.1$].) See Table 5.7 for the loadings of a factor analysis with promax rotation. As can be seen from the table above, factor 1 includes mostly Aloof items, factor 2 Rigid items, and factor 3 Pragmatic Language items.

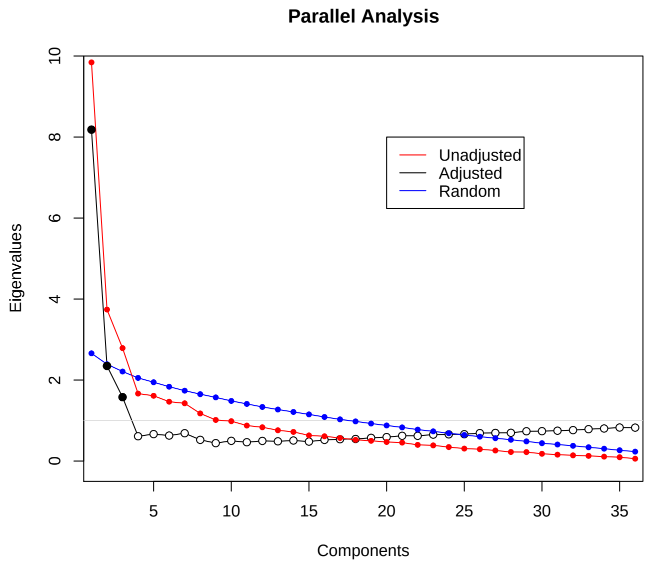

Figure 5.2: Parallel analysis of BAPQ: unadjusted, adjusted, and random eigenvalues. Suggested components to retain are marked with solid points.

Table 5.7: Loadings for a maximum likelihood factor analysis with promax rotation on our BAPQ data, for items loading ≥ 0.3 on a factor (reported to one decimal place). Original subscale placement for each item is given surrounded by square brackets. Items loading ≥ 0.3 on more than one factor are listed at the end.

	1	2	3
Factor 1			
1. [Aloof] I like being around other people	**0.8**		
9. [Aloof] I enjoy being in social situations	**1.0**		−0.2
12. [Aloof] People find it easy to approach me	**0.5**		
23. [Aloof] I am good at making small talk	**0.7**		−0.1
25. [Aloof] I feel like I am really connecting with other people	**0.7**		0.1
27. [Aloof] Conversation bores me	**0.7**		
28. [Aloof] I am warm and friendly in my interactions with others	**0.5**		0.2
31. [Aloof] I prefer to be alone rather than with others	**0.6**	0.2	
34. [Pragmatic language] I can tell when it is time to change topics in conversation	**0.5**		
Factor 2			
3. [Rigid] I am comfortable with unexpected changes in plans		**0.7**	
6. [Rigid] People have to talk me into trying something new		**0.6**	
8. [Rigid] I have to warm myself up to the idea of visiting an unfamiliar place		**0.4**	0.2
10. [Pragmatic language] My voice has a flat or monotone sound to it	0.2	**0.3**	0.2
13. [Rigid] I feel a strong need for sameness from day to day	−0.2	**0.7**	
24. [Rigid] I act very set in my ways	−0.2	**0.5**	0.1
30. [Rigid] I alter my daily routine by trying something different	0.2	**0.6**	−0.4
33. [Rigid] I like to closely follow a routine while working		**0.6**	0.1

(Continued on next page)

Table 5.7—continued from previous page

	1	2	3
Factor 3			
2. [Pragmatic language] I find it hard to get my words out smoothly	0.2		**0.6**
4. [Pragmatic language] It's hard for me to avoid getting sidetracked in conversation	−0.2		**0.6**
14. [Pragmatic language] People ask me to repeat things I've said because they don't understand	0.1	0.2	**0.4**
17. [Pragmatic language] I have been told that I talk too much about certain topics	−0.3	−0.1	**0.8**
18. [Aloof] When I make conversation it is just to be polite	0.2		**0.3**
20. [Pragmatic language] I speak too loudly or softly			**0.5**
26. [Rigid] People get frustrated by my unwillingness to bend	0.2	0.2	**0.4**
29. [Pragmatic language] I leave long pauses in conversation	0.1	−0.2	**0.6**
32. [Pragmatic language] I lose track of my original point when talking to people	−0.3	0.2	**0.7**
Cross-factor			
5. [Aloof] I would rather talk to people to get information than to socialize	**0.4**		**0.3**
7. [Pragmatic language] I am 'in-tune' with the other person during conversation	**0.4**		**0.3**
11. [Pragmatic language] I feel disconnected or 'out of sync' in conversations with others	**0.4**		**0.4**
15. [Rigid] I am flexible about how things should be done	**0.3**	**0.5**	−0.1
16. [Aloof] I look forward to situations where I can meet new people	**0.7**	**0.3**	−0.1
19. [Rigid] I look forward to trying new things	**0.4**	**0.5**	−0.3
22. [Rigid] I have a hard time dealing with changes in my routine	−0.3	**0.7**	**0.4**
35. [Rigid] I keep doing things the way I know, even if another way might be better	−0.1	**0.4**	**0.3**

5.3.3 Analysis of both

Finally we perform a factor analysis of AQ and BAPQ combined. Figure 5.3 shows a plot of the unadjusted, adjusted, and random eigenvalues, and suggests retaining four factors. See Table 5.8 for the loadings of a factor analysis with promax rotation.

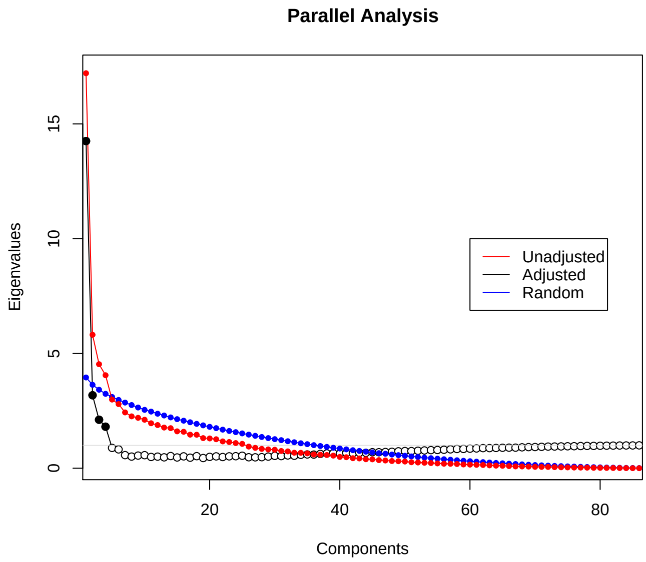

Figure 5.3: Parallel analysis of both AQ and BAPQ: unadjusted, adjusted, and random eigenvalues. Suggested components to retain are marked with solid points.

Table 5.8: Loadings for a maximum likelihood factor analysis with promax rotation on our AQ (marked A) and BAPQ (marked B) data combined, for items loading ≥ 0.3 on a factor (reported to one decimal place). Original subscale placement for each item is given surrounded by square brackets (Soc: Social Skill; AttSw: Attention Switching; AttDet: Attention to Detail; Com: Communication; Imag: Imagination; Alo: Aloof; Pra: Pragmatic Language; and Rig: Rigid). Items loading ≥ 0.3 on more than one factor are listed at the end.

	1	2	3	4
Factor 1				
A1. [Soc] I prefer to do things with others rather than on my own	**0.5**		−0.2	
A11. [Soc] I find social situations easy	**0.6**	−0.1	0.2	0.2
A15. [Soc] I find myself drawn more strongly to people than to things	**0.6**	−0.1		−0.1
A17. [Com] I enjoy social chit-chat	**0.7**	−0.2	−0.1	
A22. [Soc] I find it hard to make new friends	**0.5**		0.3	0.1
A26. [Com] I frequently find that I don't know how to keep a conversation going	**0.6**	−0.2	0.3	0.2
A38. [Com] I am good at social chit-chat	**0.8**		0.2	
A44. [Soc] I enjoy social occasions	**0.7**	0.2	−0.1	
A47. [Soc] I enjoy meeting new people	**0.7**			
B1. [Alo] I like being around other people	**0.8**		−0.1	
B5. [Alo] I would rather talk to people to get information than to socialize	**0.5**		−0.1	0.2
B9. [Alo] I enjoy being in social situations	**1.0**		−0.1	−0.2
B11. [Pra] I feel disconnected or 'out of sync' in conversations with others	**0.5**		**0.3**	0.2
B12. [Alo] People find it easy to approach me	**0.5**		0.1	
B16. [Alo] I look forward to situations where I can meet new people	**0.7**	0.2		−0.1
B23. [Alo] I am good at making small talk	**0.6**	−0.2	**0.3**	−0.1
B25. [Alo] I feel like I am really connecting with other people	**0.7**		0.2	
B27. [Alo] Conversation bores me	**0.7**		−0.1	
B28. [Alo] I am warm and friendly in my interactions with others	**0.5**		**0.3**	
B31. [Alo] I prefer to be alone rather than with others	**0.7**	0.2	−0.2	
B36. [Alo] I enjoy chatting with people	**1.0**	−0.1		−0.2

(Continued on next page)

Table 5.8—continued from previous page

	1	2	3	4
Factor 2				
A2. [AttSw] I prefer to do things the same way over and over again		**0.4**		
A16. [AttSw] I tend to have very strong interests, which I get upset about if I can't pursue	−0.2	**0.3**		0.2
A25. [AttSw] It does not upset me if my daily routine is disturbed		**0.8**	−0.2	
A34. [AttSw] I enjoy doing things spontaneously	**0.3**	**0.5**		−0.1
A40. [Imag] When I was young, I used to enjoy playing games involving pretending with other children	0.2	**0.3**		−0.2
A41. [Imag] I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc)	0.2	**0.4**	−0.2	**0.3**
A43. [AttSw] I like to plan any activities I participate in carefully	−0.2	**0.6**		0.2
B3. [Rig] I am comfortable with unexpected changes in plans		**0.8**	0.2	
B6. [Rig] People have to talk me into trying something new		**0.5**	**0.3**	
B13. [Rig] I feel a strong need for sameness from day to day	−0.2	**0.6**	0.2	
B22. [Rig] I have a hard time dealing with changes in my routine	−0.2	**0.7**	0.2	**0.3**
B24. [Rig] I act very set in my ways	−0.1	**0.5**		0.1
B30. [Rig] I alter my daily routine by trying something different		**0.5**		−0.3
B33. [Rig] I like to closely follow a routine while working		**0.6**		0.1
B35. [Rig] I keep doing things the way I know, even if another way might be better		**0.3**	0.2	0.2

(Continued on next page)

Table 5.8—continued from previous page

	1	2	3	4
Factor 3				
A14. [Imag] I find making up stories easy		**0.3**	**0.3**	
A20. [Imag] When I'm reading a story, I find it difficult to work out the characters' intentions	−0.1	0.2	**0.6**	
A27. [Com] I find it easy to 'read between the lines' when someone is talking to me	**0.3**		**0.3**	
A31. [Com] I know how to tell if someone listening to me is getting bored		0.1	**0.5**	
A32. [AttSw] I find it easy to do more than one thing at once		0.2	**0.4**	
A35. [Com] I am often the last to understand the point of a joke			**0.4**	0.2
A36. [Soc] I find it easy to work out what someone is thinking or feeling just by looking at their face	0.2		**0.5**	
A37. [AttSw] If there is an interruption, I can switch back to what I was doing very quickly			**0.6**	
A45. [Soc] I find it difficult to work out people's intentions		0.1	**0.6**	
A48. [Soc] I am a good diplomat			**0.5**	0.2
B8. [Rig] I have to warm myself up to the idea of visiting an unfamiliar place		0.3	**0.4**	
B21. [Pra] I can tell when someone is not interested in what I am saying	−0.1	0.1	**0.3**	

(Continued on next page)

Table 5.8—continued from previous page

	1	2	3	4
Factor 4				
A4. [AttSw] I frequently get so strongly absorbed in one thing that I lose sight of other things			0.1	**0.4**
A5. [AttDet] I often notice small sounds when others do not		−0.1	−0.2	**0.4**
A7. [Com] Other people frequently tell me that what I've said is impolite, even though I think it is polite				**0.5**
A18. [Com] When I talk, it isn't always easy for others to get a word in edgeways	−0.1	0.1	−0.1	**0.4**
A39. [Com] People often tell me that I keep going on and on about the same thing	−0.1	0.2	0.2	**0.5**
B2. [Pra] I find it hard to get my words out smoothly	0.2		**0.3**	**0.5**
B4. [Pra] It's hard for me to avoid getting sidetracked in conversation			0.2	**0.6**
B17. [Pra] I have been told that I talk too much about certain topics	−0.2			**0.8**
B20. [Pra] I speak too loudly or softly	0.1		0.2	**0.4**
B32. [Pra] I lose track of my original point when talking to people	−0.1	0.2	0.1	**0.7**
Cross-factor				
A13. [Soc] I would rather go to a library than a party	**0.5**	**0.3**	−0.3	
B7. [Pra] I am 'in-tune' with the other person during conversation	**0.4**	−0.1	**0.4**	
B15. [Rig] I am flexible about how things should be done	**0.3**	**0.5**	0.1	
B19. [Rig] I look forward to trying new things	**0.3**	**0.4**	0.1	−0.2
B29. [Pra] I leave long pauses in conversation	0.1	−0.3	**0.5**	**0.3**
B34. [Pra] I can tell when it is time to change topics in conversation	**0.4**		**0.3**	

5.4 Gradients of 'difficulty'

Tables 5.9 and 5.10 show the percentage of participants giving each response for AQ and BAPQ (respectively), sorted by the percentage of respondents giving especially autistic-like responses for the items. It's interesting that the autistic end of the scale is frequently endorsed for 7 out of 10 Attention to Detail items in AQ, with top of the list items 12, ‘I tend to notice details that others do not’ and 30, ‘I usually notice small changes in a situation, or a person’s appearance.’ Responses to, e.g., item 31, ‘I [don’t] know how to tell if someone listening to me is getting bored’, and 20, ‘When I’m reading a story, I find it difficult to work out the characters’ intention’, tend to stay very close to the non-autistic end. Items from the other four subscales of AQ are more evenly distributed across the continuum of difficulty.

Items from all three subscales of BAPQ are fairly evenly distributed. The items participants are most likely to endorse at the autistic end are 20, ‘I speak too loudly or softly’, and 4, ‘It’s hard for me to avoid getting sidetracked in conversation.’ The two most difficult to endorse are 27, ‘Conversation bores me’, and 28, ‘I am [not] warm and friendly in my interactions with others.’

5.5 Residualising the subscores

As a final act of correlational exploration, and to aid interpretation later, I also examined correlations between subscores and items once other subscores had been partialled out. All three subscores of BAPQ could be partially predicted from the others: Pragmatic Language [$F(2,94) = 27.3$, $p < 0.001$, $R^2 = 0.35$], Aloof [$F(2,94) = 23.1$, $p < 0.001$, $R^2 = 0.32$], and Rigidity [$F(2,94) = 17.9$, $p < 0.001$, $R^2 = 0.26$]. Four of the five subscores of AQ could be predicted from the others: Social Skills [$F(4,94) = 29.8$, $p < 0.001$, $R^2 = 0.54$], Attention Switching [$F(4,94) = 8.64$, $p < 0.001$, $R^2 = 0.24$], Communication [$F(4,94) = 25.3$, $p < 0.001$, $R^2 = 0.50$], and Imagination [$F(4,94) = 10.7$, $p < 0.001$, $R^2 = 0.28$], but not Attention to Detail [$F(4,94) = 0.5$, $p = 0.7$].

For AQ, the Social Skills score was predicted only by Communication ($t = 7.0$, $p < 0.001$) and Imagination ($t = 3.2$, $p = 0.002$); Attention Switching by Imagination ($t = 2.2$, $p = 0.03$) and Communication ($t = 1.9$, $p = 0.057$); Communication was predicted by Social Skills ($t = 7.0$, $p < 0.001$) and Attention Switching ($t = 1.9$, $p = 0.057$); and finally Imagination by Social Skills ($t = 3.2$, $p = 0.002$) and Attention Switching ($t = 2.2$, $p = 0.03$). Social Skills, Communication, and Imagination seem most tightly related.

For BAPQ, pairs of subscores always predicted the third subscore. For Aloof, Pragmatic Language ($t = 2.5$, $p < 0.001$) and Rigid ($t = 2.3$, $p = 0.02$); for Pragmatic

		Response						Response			
Item	Scale	1	2	3	4	Item	Scale	1	2	3	4
12	AD	1	14	48	37	1	SS	11	56	28	6
30	AD	8	19	44	29	22	SS	36	31	28	6
16	AS	5	30	40	25	42	I	21	48	25	6
4	AS	8	32	36	25	17	C	28	44	22	6
5	AD	15	25	35	25	18	C	32	42	20	6
43	AS	8	31	38	24	7	C	50	27	17	6
46	AS	14	26	39	20	36	SS	25	54	15	6
23	AD	10	32	39	19	13	SS	44	35	15	6
29	AD	21	29	32	18	44	SS	51	35	10	5
49	AD	19	32	31	18	2	AS	12	43	41	4
6	AD	12	32	41	15	28	AD	14	45	37	4
26	C	20	40	24	15	45	SS	14	54	28	4
14	I	18	37	32	13	15	SS	32	37	28	4
19	AD	34	28	25	13	10	AS	25	48	23	4
25	AS	28	34	27	12	24	I	31	46	19	4
38	C	23	39	26	12	32	AS	27	51	18	4
50	I	27	38	24	12	34	AS	44	37	15	4
9	AD	37	34	21	9	35	C	39	42	14	4
21	I	49	30	12	9	47	SS	46	41	9	4
11	SS	31	38	24	8	3	I	54	36	7	3
39	C	30	42	20	8	37	AS	16	43	38	2
27	C	25	41	27	7	33	C	41	38	19	2
48	SS	18	51	23	7	8	I	46	39	12	2
41	I	39	34	20	7	20	I	31	52	16	1
40	I	41	35	17	7	31	C	44	42	12	1

Table 5.9: Percentage of participants giving each responses for the items of AQ. Items have been transformed so 4 is a particularly autistic-like response and sorted in ascending order of how difficult it is to draw a response indicating presence of a particularly autistic trait. (SS: Social Skill; AS: Attention Switching; AD: Attention to Detail; C: Communication; and I: Imagination.)

		Response								Response					
Item	Scale	1	2	3	4	5	6	Item	Scale	1	2	3	4	5	6
20	P	8	18	36	10	18	10	35	R	8	31	29	17	13	2
4	P	8	26	29	17	13	8	12	A	18	31	23	16	10	2
3	R	12	22	24	24	11	7	26	R	26	41	15	6	10	2
5	A	15	29	22	14	13	6	18	A	11	22	39	17	9	2
23	A	12	22	26	22	12	6	15	R	12	31	31	16	9	2
24	R	6	18	36	24	11	6	1	A	16	39	25	15	4	2
30	R	3	12	22	41	17	5	11	P	18	43	24	9	4	2
33	R	4	19	26	30	16	5	22	R	14	40	28	12	3	2
2	P	11	28	30	15	12	5	25	A	6	25	26	24	18	1
32	P	7	41	28	13	6	5	6	R	17	35	25	15	7	1
16	A	17	32	18	22	5	5	29	P	13	40	32	7	7	1
8	R	25	24	21	14	12	4	21	P	10	43	26	17	4	1
10	P	30	34	18	8	7	4	7	P	15	37	29	12	8	0
31	A	10	22	38	21	5	4	36	A	20	34	23	16	7	0
17	P	22	41	17	9	7	3	34	P	11	38	27	18	6	0
13	R	17	28	35	12	6	3	19	R	23	43	17	12	5	0
9	A	24	37	14	17	5	3	28	A	24	41	19	12	3	0
14	P	12	37	29	15	5	3	27	A	17	38	33	10	3	0

Table 5.10: Percentage of participants giving each responses for the items of BAPQ. Items have been transformed so 4 is a particularly autistic-like response and sorted in ascending order of how difficult it is to draw a response indicating presence of a particularly autistic trait. (P: Pragmatic Language; A: Aloof; R: Rigid.)

Language, Aloof ($t = 4.46$, $p < 0.001$) and Rigid ($t = 3.4$, $p = 0.001$); and for Rigid, Pragmatic Language ($t = 3.4$, $p = 0.001$) and Aloof ($t = 2.3$, $p = 0.02$). So it appears that Pragmatic Language and Aloofness are most tightly correlated.

See Table 5.11 for the correlations for AQ and Table 5.13 for the correlations for BAPQ. Table 5.12 shows the correlations with the Attention to Detail subscore of AQ.

5.6 Screening properties

5.6.1 AQ

Although the present dataset does not record whether a clinical diagnosis was ever made (since it's a typical, not a clinical, population), it may be useful to examine how the data appear when viewing AQ as a screening tool. Table 5.14 shows the means found by Baron-Cohen, Wheelwright, Skinner, et al. (2001) for people with an ASC versus TDs, and the present sample means. As per the original design, all the means in the ASC group are higher (it turns out statistically significantly so) than the scores from the TD group. It is worth noting that the mean difference is smallest for the Attention to Detail subscore (1.4 versus between 4–5 for the other subscores).[3] Figure 5.4 shows a graphical depiction for the composite scores of AQ.

[3] Autism Diva, a blogger with a diagnosis of Asperger syndrome, writes that she is 'seriously handicapped when it comes to remembering numerals, whether they are numbers of things like the population of her home town, or just strings of numerals like a phone number.' Items of AQ in the Attention to Detail subscale assess these abilities, but her answer would reduce her autistic trait score. She continues: 'Not that the whole AQ Test hangs on those few questions about numbers and dates... but still it seems like maybe the questions were based on a stereotype. Maybe ASD people like Autism Diva are more common than realized???' Merely anecdote, but interesting nonetheless. (See http://autismdiva.blogspot.com/2007/03/hang-on-to-your-maps.html)

	Social Skill	R^2
44	I enjoy social occasions	0.30
15	I find myself drawn more strongly to people than to things	0.29
11	I find social situations easy	0.24
47	I enjoy meeting new people	0.23
22	I find it hard to make new friends	0.22
13	I would rather go to a library than a party	0.20
38	I am good at social chit-chat	0.16
1	I prefer to do things with others rather than on my own	0.14

	Attention Switching	R^2
43	I like to plan any activities I participate in carefully	0.41
25	It does not upset me if my daily routine is disturbed	0.28
16	I tend to have very strong interests which I get upset about if I can't pursue	0.27
34	I enjoy doing things spontaneously	0.21
4	I frequently get so strongly absorbed in one thing that I lose sight of other things	0.19
46	New situations make me anxious	0.19
37	If there is an interruption, I can switch back to what I was doing very quickly	0.13
2	I prefer to do things the same way over and over again	0.12
32	I find it easy to do more than one thing at once	0.11

	Communication	R^2
35	I am often the last to understand the point of a joke	0.28
31	I know how to tell if someone listening to me is getting bored	0.24
18	When I talk, it isn't always easy for others to get a word in edgeways	0.22
7	Other people frequently tell me that what I've said is impolite…	0.21
33	When I talk on the phone, I'm not sure when it's my turn to speak	0.17

	Imagination	R^2
40	When I was young, I used to enjoy playing games involving pretending…	0.31
50	I find it very easy to play games with children that involve pretending	0.20
21	I don't particularly enjoy reading fiction	0.16
14	I find making up stories easy	0.14
8	When I'm reading a story, I can easily imagine what the characters might look like	0.13
3	If I try to imagine something, I find it very easy to create a picture in my mind	0.13
20	When I'm reading a story, I find it difficult to work out the characters' intentions	0.12

Table 5.11: Item scores correlating with each AQ subscore once other subscores have been partialled out. Items scores have been included if their $R^2 > .10$.

	Attention to Detail	R^2
29	I am not very good at remembering phone numbers	0.38
6	I usually notice car number plates or similar strings of information	0.30
49	I am not very good at remembering people's date of birth	0.27
12	I tend to notice details that others do not	0.26
9	I am fascinated by dates	0.26
19	I am fascinated by numbers	0.24
23	I notice patterns in things all the time	0.23
5	I often notice small sounds when others do not	0.18
30	I don't usually notice small changes in a situation, or a person's appearance	0.13

Table 5.12: Item scores correlating with the Attention to Detail subscore of AQ. Items scores have been included if their $R^2 > .10$.

	ASC ($n = 58$)	Present sample ($n = 104$)	TD ($n = 174$)
AQ	35.8	17.5	16.4
Social skill	7.5	2.6	2.6
Attention switching	8.0	4.4	3.9
Attention to detail	6.7	5.4	5.3
Communication	7.2	2.7	2.4
Imagination	6.4	2.4	2.3

Table 5.14: Means AQ and its subscores for people with ASCs and matched TD controls in the study reported by Baron-Cohen, Wheelwright, Skinner, et al. (2001, p. 8), between which are the present sample means.

One of the original design goals of AQ was the production of a quick self-report tool for screening for ASC, for instance to aid decisions about who to refer for further assessment and therapy. Baron-Cohen, Wheelwright, Skinner, et al. (2001, p. 12) suggest a threshold score of 32 for this purpose, as 80% of the people with an ASC scored at least this highly, compared to 2% of controls. Woodbury-Smith et al. (2005) argue using a different sample

	Pragmatic Languge	R^2
17	I have been told that I talk too much about certain topics	0.39
4	It's hard for me to avoid getting sidetracked in conversation	0.37
2	I find it hard to get my words out smoothly	0.35
32	I lose track of my original point when talking to people	0.33
29	I leave long pauses in conversation	0.31
14	People ask me to repeat things I've said because they don't understand	0.20
11	I feel disconnected or 'out of sync' in conversations with others.	0.20
20	I speak too loudly or softly	0.20

	Aloof	R^2
36	I enjoy chatting with people	0.67
9	I enjoy being in social situations	0.57
1	I like being around other people	0.46
27	Conversation bores me	0.40
16	I look forward to situations where I can meet new people	0.35
25	I feel like I am really connecting with other people	0.34
23	I am good at making small talk	0.33
31	I prefer to be alone rather than with others	0.29
28	I am warm and friendly in my interactions with others	0.24
12	People find it easy to approach me	0.23
5	I would rather talk to people to get information than to socialize	0.19

	Rigid	R^2
13	I feel a strong need for sameness from day to day	0.44
3	I am comfortable with unexpected changes in plans	0.40
22	I have a hard time dealing with changes in my routine	0.36
6	People have to talk me into trying something new	0.36
33	I like to closely follow a routine while working	0.35
24	I act very set in my ways	0.27
30	I alter my daily routine by trying something different	0.26
19	I look forward to trying new things	0.22
15	I am flexible about how things should be done	0.22
35	I keep doing things the way I know, even if another way might be better	0.21
8	I have to warm myself up to the idea of visiting an unfamiliar place	0.20

Table 5.13: Item scores correlating with each BAPQ subscore once other subscores have been partialled out. Items scores have been included if their $R^2 > .10$.

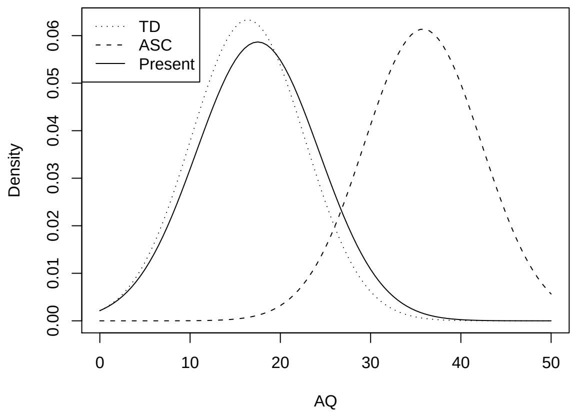

Figure 5.4: Estimates of the distributions (plotted using the Gaussian density function) of the AQ composite score for the TD and ASC groups reported by Baron-Cohen, Wheelwright, Skinner, et al. (2001), and for the present sample.

for a lower threshold, ≥ 26, for screening for ASC. This threshold correctly classified 83% of people from that population. They accept, however, that the higher threshold may be more appropriate for use in a general population as 'there may be a percentage of individuals who have many autistic traits but who do not require any clinical support (and are not seeking this).' In our university student sample, 12% of the participants had a score ≥ 26 (14% of the males and 10% of the females) and only 3% had a score ≥ 32 (equal percentages of males and females), very similar to the number found by Baron-Cohen, Wheelwright, Skinner, et al. (2001). We are unaware if any have a diagnosis of an ASC. The sex ratios are not as one would expect at the upper end of the scale, but recall that we didn't sample from, e.g., mathematics, physics, or computer science—subjects tended to be studied by students with stronger autistic traits and fewer females.

5.6.2 BAPQ

BAPQ was designed to detect the broader phenotype in the relatives of people with autism. Hurley et al. (2007, p. 1685) gives thresholds for the composite score (3.15) and each of the subscores: Aloof (3.25), Rigid (3.50), Pragmatic Language (2.75). Around 29% of our sample where over threshold for Aloof, 20% for Rigid, and 48% for Pragmatic Language. For the total composite score, 28% were over the threshold. Relating back to AQ, those over threshold in BAPQ had AQ scores in the range 16–38 (M = 23, SD = 5.7), i.e., around the mean AQ score and above.

5.7 Summary

It is common in the literature to explore the factor structure of self-report scales in an attempt to discover underlying latent traits which are generating the manifest questionnaire responses. One problem is that correlations between items say little about possible causal factors generating the correlations. Items scores may be correlated because questions are basically identical, but phrased in very subtly different ways. They may be correlated because of a deeper underlying causal process; perhaps one item loading on a factor is a good indicator of a cognitive trait and another item is a good indicator of a social consequence of possessing that trait. Factor analysis cannot uncover these distinctions. It is however interesting to see what items tend to go together as these give hints for possible causal processes.

There appears to be a well known problem with CFA. One would like to use EFA and then in a separate sample use CFA to test the structure found. This often doesn't happen, even if CFA is used on the same sample as was used for EFA—a fact that is well known in the literature (see Ferrando & Lorenzo-Seva, 2000, p. 301). The story for BAPQ is simplest: the Cronbach's α's were high and the factor structure obtained via parallel analysis and EFA was very similar to the original subscale structure. For AQ, only the α's for the overall score and Social Skills subscore were good. Using unit-weighted sumscores from the factors discovered by Austin (2005) doesn't improve the α's by much. It is more likely we will discover correlations with scores which have high α's. Given the uncertainty of the factor analysis results, and for ease of interpretation with respect to the existing literature, all the original subscales will be used and not factor scores obtained from the present sample.

One interesting finding is the lack of significant difference in AQ between sexes. On closer inspection, the sex effect was tiny anyway, and large studies (e.g., one with 1000 participants) also failed to find an effect, suggesting that we do not merely lack the power to detect a difference.

Chapter 6

Visual and Verbal-Analytic Reasoning Via Raven's Matrices

> The Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) comes in a big heavy box, which is very useful to hit people on the head with, but the hypothesis that the WAIS is valid for inflicting physical injury is certainly not the kind of hypothesis we are interested in.
>
> Borsboom, Mellenbergh, and Heerden (2004, p. 1065)

Raven's Advanced Progressive Matrices (RAPM) were originally included in the data collection for the present investigation to estimate the general factor in intelligence, g (Spearman, 1904), in anticipation of the question, 'How do these individual differences relate to IQ?'[1] Firstly, what is g? The first declaration of the 'law of the Universal Unity of the Intellective Function' (as it was then named) was described as follows (Spearman, 1904, p. 273):

> Whenever branches of intellectual activity are at all dissimilar, then their correlations with one another appear wholly due to their being all variously saturated with some common fundamental Function (or group of Functions).[2]

Little g is calculated by the following process. Participants are given a series of ability tests. The tests selected, how they are scored, and their properties, necessarily influence

[1]At the time of writing, three people have asked me this question.

[2]Note the parenthetical 'group' comment! Often important in the arguments.

the g that emerges, though there is evidence that different IQ test batteries produce the same g (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004). Principal component analysis or factor analysis is applied to the test scores, and the first component/factor extracted to form a score by computing a weighted sum of each individual test score. This g score represents the shared variance between the tasks. From someone's g score, and a knowledge of the loadings, the individual ability scores can be predicted, with the amount of noise related to the degree to which the individual scores load on g. To accurately estimate g requires the adminstration of a large number of tests, but approximations may be found by using single tests which load highly on g; RAPM is one of these tests and is often used in the literature. It's worth noting that g has been found in other animals, e.g., rats (Anderson, 1993).

As Sternberg (2000) writes, 'Spearman's g is only one aspect of intelligence, a fact that even Spearman and other diehard believers in g have acknowledged.' Researchers in the factor analytical tradition recognise a hierarchical tree structure of components to ability, for instance including verbal and visuospatial ability, with g at the root. These more specific abilities cumulatively explain the variance that g does not. Johnson and Bouchard (2007) note that (perhaps unsurprisingly) more specific abilities often predict job success better than g. For instance image rotation predicts success in physical sciences and fine arts better than g. Of relevance to the present study, as we shall see, is their concluding note: 'What has perhaps not been recognized is that inclusion of verbal ability in assessments used to recruit individuals to those fields may actually act to impair efforts to select those with the talents most relevant to the jobs in question.' For discussions about g, theories of what it represents, the dangers of naive reification, and other results from the psychometric intelligence tradition, the reader is directed elsewhere for reviews (e.g., Deary, 2000; Chabris, 2007).

One of the distinctive features of ASC is that those affected show distinctive peaks and troughs in standard tests of ability that one would normally use to compute g. In all but one of the ten studies using Wechsler Intelligence Scales with ASC participants surveyed by F. Happé (1994), the Block Design score was highest (ranging from 8.4–13.5), and in all but one (a different one) Comprehension was the lowest (1.3–6.9). Dawson, Soulières, Gernsbacher, and Mottron (2007) replicated this finding in 38 children with an

ASC. In addition, they tested performance on Raven's Progressive Matrices (RPM). They found that whilst Verbal IQ was at around the 25th percentile (i.e., on average 25% of people would have a lower score) and Performance IQ was at around the 30th percentile, performance on RPM was much higher: at the 56th percentile. They found a similar pattern in 13 adults with an ASC (using the adult version of the task): Verbal IQ was around the 45th percentile, Performance IQ the 55th, and RPM 83rd.

Another peak in performance comes from the embedded figures task. Shah and Frith (1983) discovered that children with an ASC perform better than TD children on the Children's Embedded Figures Test, obtaining a mean score of around 21 (out of 25) versus 16 for the TD group. They also found qualitative differences in behaviour. The children could respond by either pointing to the figure or by placing a cutout shape of the target over the figure in the correct place. The three strategies of solution observed were immediate solution, spotting the appropriate place quickly after the item was displayed; visual search, spending some time looking and then pointing to the appropriate area; and searching with the cutout over the figure in which the shape was embedded. Interestingly, almost all TD children used visual search at least once, compared to only half of the children in the ASC group. There was also a trend ($p = 0.06$) for the children with ASC to display the immediate strategy more than the TD children. There were no group differences found in the proportion of times the cutout search strategy was used.

Baron-Cohen and Hammer (1997) report evidence of faster performance on embedded figures in adults with ASCs than TD adults. The groups with an ASC were a mean of around 30 seconds faster than the TD group. Of importance for the spectrum view, they also report that the parents of children with an ASC who show autistic traits, but with insufficient severity for a clinical diagnosis, were also faster on embedded figures than controls. Affected mothers had a mean response time of around 30 seconds, versus around 65 seconds for female controls; affected fathers had a mean of around 30 seconds versus around 45 for male controls. More recently, Jonge, Kemner, and Engeland (2006) replicated this response time difference in people with an ASC (26 versus 36 seconds). They also found that people with an ASC made fewer incorrect attempts than controls (mean of 0.5 versus 0.9). Fathers of children with an ASC also made fewer incorrect attempts than fathers of children with Down Syndrome (mean of 0.45 versus 0.75).

McKenna (1984) reviews evidence of strong correlations between RPM and the Embedded Figures Test, Children's Embedded Figures Test, and the Group Embedded Figures Test. In a series of 19 studies from 1966–1980, 17 reported positive correlations between $r = 0.5$ to 0.7, one reported a smaller correlation of $r = 0.15$, and there was only a single blemish in 1969 of a negative correlation. McKenna (1984) uses these data to bolster an argument that embedded figures should be considered a measure of cognitive ability and not style. For our purposes, the style versus ability argument is a distraction. More interesting is the evidence that Raven's matrices may be tapping into similar processes to those used when solving embedded figures. More recently, Meo, Roberts, and Marucci (2007, p. 367) again argue that the skills necessary for solving RPM may be related to those necessary for solving the embedded figures (though they seem unaware of empirical evidence of association).

6.1 Qualitatively different strategies in Raven's Matrices

Hunt (1974) argues that there are two qualitatively different ways to solve RPM, developing fairly formal characterisations of the two techniques: a 'Gestalt' algorithm and an 'analytic' algorithm. Both achieve similar scores in Set I of the matrices. Carpenter, Just, and Shell (1990) provide an analysis of solutions in terms of rules which distinguish high and low performers, using data from verbal protocols and eyetracking. Most interestingly for the present study, DeShon, Chan, and Weissbein (1995) sorted the rules for solution from these studies into two broad categories: *visuospatial* and *verbal-analytic*. The visuospatial rules they give (pp. 139–142) are: *superimposition*, *superimposition with cancellation*, *object addition/subtraction*, *movement*, *rotation*, and *mental transformation*. These are mostly self-explanatory. Superimposition refers to features being placed on top of each other, whereas object addition/subtraction refers to whole objects being added or removed in an jigsaw puzzle fashion, with no overlap. Movement and rotation are as you'd expect. Mental transformation is a more complex procedure involving recognising a feature in one object and using it to transform a feature in another object.

The verbal-analytic rules (p. 142) are *constant in row* (same feature across columns, i.e., within a row, different quantity as you move down the rows); *quantitative pairwise progression* (same shape, incremented number); *distribution of three values* (a shape is distributed in number, the same number, across three columns); *distribution of two values* (two shapes are distributed through a row; one shape is not there or 'categorically inconsistent' with the other features).

DeShon et al. (1995) argue that some items may be solved using rules from one class (i.e., the broader visuospatial/verbal-analytic), either of these strategies, and some items require the use of rules of both classes. (See Table 6.1 on p. 126 for the broader classifications of each item.) They used the verbal-overshadowing effect (Schooler & Engstler-Schooler, 1990), the effect that concurrent task relevant verbalisation during the processing of visuospatial stimuli impairs performance, to investigate whether their classifications relate to how people process the items. Indeed, there were differences found in the degree to which the verbalisation affected performance. Participants who were describing how they were solving the items were half as likely to correctly answer the items classified as visual compared to those who were not verbalising. Examining the plot of probabilities of correct response (p. 155), the largest difference between the verbalising and non-verbalising groups was seen for items 7, 9, 10, 22, 24, and 33.

One might wonder whether these qualitative differences in strategy were explicitly designed into the tests. I couldn't find any evidence that they were, though in the first discussion of the development of the matrices by Raven (1936), excerpts from participant self-reports indicated that some individuals were aware of differences in the strategies required for solution. Of one item, a participant—'Dr. Duncker (assistant to Köhler)', later to become famous Gestalt psychologist[3]—noted (p. 112) that 'Verbalisation helps.' Of another item he remarked that 'Parts of the figure are so similar that you cannot distinguish between them by words. Verbalisation sometimes helps, but not so here.' Self-reports can be misleading, but this is a lovely reminder that listening to what our participants tell us, as well as being important in itself in the study of people, can point to ideas for models of processes.[4]

Extensive research has been carried out on the verbal-overshadowing effect more generally. Of particular relevance for the present study is work by Ryan and Schooler (1999),

[3]Karl Duncker (1903–1940) is probably most famous for the candle problem and radiation problem. It appears that when Raven collected the data, Duncker was doing pain research with Bartlett in Cambridge. Sadly Duncker died by suicide when he was 37 after suffering from endogenous depression. See Schnall (1999) for fascinating biographical information.

[4]Another famous participant appears in Raven's thesis: Roger Penrose, aged $4\frac{2}{3}$.

who used a series of measures to test the relationship between verbal and visual ability on the strength of the verbal-overshadowing effect. Here I focus on two measures they used, embedded figures and Grade Point Average (GPA; used as a measure of verbal ability), and how they affect performance on a face recognition task under verbalising and non-verbalising conditions. For those with poorer embedded figures ability and high GPA, there was a trend to do better in the verbalisation condition (around 70% accuracy) than non-verbalisation (around 60%). For those with low GPA and high embedded figures ability, performance on the verbalising condition went down to 50% accuracy, compared to around 75% in the non-verbalising control condition. One might question the validity of using a composite score of many abilities to measure verbal ability, however.

The relationships between ASC and embedded figures, and embedded figures and Raven's matrices, led to the conjecture that there'd be a positive correlation between number of autistic traits and RAPM performance. I also expected that position on the autistic trait continuum would predict better performance on visual items than on the verbal-analytic items.

6.2 Analysis

6.2.1 Properties of composite scores

The mean score was 20 (out of 36; $SD = 4.2$), around the same as that found by Hamel and Schmittmann (2006) ($M = 20.51$, $SD = 3.87$) for their sample of Psychology students at the University of Amsterdam. See Figure 6.1 for histograms of the distribution of responses to the Raven's task.

Table 6.1 shows an analysis of the items; how many were reached by participants[5], and how difficult the items were on average. All 85 participants made it as far as and completed item 13 within the 20 minutes, but from there on there is a slow decline in the number of completed responses in the time available. See Figure 6.2 for a graphical depiction of performance.

No correlation was found between AQ and total RAPM score ($r = 0.05$, $p = 0.7$), or the number of questions answered successfully in the time ($r = 0.1$, $p = 0.36$). Nor was AQ

[5]Recall from Section 4.1.2.6, p. 81, there was a 20 minute time-limit which was not revealed to participants. Also participants could drop out at any time if they found an item was too difficult.

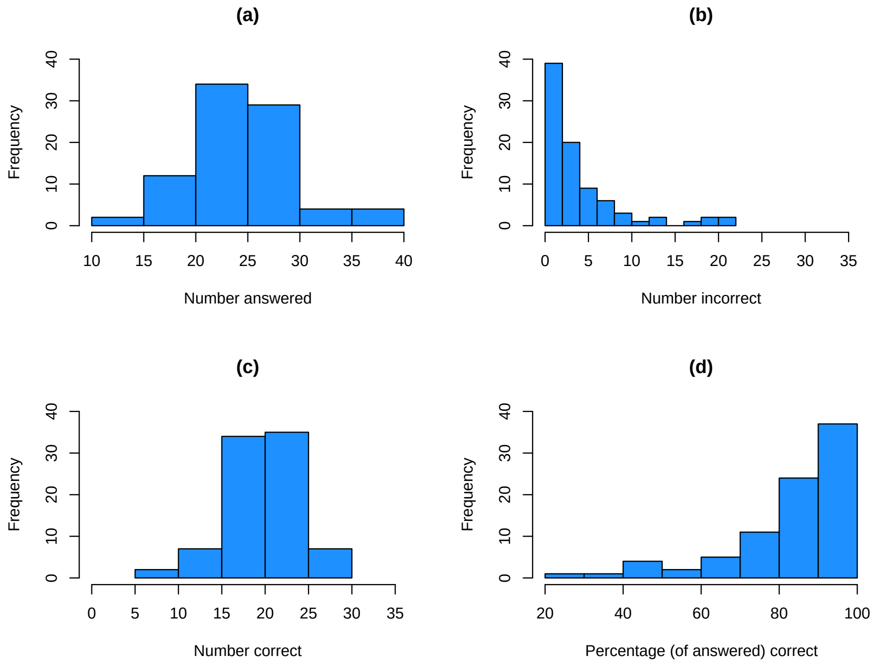

Figure 6.1: Histograms of performance in Raven's Advanced Progressive Matrices.

Item	*Survivors*	*Correct*	*Incorrect*	*% Correct*	*Problem class*
1	85	82	3	96	Analytic
2	85	74	11	87	Either
3	85	81	4	95	Visual
4	85	75	10	88	Analytic
5	85	79	6	93	Either
6	85	84	1	99	Either
7	85	75	10	88	Visual
8	85	79	6	93	Analytic
9	85	74	11	87	Visual
10	85	72	13	85	Visual
11	85	81	4	95	Visual
12	85	79	6	93	Visual
13	85	61	24	72	Analytic
14	84	77	7	92	Either
15	84	77	7	92	Uncodable
16	83	72	11	87	Visual
17	83	72	11	87	Analytic
18	77	57	20	74	Visual
19	75	59	16	79	Both
20	73	56	17	77	Both
21	71	53	18	75	Analytic
22	61	42	19	69	Visual
23	58	42	16	72	Visual
24	47	20	27	43	Visual
25	43	31	12	72	Both
26	37	22	15	59	Both
27	28	12	16	43	Analytic
28	17	6	11	35	Analytic
29	14	5	9	36	Analytic
30	12	4	8	33	Analytic
31	8	2	6	25	Both
32	8	1	7	13	Visual
33	5	3	2	60	Visual
34	5	1	4	20	Analytic
35	4	1	3	25	Both
36	4	0	4	0	Analytic

Table 6.1: Table of number of people answering each item within 20 minutes, performance on the items, and categorisations of the items according to DeShon et al. (1995).

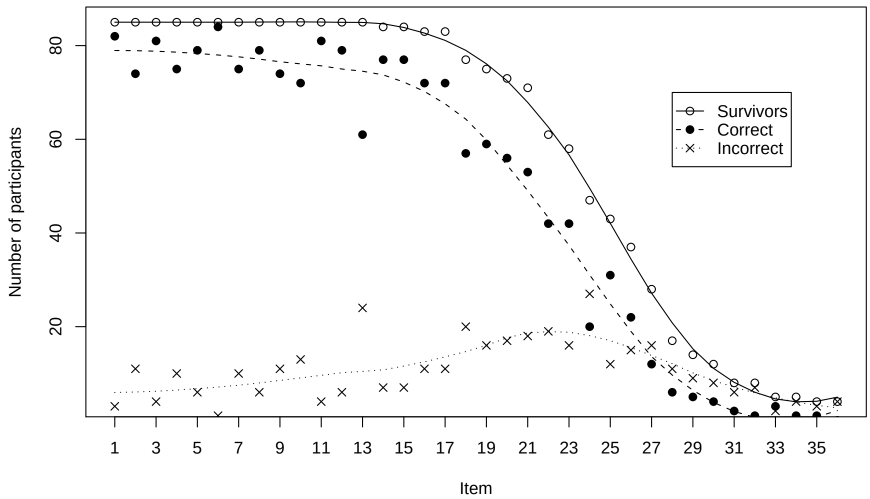

Figure 6.2: Graph of number of people reaching each item, number correct, and number incorrect. (The curves were fitted using local polynomial regression.)

associated with the number of errors ($\rho = 0.007$, $p = 0.95$). Similar (lack) of correlations were also found for BAPQ. At the subscale level, however, there were associations. For AQ, a linear regression of number correct on Social Skills ($\beta = 0.26$, $t = 2.70$, $p = 0.008$) and Attention Switching ($\beta = -0.25$, $t = -2.12$, $p = 0.04$) gave an adjusted $R^2 = 0.07$. For BAPQ, a model with Aloofness ($\beta = 1.68$, $t = 2.92$, $p = 0.005$) and (with a trend for) Rigidity ($\beta = -1.26$, $t = -1.92$, $p = 0.06$) gave an adjusted $R^2 = 0.08$. There was also a bivariate relationship between number correct and Aloofness ($r = 0.261$, $p = 0.02$; see Figure 6.3), but none with any of the AQ subscales.

6.2.2 Item-level analysis

6.2.2.1 Accuracy

Figure 6.4 (p. 129) shows the distributions of responses, split by the classification due to DeShon et al. (1995). The histograms for correctness of visual and analytic items are

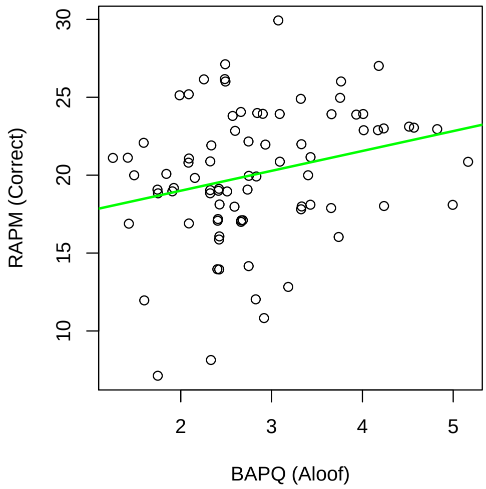

Figure 6.3: Scatterplot of the Aloofness subscale of BAPQ against number of items correct (in 20 minutes) in RAPM. (Points are jittered.)

very nearly Gaussian distributed; for problems requiring both strategies, the distribution is slightly skewed with more responses close to the lower end; for problems solvable by either type, the distribution is markedly skewed: most participants answer four correctly.

First let's check that item number predicts difficulty; a feature designed into the tests. Multilevel logistic regression was used with a varying intercept for participants, fitted using the lme4 package (Bates, Maechler, & Dai, 2008) in R. Fitted models take the form

$$\mathsf{Pr}(\hat{y} = 1) \;=\; \text{logit}^{-1}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)$$

where the β's are estimated from the data. The logit^{-1} function is shown in Figure 6.5. Since responses, j, are clustered within participants, i, we have a random intercept b_i for participants in addition to the random effect of the item level residual term, ϵ. More formally

$$\mathsf{Pr}(y_{ij} = 1) \;=\; \text{logit}^{-1}(\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_n x_{nij} + b_i + \epsilon_{ij})$$

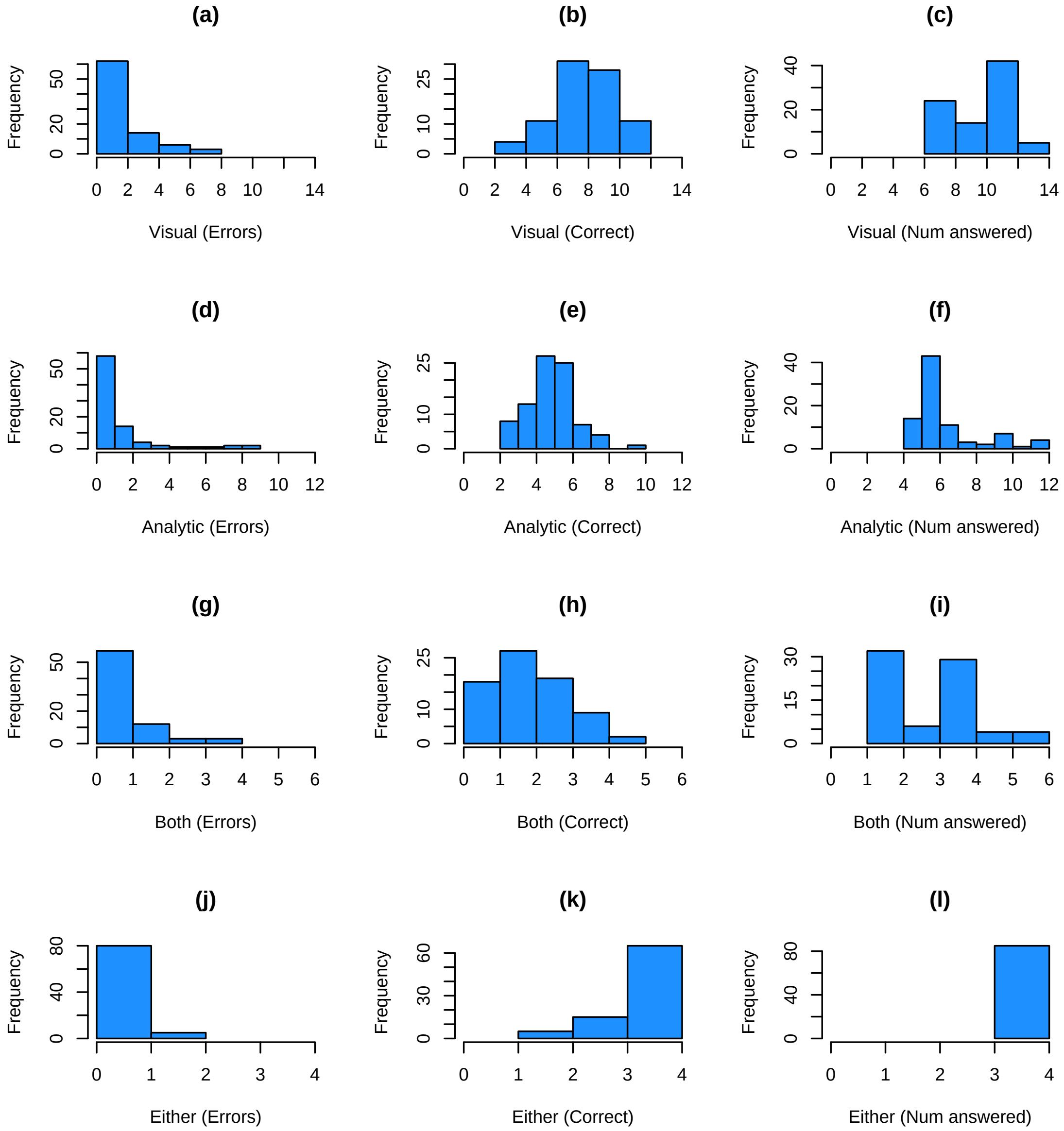

Figure 6.4: Response characteristics in RAPM when items split according to DeShon et al. (1995)

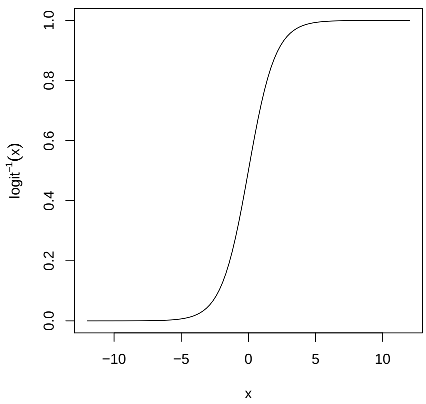

Figure 6.5: The inverse logit function. The slope is steepest, 1/4, when x is close to zero; as x increases, the slope becomes more gentle, thus the greater the increase in x required to lift the probability.

where $b \sim N(0, \sigma_b^2)$ and $\epsilon \sim N(0, \sigma_\epsilon^2)$. Think of this random-effect as being akin to a person-level residual term. For more details, see Faraway (2006); Gelman and Hill (2007).

A model predicting probability of giving a correct answer, with random intercept for participants and fixed effect of item position fitted better than the model without item position [$\chi^2(1) = 191$, $p < 0.001$]. The further into the test an item is, the less likely participants get it correct ($\beta = -0.12$, $z = -12.6$, $p < 0.001$).

We are primarily interested in relating the data back to the (modal) classifications of strategy. We do this by including an additional random intercept for items (to model participant-invariant estimates of difficulty) as well as participants (to model overall participant ability). For the participant random effect, $\sigma^2 = 1.08$, and for the item random intercept, $\sigma^2 = 1.64$. Adding a term for the class improves fit [$\chi^2(3) = 49.8$, $p < 0.001$]. The easiest problems to solve were those that could be solved by either the visual or analytic method, as one would expect. Setting the 'either' type as base category gave the following coefficients: visual $\beta = -1.4$, $z = -2.1$, $p = 0.04$; analytic $\beta = -2.1$, $z = -2.9$, $p = 0.004$; and both $\beta = -2.3$, $z = -2.9$, $p = 0.003$. I also compared visual with analytic, but found no statistically significant difference (visual as the base category, the slope for analytic was $\beta = -0.6$, $z = -1.29$, $p = 0.2$). Problems that required both strategies were only significantly different to those that could be solved using either strategy (coefficient above).

Finally let's add AQ to the model. The interaction between AQ and the class of solution was significant by the log-likelihood ratio test [$\chi^2(3) = 8.89$, $p = 0.03$], by AIC (ΔAIC $= -3$), but not BIC (ΔBIC $= 14$, and we want a negative difference). Recall that BIC penalises more than the other fit statistics for the number of parameters in the model. This split decision is a common occurrence, says the literature, so let us continue. Examining the interaction term slope estimates, 'visual' is significantly different to 'analytic' ($\beta = -0.03$, $z = -2.6$, $p = 0.01$), and to 'either' (-0.04, $z = -2.0$, $p = 0.05$). There was almost a significant difference comparing 'visual' and 'both' (visual as reference, $\beta = -0.03$, $z = -1.8$, $p = 0.07$). No other pairwise comparisons were close to significant (all $p > 0.5$).

Now let us focus on items which were solved by all participants, excluding the first item which was given in the instructions as an example (but, still 3 participants gave an incorrect response), and only on items which distinguish analytic and visual solutions. This leaves 9 items (3 of them analytic). This time the interaction was statistically significant according to all criteria [ΔAIC $= -5$, ΔBIC $= -1$, LLR $\chi^2(1) = 7.14$, $p = 0.008$]. Again the higher participants' AQ, the easier they found visual problems (visual as reference, $\beta = -0.05$, $z = -2.6$, $p = 0.009$).

Figure 6.6 shows the probability of getting the correct answer, for a model fit to all the data, i.e., including verbal, visual, either, and both. Only the verbal and visual conditions are plotted. Recall that individual differences in ability and variation in item difficulty are removed by the random effects.

To explore the effect further and facilitate future work I also fitted individual logistic regressions models for all of the visual and verbal-analytic items, using AQ as a predictor. Table 6.2 shows z-scores for the estimate of the AQ slope for each model. Of the visual part of RAPM, items 9, 22, 23, and 24 showed the largest association with AQ and of the verbal-analytic part, items 8, 13, and 28 showed the largest association, though the direction of the effect for item 28 is the opposite to that expected.

Visual		Verbal-analytic	
Item	z	Item	z
3	−0.78	1	−0.30
7	−0.78	4	−0.90
9	1.19	8	−2.27
10	0.53	13	−1.68
11	0.87	17	0.43
12	0.15	21	−0.43
16	0.36	27	−0.28
18	−0.10	28	1.71
22	2.11	29	−0.29
23	1.48	30	−0.80
24	2.38	34	0.03
32	−1.40	36	0.00
33	−1.42		

Table 6.2: The z-scores of the slope for the AQ predictor for individual logistic regression models predicting accuracy on each item. A positive z-score indicates better performance the higher AQ and a negative z-score indicates worse performance.

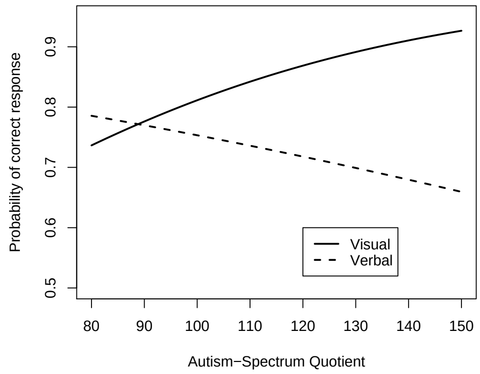

Figure 6.6: The probability of getting the correct answer on verbal and visual problems as a function of AQ, when between-subject variance in ability and between-item variance in difficulty are removed using random effects.

6.2.2.2 Reaction time

The distribution of reaction times was log-normal, as is usual for many tasks. The RTs ranged from 2.5 seconds to nearly 8 minutes, with a median of around 30 seconds and a mean of 44 seconds. For the analyses the RT was logged and a Gaussian mixed effect model fitted.

Given the reaction time results on embedded figures for TD individuals versus people with ASC, we'd expect similar results on RAPM: faster performance on visual items than verbal items, the higher one's AQ. First we analyse the data ignoring accuracy. Focussing on the important interaction between problem type and AQ, there is a significant effect [$\chi^2(3) = 9.02, p = 0.029$], however there is no difference between verbal and visual ($\beta = 0.0017$, MCMC $p = 0.42$). There is however an effect for 'both' versus 'visual': higher AQ implies slower responses for problems requiring both rule types ($\beta = 0.0085$, MCMC $p = 0.004$). A similar effect is found for 'both' versus 'analytic': higher AQ implies slower responses for problems requiring both rule types ($\beta = 0.0068$, MCMC $p = 0.025$). Finally problems which require both rules take more time than those which can be solved by either rule type, the higher AQ ($\beta = 0.007$, MCMC $p = 0.035$). To summarise, problems which require both types of rules take longer the higher AQ than any of the other problems.

Reanalysing with only correct responses (reducing the number of observations from 2058 to 1637) gave almost identical results. Again there was a significant interaction between problem type and AQ [$\chi^2(3) = 9.12, p = 0.028$]. The effect directions were identical to the models fitted to both incorrect and correct solutions so the coefficients won't be reported again.

Finally we test the three-way-interaction between AQ, type of problem, and accuracy. We detected no effect [ΔAIC = 5, ΔBIC = 21, LLR $\chi^2(3) = 1.76$, $p = 0.62$].

6.3 Discussion

6.3.1 Summary

Subscores of AQ and BAPQ, but not the composite scores, predict overall performance in the matrices. The strongest association comes from the Aloofness subscale of BAPQ. There was a small but significant effect with the subscores of AQ: reporting poorer social skills and better attention switching predicted better overall performance on RAPM. The same was found with BAPQ and aloofness and rigidity. The items consist of a series of different shapes, and require different strategies, so the association with attention switching might indicate perseveration on one particular strategy, which is likely to impair overall performance. It is not so clear, however, why poorer social skills would be associated with better performance. Perhaps those who are strongly motivated to solve these kinds of problems tend to prefer doing so over socialising with others. However this level of explanation does not translate easily to the cognitive level. The social skills subscale of AQ may best detect traits associated with weak central coherence which has been hypothesized (see Section 2.2.3) to underlie better performance on tasks like embedded figures where there are possible Gestalt effects which would impair performance, and also social skills impairment.

The most noteworthy result is that the higher an individual's AQ, the easier they find visual items and the harder they find verbal-analytic problems. I noted in the introduction that a series of studies have discovered correlations between RSPM and embedded figures.

Perhaps the processes required for solving the visual items of RAPM are similar to those needed for embedded figures. It could be argued that the visual items are just easier than the other items and this is why we found the result with AQ. To address this, I tested whether they were easier and found no difference in accuracy. I also included a random intercept for items thus attempted statistically to control for overall difficulty. This suggests that the effect is not due to overall difficulty.

One aspect of the results which is somewhat confusing, though a weaker effect than the others, is that problems which can be solved using either a verbal or a visual strategy were also harder than items requiring only a visual strategy, the higher AQ. Speculatively, this could perhaps be due to interference between attempted strategies. Or perhaps some participants with higher AQ sometimes attempted and failed to solve the items using a verbal-analytic strategy.

There are some limitations to the analysis. The main limitation of these data is that not all items in RAPM were seen by all participants. However this problem is solved to a certain extent by the fact that we perform an item-level analysis, with residuals at the level of items to control for subject-invariant between-item difficulty. Also the interaction between AQ and hypothesised item strategy is still found when attention is restricted to those items which are seen by all participants.

What kind of reasoning is used for RAPM? Participants perhaps draw inductive inferences from the first two rows and then use these inferences to transform the remaining cells into a putative conclusion—inductive in the sense that one could imagine more rows would reveal that the conclusion ought to be rejected. Perhaps a selection of deductive inference rules are available for manipulating elements of the items in various ways. Carpenter et al. (1990) argue (p. 415) that mostly goal generation and maintenance processes, deductive processes, perhaps, are required for reasoning about the items. They also note that distance from perceptual input is an important factor in the items, with for instance items easily solved by children tending to be closer to perceptual input, and more difficult items tending to be (p. 428) 'more abstract [...] in terms of dimensions and attributes.' The experiment by DeShon et al. (1995) elaborates on these comments relating to perception, with perhaps verbal-analytic processes relating more to goal generation and maintenance. Lovett, Forbus, and Usher (2007) modelled performance on the easier items of SPM using a domain-general model of analogy, which includes a sophisticated perceptual component. They note that additional stages to the model are likely to be required for solving more abstract problems. So now we have mentioned induction, deduction, and analogy. Which is it to be? A combination of the three? The story becomes even more complicated when one considers that there are deductive theories of induction and analogy. In Chapter 10 we uncover evidence of connections with the other reasoning tasks, which may help contribute to an understanding of the reasoning involved.

6.3.2 A methodological digression

Deary (2000, pp. 127–130) gives a critique of the model proposed by Carpenter et al. (1990), whose work forms an important part of the theoretical basis of the work by DeShon et al. (1995). It will be worthwhile addressing this criticism as it deals with more general concerns of how to interpret the result we have found.

Firstly, he argues that the analysis by Carpenter et al. (1990) fails to reduce a description to the 'brain's functional elements'. There are multiple levels of analysis used in the brain and behavioural sciences. It's not clear what counts as a 'functional element' in the brain. Should we be dissatisfied with any explanation which fails to relate to summed Event Related Potentials? Or vesicle release?

Secondly, he argues that 'the tests are looked to for a model of mind rather than the brain itself being interrogated'. Importantly it isn't just the tests that are examined, but rather people's *performance* when solving the tests. The key point is that features of the stimulus are examined, then used to predict variation in responses. He argues further that models of Raven's performance are 'a kind of elaborate restatement of John Raven's test building principles.' This seems true, but need not be problematic. Similarly formulae such as $s = ut + 1/2at^2$ are an elaborate and useful restatement of obvious relationships.

Deary allows that these analyses 'can be helpful in constructing and improving tests and items,' but do not, he argues, 'afford a breakthrough to a new level of description that can explain average test performance of test differences.' If the analysis improves tests, then this necessarily implies that the analysis reveals something about the workings of the brain. Importantly, if the analyses can be integrated with broader conceptualisations of cognition, then they must have some utility.

The analysis reported in this chapter shows that a characterisation of the different processes involved in solving items of Raven's matrices can be connected to variation in autistic-traits. This shows that the item-level distinctions do map to broader differences of interest.

Chapter 7

Interpreting Conditionals

'If my mother had wheels she'd be a bicycle.'

David Cameron, current leader of the UK Conservative Party
(The Independent, 2008)

7.1 Introduction

Suppose you are presented with a conditional, 'if P_1, then Q', and one atom P_1, $\neg P_1$, Q, or $\neg Q$. Modus ponens (MP) is typically drawn by around 95–100% of participants, modus tollens (MT) by 60–75%, denial of the antecedent (DA) by 50–70%, and affirmation of the consequent (AC) by 35–75% (see Evans, Newstead, & Byrne, 1993, p. 43). (See Figure 7.1 to recall these terms; remembering that they only make sense in the context of the naive classical interpretation of the natural language conditional: 'if P, then Q' $\rightsquigarrow P \Rightarrow Q$, for the classical material conditional, $\Rightarrow$.) The *suppression effect* occurs when a second conditional expression 'if P_2, then Q' is introduced to the set of assumptions and reduces the frequency with which MP, MT, DA, AC inferences are drawn. From the viewpoint of naive translation of the sentence into classical logic, this is very odd indeed.

Different classes of added conditional induce different types of suppression effects. MP and MT are typically suppressed by conditionals which express (what can be interpreted as) an additional condition for Q to hold. For instance take the sentences:

(7.1) If she has an essay to finish [P_1] then she will study late in the library [Q].

$$\frac{P \qquad P \Rightarrow Q}{Q}\ \mathsf{MP}\ \text{(classically valid)} \qquad \frac{P \Rightarrow Q \qquad \neg Q}{\neg P}\ \mathsf{MT}\ \text{(classically valid)}$$

$$\frac{\neg P \qquad P \Rightarrow Q}{Q}\ \mathsf{DA}\ \text{(classically invalid)} \qquad \frac{P \Rightarrow Q \qquad Q}{\neg P}\ \mathsf{AC}\ \text{(classically invalid)}$$

Figure 7.1: Valid inferences (modus ponens [MP] and modus tollens [MT]) and invalid inferences (denial of the antecedent [DA] and affirmation of the consequent [AC]) in classical propositional logic

(7.2) If the library stays open [P_2] then she will study late in the library [Q].

(7.3) She has an essay to finish [P_1].

The classical response would be Q, however many participants fail to draw this inference. For instance in the between-participant study by Byrne (1989), 96% of responses by participants in the simple group drew MP compared to 38% of responses from the additional group. If sentence (7.3) is replaced with:

(7.4) She will not study late in the library [$\neg Q$].

then the classical answer is now $\neg P$ by modus tollens.[1] However Byrne (1989) found 33% of participants drew this inference, compared to 92% for the simple condition.

AC and DA inferences are suppressed by conditionals which express an alternative condition for Q to hold. Consider the following sentences:

(7.5) If she has an essay to finish [P_1] then she will study late in the library [Q].

(7.6) If she has some textbooks to read [P_2] then she will study late in the library [Q].

(7.7) She doesn't have an essay to finish [$\neg P_1$].

In the experiment by Byrne (1989), 46% of participants drew DA for the simple condition, compared to only 4% in the alternative condition. To test drawing of AC, sentence (7.7) is replaced with:

[1] There are at least three different classical inferences that may be drawn: $\neg P_1$, $\neg P_2$, or $\neg P_1 \wedge \neg P_2$, but the first experiment showing this effect (Byrne, 1989) did not provide these alternatives.

(7.8) She will study late in the library [Q].

In the data by Byrne (1989), 13% of responses were AC, compared to 71% in the simple condition.

7.1.1 An alternative interpretation of the conditional and data

Stenning and van Lambalgen (2005) give an alternative interpretation of the natural language conditional, and the data, using logic programming with negation as failure (LP-NAF). LP-NAF provides an efficient way to model change of interpretation in logic. Classical logic assumes that adding premises does not change the inferences that may be drawn, a property known as monotonicity. LP-NAF, on the other hand, is a non-monotonic logic, in the sense that adding more premises may cause previously drawn inferences to be retracted. This logic naturally models the 'suppression' as dependencies between the sentences, as we shall see. First we need some definitions.

Definition 7.1 *A* ***definite clause*** *is a formula of the form* $\varphi_1 \wedge \varphi_2 \wedge \ldots \wedge \varphi_n \rightarrow \psi$*, where* φ_i *(for* $i \in 1 \ldots n$*) is (a) a propositional variable, (b) the negation of a propositional variable, or (c) verum (*$\top$*) or falsum (*$\bot$*), the true and false constants, respectively; and* ψ *is a propositional variable.*

Definition 7.2 *A* ***definite logic program*** *is a finite set of definite clauses.*

We also require is a characterisation of under what conditions ψ follows P with LP-NAF, written $P \mid\!\sim \psi$. Stenning and van Lambalgen (2005, p. 937) provide this using a translation into classical logic as follows.

Definition 7.3 *The* ***completion of program*** P*, comp(*P*) is a set, elements in which are given by translating each clause in* P *with the same consequent,* $\varphi_i \rightarrow \psi$*, into the classical disjunction of classical biconditionals:*

$$\varphi_1 \vee \varphi_2 \vee \ldots \vee \varphi_n \Leftrightarrow \psi$$

Now we can define $P \mid\!\sim \psi$ if and only if $comp(P) \models \psi$, equivalently $comp(P) \vdash \psi$ for the classical logical consequence relation, $\vdash$. There are more efficient ways than this of drawing inferences in LP-NAF, also discussed by the authors, but here we concentrate on characterizing final outputs using the more familiar world of classical logic as a metalanguage.[2]

Stenning and van Lambalgen (2005, pp. 942–945) show how commonly drawn inferences of MP, MT, AC, DA, and their suppression effects, may be modelled in this framework. Key is their translation of the English conditional 'if A, then B' into the definite clause

$$A \wedge \neg ab \rightarrow B$$

where $\neg ab$ expresses that there are no abnormalities. Examples will make clear how this applies.

7.1.1.1 Modus ponens

The simple MP is given by the program

$$\begin{array}{rcl} P \wedge \neg ab & \rightarrow & R \\ \top & \rightarrow & P \\ \bot & \rightarrow & ab \end{array}$$

The semantics of the LP-NAF means that $\top \rightarrow \varphi$ represents φ and $\bot \rightarrow \varphi$ represents $\neg\varphi$. (This is written here explicitly to make the translation transparent.) The completion set is $\{P \wedge \neg ab \Leftrightarrow R, \top \Leftrightarrow P, \bot \Leftrightarrow ab\}$. This simplifies to $\{P \Leftrightarrow R, \top \Leftrightarrow P, \bot \Leftrightarrow ab\}$, from which R follows.

The suppression case, to quote the authors' footnote 26, 'is one place where general knowledge of content enters into the selection of appropriate logical form.' They propose the following formalisation of the logical interpretation of participants who suppress MP

[2]Makinson (2005) develops various systems of non-monotonic language by systematically breaking classical logic in various ways. The reader might find it amusing to note that classical logic provers can be implemented in logic programming (e.g., Beckert & Posegga, 1995).

when an additional conditional is added:

$$\begin{aligned} P \wedge \neg ab_1 &\rightarrow R \\ Q \wedge \neg ab_2 &\rightarrow R \\ \top &\rightarrow P \\ \bot &\rightarrow ab_1 \\ \bot &\rightarrow ab_2 \\ \neg P &\rightarrow ab_2 \\ \neg Q &\rightarrow ab_1 \end{aligned}$$

Note how the two clauses provided in the task, $P \wedge \neg ab_1 \rightarrow R$ and $Q \wedge \neg ab_2 \rightarrow R$, have been made interdependent using the clauses $\neg P \rightarrow ab_2$ and $\neg Q \rightarrow ab_1$. The completion procedure gives the set:

$$\{(P \wedge \neg ab_1) \vee (Q \wedge \neg ab_2) \Leftrightarrow R,\ P,\ \bot \vee \neg Q \Leftrightarrow ab_1,\ \bot \vee \neg P \Leftrightarrow ab_2\}$$

Let's simplify. In general $\bot \vee \varphi \equiv \varphi$, which simplifies the set to:

$$\{(P \wedge \neg ab_1) \vee (Q \wedge \neg ab_2) \Leftrightarrow R,\ P,\ \neg Q \Leftrightarrow ab_1,\ \neg P \Leftrightarrow ab_2\}$$

Rewriting some more using the biconditionals, we can conclude $Q \Leftrightarrow R$.

7.1.1.2 Denial of the antecedent

Simple DA is given by the program

$$\begin{aligned} P \wedge \neg ab &\rightarrow R \\ \bot &\rightarrow P \\ \bot &\rightarrow ab \end{aligned}$$

The completion set is $\{P \wedge \neg ab \Leftrightarrow R, \bot \Leftrightarrow P, \bot \Leftrightarrow ab\}$, from which $\neg R$ follows. An 'alternative' conditional doesn't provide an obstacle to R, rather, as the label suggests, it suggests an alternative means for R to hold. The authors model this as

$$\begin{aligned} P \wedge \neg ab_1 &\rightarrow R \\ Q \wedge \neg ab_2 &\rightarrow R \\ \bot &\rightarrow P \\ \bot &\rightarrow ab_1 \\ \bot &\rightarrow ab_2 \end{aligned}$$

For this program the completion set is

$$\{P \wedge \neg ab_1 \vee Q \wedge \neg ab_2 \Leftrightarrow R, \bot \Leftrightarrow P, \bot \Leftrightarrow ab_1, \bot \Leftrightarrow ab_1\}$$

Now $\neg R$ doesn't follow but $Q \Leftrightarrow R$ does.

7.1.1.3 Modus tollens

The simple case of MT seems very similar to the MP so we try encoding analogously:

$$\begin{aligned} P \wedge \neg ab &\rightarrow R \\ \bot &\rightarrow R \\ \bot &\rightarrow ab \end{aligned}$$

The completion set is $\{(P \wedge \neg ab) \vee \bot \Leftrightarrow R, \bot \Leftrightarrow ab\}$. This simplifies to $\{P \Leftrightarrow R, \bot \Leftrightarrow ab\}$ from which $P \Leftrightarrow R$ follows, i.e., suppression when we do not want suppression as the empirical data show that people do often draw MT.

Stenning and van Lambalgen (2005) instead model MT using integrity constraints, constraints that must be satisfied by a program. This adds a computational asymmetry to the model (which is reflected in the connectionist network implementation of the logic). First, some more terminology: $?\varphi$ is a query which can have one of two values: succeeds or fails. $?\varphi$ succeeds if and only if $comp(P) \models \varphi$ and $?\varphi$ fails if and only if $comp(P) \models \neg\varphi$.

For the simple MT case, we start with the program

$$\begin{aligned} P \wedge \neg ab &\rightarrow R \\ \bot &\rightarrow ab \end{aligned}$$

and the integrity constraint that $?R$ fails. This can only be the case if at least one of P or $\neg ab$ is false. We know that $\neg ab$ is true, so R must be false: successful MT.

The additional case is represented:

$$\begin{aligned} P \wedge \neg ab_1 &\rightarrow R \\ Q \wedge \neg ab_2 &\rightarrow R \\ \bot &\rightarrow ab_1 \\ \bot &\rightarrow ab_2 \\ \neg P &\rightarrow ab_2 \\ \neg Q &\rightarrow ab_1 \end{aligned}$$

The completion set is $\{(P \wedge \neg ab_1) \vee (Q \wedge \neg ab_2) \Leftrightarrow R, \neg P \Leftrightarrow ab_2, \neg Q \Leftrightarrow ab_1\}$ which simplifies to $\{P \wedge Q \Leftrightarrow R\}$. Again we start with the integrity constraint that $?R$ fails. This is satisfiable if either P or Q is false—we do not know which so we get suppression.

7.1.1.4 Affirmation of the consequent

As for MT, AC requires integrity constraints to go through. For completeness, let's examine what happens if we try without the additional machinery:

$$\begin{aligned} P \wedge \neg ab &\rightarrow R \\ \top &\rightarrow R \\ \bot &\rightarrow ab \end{aligned}$$

The completion set is $\{P \vee \top \Leftrightarrow R\}$ from which R follows but we can't conclude anything about P. Let's try again with the following program:

$$\begin{aligned} P \wedge \neg ab &\rightarrow R \\ \bot &\rightarrow ab \end{aligned}$$

and the integrity constraint that $?R$ succeeds. This follows only if P is true: successful AC.

As for DA, suppression occurs in the presence of an ‘alternative’ conditional. This is coded as:

$$\begin{aligned} P \wedge \neg ab_1 &\rightarrow R \\ Q \wedge \neg ab_2 &\rightarrow R \\ \bot &\rightarrow ab_1 \\ \bot &\rightarrow ab_2 \end{aligned}$$

And the integrity constraint that $?R$ succeeds. For this case the completion set is $\{P \vee Q \Leftrightarrow R\}$. We know only that $P \vee Q$ must be true, not which of the disjuncts, thus the previously inferable P is now suppressed.

7.1.2 The suppression effect in autism

The suppression effect was first investigated in autism spectrum conditions by van Lambalgen and Smid (2004). They investigated performance on a range of tasks using tutorial dialogue in six participants. None of the participants suppressed MP; all 4 who drew MT in the simple condition suppressed MT in the two-conditional condition; all 4 who drew AC, suppressed; and of the 5 who drew DA, 3 suppressed. No data from a TD comparison group are reported, nevertheless this data collection gave reason to believe that MP would not be suppressed as frequently by those with an ASC compared to TD individuals.

Pijnacker et al. (In press) investigated this effect in a larger sample of young adults (11 participants with autistic disorder, 17 with Asperger syndrome, and matched controls), using a 4×3 design (inference type, e.g., MP, crossed with type of second conditional, or absence), with 10 items per condition, giving a total of 120 items.[3] The simple task always came first. Participants were asked to respond ‘yes’, ‘no’, or ‘maybe’ to conclusions phrased as questions, e.g., in the MP simple condition, the task might be to consider, given the appropriate premises: ‘Will she study late in the library?’[4] (The ‘maybe’ response is analogous to the ‘could be true or false’ response used in immediate inference—see

[3] The reader may find these details excruciating, but the subtleties turn out to matter.

[4] The questions were always positive, so an MT response was indicated by a ‘no’ response to, e.g., ‘does she have an essay?’ rather than a ‘yes’ to (something like) ‘is it not the case that she has an essay?’, which is semantically troublesome. Similarly for DA as again the response requires agreement with a negative conclusion.

Chapter 8.) Let's take the important inference types in turn. There were no significant differences in MP inferences between groups for simple items, however the ASC group showed more MP inferences when there was an additional conditional (70% vs. 50% of responses). The TDs were more likely to draw a 'maybe' inference. There were no differences for simple MT, however there was an almost statistically significant trend for less suppression of MT in presence of an additional conditional by the ASC group (62% vs. 50% of responses). Recall that van Lambalgen and Smid (2004) found everyone who drew MT in the sample case suppressed it in the additional case, however they did have a smaller sample than Pijnacker et al. (In press). There were no differences in suppression between groups for AC or DA.

7.1.3 A related effect in psychosis proneness?

Sellen, Oaksford, and Gray (2005) investigated the relationship between a measure of psychosis-proneness, the Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE; Mason, Claridge, & Jackson, 1995), and the effect of possible counterexamples for causal conditional statements in a non-clinical population. The stimuli consisted of conditionals with few or many alternative causes and few or many disabling conditions, attitudes towards which were determined experimentally by Cummins (1995). Examples are given below of each case:

(7.9) If fertilizer was put on the plants, then they grew quickly [many alternatives, many disablers].

(7.10) If Alvin read without his glasses, then he got a headache [many alternatives, few disablers].

(7.11) If the match was struck, then it lit [few alternatives, many disablers].

(7.12) If Joe cut his finger, then it bled [few alternatives, few disablers].

These were then combined with atomic propositions (e.g., 'the match was struck') to test MP, MT, AC, DA. An effect very similar to the suppression effect was found by Cummins (1995) and replicated by Sellen et al. (2005): AC and DA inferences were influenced by the number of alternative causes and MP and MT were influenced by the number of disabling conditions.

Interestingly, the authors found that score on the Impulsive Nonconformity subscale of OLIFE predicted a lower disabling conditions index, computed as the difference in number of MP/MT inferences drawn for conditionals with many disabling conditions versus few disabling conditions. The same subscale predicted a trend for lower score on the alternative causes index, computed as the difference in number of AC/DA inferences drawn for conditionals with many alternative causes versus few alternative causes.

This is a very similar effect to that found in ASC, discussed above. Also interesting is that Tsakanikos and Reed (2003) show performance on a hidden figures test was positively correlated with the same subscale of O-LIFE, Impulsive Nonconformity. People with ASC also tend to be better at hidden figures type tests (see Chapter 6), thus this is strong evidence of overlap between psychosis proneness and autistic traits.

Investigating the differences and similarities between ASCs and schizophrenia is an active area of research. Bölte, Rudolf, and Poustka (2002), for instance, used the subscales of the Wechsler tests and discovered that people with ASC did better than people with schizophrenia on the similarities subtest, whereas people with schizophrenia did better on comprehension than did people with an ASC. An examination of the other Wechsler subtests shows much overlap between ASC and schizophrenia in mean ability scores.

7.1.4 The classic conditional logic task: Wason's Selection Task

A number of different versions of the selection task have been used. In the data collection I included only the traditional task using what is sometimes called 'abstract' content as it shows most variability in response. One difference from the standard instructions is that participants were not asked only to turn the cards they really needed to turn: again in the hope of increasing between-participant variability. Mention of the selection task often elicits a groan from reasoning researchers. Sperber and Girotto (2002, p. 289) famously wrote:

> . . . much of the work done with the selection task should be considered a sunk cost in the history of the psychology of reasoning and further investments of research effort and journal pages in uses of the task should be discouraged.

	Simple	Additional
MP	76	34
Guarded MP	3	35
Silence	3	6
Other	2	9
n	84	84

Table 7.1: Modus ponens responses

Given the speed with which the task may be administered, and curiosity about how it relates to other reasoning tasks, it was included anyway in the data collection.[5]

7.2 Analysis of the Suppression Task Data

7.2.1 Coding the sentences

Recall that responses were freely typed text. In this section the coding of the sentences produced is described. The different sentences test particular inferences: some classically valid (MP, MT), some not (AC, DA). As has been shown, these responses can also be interpreted against the LP-NAF model. Participants' responses were assessed according to the degree to which they drew one of the classical inferences. Also the possible CWR responses as predicted by the LP-NAF model were recorded separately.

7.2.1.1 Modus ponens

Table 7.1 shows the frequencies of each response type. Modus ponens responses drop clearly for the additional condition, most often being replaced with a guarded modus ponens, e.g., 'so she will study late in the library, if it remains open'. The miscellaneous other cases were responses which could not easily reveal views of modus ponens, e.g., 'She will finish her essay' and 'She will not finish her essay'. Of the 76 participants who drew

	Simple	Additional
MT (no essay)	62	16
MT (library closed)		14
MT (conjunction)		6
Disjunction		23
Maybe MT (no essay)	2	1
Maybe MT (both)		0
Maybe MT (library closed)		0
Not MT	2	0
Silence	3	9
Other	15	15
n	84	84

Table 7.2: Modus tollens responses

MP for the simple condition, only 32 (42%) drew MP for the additional condition.

7.2.1.2 Modus tollens

See Table 7.2. As for MP, some responses didn't bear on whether an MT inference was drawn, e.g., the response 'Since the library doesn't stay open late she will not finish her essay'. The antecedent describes that she has an essay to finish, not whether she finishes it. Whereas 'she has finished her essay' does bear on MT, as does 'she has no essay'. Twenty-three participants responded with (something like) 'She has no essay to finish or the library closes early'. This was recorded as a disjunctive response. Of the 62 participants who drew MT for the simple condition, 4 (6%) drew the conjunctive MT for the additional condition, 11 (18%) drew MT and concluded the library was closed, and 15 (25%) drew MT to infer that she has no essay to finish. There were 20 participants (32%) who drew the disjunctive, that either she has no essay or the library is closed.

[5]To become a believer again in the selection task, try asking a bunch of undergraduate philosophy students who recently have taken a course in classical logic to solve the problem in the presence of their

	Simple	Alternative
DA	58	17
Maybe DA	11	0
Guarded DA (doesn't study late)	1	1
Guarded DA (studies late)	0	38
Not DA	0	5
Maybe Not DA	0	3
Silence	10	7
Other	4	13
n	84	84

Table 7.3: Denial of the antecedent responses

7.2.1.3 Denial of the antecedent

Table 7.3 shows the response types. Nearly 70% of participants drew DA for the simple case, with the majority of the others drawing a 'maybe' DA or not responding. There was a drop in DAs for the alternative condition with most responses now being guarded, e.g., 'well then if she has some textbooks to read then she she [sic] will study late in the library' and 'she might study late in the library if she has textbooks'. Of the 58 who drew DA in the simple condition, 25 (around 40% of these) drew a guarded responses in the alternative condition, and 15 (around 25%) again drew DA.

7.2.1.4 Affirmation of the consequent

See Table 7.4 for the responses. For the simple case, the most common response was 'she has an essay to finish'. For the alternative case, most participants responded with (something like) 'she has an essay to finish or some textbooks to read' or 'she either has an essay to finish or textbooks to read or even both'. Of the 62 who drew DA in the simple condition, 36 (around 60%) of these drew the disjunctive; the next majority with

tutor, preferably a tutor who believes they'll provide the classical competence answer.

	Simple	Alternative
AC (Essay)	62	2
AC (Both)		9
Maybe AC	5	0
Not AC	1	0
Not AC (Disjunction)		43
Silence	5	14
Other	11	14

Table 7.4: Affirmation of the consequent responses

10 (15%) was silence; and only 9 (15%) drew some form of DA.

7.2.2 Exploring the closed world reasoning model

For the simple cases of MP and MT, classical interpretations are indistinguishable from CWR: 90% of people drew MP and 73% of people drew MT. However for the simple cases of DA and AC the two reasoning types are distinguishable. For DA, 13% of responses were consistent with a classical interpretation and 68% were consistent with CWR (i.e., they drew DA when classically they shouldn't). For AC, 6% of responses were classical and 73% were consistent with CWR (again they drew DA when classically they shouldn't).

To aid the reader, a summary of Section 7.1.1 is reiterated here. The following inferences are predicted by CWR for the additional and alternative cases. Suppose the two conditionals provided are:

If P, then R
If Q, then R

Then in the MP additional case and DA alternative case, CWR will result in *if Q then R*; MT (additional) and AC (alternative) will give *P or Q*. In the present dataset, the proportions were: MP 41%, MT 23%, DA 45%, AC 51%.

How consistent are people cross-item in having a CWR interpretation? To the best of my knowledge this has not previously been analysed in the literature. Figure 7.2 shows

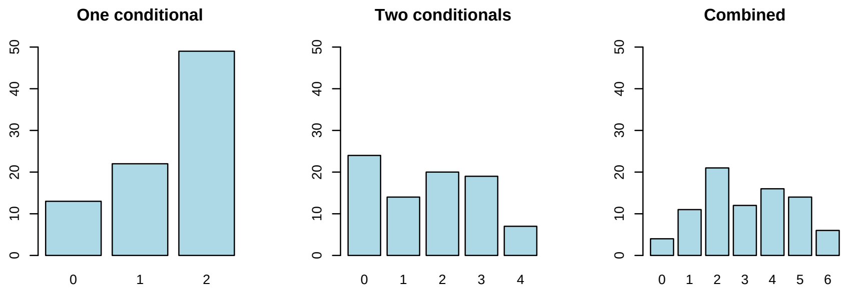

Figure 7.2: Distribution of number of CWR interpretations overall for one conditional (simple AC and DA), two conditional (MP, MT, DA, AC), and combined score.

the distribution of the number of interpretations. For one conditional, most people are consistent with CWR. For two conditionals, there are peaks at 0, 2, and 3. Table 7.5 gives contingency tables showing all two-way relationships in CWR for the additional and alternative cases. The most common two inferences drawn together by the same person were DA and AC (around a third of participants), both of which have alternative premises. The least common pair was MT and DA (16% of participants), a backwards negative premise with an additional premise and a forward positive premise with an alternative premise. MP and DA (both forward) were drawn by 27% of participants and MT and AC (both backwards) were drawn by 24% of participants (the largest association with MT).

Correlations are more revealing (see Table 7.6). The strongest correlation is between MT and AC, both reverse inferences. The two simple inferences (DA and AC) are correlated as are the two alternative (again DA and AC) items. Finally observe the correlation between the two forward inferences (DA and MP). The weakest correlation (tetrachoric $r = 0.02$) is between MP (forward, additional) and simple AC (reverse, simple). Also the correlations between simple DA and the two alternative (DA and AC) items is low (tetrachoric r's $= -0.1$ and 0.1 respectively).

	AC					
DA	no	yes				
yes	13	32				
no	36	19				
			DA			
MP	no	yes	no	yes		
yes	14	27	16	26		
no	35	24	39	19		
					MP	
MT	no	yes	no	yes	no	yes
yes	4	24	12	16	11	17
no	45	27	43	30	48	25

Table 7.5: Percentages of people drawing each possible pair of inferences for the four inference types with two conditional premises. MP and MT have an additional premise whereas DA and AC have an alternative premise. View this as a bivariate correlation matrix for frequencies. The diagonal within each 2×2 cell with the largest numbers indicates the direction of the correlation; the difference between the numbers in this diagonal versus the surrounding cells is indicative of the strength of the correlation.

	MP (add)	MT (add)	DA (alt)	AC (alt)	DA (simp)
MT (add)	.39				
DA (alt)	.46	.23			
AC (alt)	.38	.70	.54		
DA (simp)	.16	.44	−.11	.11	
AC (simp)	.02	.36	.27	.38	.57

Table 7.6: Tetrachoric correlations between the six binary variables representing whether the responses were consistent with CWR.

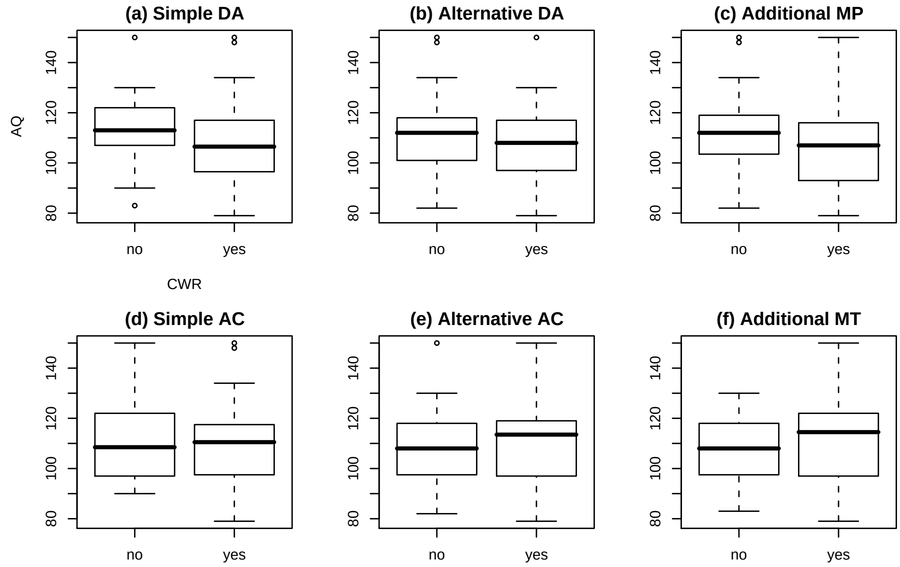

Figure 7.3: Boxplots of AQ scores as a function of whether or not each participant's inference was consistent with CWR. Plots (a) to (c) are the forward inferences and (d) to (f) the reverse inferences.

7.2.3 Autistic traits

Do these characterisations of the interpretation reasoned to explain any of the variation in autistic-like traits? Examination of boxplots of AQ against each of the CWR inferences (Figure 7.3) revealed two distinct patterns: those who are consistent with CWR for forward inferences have lower median AQ and those who are consistent with CWR for reverse inferences have higher median AQ.

Two scores were calculated to represent these patterns: CWR forward and reverse. (See Figure 7.4 for their distribution.) These were positively correlated ($\rho = 0.45$, $p < 0.001$). A linear regression of AQ on the two scores [$F(2, 79) = 5.31$, $p = 0.01$, adjusted $R^2 = 0.09$] supported the eyeballing: a higher CWR forward score predicted lower AQ ($\beta = -5.6$, $SE = 1.9$, $p = 0.003$) and a higher CWR reverse score predicted (a not quite statistically significant trend for) higher AQ ($\beta = 3.2$, $SE = 1.7$, $p = 0.07$).

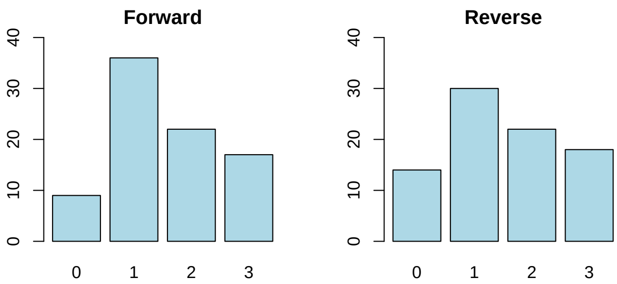

Figure 7.4: Histograms of CWR forward and reverse scores.

There is evidence of differential involvement at the subscale level for both AQ and BAPQ. To explore this, a series of linear regressions were fitted with each of the subscales as outcome variable. (See Table 7.7 for the parameter estimates.) For AQ, only social skills [$F(2, 81) = 6.71$, $p = 0.0021$, adjusted $R^2 = 0.12$] and communication [$F(2, 81) = 4.97$, $p = 0.0091$, adjusted $R^2 = 0.09$] could be predicted by the CWR scores.[6] The largest effect found was for the model with the pragmatic language part of BAPQ as outcome variable [$F(2, 81) = 8.65$, $p < 0.001$, adjusted $R^2 = 0.16$]. See Figure 7.5 for a boxplot of the CWR forward score against pragmatic language.

These data point to a slightly different result to that found by Pijnacker et al. (In press), which calls for some discussion. The summary of their result is that those in the ASC group were less likely to be consistent with the closed-world reasoning model for MP and MT additional inferences and there was no difference between the groups for DA and AC alternative. The summary of our result is that forward and reverse inferences is the biggest distinction: better social skills predict fewer CWR-consistent reverse inferences; better pragmatic/communication skills predict more CWR-consistent forward inferences.

One vital difference is that in our experiment, participants were presented with one or two conditionals and an atom, and were asked to respond with what follows using a textbox. Pijnacker et al. (In press) presented one or two conditional and then a question, the latter always positively worded, and the response was one of ‘yes’, ‘no’, or ‘maybe’. The positive wording might affect the range of strategies that could be utilised by the

[6]The F’s < 1.2 and p’s > 0.3 for the other variables.

(a) AQ (Social skills)

	β	SE	t	p
CWR forward	−2.1	0.6	−3.3	0.001
CWR reverse	1.7	0.6	2.9	0.005

(b) AQ (Communication)

	β	SE	t	p
CWR forward	−1.8	0.6	−3.2	0.002
CWR reverse	0.7	0.5	1.4	0.17

(c) BAPQ (Pragmatic Language)

	β	SE	t	p
CWR forward	−0.4	0.1	−4.1	< 0.001
CWR reverse	0.1	0.1	1.5	0.15

(d) BAPQ (Aloof)

	β	SE	t	p
CWR forward	−0.2	0.1	−1.3	0.19
CWR reverse	0.2	0.1	2.2	0.04

(e) BAPQ (Rigid)

	β	SE	t	p
CWR forward	−0.2	0.1	−2.3	0.02
CWR reverse	0.1	0.1	1.5	0.14

Table 7.7: A table of linear regressions of each subscore of AQ and BAPQ, variance in which could be explained by CWR forward and reverse scores.

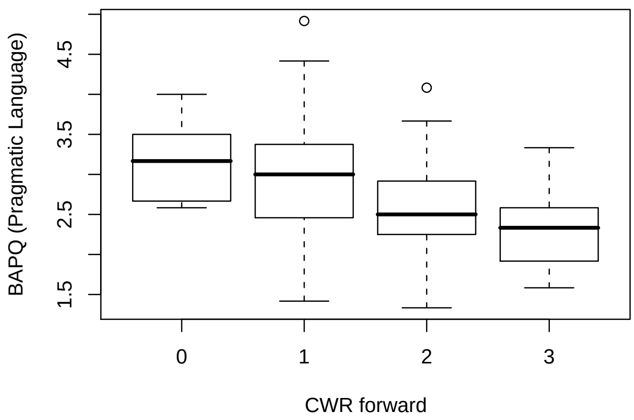

Figure 7.5: Boxplot of CWR forward score against the pragmatic language subscale (higher score implies more impairment) of BAPQ.

participants. Another difference is that Pijnacker et al. (In press) presented ten items for every condition using different content, which we were unable to do as we were trying to get through a large number of tasks. They also used a block design, with simple items always first, whereas we presented the items in (pseudo)random order. At the item level, they have more power; however they have fewer participants.

What does this mean? It appears that (reporting) good language pragmatics and communicative competence predicts a stronger tendency to integrate the sentences for forward inferences in the manner modelled by the LP-NAF; with the general idea that people with stronger autistic traits are less likely to use credulous reasoning. For the reverse inferences there is evidence that more social impairment predicts more CWR. This would suggest that people with social impairment are particularly good at doing diagnostic reasoning in a closed world context.

For the two-conditional cases, forward inferences, drawn more frequently by lower autistic trait people, are implications, whereas the reverse inferences, drawn more frequently by higher AQ people, are disjunctions. This might point to an explanation for the difference. It is possible that the response 'P or Q' should be interpreted '$\vdash P$ or $\vdash Q$' rather than '$\vdash P \vee Q$'. Also note that in the LP-NAF model, the forward and reverse inferences are distinguishable: the reverse inferences require integrity constraints whereas the forward do not.

7.3 Analysis of the Selection Task Data

Recall that participants were presented with the text:

> Below there are four cards. Each has a letter on one side and a number on the other side. As you can see, some have the letter side showing, some have the number side showing.

Below this were four cards showing 'A', 'K', '4', and '7'. Participants were told:

> I claim that if there is a vowel on one side then there is an even number on the other side. Type the cards you'd need to turn to work out if my claim is correct.

In this section I abbreviate 'there is a vowel on one side' as P and 'there is an even number on the other side' as Q.

7.3.1 Distribution of responses

Ignoring order of selection, there are $2^4 = 16$ different combinations of card choices. Table 7.8 shows the responses made. The majority of participants (37%) opted for P and Q; only 12% were consistent with a strict classical interpretation, turning exactly those cards of which Wason would have approved. Relaxing the criteria slightly, including participants who select other cards in addition to P and $\neg Q$ increases the number of classically competent participants to 29%. This is mostly due to participants who select all four cards (15% of the total number of participants). Examining the individual cards, P was the most frequently selected (80% of participants) and $\neg P$ the least frequently (29%). Frequencies were similar to those found in a fairly recent study by Stenning and van Lambalgen (2004, p. 516), though the present sample had slightly more classical responses (12% versus 4%), fewer P,Q responses (37% versus 52%), and more all card responses (15% versus 8%). The instructions used in the present experiment may have caused this difference, especially the increased turning of all cards since the instructions did not try to limit the number of cards people turned, though it is difficult to see why the instructions might have made people more consistent with the classical competence model.

Although card presentation order was not randomised (they were presented in the order $\langle P, \neg P, Q, \neg Q\rangle$, left-to-right), let's briefly examine response orders. Of the participants who chose P and Q, all but one selected them in the order $\langle P, Q\rangle$. The only other order effect was found for choices of all four cards: 8 chose the cards in order left to right $\langle P, \neg P, Q, \neg Q\rangle$; 4 chose $\langle P, Q, \neg P, \neg Q\rangle$, and finally 1 chose in the order $\langle \neg P, P, \neg Q, Q\rangle$. Given the negligible variation, order will not be considered further for this task.

7.3.2 Exploratory analysis relating to autistic traits

A series of exploratory backward stepwise regressions were run to investigate whether severity of autistic traits could be predicted by the selection task performance. Card-choices, coded as four binary variables, were used as predictors, and the autistic traits scores used as outcome variables. Akaike's information criterion was used for variable selection (AIC; see Venables & Ripley, 2002, pp. 175–177). Nothing was found for the composite AQ and BAPQ scores, even allowing 4-way interactions (smallest $p > 0.2$). There were results for subscores, however. For BAPQ, the models with outcome variable Pragmatic Language [$F(8, 75) = 2.12$, $p = 0.04$, adjusted $R^2 = .10$] and Aloofness [$F(1, 79) = 4.31$, $p = 0.04$, adjusted $R^2 = .04$] were significant fits to the data. For AQ, the models with outcome variables Attention to Detail [$F(5, 78) = 2.52$, $p = 0.04$, adjusted $R^2 = .08$] and Imagination [$F(1, 81) = 6.64$, $p = 0.01$, adjusted $R^2 = .06$] were significant fits. See Table 7.9 for the model estimates.

The clearest effect for Pragmatic Language was that making a classical response was associated with more pragmatic language difficulties. More aloofness implied a higher frequency of choosing $\neg Q$, the card choice of the two classical responses which is least frequently made. Interestingly, better attention to detail was associated with a more frequent selection of the modal response of turning P and Q. Finally reporting better imagination implied more frequent selection of the Q card, part of the modal response pair and the second most frequent card choice.

Number turned		*Card(s) turned* P	$\neg P$	Q	$\neg Q$	n	%	*S & vL* (%)
0						4	5	7
1					✔	1	1	3
				✔		3	4	6
			✔			3	4	
		✔				9	11	8
2				✔	✔	1	1	
			✔		✔	1	1	2
			✔	✔		3	4	6
		✔			✔	10	12	4
		✔		✔		31	37	52
		✔	✔			3	4	
3			✔	✔	✔	0	0	
		✔		✔	✔	1	1	1
		✔	✔		✔	0	0	
		✔	✔	✔		1	1	
4		✔	✔	✔	✔	13	15	8
	n	68	24	53	27			
	%	81	29	63	32			

Table 7.8: How many people made each choice on the selection task (total $n = 84$). The right-most column gives frequencies reported by Stenning and van Lambalgen (2004, p. 516) for their 108 participants.

These selection task results should be interpreted with caution as the exceedingly exploratory technique of stepwise regression was used.

7.4 Summary

The association between performance on the suppression task and AQ is more good news for the continuum view of ASC. It was discovered that the competence model characterised well differences in responses as a function of autistic traits, especially the notion of forward versus reverse inference. This is a slightly different result to that found in the literature for ASC. Social interaction competence and pragmatic language ability both related to interpretation, and it appears they are separable. Inferences on Wason's selection task were also found to relate to autistic-like traits.

(a) BAPQ (Pragmatic Language)

	β	SE	t	p	
(Intercept)	3.3	0.3	11.9	< 0.001	***
P	-0.8	0.3	-2.6	0.01	*
$\neg P$	0.0	0.3	0.1	0.91	
Q	-0.5	0.4	-1.4	0.18	
$\neg Q$	-0.8	0.5	-1.7	0.10	†
$P \times Q$	0.7	0.4	1.7	0.09	†
$P \times \neg Q$	1.0	0.5	2.1	0.04	*
$\neg P \times \neg Q$	0.9	0.5	1.8	0.08	†
$Q \times \neg Q$	-0.9	0.5	-1.9	0.06	†

(b) BAPQ (Aloof)

	β	SE	t	p	
(Intercept)	2.7	0.1	23.0	< 0.001	***
$\neg Q$	0.4	0.2	2.1	0.04	*

(c) AQ (Attention to Detail)

	β	SE	t	p	
(Intercept)	28.5	1.6	18.3	< 0.001	***
P	-4.5	1.7	-2.7	0.01	**
$\neg P$	-3.2	1.8	-1.7	0.09	†
Q	-3.8	2.2	-1.7	0.09	†
$P \times Q$	6.3	2.3	2.7	0.01	**
$\neg P \times Q$	3.2	2.2	1.5	0.14	

(d) AQ (Imagination)

	β	SE	t	p	
(Intercept)	21.1	0.7	30.5	< 0.001	***
Q	-2.3	0.9	-2.6	0.01	*

Table 7.9: A table of linear regressions of each subscore of AQ and BAPQ, variance in which could be explained by the cards selected. Note: $\times$ represents an interaction (coded as a product). † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Chapter 8

‘Immediate inference’: one premise monadic quantifier reasoning

8.1 Introduction

The notion of ‘immediate inference’ is quite old. Thompson (1878, p. 474) provides an idea of the sense in which the term has been used:

> Where the mind passes directly from one object to a second, identifying the one with the other, the process is commonly spoken of as Immediate Inference; where it arrives at its conclusion only through the intervention of a third or mediate object, the proceeding is denominated Mediate Inference.

The particular ‘immediate inference’ task considered in this chapter concerns problems like: assume *some A are B* is true; is *some B are A* true, false, or could it be either true or false? Let’s write this *some A are B* $\models_?$ *some B are A*. Thompson continues:

> Although these distinctions have been very generally made in treatises on logic, I conceive them to be highly objectionable, for all inference [...] is mediate, that termed immediate being only relatively so—relatively simple and direct as compared with the other.

Not everyone agrees. Robinson (1929, p. 253), for instance, argued around fifty years later that, ‘Immediate inference, as described in formal logic, is not inference at all.’ He goes on to argue:

> Although we may *say* no more than that All A is B, if we know anything at all we necessarily know, in one and the same apprehension, both that All A is B and that Some B is A.

Immediate inference has been studied extensive in the psychology literature. Begg and Harris (1982), for instance, performed a series of experiments using the task, and concluded that participants 'do not play the logic game, but play a language game instead.' Newstead and Griggs (1983) performed a similar series of studies, and noted that 'there seem to be substantial and quite consistent individual differences in the way in which quantifiers are interpreted. [. . .] Clearly such findings render general model building in this area a rather dangerous exercise.' Good news for the present thesis! Much work has been done on the properties of representations of the sentences used in the task. Recently, for instance, Politzer et al. (2006) exposited correspondence between a graphical and an algebraic representation of quantifiers, using Gergonne circles.

The main interest of this task is that it reveals many interpretable individual differences, for instance related to language pragmatics (our guiding credulous versus sceptical distinction), and also that it relates to the inferences drawn on categorical syllogisms (discussed in Chapters 9 and 10). In this chapter we study internal properties of immediate inference and relationships with autistic traits.

8.2 Rashness and Hesitancy

8.2.1 On existence

One of the delights of the psychology of reasoning is its rich historical context, extending into philosophy and logic. Often it is assumed in psychology that *all* implies *some* and *no* implies *some. . . not* (though not always; e.g., Johnson-Laird & Bara, 1984; Chater & Oaksford, 1999). This is not so in the recent literature on logic (Boolos, 1984). Even a century ago, this assumption was not always made. For instance Peirce (1880, p. 23) writes:

> . . . it is usually understood that affirmative propositions imply the existence of their subjects, while negative ones do not. Accordingly, it is said that there is an immediate inference from A [all] to I [some] and from E [no] to O [some. . . not]. But in the sense assumed in this paper, universal propositions do not, while particular propositions do, imply the existence of their subjects.

			Response		
Premise	*Conclusion*	*Order*	*True*	*False*	*Either*
All A are B	All A are B	In	Classical	—	Hesitant
All A are B	Some A are B	In	Classical	—	Hesitant
No A are B	No A are B	In	Classical	—	Hesitant
No A are B	Some A are not B	In	Classical	—	Hesitant
Some A are B	Some A are B	In	Classical	—	Hesitant
Some A are not B	Some A are not B	In	Classical	—	Hesitant
All A are B	Some B are A	Out	Classical	—	Hesitant
No A are B	No B are A	Out	Classical	—	Hesitant
No A are B	Some B are not A	Out	Classical	—	Hesitant
Some A are B	Some B are A	Out	Classical	—	Hesitant

Table 8.1: Classification of responses for immediate inference (where the classical response is *true*)

(We shall see this issue of existential supposition, or the no-empty-sets assumption, again in Section 9.2, p. 185.)

8.2.2 Classifications

Stenning and Cox (2006) categorised a response as *rash* if (with existential presupposition) the antecedent and consequent are classically logically independent[1] and the participant responds *true* or *false*, and *hesitant* if classically the proposition is *true* or is *false* but the participant responds *either*. They classified items as *in-place* if the consequent term-order matched that in the antecedent and *out-of-place* if not. From this further classification of rash in-place and out-of-place, and hesitant in-place and out-of-place were formed. Tables 8.1–8.3 show how responses are classified, and the classically correct responses. As Table 8.3 highlights, being rash is equivalent to drawing an inference which is classically incorrect for problems which are logically independent.

[1]There is a model M_1 such that $\varphi \models_{M_1} \psi$ and a model M_2 such that $\varphi \not\models_{M_2} \psi$.

			Response		
Premise	*Conclusion*	*Order*	*True*	*False*	*Either*
All A are B	No A are B	In	—	Classical	Hesitant
All A are B	Some A are not B	In	—	Classical	Hesitant
No A are B	All A are B	In	—	Classical	Hesitant
No A are B	Some A are B	In	—	Classical	Hesitant
Some A are B	No A are B	In	—	Classical	Hesitant
Some A are not B	All A are B	In	—	Classical	Hesitant
All A are B	No B are A	Out	—	Classical	Hesitant
No A are B	All B are A	Out	—	Classical	Hesitant
No A are B	Some B are A	Out	—	Classical	Hesitant
Some A are B	No B are A	Out	—	Classical	Hesitant

Table 8.2: Classification of responses for immediate inference (where the classical response is *false*)

Refer to Figure 8.1 for the distribution of scores found in the present sample. As can be seen, just over 60 participants were not hesitant at all for in-place problems. Around 10 were hesitant for one item, in-place. The range of scores was 0–4. The theoretical maximum is 12, or 8 if we exclude items of the form $\varphi \models_? \varphi$. For hesitant out-of-place, again the modal response is 0 (around 40 participants), with a second peak around 2 ($\sim$10 participants). The range is 0–8 (and the maximum possible is 8), though the numbers drop off dramatically after 4. Taking now rashness in-place, the modal response is 4 (around 60 participants)—the maximum possible. Finally rashness out-of-place shows the most variation. The biggest peak is at 8 (the maximum possible), around 20 participants, with a smaller peak at 4 and at 2 (10 participants each). Cronbach's α was adequate for rash in-place (0.72), good for rash out-of-place (0.84) and hesitant out-of-place (0.80), but poor for hesitant in-place (0.47).

In the present sample, hesitancy out-of-place and rashness out-of-place were negatively correlated ($r = -0.45, p < 0.001$) with a similar coefficient to that found by Stenning and

			Response		
Premise	*Conclusion*	*Order*	*True*	*False*	*Either*
Some A are B	All A are B	In	Rash	Rash	Classical
Some A are B	Some A are not B	In	Rash	Rash	Classical
Some A are not B	No A are B	In	Rash	Rash	Classical
Some A are not B	Some A are B	In	Rash	Rash	Classical
All A are B	All B are A	Out	Rash	Rash	Classical
All A are B	Some B are not A	Out	Rash	Rash	Classical
Some A are B	All B are A	Out	Rash	Rash	Classical
Some A are B	Some B are not A	Out	Rash	Rash	Classical
Some A are not B	All B are A	Out	Rash	Rash	Classical
Some A are not B	No B are A	Out	Rash	Rash	Classical
Some A are not B	Some B are A	Out	Rash	Rash	Classical
Some A are not B	Some B are not A	Out	Rash	Rash	Classical

Table 8.3: Classification of responses for immediate inference (where the classical response is *either*)

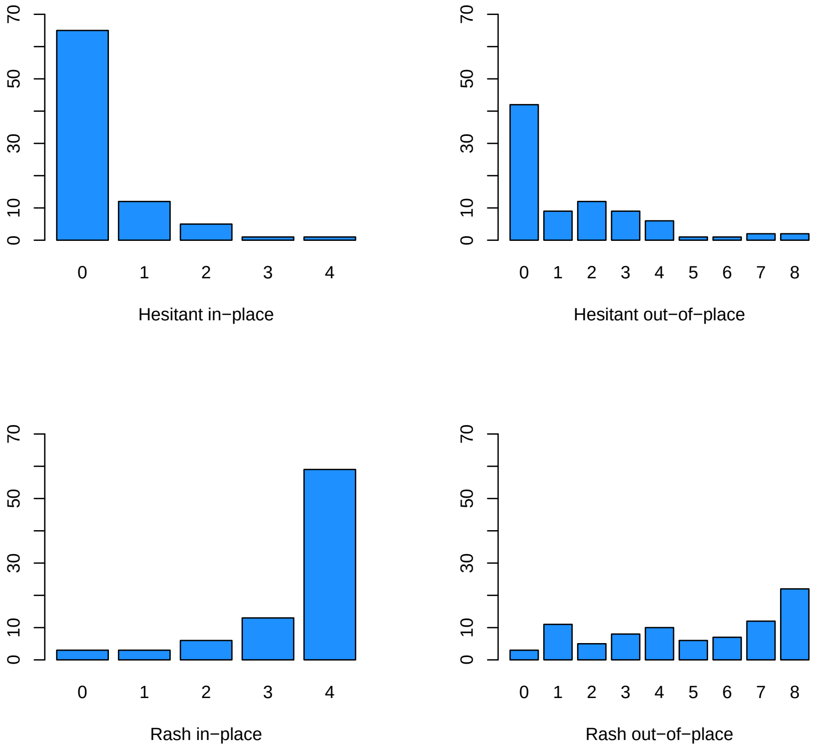

Figure 8.1: Histograms of immediate inference responses in terms of rashness and hesitancy, in and out of place.

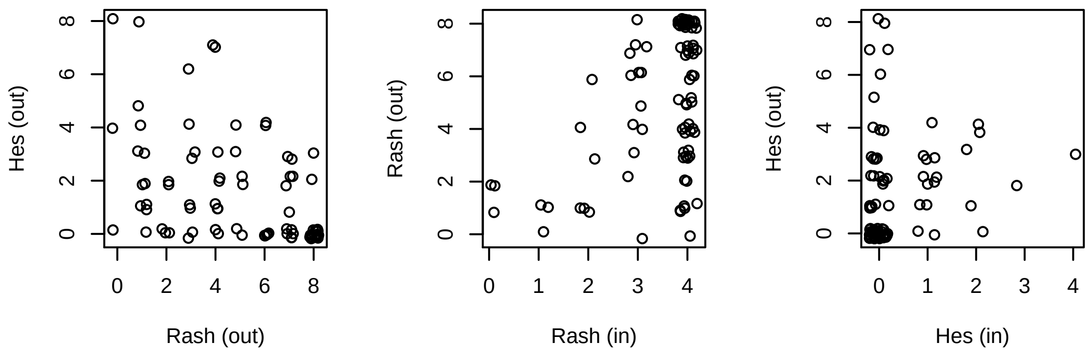

Figure 8.2: Jittered plots of statistically significant bivariate correlations between the immediate inference scores.

Cox (2006) ($r = -0.50$). Also rashness in-place was positively correlated with rashness out-of-place ($r = 0.47, p < 0.001$), again with a coefficient very similar to that found by Stenning and Cox (2006) ($r = 0.40$). We repeat the analysis using Spearman's ρ because the distributions are clearly not Gaussian. Now we find an additional but weak positive correlation between hesitancy in-place and out-of-place ($\rho = 0.32$, $p \approx 0.003$). Again there was a negative correlation between hesitancy out-of-place and rashness out-of-place ($\rho = -0.50$, $p < 0.001$), and a positive correlation between rashness in-place and out-of-place ($\rho = 0.41$, $p < 0.001$). See Figure 8.2 for jittered bivariate plots.

8.2.3 Relationships with autistic-traits in my data

Table 8.4 shows Spearman's ρ correlations between the AQ subscores and the immediate inference scores. Only attention to detail was significantly (for $\alpha = 0.05$) associated with anything in immediate inference (hesitant out-of-place). The table nonetheless shows some hints of associations for $\alpha = 0.1$; in the spirit of exploration these are marked in bold.

None of the BAPQ variables (see Table 8.5) show statistically significant bivariate correlations with the immediate inference scores. The smallest p-value ($p = 0.13$) was

		Hesitant		*Rash*	
		In	Out	In	Out
Social Skills	ρ	0.08	-0.01	-0.01	-0.14
	p	0.46	0.95	0.9	0.22
	n	82	82	82	82
Attention Switching	ρ	0.02	**$-$0.19**	0.07	0.07
	p	0.88	0.1	0.55	0.53
	n	81	81	81	81
Attention to Detail	ρ	-0.01	**$-$0.21**	-0.03	0.11
	p	0.96	0.05	0.76	0.33
	n	82	82	82	82
Communication	ρ	0.05	0.02	-0.05	**$-$0.19**
	p	0.67	0.88	0.68	0.08
	n	82	82	82	82
Imagination	ρ	0.07	-0.04	-0.08	**$-$0.2**
	p	0.54	0.7	0.46	0.08
	n	81	81	81	81

Table 8.4: Spearman’s ρ correlations between AQ subscores and immediate inference with pairwise deletions of missing values. All coefficients with $p \leq 0.1$ (two-tailed) are marked in bold.

		Hesitant		*Rash*	
		In	Out	In	Out
Aloof	ρ	0.07	0.07	0.00	−0.15
	p	0.54	0.56	0.98	0.18
	n	79	79	79	79
Pragmatic Language	ρ	0.06	−0.02	−0.10	−0.12
	p	0.60	0.88	0.36	0.29
	n	82	82	82	82
Rigid	ρ	0.04	−0.17	0.07	0.04
	p	0.73	0.13	0.56	0.73
	n	80	80	80	80

Table 8.5: Spearman's ρ correlations between BAPQ subscores and immediate inference with pairwise deletions of missing values.

between hesitancy out-of-place and rigidity.

We also explored multivariate relationships. The distribution of the rashness and hesitancy variables suggests the residuals of a linear regression will not show a Gaussian distribution, so proportional odds logistic regression models were fitted using the polr function of the MASS package (see Venables & Ripley, 2002, pp. 204–205), implemented in R (R Development Core Team, 2008). These models basically amount to the simultaneous estimation of $k - 1$ logistic regressions for an ordinal variable with k levels. Consider the case where the dependent variable has three values: 0, 1, and 2. The fitted functions are of the form:

$$\mathsf{Pr}(\hat{y} < 1) = \text{logit}^{-1}(\alpha_1 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n) \quad (8.1)$$

$$\mathsf{Pr}(\hat{y} < 2) = \text{logit}^{-1}(\alpha_2 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n) \quad (8.2)$$

where α_1 and α_2 are $(k - 1)$ estimates of intercepts and the β's are constant across the equations. The resulting model is sufficient to calculate the probability of any particular value.

			95% CIs			
	β	*SE*	2.5%	97.5%	t	
Social Skills	−0.03	0.06	−0.14	0.09	−0.46	
Attention Switching	0.14	0.06	0.03	0.25	2.43	⋆
Attention to Detail	0.07	0.05	−0.02	0.17	1.55	
Communication	−0.04	0.07	−0.17	0.09	−0.63	
Imagination	−0.15	0.06	−0.28	−0.03	−2.40	⋆

Table 8.6: Regressing rashness out-of-place on the five subscores of AQ. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

For AQ, the only result came from regressing rashness out-of-place on the five subscores. Compared to the null model, this model was a significant improvement in fit [likelihood ratio $\chi^2(5) = 13.3$; $p = 0.02$]. The slopes for Attention Switching ($\beta = 0.14$) and Imagination ($\beta = -0.15$) are both significantly different to zero as both ends of the 95% confidence intervals are on the same side of and not equal to zero: so impairments in attention switching are associated with more rashness out-of-place and impairment in imagination is associated with less rashness out-of-place, i.e., being more classical. (See Table 8.6 for all the coefficients.)

For BAPQ the only result came from regressing rashness in-place on the three subscores. According to the likelihood ratio test, the model just failed to fit significantly better than the null model [$\chi^2(3) = 7.35$, $p = 0.06$]; we continue anyway as there wasn't much variance in the data (the majority of participants are rash) but there could be still something of interest. In the presence of the other two predictors, reporting problems with language pragmatics is associated with more classical responses, less rashness, in-place ($\beta = -1.10$). The other predictors didn't differ significantly from zero. (See Table 8.7 for the coefficients.)

It appears that the in-place logically independent problems are the best predictor of pragmatic language ability, as the concept is understood by participants themselves, once aloofness and rigidity have been partialled out. These items include the two most often

			95% CIs			
	β	SE	2.5%	97.5%	t	
Aloofness	0.46	0.36	−0.22	1.21	1.28	
Pragmatic Language	−1.10	0.43	−1.99	−0.25	−2.53	⋆
Rigidity	0.71	0.41	−0.08	1.53	1.74	

Table 8.7: Regressing rashness in-place on the three subscores of BAPQ. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

considered related to Grice's conversational maxims, i.e., uttering *some A are B* implies that the speaker knew that not *all A are B* and that uttering *some A are not B* implies that the speaker knew that at least *some A are B*. Something different happens when the conclusion term-order is reversed: self-report questions which arguably refer to more traditional executive functions like attention switching begin to influence the inferences that people draw.

8.3 Delving deeper

This section reports a more detailed analysis of the immediate inference response classifications to check that no interesting sources of variance are being obscured. The following aspects of the scores are worth highlighting.

1. Four of the hesitancy in-place responses refer to problems of the form $\varphi \models_? \varphi$, for which there is (unsurprisingly) a minuscule amount of variance in the data.

2. Scalar implicature (for instance that *some* is often interpreted as implying *some... not*) is merged in with the rashness sumscores.

3. Existential presupposition is ignored and could perhaps be detected by whether a participant reasons that *all A are B* $\models$ *some A are B* or *no A are B* $\models$ *some A are not B*. Classically since *some A are B* $\equiv$ *some B are A* and *no A are B* $\equiv$ *no B*

are A, we might also expect assumptions of empty-sets to be detected by *all A are B* $\models$ *some B are A* or *no B are A* $\models$ *some A are not B*.

4. Commutativity and scalar implicature are combined in the out-of-place sumscores.

5. Some response types are not distinguished by the rash and hesitancy scores. Responding false for a classically true logically dependent problem in not counted; nor is responding true for a classically false logically dependent problem. (See cells marked with an dash in Tables 8.1–8.3.)

8.3.1 Item classification adequacy

Beginning then with rashness (Table 8.9), overall *some/some... not* $\models_?$ *all/no* problems tend to give either *false* or *either* responses; *all/no* $\models_?$ *some/some... not* problems tend to give either *true* or *either* responses—so actually the rashness score characterises items responses well (we don't yet know if the item *groupings* are meaningful). The exceptions are:

- *all A are B* $\models_?$ *all B are A*.
- *all A are B* $\models_?$ *some B are not A*.

For both these problems, all three responses (*true*, *false*, *either*) are made by a non-negligible number of participants, 14–45 in each cell. This is similar to the following hesitancy out-of-place problems:

- *no A are B* $\models_?$ *no B are A*.
- *no A are B* $\models_?$ *some B are not A*.

One might expect an association between these responses. There is indeed an association between commutativity of all and no (Fisher's exact test $p = 0.01$), but none between *all A are B* $\models_?$ *some B are not A* and *all A are B* $\models_?$ *some B are not A* (Fisher's $p = 0.8$). See Table 8.8 for counts.

For hesitancy, items which are classically *true* (Table 8.11) tend to be given the response *true* or *either*, except:

no A are B $\models_?$ *no B are A*	*all A are B* $\models_?$ *all B are A* E	T	F
E	16	1	2
T	23	22	10
F	6	2	2

no A are B $\models_?$ *some B are not A*	*all A are B* $\models_?$ *some B are not A* E	T	F
E	11	5	6
T	15	11	17
F	7	5	7

Table 8.8: Contingency tables for responses (a) *all A are B* $\models_?$ *all B are A* against *no A are B* $\models_?$ *no B are A* and (b) *all A are B* $\models_?$ *some B are not A* against *no A are B* $\models_?$ *some B are not A*.

			Rash		*Classical*
Premise	*Conclusion*	*Order*	*True*	*False*	*Either*
Some A are B	All A are B	In	3	69	12
Some A are not B	No A are B	In	1	72	11
Some A are not B	Some A are B	In	69	4	11
Some A are B	Some A are not B	In	68	4	12
All A are B	All B are A	Out	25	14	45
Some A are B	All B are A	Out	2	44	38
Some A are not B	All B are A	Out	3	48	33
Some A are not B	No B are A	Out	2	57	25
Some A are not B	Some B are A	Out	55	3	26
All A are B	Some B are not A	Out	21	30	33
Some A are B	Some B are not A	Out	54	2	28
Some A are not B	Some B are not A	Out	52	5	27

Table 8.9: Count of participants' ($n = 84$) responses for each item contributing to a rashness score.

no A are B $\models_?$ *some A are not B*	*all A are B* $\models_?$ *some A are B* E	T	F
E	3	2	4
T	2	38	5
F	3	8	19

Table 8.10: Attempt to separate classical correctness with and without existential presupposition and scalar implicature

- *all A are B* $\models_?$ *some A are B.*
- *no A are B* $\models_?$ *some A are not B.*

True is likely to indicate classical correctness with no-empty-sets; *false* is likely to indicate a scalar implicature effect; and *either* indicates classical correctness without the no-empty-sets assumption. There is an association between rows and columns in the contingency table (Fisher's $p < 0.001$), however only for true (classical with no-empty-sets) and false (scalar implicature) responses.

For in-place problems for which the classical response is *false* (Table 8.12), remarkably[2] almost all (81–83 out of 84) participants give the classical response. There is a little more variation for *out-of-place* problems for which the classical response is *false*:

- *all A are B* $\models_?$ *no B are A.*
- *no A are B* $\models_?$ *some B are A.*

Again, though, the *either* and *false* responses cover the vast majority of the variance, so these responses will not be further analysed.

[2] I am often surprised when people are actually consistent with classical logic!

			Classical		Hesitant
Premise	*Conclusion*	*Order*	*True*	*False*	*Either*
All A are B	All A are B	In	83	0	1
No A are B	No A are B	In	83	1	0
All A are B	Some A are B	In	48	28	8
Some A are B	Some A are B	In	78	0	6
No A are B	Some A are not B	In	45	30	9
Some A are not B	Some A are not B	In	82	1	1
No A are B	No B are A	Out	55	10	19
All A are B	Some B are A	Out	56	6	22
Some A are B	Some B are A	Out	63	3	18
No A are B	Some B are not A	Out	43	19	22

Table 8.11: Count of participants' ($n = 84$) responses for each item contributing to a hesitancy score, where the classically correct answer is true.

				Classical	Hesitant
Premise	*Conclusion*	*Order*	*True*	*False*	*Either*
No A are B	All A are B	In	1	83	0
Some A are not B	All A are B	In	1	82	1
All A are B	No A are B	In	2	81	1
Some A are B	No A are B	In	1	83	0
No A are B	Some A are B	In	1	83	0
All A are B	Some A are not B	In	0	82	2
No A are B	All B are A	Out	2	73	9
All A are B	No B are A	Out	5	72	7
Some A are B	No B are A	Out	1	72	11
No A are B	Some B are A	Out	9	58	17

Table 8.12: Count of participants' ($n = 84$) responses for each item contributing to a hesitancy score, where the classically correct answer is false.

8.3.2 Item grouping coherence

To summarise, no substantial gains result from considering responses not captured by the rash and hesitant characterisations. Finally it may be revealing to consider the correlation structure of the items. Table 8.13 shows the loadings resulting from an exploratory factor analysis on the tetrachoric correlation matrix with varimax rotation (items with little response variance were first removed). Factor 1 nicely captures the inversely correlated hesitant and rash out-of-place items. Factors 2 and 4 appear to characterise scalar implicature: factor 2 captures refusal to allow a universal to follow from an existential and factor 4 captures refusal to allow an existential to follow from a universal. Factor 3 is interesting, mostly capturing hesitant out-of-place problems where the premise is a *no* quantifier.

8.3.3 Response time

A series of studies have investigated effects related to the rash in-place problems, so let's focus here on those. Bott and Noveck (2004) investigated the time course of responses, which would be worth investigating further. Their experimental setup is slightly different to that which generated the present data. Items were propositions like 'some trout are fish' and 'all monkeys are mammals'. Quantifiers *all* and *some* were used and sentences chosen which were classically true or false. The key 'pragmatic' condition consisted of some sentences which would be true if given a classical interpretation and false if given an interpretation of some but not all.[3] The participants' preferred interpretation was not examined, but rather participants were instructed in two conditions, during one of which they had to take a pragmatic interpretation ('understand it as some but not all') and the other a classical logic interpretation ('understand it as some and possibly all'). Oddly, participants seemed to find the 'some but not all' interpretation more difficult. As seen above, our participants by default opted for this interpretation. The authors also found that responses were slower for the pragmatic interpretation.

Let's complete the investigation of this in our data for response times. Figure 8.3 shows the distribution of response times for each condition. For three out of the four problems,

[3] Two details are worth noting: the experiment was in French; and words were flashed on the screen one at a time.

Type	Q_1	Q_2	*Order*	1	2	3	4
Hes	A	S	AB				0.6
Hes	N	Sn	AB				0.9
Hes	S	S	AB			−0.9	
Hes	A	N	BA	−0.7	0.7		
Hes	A	S	BA	−0.6			0.6
Hes	N	A	BA	−0.6		0.4	
Hes	N	N	BA	−0.5		0.9	
Hes	N	S	BA	−0.4		0.9	
Hes	N	Sn	BA	−0.4		0.6	0.7
Hes	S	N	BA	−0.6			
Hes	S	S	BA	−0.7			
Rash	S	A	AB		0.8		
Rash	S	Sn	AB		0.9		
Rash	Sn	N	AB		0.8		0.5
Rash	Sn	S	AB		0.8	0.4	
Rash	A	A	BA	0.7			
Rash	A	Sn	BA	0.8	0.5		
Rash	S	A	BA	0.4	0.6		
Rash	S	Sn	BA	0.5	0.6		
Rash	Sn	A	BA	0.7			
Rash	Sn	N	BA	0.8			
Rash	Sn	S	BA	0.6	0.5		
Rash	Sn	Sn	BA	0.8			

Table 8.13: Factor analysis of item responses, coded as whether they are hesitant/rash. Absolute loadings < 0.4 have been omitted. (A: *all*; N: *no*; S: *some*; Sn: *some. . . not*.)

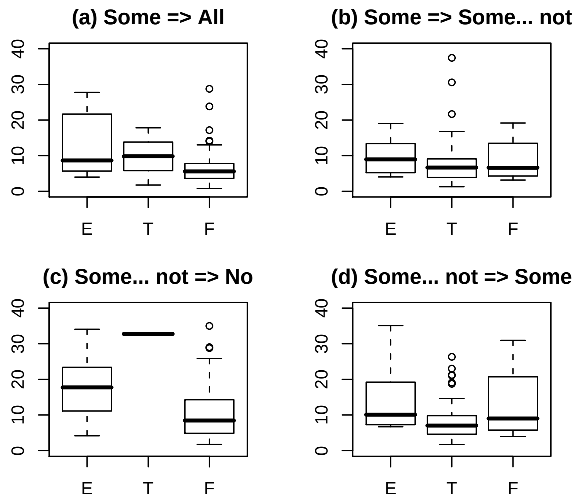

Figure 8.3: Response times (in seconds) for each of the rash in-place problems, as a function of the response given. (T: true; F: false; E: either true or false.) Recall that the classical response is *either*.

those who were rash (taking a ‘pragmatic’ interpretation, one might argue) tended to be faster: by around 6 (SE = 1.78) seconds for *some* ⇒ *all* $[F(2, 81) = 5.81$, $p = 0.004$, $R^2 = .10]$; 7 (SE = 2.78) seconds for *some. . . not* ⇒ *no* $[F(2, 81) = 6.28$, $p = 0.003$, $R^2 = .11]$ and 5 (SE = 2.39) seconds faster for *some. . . not* ⇒ *some* $[F(2, 81) = 8.69$, $p < 0.001$, $R^2 = .16]$. There was no for effect for *some* ⇒ *some. . . not* $[F(2, 81) < 1]$.

Two of these effects were moderated by (composite) autistic traits; here I focus on the BAPQ subscores. The full model for *some* ⇒ *all* was a good fit to the data $[F(10, 69) = 3.05$, $p = 0.002$, $R^2 = .21]$. For those who were rash, higher Rigidity was associated with a slower response ($\beta = 7.32$, $SE = 2.78$, t = 2.6, $p = 0.01$). There was a large effect for *some* ⇒ *some . . . not* $[F(11, 68) = 6.17$, $p < 0.001$, $R^2 = .42]$. For those were rash, Pragmatic Language impairment predicted a slower response ($\beta = 20.4$, $SE = 6.1$, $t = 3.4$, $p = 0.001$) and Aloof predicted a faster response ($\beta = -19.4$, $SE = 3.4$, $t = -5.7$, $p < 0.001$).

8.4 Summary

The most noteworthy result here is that self-reported and general pragmatic language difficulties predict a pragmatic-language effect for in-place logically independent problems (i.e., a rash response). These are the problems typically viewed as testing Gricean implicature. Also interesting is that for these problems, the pragmatic response was faster—the opposite to the result found by Bott and Noveck (2004). Clearly something odd happens to interpretation when people are given in natural language what amounts to a classical specification of a non-classical concept. Gricean interpretation was simulated in a classical logic interpretation, explaining the slower response. Autistic traits also moderate the speed of the responses in our data, with signs that those with Pragmatic Language impairment do not find the Gricean interpretation as effortless as those who report being adept with pragmatic language.

Out-of-place problems are more complex. Impairments in attention switching are associated with being more rash on out-of-place problems, as is better imagination. Perhaps those who report problems with attention switching find it difficult to change strategy as each problem is displayed. It's not clear, however, why this would push them towards a definite *true* or *false*. A related result found by De Neys and Schaeken (2007) was that when people are under a cognitive load (presumably related to attention switching), they tend to be *less* likely to draw rash-like inferences—the opposite to what our individual differences result (albeit with self-report) would predict. The imagination subscore could perhaps be associated with a preference for representing the premises in such a way that information regarding order is lost, e.g., using a representation akin to a diagram, or revealing a tendency for credulous reasoning, i.e., related to deeper discourse processing, which would be consistent with the items in the subscale which relate to story comprehension and pretend play. See Section 10.4.1 later when we interpret the data along with RAPM performance where the plot thickens!

Together this is clear evidence that there are qualitatively different patterns of response and that they are moderated by broader dimensions of individual difference.

Chapter 9

Categorical syllogisms: integrating two quantified premises

9.1 Introduction

This chapter describes analyses of categorical syllogisms, labelled the *Aplysia* of reasoning research (Knauff, 2007), and relationships with autistic traits. Of all the tasks participants were asked to solve, the 64 syllogisms were considered the hardest. As the reader may recall from Section 3.1 (pp. 38 ff.), there are many ways to characterise the syllogism, and they are difficult to distinguish empirically. The approach taken in this chapter follows work by Stenning and colleagues on representations which abstract across many of the models, focussing on the logical features of the task which may influence performance.

9.2 Classical competence

This section summarises classical competence scores for our sample. First we need the classical answers. Some premise pairs have more than one conclusion. What is assumed to be a correct answer depends on whether existential presupposition (∃P), or no-empty-sets, is assumed, that is, whether it is assumed there is at least one A, one B, and one C.[1] With ∃P, 27 of the premises have a valid conclusion from the list provided to

[1] There is much confusion on this point in the literature. Roberts (2005) for instance makes an unnecessary appeal to ∃P when introducing syllogisms via the argument *all B are A*, *all C are B*, therefore *all*

participants; without ∃P, 22 have a valid conclusion. Some problems which may be solved without ∃P have additional solutions if the assumption is made. Figures 9.1–9.4 show the classically correct responses, grouped by how many conclusions there are for the premises, and indicating which conclusions require ∃P.

The mean score without ∃P is 19 (range 3–50, median 16) and with ∃P the mean is 21 (range 3–53, median 18), indicative of a general tendency to over-infer (with respect to classical logic). Figure 9.1 shows histograms of the number of classically correct responses we found in our sample, split into problems with valid conclusions and problems where the classical response is no valid conclusion. Since when summing across problems ∃P makes little difference, we assume it holds. The mean score for VC problems is 14 ($SD =$ 4.18, range 3–23). For NVC problems the modal scores lie between 0 and 5, with a long tail down to zero above this interval.

C are A. This is a valid syllogism even if there are no *A*'s, *B*'s, and *C*'s. Perhaps some representations make the empty sets problem more salient than others.

Premise 1	Premise 2	Conclusion
All B are A	No B are C	Some A are not $C_{\exists}$
All B are A	No C are B	Some A are not $C_{\exists}$
No A are B	All B are C	Some C are not $A_{\exists}$
No B are A	All B are C	Some C are not $A_{\exists}$
All A are B	Some C are not B	Some C are not A
All B are A	Some B are not C	Some A are not C
Some B are not A	All B are C	Some C are not A
Some A are not B	All C are B	Some A are not C
No A are B	Some B are C	Some C are not A
No A are B	Some C are B	Some C are not A
No B are A	Some B are C	Some C are not A
No B are A	Some C are B	Some C are not A
Some A are B	No B are C	Some A are not C
Some B are A	No B are C	Some A are not C
Some A are B	No C are B	Some A are not C
Some B are A	No C are B	Some A are not C

Table 9.1: Premise pairs with one classically correct conclusion ($_{\exists}$ marks conclusions which require existential presupposition)

Premise 1	Premise 2	Conclusions
All B are A	All B are C	Some A are C$_{\exists}$
		Some C are A$_{\exists}$
All B are A	Some B are C	Some A are C
		Some C are A
All B are A	Some C are B	Some A are C
		Some C are A
Some A are B	All B are C	Some A are C
		Some C are A
Some B are A	All B are C	Some A are C
		Some C are A

Table 9.2: Premise pairs with two classically correct conclusions ($_{\exists}$ marks conclusions which require existential presupposition)

Premise 1	Premise 2	Conclusions
All A are B	All B are C	All A are C
		Some A are C$_{\exists}$
		Some C are A$_{\exists}$
All B are A	All C are B	All C are A
		Some A are C$_{\exists}$
		Some C are A$_{\exists}$

Table 9.3: Premise pairs with three classically correct conclusions ($_{\exists}$ marks conclusions which require existential presupposition)

Premise 1	Premise 2	Conclusions
All A are B	No B are C	No A are C
		No C are A
		Some A are not C$_\exists$
		Some C are not A$_\exists$
All A are B	No C are B	No A are C
		No C are A
		Some A are not C$_\exists$
		Some C are not A$_\exists$
No A are B	All C are B	No A are C
		No C are A
		Some A are not C$_\exists$
		Some C are not A$_\exists$
No B are A	All C are B	No A are C
		No C are A
		Some A are not C$_\exists$
		Some C are not A$_\exists$

Table 9.4: Premise pairs with four classically correct conclusions ($_\exists$ marks conclusions which require existential presupposition)

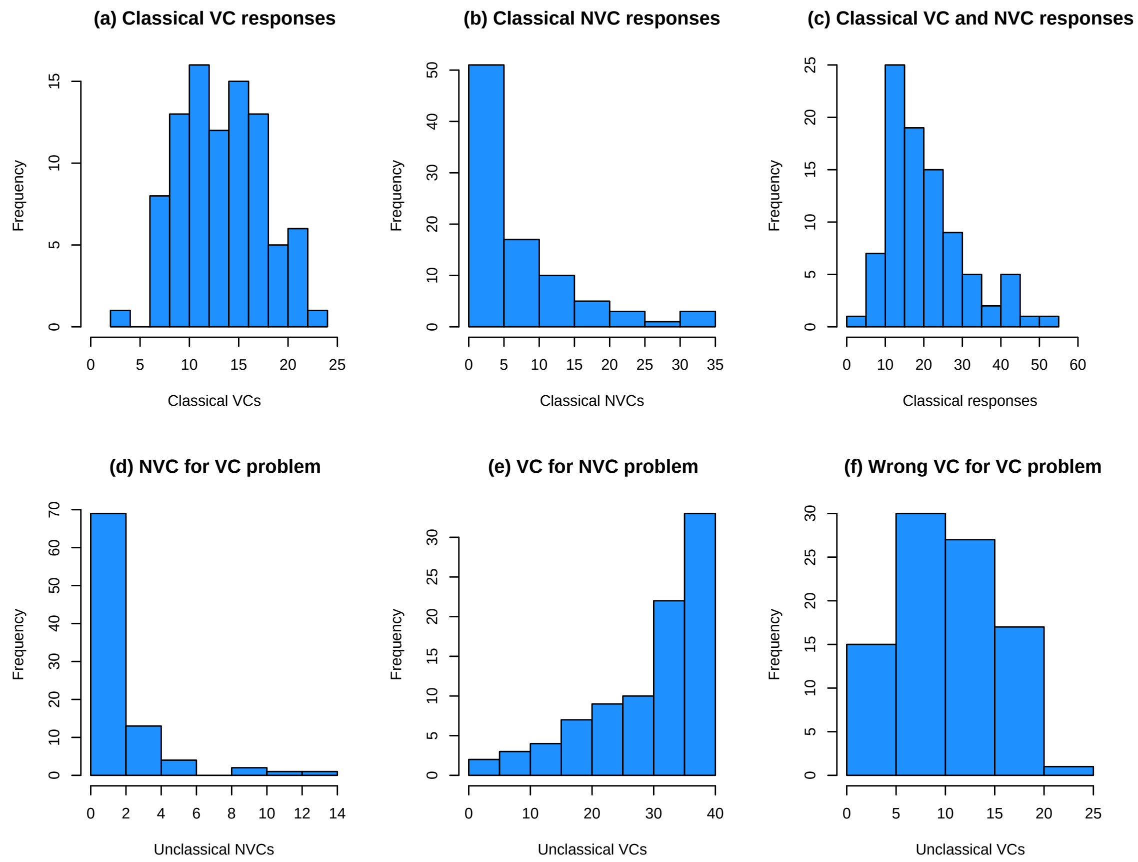

Figure 9.1: Histograms of classical correct and incorrect responses: (a) correct responses for problems with valid conclusions; (b) correct NVCs for problems with no valid conclusion; (c) classically correct responses for all problems; (d) NVCs when a valid conclusion is classically correct; (e) VC when NVC is classically correct; (f) incorrect VC for a VC problem.

The distribution of scores obtained by a sample randomly pushing buttons was simulated and is displayed in Figure 9.2. Visual inspection of the distributions suggests that in general people are not just guessing randomly for problems for which NVC is the classical solution: they are willfully trying to do something else. Analytically, the expected value of a discrete random variable X is given by

$$\sum_x xP(X = x)$$

(e.g., Grimmett & Welsh, 1986, p. 29). For NVC problems, the expected value is

$$37 \cdot 1/9 = 4.11$$

For VC problems, the value is

$$(16 \cdot 1/9) + (5 \cdot 2/9) + (2 \cdot 3/9) + (4 \cdot 4/9) = 5.33$$

9.3 Extending the source-founding model

The source founding model (Stenning & Yule, 1997) is an algorithm for the construction of a critical individual, and then from that a conclusion for syllogistic problems. It may be viewed as a competence model, the details of which provide clues for variation in the strategies utilised during performance. See Figure 9.3 for a description of the algorithm. The sequential steps are first to look for a unique existential premise; if there is one, then that is chosen as the starting point. Then essentially modus ponens or modus tollens are used draw conclusions. Universals are treated specially, but the main difference is the assumption of the antecedent of the source premise, and then when finally drawing a conclusion, keeping track of whether the source was an existential or a universal. The algorithm may be viewed as a theorem prover for a subset of natural deduction for the sentential calculus, with guidance from above on how to ensure existentials do not erroneously become universals.

In this section, I will describe a series of statistical implementations of the model and interactions with autistic-like traits. The general idea is that rather than implementing a monolithic process model computationally, using it to generate predictions, and then

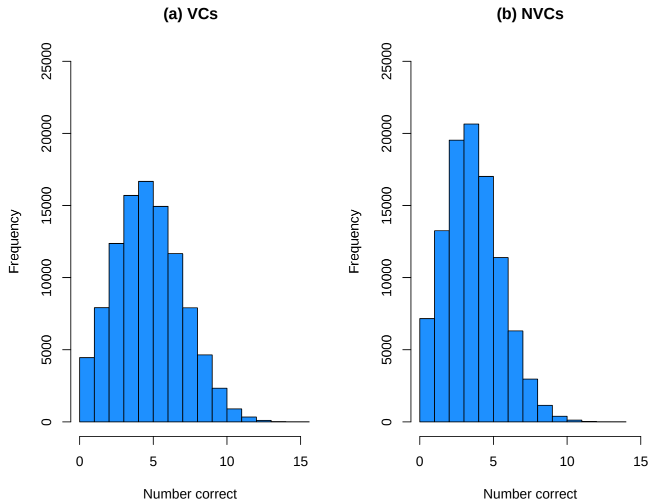

Figure 9.2: Histograms of chance performance in syllogisms, from 1×10^5 simulated trials. (a) shows the distribution of scores for problems with a valid conclusion; (b) shows the distribution where the correct answer is none of the listed options. For simplicity we assume existential presupposition, so there are 27 premise pairs for the VC problems and 37 for the NVC problems.

	% of responses		
Figure	AC	CA	NVC
ABBC	61	28	12
ABCB	48	34	18
BABC	54	33	13
BACB	43	44	13

Table 9.5: Proportion of term orders (including no conclusion) for the four figures in syllogisms

comparing fit to the data, e.g., individually for each participant, a statistical model is implemented which embodies the main predictions of the process model. Then the components of the model can be tested for their predictive power, and differential effects teased apart as a function of individual difference parameters. Families of models then span the continuum from top-down to bottom-up exploration guided by the top-down theory.

9.3.1 Predicting term-order

Firstly, let's examine the data to search for effects predicted by the source-founding model. Averaging across all items and participants, 52% were *AC*, 35% were *CA*, and 14% were *NVC*, showing a clear overall preference to maintain the premise order and also to draw a conclusion.

One of the earliest results relating to order is the that figure, the premise term-orders, affects a number of features of response. Störring (1908) discovered that figure affects response times (see Politzer, 2004): *ABBC* problems are solved faster than *BACB* problems. Figure affects the term order of response (Johnson-Laird & Steedman, 1978). The largest effect found is between figures *ABBC* and *BACB*. From their 20 participants, they found that, of the valid conclusions, 71% in figure *ABBC* were *AC* conclusions, whereas in figure *BACB* the order flipped and 70% were *CA* conclusions. For now we ignore classical competence. In our data ($n = 90$) we see the percentage of responses in Table 9.5. The biggest difference in term order is between *ABBC* and *BACB* as with Johnson-Laird and

1. Seek a unique existential premiss
   (a) If there are two, then respond NVC and quit.
   (b) If there are none, then go to 2.
   (c) If there is a unique one, make it the source premiss and go to 3
2. Seek a unique universal-premiss end-term subject
   (a) If there are none, choose an arbitrary source premiss. Go to 3.
   (b) If there are two, reconfigure a 'No' premiss [i.e., take advantage of the classical equivalence *no A are B* ≡ *no B are A*]. On failing, conclude NVC.
   (c) If there is now a unique one, choose its premiss as the source premiss and go to 3.
3. If the source premise is existential, then take its two terms as the first two clauses of the individual description. If the source premiss is universal, assume its antecedent. Apply modus ponens and conjoin the consequent to the antecedent and make the first two clauses of the individual description.
4. Compare middle terms
   (a) If a source middle term matches (with regard to negation) the antecedent middle term of the conditional premiss, apply modus ponens, and conjoin consequent term to individual description. Go to 5.
   (b) If the source middle term mismatches (with regard to negation) with the conditional consequent middle term, apply modus tollens to the conditional premiss, and conjoin consequent term to individual description. Go to 5.
   (c) ELSE conclude NVC and quit.
5. Individual description is now complete.
6. Draw Abstract conclusion from individual description:
   (a) Delete B conjunct from individual description. Quantify existentially for an existential conclusion (reordering any positive conjunct into subject position).
   (b) If clause 2c was satisfied, then there is a universal conclusion with the source premise end-term as subject.

Figure 9.3: The source-founding model (Stenning & Yule, 1997, p. 121)

		% of responses		
Q_1	Q_2	AC	CA	NVC
All	All	69	21	10
All	No	73	22	5
All	Some	57	39	4
All	Some...not	60	34	6
No	All	33	60	8
No	No	48	12	40
No	Some	30	61	9
No	Some...not	26	58	16
Some	All	40	57	3
Some	No	72	19	9
Some	Some	62	16	23
Some	Some...not	62	17	21
Some...not	All	40	55	5
Some...not	No	64	22	14
Some...not	Some	38	43	19
Some...not	Some...not	51	19	30

Table 9.6: Proportion of term orders (including no conclusion) for the various combinations of quantifiers. Q_1: first quantifier; Q_2: second quantifier.

Steedman (1978), except the present dataset shows no symmetry. There appears to be a jump in the proportion of NVCs for $ABCB$ problems.

Table 9.6 shows the proportion of ACs and CAs for each of the quantifiers. As can be seen, when offered an existential premise (*some* or *some...not*) and *no*, participants are clearly pushed to source from the existential premise. There appears, to be a preference, averaging across all responses, to source from a unique (positive) universal premise. However relative to the cases were there are two universals, participants do appear to source more from an existential in the second premise (e.g., 39% CA for *all-some*, versus 21% for

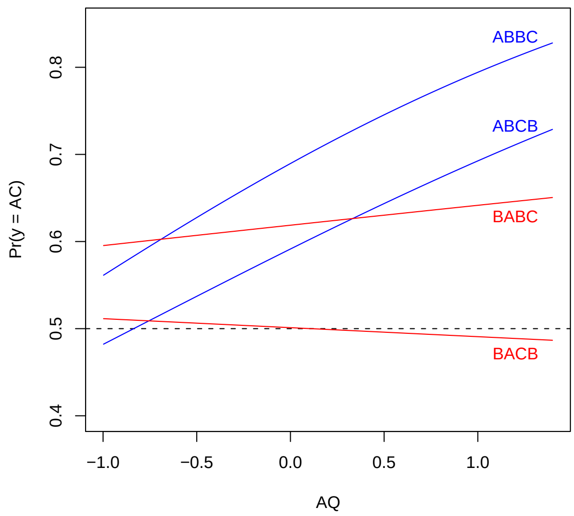

Figure 9.4: The relationship between AQ and the probability of an AC conclusion for each figure in syllogisms

all-all). This is not the case when the existential is in the first premise.

Initially a series of exploratory models were fitted to relate term-order and AQ. Multilevel logistic regression was used with a varying intercept for participants, fitted using the lme4 package (Bates et al., 2008) in R. Table 9.7 gives fit statistics for the models. It can be challenging to understand the direction of effects in the models by examining only coefficients so let's tease the effects apart using graphs. Figure 9.4 shows the relationship between AQ and the probability of an AC conclusion for each of the figures in the model where the only predictors are figure, AQ, and interactions between. Some authors (e.g., Geurts, 2003, p. 230) argue that figure is related to topic continuity, so one would expect autistic-like traits, related to conversation and socialising, to moderate the effect of figure on conclusion term orders. And indeed they do: people with higher AQ tend to be pushed towards drawing an AC conclusion for *ABBC* compared to *BACB* problems. Given that high AQ implies impairment in language pragmatics, it is difficult to argue that the effect

Terms added	*Df*	*AIC*	*BIC*	ℓ	χ^2	χ^2 *Df*	Pr($> \chi^2$)
0. (Null)	2	6487	6500	−3241			
1. Figure	5	6397	6429	−3193	95.8	3	< 0.000001
2. AQ	6	6207	6246	−3098	191.6	1	< 0.000001
3. Figure × AQ	9	6198	6256	−3090	14.9	3	0.002
4. $Q_{i\in\{1,2\}}$	15	5735	5832	−2853	475.2	6	< 0.000001
5. Q_i × AQ	21	5728	5864	−2843	19.2	6	0.004
6. Figure × Q_i	39	5697	5949	−2809	67.0	18	< 0.000001
7. Figure × Q_i × AQ	57	5716	6085	−2801	16.5	18	0.6

Table 9.7: Fit statistics for a series of nested models predicting term-order in syllogisms, each of which has a random intercept per participant. Each row gives fit statistics for the model with the specified terms added to the previous model. Q_i expands to terms for the two quantifiers; ℓ is the log-likelihood; χ^2, the log-likelihood ratio. Adding the final three-way interaction (Figure × Q_i × AQ) does not significantly improve the model fit.

is a result of language pragmatics processes. (Later we will tease apart the unique variance related to the components of AQ and BAPQ.)

Following this, predictors describing location of the quantifiers was added. Again there were interactions between these predictors and AQ. Finally we added interactions between the quantifiers and figure. This improved the fit according to AIC and LLR tests, though not BIC, which penalises more than the other two tests for the number of parameters in the model. Adding the three-way interaction between figure, quantifiers, and AQ failed to improve the fit of the model.

Table 9.8 shows the model coefficients for model 6; here for ease of interpretation AQ is centred on zero and scaled to have $SD = 0.5$ (see Gelman, To appear).

Table 9.8: Multilevel model predicting term order (probability of a *CA* conclusion) from figure, the quantifiers present, and AQ (centred and scaled so $SD = 0.5$). (4751 observations; 87 participants; random intercept per participant variance 0.283.)

	β	SE	z	p
(Intercept)	−1.00	0.18	−5.48	< 0.000001
Figure ABCB	0.71	0.24	2.93	0.0034
Figure BABC	0.41	0.24	1.71	0.087
Figure BACB	1.32	0.23	5.64	< 0.000001
AQ	−0.63	0.24	−2.62	0.0089
No_1	1.57	0.19	8.19	< 0.000001
Some_1	0.51	0.20	2.61	0.0091
Some…Not_1	0.50	0.20	2.56	0.01
No_2	−1.21	0.20	−6.06	< 0.000001
Some_2	−0.23	0.17	−1.31	0.19
Some…Not_2	−0.69	0.19	−3.72	0.0002
ABCB × AQ	0.13	0.20	0.65	0.52
BABC × AQ	0.51	0.20	2.58	0.0099
BACB × AQ	0.67	0.19	3.52	0.00043
AQ × No_1	0.01	0.19	0.08	0.94
AQ × Some_1	−0.42	0.19	−2.17	0.03
AQ × Some…not_1	−0.37	0.19	−1.96	0.05
AQ × No_2	−0.039	0.20	−0.19	0.85
AQ × Some_2	0.46	0.18	2.59	0.0095
AQ × Some…not_2	0.29	0.18	1.57	0.12
ABCB × No_1	−0.68	0.26	−2.57	0.01
BABC × No_1	0.12	0.27	0.46	0.65
BACB × No_1	−0.84	0.26	−3.25	0.0012
ABCB × Some_1	−1.02	0.27	−3.79	0.00015
BABC × Some_1	0.03	0.27	0.11	0.92
BACB × Some_1	−0.87	0.26	−3.32	0.0009

(Continued on next page)

Table 9.8—continued from previous page

	β	SE	z	p
ABCB × Some…not$_1$	−0.34	0.27	−1.27	0.2
BABC × Some…not$_1$	0.73	0.27	2.67	0.0076
BACB × Some…not$_1$	−0.22	0.26	−0.86	0.39
ABCB × No$_2$	0.17	0.28	0.61	0.54
BABC × No$_2$	−0.63	0.28	−2.25	0.024
BACB × No$_2$	−0.16	0.27	−0.59	0.56
ABCB × Some$_2$	0.28	0.25	1.15	0.25
BABC × Some$_2$	−0.28	0.25	−1.13	0.26
BACB × Some$_2$	−0.10	0.24	−0.42	0.67
ABCB × Some…not$_2$	0.66	0.26	2.55	0.011
BABC × Some…not$_2$	−0.38	0.26	−1.46	0.15
BACB × Some…not$_2$	0.31	0.25	1.22	0.22

Figure 9.5 (p. 201) shows the probabilities of an AC term-order predicted by the fixed-effects of the multilevel model shown in Table 9.8 for a series of quantifier pairs, concentrating on those pairs of relevance to the source-founding model. The graphs finesse between-participant variation in the intercept. This poses difficulties for interpretation, but they may honestly be viewed as illustrating group mean effects and also differences in response between conditions, e.g., between the different figures and, cross graph, different quantifiers.

For problems where the first premise contains a universal quantifier and the second an existential, AQ score is positively correlated with the difference in probability of AC between figures $ABBC$ and $BACB$. So for these quantifier configurations, the higher a reasoner's AQ, the more figure is driving term-order. Consider now figure $BACB$, which, if figure is driving inference, should be more likely to generate a CA conclusion. It can be seen that higher AQ implies a tendency to source from the premise which contains the existential quantifier.

Each of the included item types combine an existentially quantified premise with a universally quantified premise. From a discourse point of view these are interesting as existentials may be viewed as introducing individuals to the discourse. Also from a logical perspective, universally quantified statements may be viewed as devices for transforming facts about one kind of individual into facts about other kinds of individual.

Figure 9.6 shows the model's predictions of term-order for the other quantifiers. One noticeable effect is when faced with the quantifiers *no-all* when the first premise term-order is AB, higher AQ implies more a preference to source from the *no* premise.

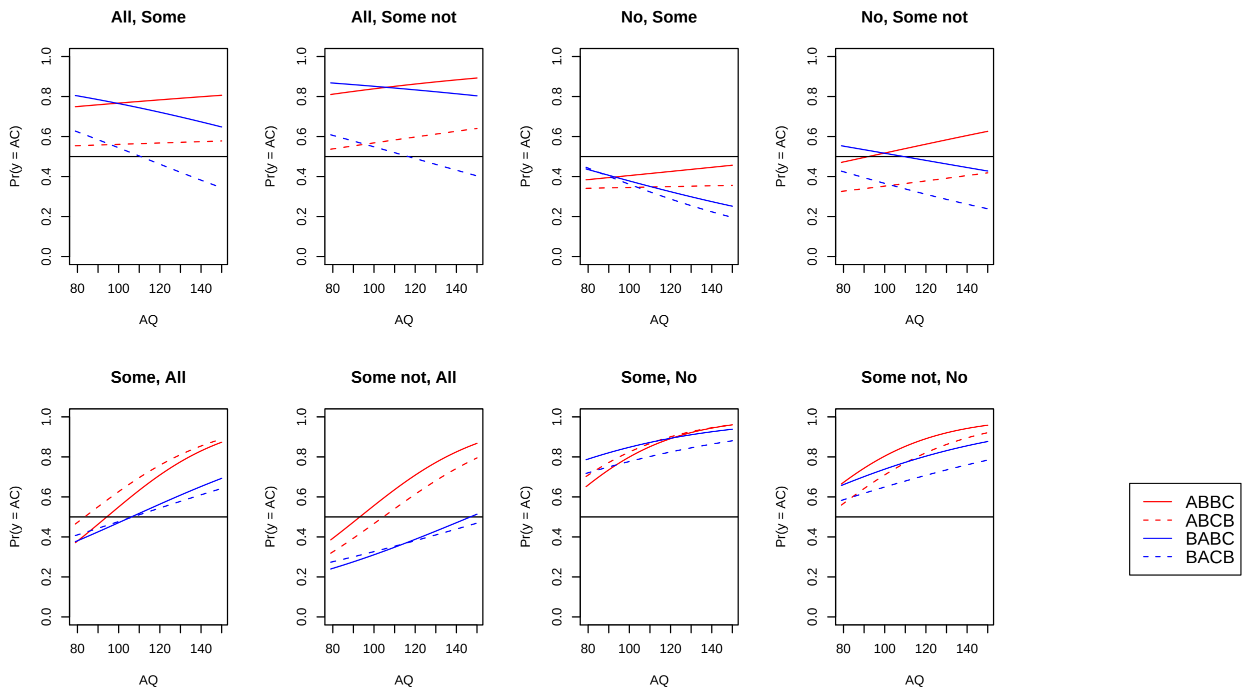

Figure 9.5: Term-order predicted by the fixed-effects of the multilevel model shown in Table 9.8 for premise pairs with one existential and one universal premise.

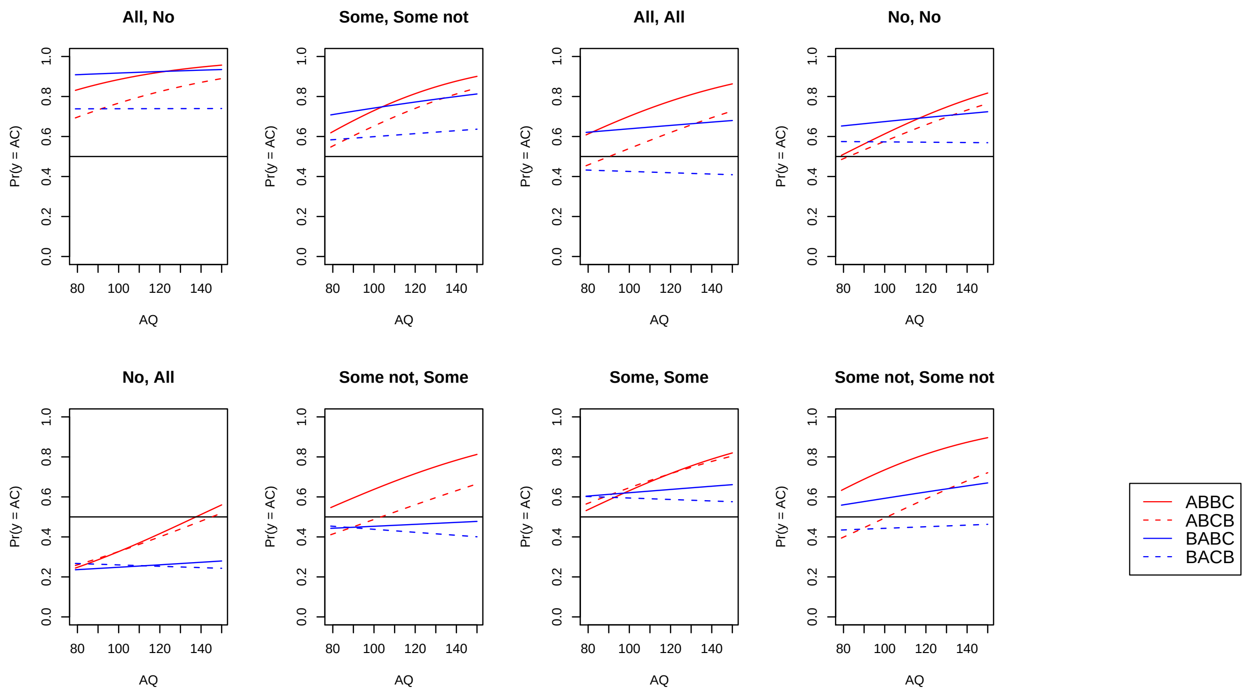

Figure 9.6: Term-order predicted by the fixed-effects of the multilevel model shown in Table 9.8 for premise pairs with both existential or both universal premises.

Premise order

Q_1	**A**	B
Q_2	B	C
Q_3	**A**	C

Existential end-term

all	B	A
some	B	**C**
some	**C**	A

Premise term-order

Q_1	**A**	B
Q_2	B	**C**
Q_3	**A**	**C**

Q_1	B	**A**
Q_2	**C**	B
Q_3	**C**	**A**

Figure 9.7: A simplified model. Bold terms drive the construction of the critical individual.

These complex models demonstrate that something of interest is going on with syllogisms and the self-report autistic-traits, but perhaps we can produce a more easily interpretable model by ascending to the theory. Figure 9.7 shows a simpler model of the forces guiding the inferences drawn when a conclusion is drawn.

First a model was fitted to examine the effect of premise order, premise term-order (1 for *ABBC*, −1 for *BACB*, 0 otherwise), existential position (1 if there is a unique existential in the first premise, −1 if there is a unique existential in the second premise, 0 otherwise), and the pull towards positive premises (1 if there is a unique positive quantifier in the first premise, −1 if there is a unique positive quantifier in the second premise, 0 otherwise). Probability of an AC conclusion was predicted with a random intercept for participants.

There were statistically significant effects of sourcing from an existential [ΔBIC $= -31$, $\chi^2(1) = 39.8$, $p < 0.001$], a drive to maintain subject and predicate end-term orders in the conclusion order [ΔBIC $= -89$, $\chi^2(1) = 98.2$, $p < 0.001$], and a preference for sourcing from a premise with a positive quantifier [ΔBIC $= -347$, $\chi^2(1) = 356$, $p < 0.001$]. Also the mean tendency was to prefer AC conclusions, indicating an effect of premise order [intercept $= 0.45$, Wald $z = 7.1$, $p < 0.001$].

Examining now effects of AQ, sourcing from an existential is moderated by AQ [ΔBIC

	β	SE	z	p	
Premise order	0.461	0.0638	7.2	< 0.001	***
Source from unique existential	0.273	0.0444	6.2	< 0.001	***
Keep premise term-order	0.433	0.0455	9.5	< 0.001	***
Source from unique positive	0.827	0.0463	17.9	< 0.001	***
AQ (centred)	0.010	0.0043	2.3	0.02	*
Source exist $\times$ AQ	0.014	0.0031	4.5	< 0.001	***
Keep term-order $\times$ AQ	0.011	0.0032	3.5	< 0.001	***

Table 9.9: Moderating effects of AQ on the term-order predictions of the source-founding model. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

$= -12$, $\chi^2(1) = 20.4$, $p < 0.001$], as is tendency to maintain premise term-order [ΔBIC $= -4$, $\chi^2(1) = 12.3$, $p < 0.001$], but not polarity [ΔBIC $= 8$, $\chi^2(1) = 0.4$, $p = 0.5$]. Higher AQ also predicts an overall tendency to draw an AC conclusion, i.e., maintaining the premise order. See Table 9.9 for fixed-effects estimates.

Now we have simplified the model, it is feasible to explore whether effects of the subscores of AQ and BAPQ may be fractionated. Beginning first with BAPQ (for pragmatic reasons: fewer variables). There was a hint of an additional association with quantifier polarity and rigidity [ΔAIC $= -1$, ΔBIC $= 5$, $\chi^2(1) = 3$, $p = 0.083$], though the BIC estimate suggests evidence of overfitting. Simplifying by removing the three interaction terms related to polarity has no statistically significant effect on fit [ΔBIC $= 20$, $\chi^2(3) = 4.5$, $p = 0.22$]. The final model chosen is presented in Table 9.10. This exposes an interesting phenomena: rigidity predicts a greater preference for sourcing from the existential and for maintaining premise term-order, however pragmatic language impairment (but not aloofness) predicts more sourcing from the existential and being aloof (but not pragmatic language impairment) predicts a greater tendency to maintain the premise-term order. This statistical model suggests that an increased tendency to source from an existential is not related to pragmatic language competence, but rather to an independent ecthetic ability. (Though there is still an overall mean effect that people tend to source from the

	β	SE	z	p	
Premise order	0.482	0.0661	7.3	< 0.001	***
Source from unique existential	0.290	0.0454	6.4	< 0.001	***
Keep premise term-order	0.444	0.0467	9.5	< 0.001	***
Source from unique positive	0.857	0.0472	18.2	< 0.001	***
Aloof	0.083	0.0936	0.9	0.374	
Pragmatic Language	0.093	0.1202	0.8	0.438	
Rigid	−0.009	0.1011	−0.1	0.931	
Source exist. × Aloof	−0.014	0.0645	−0.2	0.831	
Source exist. × Prag. Lang.	0.241	0.0823	2.9	0.004	**
Source exist. × Rigid	0.141	0.0698	2.0	0.043	*
Premise term-order × Aloof	0.200	0.0662	3.0	0.003	**
Premise term-order × Prag. Lang.	−0.075	0.0859	−0.9	0.382	
Premise term-order × Rigid	0.147	0.0721	2.0	0.042	*

Table 9.10: Moderating effects of BAPQ subscores on the term-order predictions of the source-founding model. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

existential; this is moderated by pragmatic language ability.) It also suggests again that the effect of figure is not related to pragmatic language.

Let's do the same again for AQ's subscores. The pattern is practically the same as for BAPQ, except it appears that the interactions with premise term-order have vanished. The model *without* these five interactions fits significantly worse [ΔAIC $= -4$, ΔBIC $=$ 29, $\chi^2(5) = 13$, $p = 0.02$] (though BIC hints at overfitting). Also all five estimates point in the same direction. Note that now the interaction between sourcing with the positive and Attention Switching (close to Rigidity in BAPQ) is statistically significant. See Table 9.11 for the fixed-effect estimates.

	β	SE	z	p	
Premise order	0.471	0.0628	7.5	< 0.001	***
Source from unique existential	0.285	0.0448	6.4	< 0.001	***
Keep premise term-order	0.440	0.0459	9.6	< 0.001	***
Source from unique positive	0.828	0.0467	17.7	< 0.001	***
Social Skills	0.027	0.0178	1.6	0.122	
Attention Switching	0.008	0.0176	0.5	0.639	
Attention to Detail	0.032	0.0155	2.1	0.039	*
Communication	−0.000	0.0203	−0.0	0.980	
Imagination	−0.011	0.0194	−0.6	0.570	
Source exist. × Social Skills	0.004	0.0128	0.4	0.729	
Source exist. × Attention Switching	0.007	0.0127	0.6	0.575	
Source exist. × Attention to Detail	0.027	0.0113	2.4	0.018	*
Source exist. × Communication	0.047	0.0147	3.2	0.001	**
Source exist. × Imagination	−0.013	0.0140	−1.0	0.344	
Premise term-order × Social Skills	0.017	0.0131	1.3	0.187	
Premise term-order × Attention Switching	0.011	0.0130	0.9	0.384	
Premise term-order × Attention to Detail	0.003	0.0116	0.3	0.779	
Premise term-order × Communication	0.010	0.0152	0.7	0.495	
Premise term-order × Imagination	0.005	0.0142	0.4	0.724	
Source positive × Social Skills	−0.027	0.0132	−2.1	0.038	*
Source positive × Attention Switching	0.037	0.0133	2.8	0.006	**
Source positive × Attention to Detail	0.013	0.0118	1.1	0.262	
Source positive × Communication	−0.004	0.0152	−0.3	0.783	
Source positive × Imagination	0.017	0.0145	1.2	0.245	

Table 9.11: Moderating effects of AQ subscores on the term-order predictions of the source-founding model. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

	β	SE	z	p	
(Intercept)	−2.240	0.2074	−10.8	< 0.001	***
Both negative quantifiers	0.793	0.0511	15.5	< 0.001	***
Both existentials	0.626	0.0512	12.2	< 0.001	***
Columnar term-order	0.196	0.0481	4.1	< 0.001	***
AQ (centred)	−0.001	0.0138	0.0	0.97	
Columnar × AQ	0.011	0.0031	3.5	< 0.001	***

Table 9.12: Moderating effects of AQ on the NVC predictions of the source-founding model. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

9.3.2 Predicting whether a conclusion is drawn

Another aspect of the source-founding model which hasn't yet been investigated for individual differences is what drives people to draw a conclusion at all. As discussed in Chapter 8, there is a range of individual differences in the degree to which people are rash (drawing a conclusion when classically it is not valid to do so) or hesitant (withholding an conclusion when classically there is one to draw) so there is reason to believe something of this nature may be occurring for syllogisms as well.

A series of models were fitted first to examine overall preference for drawing NVC, with predictors of whether both premises have negative quantifiers (1 if yes, −1 otherwise), both existential (1 if yes, −1 otherwise), and whether the premise term-orders are columnar, i.e., *ABCB* or *BABC*. All of these predictors should increase the probability of an NVC, and indeed they did. The largest effect was for the quantifier polarity [ΔBIC = −245, $\chi^2(1) = 247$, $p < 0.001$], followed by the presence of two existential quantifiers [ΔBIC = −171, $\chi^2(1) = 174$, $p < 0.001$], and then finally a columnar figure [ΔBIC = −19, $\chi^2(1) = 21.2$, $p =< 0.001$]. There is also a preference for drawing a conclusion rather than responding NVC [intercept = −2.18, Wald $z = -10.7$, $p < 0.001$].

Now examining interactions with AQ, there was a moderating effect with columnar figure [ΔBIC = −3, $\chi^2(1) = 12.3$, $p < 0.001$]—people with higher AQ were more likely to draw NVC for columnar orders—but none with quantifier polarity or the presence of two existential premises [ΔBIC = 17, $\chi^2(2) = 0.12$, $p = 0.9$], so a model without the latter two predictors was chosen. See Table 9.12 for the fixed-effect estimates.

As previously, let's investigate whether there are contributions of the BAPQ subscores. Taking interactions between the source-founding parameters and triplets of BAPQ subscores, there was an effect for columnar term-orderings [ΔAIC = −4 ΔBIC = 16, $\chi^2(3) = 10.2$, $p = 0.017$], though not by BIC; a trend to an effect of two existentials [ΔAIC = −1 ΔBIC = 19, $\chi^2(3) = 6.7$, $p = 0.082$], and no effect for quantifier polarity as before [ΔAIC = 5, $\chi^2(3) = 0.69$, $p = 0.88$]. See Table 9.13 for the fixed-effect estimates. Being more rigid predicts a higher probability of drawing a conclusion when there are two existentials, a result which is difficult to interpret. People reporting better pragmatic language tend to be more likely to draw a conclusion for columnar term-orders, and there's a trend that they are more likely to draw a conclusion for two-existential premises. These results may be seen as signs of pragmatic language influences in syllogistic reasoning. It is important to note that the term 'pragmatic language' as used in BAPQ is quite general. Hurley et al. (2007, p. 1681) write:

> pragmatic language problems refer to deficits in the social aspects of language, resulting in difficulties communicating effectively or in holding a fluid, reciprocal conversation.

Items include: 'I am "in-tune" with the other person during conversation', 'It's hard for me to avoid getting sidetracked in conversation'. These are traits that people tend to be particularly aware of in themselves, at the personal-level, rather than performance measures at the level of, say, the field of experimental pragmatics. They are also traits that people tend to notice about others. Cognitively the question is, what sub-personal level processes support these behaviours?

It is interesting that the grammar of the premises interacts with pragmatic language effects. As we discuss later, order is very important in language. Order is also important in the source-founding model, e.g., characterising how premise term-orders relate to conclusion term-orders chosen by participants. For the columnar figures, the term-order does not provide a clue for how to construct critical individuals; despite this, those who report being good at pragmatic language tend to be more likely to draw a conclusion, showing a particularly strong force to reason credulously.

Finally let's investigate AQ. Taking groups of interactions, each source-found predictor

	β	SE	z	p	
(Intercept)	-2.228	0.209	-10.7	< 0.001	***
Both negative quantifiers	0.787	0.052	15.2	< 0.001	***
Both existentials	0.613	0.052	11.7	< 0.001	***
Columnar term-order	0.190	0.049	3.9	< 0.001	***
Aloof	0.055	0.295	0.2	0.851	
Pragmatic Language	0.602	0.369	1.6	0.103	
Rigid	-0.366	0.317	-1.2	0.248	
Existentials $\times$ Aloof	0.054	0.071	0.8	0.449	
Existentials $\times$ Prag. Lang.	0.141	0.083	1.7	0.089	†
Existentials $\times$ Rigid	-0.159	0.078	-2.0	0.042	*
Columnar $\times$ Aloof	-0.015	0.067	-0.2	0.826	
Columnar $\times$ Prag. Lang.	0.156	0.076	2.1	0.040	*
Columnar $\times$ Rigid	0.077	0.074	1.0	0.299	

Table 9.13: Moderating effects of BAPQ subscores on the NVC predictions of the source-founding mode. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

and the five subscores of AQ, there were effects of existentials [ΔAIC $= -19$, ΔBIC $= 14$, $\chi^2(5) = 28.5$, $p < 0.001$] and columnar figure [ΔAIC $= -5$, ΔBIC $= 28$, $\chi^2(5) = 14.4$, $p = 0.013$], though not by BIC, but no effect even with AIC for polarity [ΔAIC $= 6$, $\chi^2(5) = 4$, $p = 0.5$]. See Table 9.14 for the fixed-effect estimates. The patterns for the AQ and BAPQ subscore were very similar, except note that the unique variance from the Communication subscore doesn't quite meet significance at the 0.05 level, whereas that for Pragmatic Language in BAPQ did. Also the predictor for Attention Switching from AQ is a much better predictor of the increased tendency to draw a conclusion when there are two existentials than is Rigidity from BAPQ.

Recall that we also found this association between Attention Switching and classical rashness for immediate inference (Section 8.2.3). Given the effect size, and the consistency cross-task, this is evidence that the effect is not merely noise. The closest result I can find in the literature relates to an experiment on concurrent random-tapping (conceivably related to attention switching) and syllogistic performance by Gilhooly et al. (1993). They hypothesised that concurrent-tapping, thought to impair performance on the (rather monolithic) central executive component of working memory, pushes people towards random guessing. Clearly the immediate inference result does not point to random guessing, but rather systematic avoidance of the 'could be true or false' option. It is difficult to see how to relate attention switching to the credulous versus sceptical distinction, but perhaps the Type 1 versus Type 2 distinction is of relevance. Rigidity and attention switching might relate to Type 2, effortful, processes, and pragmatic-type responses might relate to Type 1 effortless, processes. This is contrary to what De Neys and Schaeken (2007) argue, based on their dual-task experiment.

9.3.3 An analysis of integration strategies

One way to solve syllogisms is derived from the observation that the quantifiers can be eliminated, after which only sentential logic rules are required (see Section 3.1.1.2, p. 46). See Figure 9.15 for the rules that apply to each syllogistic conjecture, here focusing on premises where exactly one of the pair is an existential. The non-classically valid DA and AC rules are included, as they were for conditional reasoning (Chapter 7). The most common operation is MP (12 items), followed by MT and AC (both 8), and the

	β	SE	z	p	
(Intercept)	−2.2716	0.204	−11.1	< 0.001	***
Both negative quantifiers	0.8020	0.052	15.6	< 0.001	***
Both existentials	0.5781	0.054	10.8	< 0.001	***
Columnar term-order	0.2016	0.050	4.0	< 0.001	***
Social Skills	0.0159	0.057	0.3	0.780	
Attention Switching	−0.0946	0.057	−1.7	0.096	†
Attention to Detail	−0.0898	0.050	−1.8	0.074	†
Communication	0.0549	0.064	0.9	0.392	
Imagination	0.0405	0.062	0.7	0.512	
Existentials × Social Skills	0.0214	0.015	1.5	0.148	
Existentials × Attention Switching	−0.0739	0.015	−5.0	< 0.001	***
Existentials × Attention to Detail	−0.0262	0.014	−1.8	0.069	†
Existentials × Communication	0.0274	0.016	1.7	0.082	†
Existentials × Imagination	0.0183	0.016	1.2	0.241	
Columnar × Social Skills	−0.0053	0.014	−0.4	0.704	
Columnar × Attention Switching	0.0131	0.013	1.0	0.329	
Columnar × Attention to Detail	0.0063	0.013	0.5	0.635	
Columnar × Communication	0.0262	0.015	1.8	0.073	†
Columnar × Imagination	0.0163	0.015	1.1	0.270	

Table 9.14: Moderating effects of AQ subscores on the NVC predictions of the source-founding model. † $p < 0.1$; ⋆ $p < 0.05$; ⋆⋆ $p < 0.01$; ⋆⋆⋆ $p < 0.001$.

least common is DA (only 4 items). There is an equal split on forward and reverse inferences, however. Four of the syllogisms have what Stenning and Yule (1997) term U-conclusions:[2] *some not-A are not-C* and *some not-C are not-A*. These could not be drawn in my experiment. Charles Lutwidge Dodgson (a.k.a. Lewis Carroll) also investigated these conclusions, for instance noting that all Euler Diagrams assert *some not-x are not-y* (Dodgson, 1958, p. 174). Of course in modern logic, these conclusions are also easily expressible, e.g., as $\exists x.\neg A(x) \wedge \neg C(x)$.

First the modelling proceeded by trying to explain group average inferences. Predictors were included of whether a *no* quantifier was present, which has a very special morphology, the polarity of the inference required (positive for MP and AC, negative for DA and MT), the direction of the required inference (forward for MP and DA, reverse for AC and MT), and also a predictor was included for whether (it appears!) a double negative is required, as it is if the DA inference is drawn for the premise pairs ⟨*no B are A, some C are not B*⟩ and ⟨*some A are not B, no B are C*⟩. All were given a dummy coding. The outcome variable was whether the predicted *some* or *some... not* inference was drawn. For this model, it was assumed that *some* commutes and *some... not* does not.

[2] The U being a continuation of the A, E, I, and O Latin mnemonics of the quantifier names (mentioned in passing in Section 8.2.1, p. 166).

(a)

All A are B		No A are B		All B are A		No B are A	
Some B are C	AC	Some B are C	MT	Some B are C	MP	Some B are C	MP
Some B are not C	AC	Some B are not C^U	MT	Some B are not C	MP	Some B are not C^U	MP
Some C are B	AC	Some C are B	MT	Some C are B	MP	Some C are B	MP
Some C are not B	MT	Some C are not B	AC	Some C are not B	DA	Some C are not B	DA

(b)

All B are C		No B are C		All C are B		No C are B	
Some A are B	MP	Some A are B	MP	Some A are B	AC	Some A are B	MT
Some A are not B	DA	Some A are not B	DA	Some A are not B	MT	Some A are not B	AC
Some B are A	MP	Some B are A	MP	Some B are A	AC	Some B are A	MT
Some B are not A	MP	Some B are not A^U	MP	Some B are not A	AC	Some B are not A^U	MT

Table 9.15: Syllogisms (with one existential and one universal quantifier) with classical modus ponens (MP) and modus tollens (MT) solutions and also classically invalid denial of the antecedent (DA) and affirmation of the consequent (AC). (a) shows problems where the universal is in the first premise and (b) where the universal is in the second premise. ^{U} marks syllogisms with U-conclusions.

	β	SE	z	p	
(Intercept)	-0.9	0.15	-6.1	< 0.001	***
Contains *no*	0.53	0.16	3.3	0.001	**
Positive	2.05	0.14	14.4	< 0.001	***
Reverse	-0.27	0.1	-2.6	0.009	**
Double negative	-0.41	0.22	-1.9	0.057	†
Contains *no* × Positive	-2.37	0.21	-11.4	< 0.001	***

Table 9.16: Model coefficient for aspects of the integration strategies affecting inferences drawn. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

A first attempt to fit all three-way interactions failed to fit. The most complicated model to fit contained all two-way interactions plus a main effect of double-negative. There was no significant interaction between *no* and direction [ΔAIC $= 1$, $\chi^2(1) = 0.6$, $p = 0.45$], or between polarity and direction [ΔAIC $= 2$, $\chi^2(1) = 0.27$, $p = 0.6$], so these predictors were excluded. See Table 9.16 for the model estimates for the final model selected. As can be seen, positive inferences were more likely to be drawn (this was the largest effect), except when the universal premise was *no*. Reverse inferences were less likely to be drawn.

A first attempt to model the effect of AQ failed, so I reverted first to the subscores of BAPQ. The first model fitted contained the most complicated model of sample average performance and interactions with the three subscores of BAPQ. Six terms could be removed: those representing three-way interactions between *no*, polarity, and the BAPQ subscores [ΔAIC $= 3$, $\chi^2(3) = 2.9$, $p = 0.4$], and three-way interactions between *no*, direction, and the BAPQ subscores [ΔAIC $= 5$, $\chi^2(3) = 1.1$, $p = 0.8$]. Nor were there interactions with BAPQ and variables for double negatives [ΔAIC $= 2$, $\chi^2(3) = 3.8$, $p = 0.3$] or with BAPQ and *no* [ΔAIC $= 4$, $\chi^2(3) = 2$, $p = 0.6$]. The final model selected is shown in Table 9.17. The main remarkable effects were that people who report being good at Pragmatic Language tend to be more likely to draw positive reverse inferences, i.e., affirmation of the consequent, and reporting being more Rigid was associated with making fewer reverse inferences overall.

	β	SE	z	p	
(Intercept)	−1.53	0.68	−2.3	0.025	*
Contains *no*	0.51	0.33	1.6	0.120	
Positive	2.40	0.69	3.5	< 0.001	***
Reverse	−0.04	0.75	−0.1	0.957	
Double negative	−0.44	0.41	−1.1	0.287	
Prag. Lang.	−0.13	0.25	−0.5	0.597	
Rigid	0.39	0.22	1.8	0.079	†
Aloof	−0.07	0.20	−0.3	0.731	
Contains *no* × Positive	−2.37	0.29	−8.1	< 0.001	***
Contains *no* × Reverse	0.13	0.25	0.5	0.597	
Positive × Reverse	0.76	0.92	0.8	0.408	
Positive × Prag. Lang.	0.04	0.25	0.2	0.870	
Positive × Rigid	−0.31	0.22	−1.4	0.164	
Positive × Aloof	0.16	0.21	0.8	0.431	
Reverse × Prag. Lang.	0.43	0.26	1.6	0.102	
Reverse × Rigid	−0.50	0.24	−2.1	0.037	*
Reverse × Aloof	−0.01	0.22	−0.1	0.947	
Positive × Reverse × Prag. Lang.	−0.88	0.33	−2.7	0.008	**
Positive × Reverse × Rigid	0.43	0.29	1.5	0.145	
Positive × Reverse × Aloof	0.20	0.27	0.8	0.448	

Table 9.17: Model coefficient for aspects of the integration strategies affecting inferences drawn and moderating effects of BAPQ subscores. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

9.4 Summary

Once one ceases to analyse everything with respect to classical logic scoring and instead focuses on features like term-order and whether a conclusion is drawn, the models become more complex. But it is clear from a range of studies that this is what needs to be done to characterise the inferences that people draw and how the processes responsible for these inferences are more general than the syllogistic.

A first attempt to relate syllogisms and AQ in a purely exploratory fashion resulted in a massive and difficult to interpret model, but reassuringly at least one with huge effects. Following this we examined a series of variations on the source-founding model, inspired by the approach of Stenning and Cox (2006). The present approach uses random effects at the level of individuals as well as residuals and I have tried to emphasise simplicity, following the mantra that all models are false, so it's a good idea to begin with comprehensibly false models.[3]

I viewed the source-founding model in three ways: predicting term-order, simplifying the model and extending to include autistic-traits; predicting whether people draw a conclusion at all; and finally predicting the four main types of inferences, MP, MT, DA, and AC. As the results show, self-reported autistic-traits do indeed moderate the parameters of the source-founding model. Interestingly, it appears that the different components of both AQ and BAPQ differentially moderate the parameters of the process model. For instance those who report problems with pragmatic language tend to be more likely to source from an existential (i.e., use ecthetic proof). This suggests that sourcing from the existential does not reflect a discourse processing mode of reasoning. Those who report being aloof tend to be more influenced by the premise term-orders and draw a conclusion maintaining this order, perhaps indicative of a more stimulus-dependent bottom-up mode of processsing. Similar differences and commonality between the components of AQ and BAPQ are found for the other models.

A range of cognitive processes could be at work here. Overall cognitive architecture is likely to play a role, for instance moderating the degree of gisting taking place, dependence on stimulus-oriented as opposed to stimulus-independent processes, and thus affecting

[3]Statisticians usually attribute the maxim 'All models are wrong but some are useful' to George E. P. Box, though presumably he wasn't the first to express the general sentiment.

likelihood of maintaining subject-predicate ordering. Finally there were some signs of inferences related to what it is thought the interlocutor was trying to convey.

Chapter 10

Unity and Diversity of Reasoning Processes

Up until now we have encountered many details of the individual tasks and their associations with self-report autistic traits. There are some hints from the results and also from the wider literature that there may be something in common between aspects of the tasks. In the present final results chapter we investigate this. The guiding question is whether the idea from logic of reasoning *to* and *from* interpretations enables us to make greater sense of the inferences that people draw. This chapter draws upon the characterisations explored in previous chapters and seeks to tie together the various tasks we have considered.

10.1 Connecting immediate inference and syllogisms

10.1.1 Classical competence connections

Five scores were used from syllogisms (see Figure 9.1, p. 190, for histograms of their distribution). Recall these were: number classically correct for NVC problems (Correct NVC); classically correct for VC problems (Correct VC); times NVC was drawn for a VC problem (Incorrect NVC for VC); times a VC was drawn for an NVC problem (Incorrect VC for NVC); and number of times an incorrect VC was drawn for a VC problem (Incorrect VC for VC).

From immediate inference, four scores were used (see Section 8.2.2, p. 167, to recall details of their properties): rash (concluding *true* or concluding *false* for logically independent problems) and hesitant (concluding *could be true or false* for logically dependent problems). These characterisations are further subdivided into in-place and out-of-place, referring to whether the conclusion term-order matches that of the premises. An unrash response is classical. It is possible, though rare, to be neither hesitant nor classical for logically dependent problems.

Since not all of the scores are Gaussian distributed, Spearman's ρ is used throughout. (Exact p-values could not always be computed because of ties after ranking.) A number of correlations were found. Beginning with classical misses, drawing the wrong VC conclusion in syllogisms for problems where another VC would have been correct was positively correlated with hesitancy in-place ($\rho = 0.41$, $p < 0.001$), rashness in-place ($\rho = 0.29$, $p \approx 0.007$), and rashness out-of-place ($\rho = 0.22$, $p \approx 0.04$). Wrongly drawing a VC for NVC problems (i.e., being rash in syllogisms) was positively correlated with both rashness in-place ($\rho = 0.35$, $p \approx 0.001$) and out-of-place ($\rho = 0.33$, $p \approx 0.002$), as one would have hoped.

Giving the classical answer for VC problems was negatively correlated with hesitancy in-place ($\rho = -0.44$, $p < 0.001$) and rashness in-place ($\rho = -0.26$, $p \approx 0.02$). Giving the classical answer for NVC problems was negatively correlated with rashness in-place ($\rho = -0.34$, $p \approx 0.002$) and out-of-place ($\rho = -0.34$, $p \approx 0.002$).

The following section explores these relationships in more detail in the context of the statistical instantiations of the source-founding models.

10.1.2 Source founding with hesitancy and rashness

I have stalled explanation of the full model by Stenning and Cox (2006) until now as it takes some time to absorb. The final model fitted by the authors consisted of 28 predictors. The key additional aspects to consider now are the interactions with the hesitancy and rashness classifications from the immediate inference task. A range of interactions are present. Hesitancy out-of-place moderates the effect of premise term-order and the presence of a *no* premise. Rash out-of-place also moderates the effect of a *no* premise. Finally rash in-place moderates the effect of a *some... not* premise, which in turn is moderated by premise term-order.

First let's try the simplified model explained earlier which focuses on the main aspects of the process model predicting conclusion term order—premise order, premise term-order, existential location, and premise polarity—in an attempt to reduce the number of free parameters. These parameters capture the main features of the syllogisms which interacted previously with rashness and hesitancy. It's worth mentioning in passing that Stenning and Cox use a model which assumes all responses are independent, whereas the current models allow participants to have varying intercepts, taking into consideration the empirical fact that participants vary in their dependence on premise order. Table 10.1 shows the fixed-effects estimates for the final model chosen. The two largest effects were interactions with hesitancy out of place which predicted a reduced tendency to source from a unique existential premise in syllogisms (Stenning and Cox do not have a term in their statistical model for this effect) and also reduced the trend to source from a unique positive premise (the closet analogue of this in Stenning and Cox's model is a two-way interaction between figure and *no*). A decreased tendency to maintain premise term-orders was associated with increased rashness out-of-place, which would be consistent with a view that rash out-of-place reasoners are using a more order-invariant representation of the problems. Rash out-of-place does appear in the model by Stenning and Cox, however only in connection with the location of a unique *no* premise.

Now again with the NVC predictions. Given the results reported in the previous section showing correlations between immediate inference rashness and syllogistic rashness, one would hope the detailed model to highlight just where these relationships are turning up. First a model was fitted with all possible interactions between the logical predictors, and their interactions with the four immediate inference scores. This was then simplified hierarchically, removing non-significant predictors: all four-way interactions [ΔAIC = 4, $\chi^2(4) = 4.2$, $p = 0.4$] and then three-way interactions of existentials, columnar term-order, and the four immediate inference scores [ΔAIC = 4, $\chi^2(4) = 4.3$, $p = 0.4$]. The fixed-effect estimates for the final model selected is shown in Table 10.2. By far the biggest interaction is between rashness in-place and existentials: more rash on immediate inference implies drawing more conclusions, a kind of syllogistic classical rashness, which seems clear evidence of a cross-task notion of credulous interpretation. The combination of

	β	SE	z	p	
(Intercept)	0.336	0.263	1.3	0.203	
Source from unique existential	0.603	0.193	3.1	0.002	⋆⋆
Keep premise term-order	0.785	0.217	3.6	< 0.001	⋆⋆⋆
Source from unique positive	1.646	0.222	7.4	< 0.001	⋆⋆⋆
Hesitant (in-place)	0.139	0.088	1.6	0.113	
Hesitant (out-of-place)	−0.061	0.037	−1.6	0.102	
Rash (in-place)	0.074	0.076	1.0	0.330	
Rash (out-of-place)	−0.012	0.032	−0.4	0.712	
Source exist. × Hes. (in)	−0.123	0.061	−2.0	0.045	⋆
Source exist. × Hes. (out)	−0.094	0.026	−3.7	< 0.001	⋆⋆⋆
Source exist. × Rash. (in)	0.007	0.054	0.1	0.902	
Source exist. × Rash. (out)	−0.037	0.022	−1.7	0.096	†
Premise term-order × Hes. (in)	−0.073	0.064	−1.2	0.249	
Premise term-order × Hes. (out)	−0.030	0.026	−1.1	0.257	
Premise term-order × Rash. (in)	−0.006	0.059	−0.1	0.925	
Premise term-order × Rash. (out)	−0.049	0.023	−2.2	0.031	⋆
Source positive × Hes. (in)	−0.002	0.065	0.0	0.980	
Source positive × Hes. (out)	−0.097	0.027	−3.7	< 0.001	⋆⋆⋆
Source positive × Rash. (in)	−0.115	0.060	−1.9	0.056	†
Source positive × Rash. (out)	−0.047	0.023	−2.0	0.043	⋆

Table 10.1: Predicting term-order from source-founding parameters and immediate inference classifications. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

two negative premises and two existentials or two negatives and columnar figure tends to push more rash out-of-place people towards drawing a conclusion (or phrased differently, doesn't seem to put them off drawing a conclusion). Interestingly, two negative existentials (or negative universals or existentials, with columnar figure), is finally enough to push rash-out-place reasoners towards not drawing a conclusion.

Finally, let's investigate the model of forward/reverse positive/negative inferences for when there is one existential and one universal premise. The final model chosen is displayed in Table 10.3. Those who were hesitant out-of-place were less likely to draw the predicted inferences in the presence of a *no* quantifier. This could be because half of the hesitant out-of-place problems include *no* quantifiers, and the interpretation of *no* is carrying between tasks. Both those who are hesitant in-place and rash in-place were more likely to draw reverse positive (i.e., affirmation of the consequent; AC) conclusions.

10.2 Conditional reasoning and quantifier reasoning

There is good reason to believe that there will be some overlap in interpretative processes between conditional reasoning problems and problems involving universally quantified statements present in both immediate inference and categorical syllogisms. Quantified statements *all A are B* may be thought of as expressing *if it is an A then it is a B*. For instance, *all humans are animals* may be expressed *if it is a human then it is an animal.* The empirical question is, then, if there is any commonality between interpretations.

Recall that in the key condition of the suppression task presenting two conditionals were presented:

If P, then R
If Q, then R

In the MP additional case and DA alternative case, CWR will result in *if Q then R*; MT (additional) and AC (alternative) will give *P or Q*.

10.2.1 One-premise and suppression

Table 10.4 shows coefficient estimates for four logistic regressions of each type of suppression on the immediate inference characterisations. As can be seen, there were effects for

	β	SE	z	p	
(Intercept)	−2.88	0.8	−3.6	< 0.001	***
Both negative quantifiers	4.26	0.58	7.3	< 0.001	***
Both existentials	5.20	0.52	10.1	< 0.001	***
Columnar term-order	1.16	0.42	2.8	0.005	**
Hes. (in)	−0.20	0.32	−0.6	0.544	
Hes. (out)	0.05	0.13	0.4	0.698	
Rash. (in)	−0.40	0.25	−1.6	0.109	
Rash. (out)	0.03	0.12	0.3	0.777	
Both negative quantifiers × Both existentials	−3.90	0.9	−4.3	< 0.001	***
Both negative quantifiers × Columnar term-order	−1.65	0.69	−2.4	0.017	*
Both existentials × Columnar term-order	−0.24	0.28	−0.9	0.39	
Both negative quantifiers × Hes. (in)	0.20	0.25	0.8	0.405	
Both negative quantifiers × Hes. (out)	−0.17	0.1	−1.6	0.101	
Both negative quantifiers × Rash. (in)	−0.20	0.19	−1.1	0.288	
Both negative quantifiers × Rash. (out)	−0.23	0.09	−2.6	0.01	*
Both existentials × Hes. (in)	−0.32	0.22	−1.5	0.144	
Both existentials × Hes. (out)	0.05	0.08	0.6	0.574	
Both existentials × Rash. (in)	−0.73	0.16	−4.7	< 0.001	***
Both existentials × Rash. (out)	−0.20	0.07	−2.7	0.006	**
Columnar term-order × Hes. (in)	−0.03	0.22	−0.2	0.874	
Columnar term-order × Hes. (out)	−0.20	0.08	−2.6	0.009	**
Columnar term-order × Rash. (in)	0.21	0.14	1.5	0.126	
Columnar term-order × Rash. (out)	−0.19	0.07	−2.8	0.006	**
Negatives × Existentials × Columnar term-order	1.44	0.46	3.1	0.002	**
Negatives × Existentials × Hes. (in)	0.13	0.33	0.4	0.683	
Negatives × Existentials × Hes. (out)	0.06	0.13	0.5	0.655	
Negatives × Existentials × Rash. (in)	0.21	0.26	0.8	0.426	
Negatives × Existentials × Rash. (out)	0.25	0.11	2.2	0.028	*
Negatives × Columnar × Hes. (in)	−0.19	0.30	−0.6	0.534	
Negatives × Columnar × Hes. (out)	0.18	0.12	1.5	0.145	
Negatives × Columnar × Rash. (in)	−0.14	0.22	−0.7	0.512	
Negatives × Columnar × Rash. (out)	0.26	0.11	2.4	0.015	*

Table 10.2: Moderating effects of immediate inference classifications on the NVC predictions of the source-founding model. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

	β	SE	z	p	
(Intercept)	−0.24	0.67	−0.4	0.726	
Contains *no*	−0.36	0.39	−0.9	0.364	
Double negative	−0.65	1.02	−0.6	0.521	
Positive	2.42	0.74	3.3	0.001	⋆⋆
Reverse	1.44	0.77	1.9	0.062	†
Hes. (in)	−0.18	0.26	−0.7	0.497	
Hes. (out)	−0.14	0.11	−1.2	0.224	
Rash. (in)	−0.01	0.20	−0.1	0.944	
Rash. (out)	−0.06	0.09	−0.7	0.506	
Positive × Reverse	−3.3	0.93	−3.5	< 0.001	⋆⋆⋆
Contains *no* × Hes. (in)	0.09	0.14	0.7	0.511	
Contains *no* × Hes. (out)	−0.14	0.06	−2.4	0.017	⋆
Contains *no* × Rash. (in)	−0.07	0.11	−0.6	0.565	
Contains *no* × Rash. (out)	−0.09	0.05	−1.9	0.053	†
Double negative × Hes. (in)	0.12	0.37	0.3	0.742	
Double negative × Hes. (out)	0.32	0.16	2.0	0.044	⋆
Double negative × Rash. (in)	0.19	0.30	0.6	0.522	
Double negative × Rash. (out)	0.15	0.13	1.2	0.252	
Positive × Hes. (in)	−0.37	0.28	−1.3	0.185	
Positive × Hes. (out)	0.16	0.12	1.4	0.169	
Positive × Rash. (in)	−0.31	0.22	−1.4	0.162	
Positive × Rash. (out)	0.04	0.09	0.4	0.706	
Reverse × Hes. (in)	−0.3	0.31	−1	0.323	
Reverse × Hes. (out)	0.13	0.13	1.1	0.295	
Reverse × Rash. (in)	−0.3	0.23	−1.3	0.186	
Reverse × Rash. (out)	0.08	0.10	0.8	0.455	
Positive × Reverse × Hes. (in)	0.82	0.35	2.4	0.019	⋆
Positive × Reverse × Hes. (out)	−0.14	0.15	−0.9	0.351	
Positive × Reverse × Rash. (in)	0.64	0.28	2.3	0.02	⋆
Positive × Reverse × Rash. (out)	0.02	0.12	0.2	0.866	

Table 10.3: Moderating effects of immediate inference classifications on the integration strategy predictions. † $p < 0.1$; ⋆ $p < 0.05$; ⋆⋆ $p < 0.01$; ⋆⋆⋆ $p < 0.001$.

DA, MT, and AC, but not MP: The first observation of note is that it is the in-place immediate inference problems that predict the forward inference (the atom is the negation of the antecedent) DA, whereas it is out-of-place problems that predict the reverse inferences (the atom is the consequent or its negation) MT and AC. A hesitant response is saying 'can't tell' when the classical response is true or is false. This is similar to CWR-DA as for DA a guarded response will be provided: *if Q then R*, except that of course DA is not classically valid. Being less rash out-of-place implies being more classical on this immediate inference problems. This is particularly unusual as this appears to be associated with not interpreting MT problems classically, instead drawing P *or* Q. Also being less hesitant out-of-place implies more MT-CWR. Finally being less rash out-of-place (more classical) is associated with drawing a disjunction P *or* Q. This could viewed as evidence that the key latent force here is not some general, interpretation invariant, ability at classical logic. Another explanation, alluded to earlier, is that the response 'P or Q' means something like '$\vdash P$ or $\vdash Q$' and not '$\vdash P \vee Q$' (or perhaps, 'either you mean P or you mean Q' as opposed to 'you mean P *or* Q').

10.2.2 Two-premises and suppression

In Section 7.2.3 (pp. 155 ff.) I introduced forward and reverse CWR scores, two scores which represent how consistent someone is with the closed world reasoning model proposed by Stenning and van Lambalgen (2005). Here let's investigate connections with the source-founding model. First for term-order, interactions with the three parameters of the model and the suppression scores were fitted. There was no interaction with the polarity predictors [ΔAIC = 3, $\chi^2(2) = 1.1$, $p = 0.6$]. See Table 10.5 for the model coefficients. There were effects for the other two predictors, with premise term-order and premise order showing the largest effects. Performing CWR for the forward inferences was associated with less of a premise order effect and less of a an effect of premise term-order. If forward CWR is interpreted as a credulous style of reasoning, then this, together with the pragmatic language result reported in Chapter 9, again points to sourcing from the existential as not being a discourse processing style of inference. Performing CWR for reverse inferences was associated with more of a tendency to maintain premise term-order. It's unclear why this might be. Speculatively, perhaps those who are more affected by

(a) MP additional CWR

	β	*SE*	z	p	
(Intercept)	−0.21	0.84	−0.25	0.80	
Hesitant (in-place)	0.09	0.30	0.31	0.76	
Hesitant (out-of-place)	−0.03	0.13	−0.20	0.84	
Rash (in-place)	−0.05	0.25	−0.21	0.83	
Rash (out-of-place)	0.01	0.11	0.12	0.90	

(b) MT additional CWR

	β	*SE*	z	p	
(Intercept)	−0.47	0.98	−0.48	0.63	
Hesitant (in-place)	0.15	0.33	0.45	0.65	
Hesitant (out-of-place)	−0.35	0.18	−1.97	0.05	⋆
Rash (in-place)	0.41	0.30	1.35	0.18	
Rash (out-of-place)	−0.31	0.13	−2.46	0.01	⋆

(c) DA alternative CWR

	β	*SE*	z	p	
(Intercept)	−0.48	0.89	−0.54	0.59	
Hesitant (in-place)	1.18	0.46	2.56	0.01	⋆
Hesitant (out-of-place)	−0.10	0.14	−0.74	0.46	
Rash (in-place)	0.20	0.27	0.74	0.46	
Rash (out-of-place)	−0.12	0.11	−1.06	0.29	

(d) AC alternative CWR

	β	*SE*	z	p	
(Intercept)	0.83	0.86	0.96	0.34	
Hesitant (in-place)	0.06	0.31	0.18	0.86	
Hesitant (out-of-place)	−0.12	0.13	−0.90	0.37	
Rash (in-place)	0.17	0.26	0.65	0.52	
Rash (out-of-place)	−0.24	0.11	−2.11	0.04	⋆

Table 10.4: Predicting the four types of 'suppression' from immediate inference. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

	β	SE	z	p	
(Intercept)	0.493	0.064	7.8	< 0.001	***
Source from unique existential	0.258	0.045	5.7	< 0.001	***
Keep premise term-order	0.440	0.046	9.5	< 0.001	***
Source from unique positive	0.842	0.047	17.8	< 0.001	***
Forward CWR	−0.209	0.076	−2.8	0.006	**
Reverse CWR	0.094	0.070	1.4	0.178	
Source exist. × Forward CWR	−0.112	0.054	−2.1	0.037	*
Source exist. × Reverse CWR	−0.004	0.049	−0.1	0.928	
Premise term-order × Forward CWR	−0.011	0.055	−0.2	0.845	
Premise term-order × Reverse CWR	0.141	0.051	2.8	0.005	**

Table 10.5: Moderating effects of suppression task CWR scores on the term-order predictors in the source-founding model. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

term-order (recall, we previously hypothesised they might be using more bottom-up processing) are being guided by the surface form of the suppression task to drawing reverse CWR.

Let's try again with NVC predictions. There were no effects of columnar premise term-order [ΔAIC = 2, $\chi^2(2) = 1.4$, $p = 0.5$] or polarity [ΔAIC = 3, $\chi^2(2) = 0.15$, $p = 0.9$]. The model with these predictors removed is shown in Table 10.6; the crucial remaining interaction is that forward CWR is associated with more conclusion drawing for problems where both premises are existentially quantified, i.e., consistent with a credulous interpretation of the task.

10.3 Conditional reasoning

I also tried to connect Wason's selection task and the suppression task. An obvious first place to look is the MP, MT, AC, and DA simple conditional reasoning, and their connections with the four card turns. This was unsuccessful, using logistic regression

	β	SE	z	p	
(Intercept)	−2.483	0.474	−5.2	< 0.001	***
Both negative quantifiers	0.797	0.052	15.2	< 0.001	***
Both existentials	0.817	0.119	6.9	< 0.001	***
Columnar term-order	0.196	0.049	4.0	< 0.001	***
Forward CWR	−0.031	0.262	−0.1	0.907	
Reverse CWR	0.166	0.244	0.7	0.497	
Both existentials × Forward CWR	−0.129	0.063	−2.0	0.041	*
Both existentials × Reverse CWR	0.038	0.060	0.6	0.531	

Table 10.6: Moderating effects of suppression task CWR scores on the NVC predictors in the source-founding model. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

predicting each card turn separately (greatest $\chi^2(4) = 4.9$, p's > 0.3). There was one link between the CWR scores of the suppression task and turning the $\neg Q$: the model with both forward and reverse CWR predictors fitted better than the null model ($\chi^2(2) = 6.7$, $p = 0.035$): tendency for reverse CWR inference drawing predicted a greater tendency to turn the $\neg Q$ card ($z = 2.1$, $p = 0.04$). This is not a massive effect size, so should be interpreted with caution, though again it is interesting that (seemingly) reverse CWR is associated with a more classical inference! The forward CWR score was unrelated ($z = 0.4$, $p = 0.7$).

10.4 Raven's matrices and quantifier reasoning

In Chapter 6 (pp. 119 ff.) we explored qualitative differences in the cognitive processes that people seem to require for items on Raven's Advanced Progressive Matrices. In this section we examine relationships with quantifier reasoning.

10.4.1 Immediate Inference

No correlations were found between the composite RAPM score and the four scores characterising immediate inference (p's from 0.12 to 0.56), i.e., it appear that the best estimate in the dataset of g does not correlate (at least with a large effect size) with immediate inference. A negative correlation was found between performance on the items classified as visual, and rashness out-of-place (Spearman's $\rho = -0.30$, $p = 0.006$). No other correlations were found ($|\rho|$'s < 0.1, p's > 0.35). This correlation means that the better you are at visual RAPM problems, the less rash out-of-place you are. Recall from Section 8.2.3 that from AQ, better imagination implied more rashness. What's going on here? It turns out that there is a positive correlation between imagination impairment and the visual RAPM performance ($\rho = 0.23$, $p = 0.04$), evidence that the self-report notion of imagination ought not be mapped directly to the visual processing required for RAPM items. Score for imaginative adeptness relates to (self-reported claims) to be able to play imaginative games as a child, make up stories, visualise characters in stories. This might be interpretable as a more top-down process, whereas the visual items of RAPM require a more bottom-up process. If we follow this line or argument through then in fact visual performance might not be indicative of some visual method to represent the immediate inference problems, but rather indicative of better ability to manipulate the linguistic terms.

10.4.2 Syllogisms

There is some evidence that categorical syllogisms differentially involve spatial and verbal processes. Ford (1995) and Bacon et al. (2003) discovered that external representation use preference could affect the inferences people draw. Using exploratory factor analysis, Ando et al. (2006) found that the first two factors explaining most of the variance in classical logic ability were related to the quantifiers of the premises and unrelated to content and modality. Let's write a syllogistic conjecture as Q_1-Q_2/Q_3, where Q_1 and Q_2 are the quantifiers of the premises and Q_3 is the quantifier of the classically valid conclusion. The first factor loaded mostly on *all-all/all*, *all-some/some*, *some-all/some*, *all-no/no*, and *no-all/no* problems and the second factor loaded mostly on *no-some/some... not* problems (i.e., only when the *no* premise came first). (Loadings were all positive and the factor scores were setup to correlate positively with classical accuracy.) The authors found that the first factor score correlated mostly with a measure of verbal IQ ($r \approx .3$ for verbal versus $r \approx .2$ for spatial) and the second score with spatial IQ ($r \approx .4$ for spatial versus $r \approx .2$ for verbal), although it is not reported whether this difference in correlations is statistically significant.

Let's try to repeat Ando et al.'s analysis, using the verbal analytic and visual subscores of RAPM. The hypothesis is that the verbal analysis and visual subscores of RAPM will relate to Ando et al.'s analysis of syllogisms. Multilevel logistic regression was used, with classical logical accuracy as the dependent variable. The predictor was coded as one variable: -1 if the item was a memory of Ando et al.'s verbal class, 1 if the spatial class, and 0 otherwise (i.e., all the items were included so we can estimate overall classical ability). Firstly, there was no overall effect of the RAPM analytic and visual scores on classical competence [$\Delta\text{AIC} = 3$, $\chi^2(2) = 0.8$, $p = 0.7$]. Examining the key interactions, there was no moderating effect of the visual score [$\Delta\text{AIC} = 1$, $\chi^2(2) = 0.7$, $p = 0.4$]; there was an interaction with the verbal analytic score [$\Delta\text{AIC} = -6$, $\Delta\text{BIC} = -1$, $\chi^2(1) = 9.5$, $p = 0.002$]: higher score is associated with poorer performance on spatial syllogistic items compared to verbal items [$\beta = -0.16$, $SE = 0.05$, $z = -3.1$]. So it appears that the visual component of RAPM does not relate to the same concept of spatial as consider by Ando et al., however the verbal-analytic component of RAPM does relate to their concept of verbal, as hypothesised.

10.5 Summary

A series of connections were established between categorical syllogisms and immediate inference. As seen in Chapter 9, there are three main ways to characterise the strategies in syllogisms in terms of the effect the stimulus features have on the inferences drawn and how these effects are moderated. Clearly these are related, especially the models of integration strategies (MP, MT, AC, DA) with the other two models of term-order prediction and NVC prediction. One satisfying result worth highlighting is the relationship between rashness in immediate inference and syllogisms, manifested by the tendency to draw conclusions for two existentially quantified premises in syllogisms. This is a clear sign that rash in-place participants, and recall these were the participants who self-reported that they are particularly good at language pragmatics, are treating the syllogisms as some kind of credulous communicative act.

Another novel result is the relationship found between the suppression task and quantifier reasoning. For premise order, this result is somewhat counter-intuitive. One might expect forward closed world reasoning to be associated with more of an effect of premise order: texts tend to introduce topic and then cumulatively built on those topics in an ordered fashion. The opposite is the case. Perhaps less of an effect of premise order is indicative of more integrative processes. And one can see how this might make sense. Those who draw classical modus ponens for the additional case of the suppression task are perhaps drawing conclusions using as few premises as possible, so they might be more affected by the order in which information is presented. The ease with which those with more autistic traits do reverse CWR seems consistent with this hypotheses.

Finally I would like to highlight the connection between quantifier reasoning and the subscores of Raven's matrices. Surprisingly, the verbal analytic part of RAPM does the work of bridging the gap, not the visual part. Again this is consistent with the observation that despite appearing superficially to be a visuospatial task—look, it has pictures!—RAPM loads highly on g, which represent the shared variance between many different operationalisations of intelligence, an important aspect of which must be verbal-analytic reasoning, not only manipulation of pictures. Verbal analytic reasoning is not synonymous with linguistic reasoning. Perhaps the items characterised by Ando et al. (2006) as being verbal in fact require more use of executive functions such as sequencing and planning.

Chapter 11

Conclusions

> 'Why this is a superior test to the Rorschach,' the seated deputy interrupted, producing the next drawing, 'is that it is not interpretive; there are as many wrongs as you can think up, *but only one right*. The right object that the U. S. Department of Psych-Graphics drew into it and certified for it, for each card; that's what's right, because it is handed down from Washington.'
>
> From *A Scanner Darkly* by Philip K. Dick (1977/2006)

I set out to explore whether autistic-like traits, as assessed by self-report in a typically developing population, predict features of the inferences that people draw, and to assess the extent to which interpretations carry between the various tasks within individuals. As Stenning and Cox (2006) write:

> Exploratory results are messy and complicated. [. . .] Such messy explorations are necessary if premature dismissals of theories are to be avoided.

And indeed we have seen some of this mess in the preceding chapters. In this chapter we abstract a little from the details of the results and, recalling the overall goals of the expedition, interpret the results more generally.

At the outset, I assumed that it was meaningful to study reasoning using concepts from logic: interpretation precedes derivation, and logics ought not to be treated as if they only provide solutions to reasoning exams, but rather give ideas for putative properties of the many kinds of inferences that people draw and the mechanisms for drawing them. Participants, I assumed, would bring to bear deductive processes they use outside the lab when confronted with reasoning tasks. This was not to be a study of how well the participants absorbed Logic 101 (though this is an important question too). The thesis opened with a review of the well known triad of theories of ASC: theory of mind impairment, executive dysfunction, weak central coherence. The hypothesis was that these, through self-reported autistic-like traits, would relate to the inferences that people draw.

Logically, interpretation is a very general concept involving the selection of a logic and a mapping between the task and logic—more general than whether 'some' means 'some but not all' or whether 'if' means 'if and only if'. Importantly the goal of the computation also has to be interpreted in addition to the correspondences set between explicit linguistic aspects of the task. Is the goal to find a counterexample? Prove the conclusion is true in all models of the premises? Find a model where the conclusion is true? What aspects of the context are allowed to influence the derivational processes? All reasoning tasks involve interpretation to some degree. Even tasks which appear to test interpretations themselves still involve both an interpretative and a derivational component. There is no pure test of interpretation or pure test of derivation. However there is a continuum of ambiguity of interpretation. Presumably a sentence like '$2 + 3 = 4$' is easier to interpret than is a sentence like:

> Can I say 'bububu' and mean 'If it doesn't rain I shall go for a walk'? (Wittgenstein, 1953/2001)

For the former, an arbitrary individual equipped with Hindu-Arabic arithmetic is extremely likely to come to an understanding that is consistent with many others' interpretation. For the latter some will respond, 'No, actually you can't' whereas others will build a career analysing interpretations of such utterances. The continuum of ambiguity has been seen in the present thesis. At one end of the continuum lies the conclusion that *all A are B* cannot be true if *no A are B* is true, drawn by 83 out of 84. Closer to the other lies interpretations of the text:

> If she has an essay to finish then she will study late in the library.
> If the library stays open then she will study late in the library.
> She will not study late in the library.

There is quite a spread of responses: 16 out of 84 conclude that she has no essay; 14 conclude that the library is closed; 23 conclude that either she has no essay or the library is closed. (Interpretation or derivation? See later in this chapter for a discussion.)

The following sections give a recap of the key results found, interpret them more generally with respect to the theories of autism and our guiding distinction between credulous and sceptical reasoning, discuss what we have learned about interpretation and derivation, and finally point to directions for the future.

11.1 Summary and discussion of key results

There was evidence that the credulous versus sceptical interpretation distinction does indeed relate to the inferences that people draw. For instance take the observation that people who consider themselves more adept with pragmatic language tend to be more likely to be rash in-place, i.e., over-inferring classically, seemingly to an intended model. Further note the correlation between rashness in-place for immediate inference problems and the tendency to be syllogistically rash for problems where there are two existential premises, which is again indicative of a goal to integrate sentences and infer an intended model rather than reason sceptically. Forward CWR on the suppression task, an integrative style of inference, also predicted a greater tendency to integrate two existential syllogistic premises. Together this is evidence that a goal to reason credulously carries between these three tasks.

In conversation it is common to draw conclusions, make utterances related to these conclusions, and then later make repairs as further inferences are drawn (e.g., Schegloff, Jefferson, & Sacks, 1977). As Benthem (2008) argues, the dynamic notion of correction is more important than is a static notion of correctness. Presumably the force to draw an inference varies as a function of the extent to which actions taken as a consequence of the inference drawn may be undone.

The issue of generality and specificity recurs often in psychology: the general factor in intelligence (Spearman, 1904)—an extreme case where a single continuous variable is thought to explain (in the statistical not causal sense!) around half the variance in cognitive ability; the unity and diversity of executive function (e.g., Duncan et al., 1997; Friedman et al., 2008); the extent to which characteristics associated with ASC can be fractionated (e.g., F. Happé & Ronald, In press); and here interpretation in reasoning. Bonnefon, Eid, et al. (2007, p. 9) write:

> We believe [...] that most of the interpretative processes that influence con-

> ditional reasoning are task specific; and thus, that subpopulations of reasoners on a conditional reasoning task will not readily map onto, for example, subpopulations of reasoners on the selection task.

The present thesis contributes further evidence of aspects of interpretation that are homogenous cross-task, and the details of that homogeneity. There are mappings between the tasks, including an exploratory result predicting a greater tendency to turn $\neg Q$ in the selection task as a function of reverse CWR score on the suppression task, and mappings across to quantifier reasoning from conditional reasoning. I cannot agree, therefore, that most of the interpretative processes in conditional reasoning are task specific.

As reported previously, Pijnacker et al. (In press) found that people with ASC were less likely to use CWR for forward MP additional inferences than were TDs, and they found a trend to an effect of less CWR for MT for the ASC than the TD group. There were no differences found for AC or DA. In the present sample, CWR was moderated by autistic traits, however the form of the effect is different to that found by Pijnacker et al. Participants with more autistic-like traits were less likely to use CWR for forward inferences and were more likely to use CWR for reverse inferences. The forward inferences related more to pragmatic language and the reverse inferences to social interaction and rigidity. Clearly some kind of cognitive processes related to autistic traits are being detected by performance on the suppression task. This is good news for the continuum view of the autism spectrum, i.e., that the traits are present in milder quantities even in typically developing populations, but the details of the effect haven't quite been pinned-down. One crucial difference between experiments was that in the present experiment participants were free to type any response they desired, and were not limited to respond yes, no, or maybe to putative conclusions as in the experiment run by Pijnacker et al. To further make sense of these phenomena will require allowing participants more freedom in their responses.

There is evidence that forces other than the credulous/sceptical distinction are driving the inferences that people draw. Take for instance the force to maintain premise term-order in syllogistic conclusion term-order. One might have expected this to relate to credulous interpretation, and for instance the pragmatic language impaired to show less of an effect. This was not the case. In fact pragmatic language did not predict the effect of term-order, but aloofness and rigidity did—both in the direction of more autistic-like responses predicting more of an effect of term-order. The best interpretation of this effect is that the premise-term order acts as a syllogistic implementation of the embedded figures effect. Perhaps those who report being more aloof and rigid are more likely to be engaging in stimulus-oriented, bottom up processing.

Differences between perceptual processing tendencies are also revealed by performance on Raven's Advanced Progressive Matrices (RAPM). The main result was that the higher an individual's AQ, the easier they find visual items and the harder they find verbal-analytic problems (recall that item classifications were due to DeShon et al., 1995). Initially RAPM were included only to estimate g, but even tests of psychometric intelligence show evidence of qualitative individual differences. There are links too between RAPM and quantifier reasoning: performance on the visual RAPM items predicted less rashness out-of-place and the verbal-analytic score predicted better performance on syllogistic items classified by Ando et al. (2006) as being verbal compared to performance on visual items. These results are difficult to reconcile with a credulous versus sceptical distinction, but may perhaps be considered to reveal properties of the cognitive representations into which the reasoning task may be interpreted, and the availability of planning processes for drawing inferences.

There is a strong theme of order-effects across the tasks. Both immediate inference and categorical syllogism were shown to relate to CWR on the suppression task. For instance forward CWR was associated with in-place immediate inference and reverse CWR was associated with out-of-place immediate inference, showing the subject-premise and antecedent-consequent representations relate cross-task. This is an extension of results by Stenning and Cox (2006) who showed that subject-predicate and premise order were important in immediate inference and syllogisms. Order-effects have a long history in studies of language. Girotto, Mazzocco, and Tasso (1997, p. 24) note that:

> ...order effects, far from being an oddity, are what one expects at all times in communication. From a pragmatic point of view, any newly communicated information is processed in the context of whatever information has been previously processed, so that inferences will vary when the order of presentation, and therefore context, is altered.

Order relates to more general cognitive processes than those supporting language: planning and executing actions and working memory for ordered structures like telephone numbers are other examples. Effects of order are not studied only by cognitive scientists. Literary uses of order are well known, for instance in the *chiasmus*, a crosswise-ordering rhetorical device (e.g., Lund, 1930).

For the self-reported autistic-like traits, contrary to Baron-Cohen, Wheelwright, Skinner, et al. (2001), we found no difference in AQ between males and females, a lack of an effect that has been replicated in other studies (e.g., Hurst, Mitchell, et al., 2007; Hurst, Nelson-Gray, et al., 2007), though sex differences have also been found in other studies (e.g., Austin, 2005). Clearly this is of interest given that males are more likely than females to be diagnosed with ASC, so one might expect, following the continuum hypothesis, males to have higher AQ scores. Another pair of autistic-like trait measures from Baron-Cohen's group, the systemizing quotient (SQ) and empathising quotient (EQ), show greater differences in score between males and females, with males tending to have higher SQ and females higher EQ (Baron-Cohen et al., 2003). The original SQ was later revised as (Wheelwright et al., 2006, p. 49):

> items [. . .] were drawn primarily from traditionally male domains. To counter this, new items were added to the SQ to create the SQ-Revised (or SQ-R), including more items that might be relevant to females in the general population.

The new version did again show higher scores for males. It is difficult to tell if this is because SQ is still just a good test of maleness. There are also some intriguing results on sex differences in the opposite direction to that expected, for instance with Russell, Tchanturia, Rahman, and Schmidt (2007) finding a male advantage on a test of theory of mind on which, given its relationship to empathy, females were expected to show better performance. Another test of sex differences was performed by Voracek and Dressler (2006) who investigated relationships between digit ratio (2D:4D)—thought to indicate permanent effects of prenatal testosterone on the brain—and AQ, SQ, and EQ. They found sex differences for 2D:4D, EQ, SQ, and AQ (though the AQ difference was again small), all in the expected direction, however 2D:4D was unrelated to EQ, SQ, and AQ. Clearly the over representation of males in ASC groups is still not well understood.

My results provide evidence that deductive reasoning, suitably broadly conceptualised, does indeed relate to autistic traits. The cognitive theories of autism are at a fairly early stage of development, and much work needs to be done to examine the details of performance, to which this thesis contributes for non-clinical populations. However it is helpful to step back from the empirical details and recall the general idea of the triad of theories for the triad of impairments (and additional peaks of advantage) in ASC.

The imputation of others' mental states seems important when drawing inferences about the intended meaning of discourse, and, as reviewed earlier, people with ASC have difficulty inferring intentions behind discourse. To see how theory of mind could apply to the tasks used in the data collection, one has to remember the whole setup for the experiment. Although answering questions on a computer screen, the participant would have known that, indirectly, they were answering questions chosen by me. Some questions made reference to other people, e.g., a girl in a library. Others, such as immediate inference and syllogisms, could be viewed as (me, some unknown author?) trying to convey an intended model. Inferring that a speaker means 'some but not all' on uttering 'some' depends upon the reasoner's inferences about the speaker's knowledge, moderated by any uses of irony (cf. the example by Noveck & Reboul, 2008, of a teacher saying, 'some of you passed the exam', despite knowing that all passed). Finally it's important to bear in mind that the participant typically will not have used their cognitive processes for solving explicit psychology tasks about reasoning, but rather will have been out in the world drawing inferences about massive quantities of speech and text, thousands of mini facial convulsions. It is unlikely the processes developed out there will have ceased to function completely in the lab.

Executive functions are involved in many aspects of cognition, as already outlined. Worth commenting on here is the often discussed distinction of System 1 and System 2 or Type 1 and Type 2 processes (e.g., Evans, 2008). Type 1 processes are fast and feel automatic, Type 2 processes are slow and feel effortful. This seems very closely related to the distinction between action/thought schemas and the Supervisory Attention System (e.g., Shallice, 1982). What can we say about a process that is fast? As we saw in the analysis of immediate inference, how fast one is depends not only on the type of inference drawn, but also on deeper traits of the person. As an organising framework, the notion of expertise and what influences expertise during development is likely to be important. If participants have not seen the particular lab tasks before then this expertise necessarily comes from outside the lab: linguistic interpretation, perceptual ability, algebra perhaps?

Finally from the triad of theories, central coherence, the extent to which people tend to try to integrate information across many different processes, is arguably an important factor in the inference people draw. The more general notion of top-down and bottom-up processing, or the continuum of stimulus-oriented versus stimulus-independent thought (e.g., P. W. Burgess, Gilbert, & Dumontheil, 2007; P. W. Burgess, Dumontheil, & Gilbert, 2007), seems to capture some of the patterns of data that we saw in the present thesis.

As reviewed in Section 2.2.4 (pp. 20 ff.), there are many ways in which the triad of theories may overlap, but why stop there? Reasoning is clearly not separate from other aspects of cognitive function. Neither should be the theories used to explain how it functions. This is not to say that inferential behaviour should be *reduced* to products from, e.g., Working Memory and Executive Function Plc, but rather correspondences should be sought between the various theories. There are already some signs of this: Stenning and van Lambalgen (2005) give a logical model of working memory with a mapping to a connectionist network; Stenning and van Lambalgen (2007) give a logical model of theory of mind and executive function. The empirical results reported in this thesis provide further constraints on the form such modelling efforts should take.

One key issue is whether it is valid to consider the traits in TD individuals as being 'autistic' or even 'autistic-like' when by definition the individuals do not have ASC. Barbeau, Mendrek, and Mottron (2008) comment on this, responding to an article relating a version of AQ adapted for children and fetal testosterone. They write:

> . . . what is actually measured, are male traits in a normal population rather than a correlation between testosterone and autism. The measured traits cannot be labelled autistic since the children are not autistic, they are typical traits found in typical individuals at different levels, together with all the normal behavioural traits. Finding a correlation between those traits and testosterone in the normal population does not imply that their presence in autism results from the same causal pathways. [. . .] The autistic brain functions differently, [. . .] but we should consider that it might actually function in its own unique way.

Consider the dimension of height, which shows variation in typically developing populations, but in extremes is also associated with a range of clinical conditions including endocrine disorders (see Kant, Wit, & Breuning, 2005). Barbeau et al. produce an argument akin to that of saying that 'endocrine disorder induced tallness' traits cannot be attributed to people who are not tall and who do not have an endocrine disorder—an argument with which I agree across in the analogy. There must, however, be overlap in causal processes between growth to typical height and growth to an extreme clinical condition driven height, e.g., related to bone structure development. Perhaps we ought then use analogies of height, and the mechanisms producing a given height, back across in the world of ASC. This would amount to referring directly to details of traits found in ASC, e.g., self-reported personality traits and details of cognitive function, and, as already discussed, it was possible to fractionate these individual traits. I would argue in favour of defence of the label 'autistic-like traits' as merely shorthand for the class of traits which are of relevance to a study of ASC, so long as it is emphasised that there is overlap between conditions, for instance between ASC, psychosis, and anorexia. As more is known about the purer dimensions of importance, then it becomes easier to move instead towards discussion of these.

As outlined in Sections 2.4.2 (pp. 27 ff.) and 2.4.3 (pp. 29 ff.), there is a growing body of evidence of continuums of traits related to ASC in unaffected relatives of people with ASC and also in the general population with (as yet) no children with ASC. Self and other reported autistic-like personality traits covary with cognitive function in a way that is consistent with the continuum hypothesis of ASC, so there is accumulating evidence that function in ASC is not unique compared to TD. The underlying causes of this covariation is still under debate. Although there is evidence that impairment in social interaction and field independence covary in both TD and ASC populations, which has been used in favour of the continuum hypothesis, there is also evidence that the two traits are separable. Davidoff, Fonteneau, and Fagot (2008) report evidence that members of a remote tribe, the Himba, are particularly good at the Navon task, performance on which appears to require ignoring perceptual context. Good performance on Navon ought to be associated with poor social functioning, but this is not the case with the Himba, who are very social. Rather it appears that their particular perceptual strategy has been influenced by a need to distinguish between herd animals. So there are ways of becoming field-independent which have no impact on social function. Perhaps cross-cultural studies on cognitive function related to autistic-like traits might reveal the nature of the causes.

Finally I would like to draw the reader's attention to Appendix D (pp. 259 ff.), a combinatorial exploration of first-order models of syllogisms, about which I will say nothing more here (but do have a look).

11.2 Interpretative and derivational knots (again)

Given the present results, where do we draw the boundary between interpretation and derivation? If immediate inference is viewed as a test of interpretation, then the story is easy: a mapping from immediate inference performance to performance on other tasks exposes the interpretative component. A faithful notion of interpretation requires that immediate inference includes both interpretive and derivational components. Intuitively it looks like syllogisms require 'more' derivation than immediate inference, but intuitions can sometimes lead one astray! It could well be possible for two participants to reason to the same interpretation but differ in terms of derivational strategy when drawing inferences. Take two participants who both interpret *some A are B* as $\exists x.A(x) \wedge B(x)$ in the same logic. One of the reasoners' derivational machinery could feasibly more readily access a rule[1] that $\varphi \wedge \psi \equiv \psi \wedge \varphi$, and thus be more likely to draw the inference $\exists x.A(x) \wedge B(x) \vdash \exists x.B(x) \wedge A(x)$. The other reasoner for whom the equivalence rule is less available might not find this inference as easy to draw. This kind of argument has been seen before in Section 3.3 (pp. 62 ff.), where we saw how, e.g., Rijmen and Boeck (2003) argued that some reasoners may not have mastered MT.

Similarly it is possible for people with two different interpretations to reason to the same conclusion. Let's explore this to see where it goes. Take the example of concluding that *some A are not B* follows from *some A are B*. Here are two ways of doing so, for both using a model theoretic and a proof theoretic approach (so as not to invoke more model versus rule misery). For interpretation 1, use the mapping

$$\textit{Some A are B} \quad \rightsquigarrow \quad \exists x.A(x) \wedge B(x)$$

It is possible to conclude by a model search mechanism that there exists a model, M, such that $\exists x.A(x) \wedge B(x) \models_M \exists x.A(x) \wedge \neg B(x)$, and that this model was the intended model. Alternatively prove that $\exists x.A(x) \wedge B(x), \exists x.A(x) \wedge \neg B(x) \nvdash \bot$ and conclude

[1] See later for a brief meditation on levels of analysis.

from the discovery of consistency that the intended conclusion is $\exists x.A(x) \wedge \neg B(x)$. For interpretation 2, use the mapping

$$\textit{Some A are B} \quad \rightsquigarrow \quad (\exists x.A(x) \wedge B(x)) \wedge (\exists x.A(x) \wedge \neg B(x))$$

Let this sentence be the single member of set Γ. By conjunction elimination it is possible to conclude $\Gamma \vdash \exists x.A(x) \wedge \neg B(x)$. Alternatively conclude by a model search procedure that $\Gamma \models \exists x.A(x) \wedge \neg B(x)$. Or perhaps the interpretation of *some A are not B* is the same as that for *some A are B*:

$$\textit{Some A are not B} \quad \rightsquigarrow \quad (\exists x.A(x) \wedge B(x)) \wedge (\exists x.A(x) \wedge \neg B(x))$$

Trivially then one follows from the other.

Although I have used classical logic here, I claim that the arguments transfer to other logics. It is not easy empirically to distinguish between interpretation and derivation! This is not an attack on interpretation. It is a truism that interpretation precedes derivation when a logical approach is taken. It does imply that when focussing on the details of a logical model in a world where interpretations of discourse may vary, there can be many different reasons for drawing an inference. Logically a preference to draw 'pragmatic' inferences may be due to a classical interpretation of the discourse and an explicit search for a model where a sentence is true (nothing about classical logic states that one must use it to reason about all models), or an explicit inference to a less obvious mapping from natural language to logical form with valid classical inference drawing within that interpretation.

One way to untie this knot may be to reconsider levels of analysis. Marr (1982) famously provided three levels of explanation:

1. The computational theory, specifying the goals of the computation.

2. Representation and algorithm, giving a representation of the input and output and the algorithm which transforms one into the other.

3. The hardware implementation, how algorithm and representation may be physically realised.

The idea is that all the levels need to be considered, and mappings between the levels, in order accurately to model cognition. When studying communicative organisms, we also have the issue of what sense to make of self-report personal-level descriptions. Even restricting attention to cognitive-level mechanistic descriptions can lead to many more levels than three: think about the various levels of detail, proximity to the underlying neural machinery, and below to a quantum physical level of explanation.

Let's take a break from our own present struggle and examine struggles that other researchers have had with levels. For instance, take the axiomatic approach to describing the inferences people draw. Da Silva Neves, Bonnefon, and Raufaste (2002) write, of a set of properties they hypothesised that reasoners' inferences ought to satisfy:

> ...even if we expect human inference to corroborate these properties, we know of no sufficient reason to think that lay reasoners would recognize any rationality postulate as valid, neither that they would conscientiously use them to guide their reasoning. [...] we assume that human inference is constrained by knowledge organisation in memory and that its formal properties emerge from a spreading activation process operating directly on knowledge structures.

Politzer et al. (2006, Footnote 2) write that they 'do not test and are not committed to the notion that people have internal representations in the form of Gergonne (1817) diagrams.' Stenning and van Lambalgen (2008, p. 39) write: 'Of course we agree that subjects don't "know these logics" just in the sense that they don't know the grammar of English, but they *do* know these logics just in the sense that they *do* know the grammar of English.'

An example of an instance of level confusion is provided by Rips (1994, p. 349) when he attacks a model of reasoning based on Euler Circles:

> There is little doubt that people can form mental images of Euler circles if given sufficient training with them; however the model is directed not at experts but at the intuitions of naive subjects.

Rips does not appear to have considered the possibility that some aspect of reasoning processes of which participants are unaware could be *put into correspondence with* a system of Euler Circles.

Earlier we saw the result of assuming participants' interpretation and reasoning processes can be directly changed by communicating a theorist's logical description. Bott and Noveck (2004), using a classical logic model of Gricean interpretation, $(\exists x.A(x) \wedge B(x)) \wedge (\exists x.A(x) \wedge \neg B(x))$, attempted directly to explain what usually comes naturally to participants using the classical logic language. The result seemed to be that easy inferences became difficult. This (alone) does not imply that the theorist's model of the utterance interpretation is incorrect. As a thought experiment, consider the roboticist's model of how to walk versus how people learn to walk. Clearly that a typical expert at walking is unlikely to understand the maths of the roboticist's model does imply that the model is invalid.

We struggled with a superficial sketch of the representation and algorithms people might possess for logical inference. Retreating a level to the computation theory might be helpful. Some participants are trying to reason to a likely interpretation which, if later proved to be wrong, can be retracted. Others are trying to reason to an interpretation which is robust to all kinds of adversary. It is interpretation in this more general sense which is distinguishable empirically, the goals the reasoners have—this carries between the tasks. As results related to, e.g., attention switching and rigidity have demonstrated, there are complex interactions between one's executive control processes and interpretative and possibly derivational inferences. I am in agreement with De Neys and Schaeken (2007) that still much work needs to be done to get to grips with these issues. It is unlikely that progress will be made unless researchers move beyond 'logical' versus 'pragmatic' distinctions.

11.3 Future work

There is much scope for theoretical work to flesh out explanations of why inferences relate cross-task in the way that they do, and what sorts of process may be influencing the decision to opt for credulous rather than sceptical interpretation. As I tried to sketch in the previous section, one needs to be careful when describing mechanisms. It might be particularly revealing to implement the same model in different ways and explicitly show mappings between the models. The equivalence class of models is what matters, where the equivalence relation is defined, e.g., by general goals of the reasoners and aspects of behaviour which may be examined experimentally. Theories of the different goals reasoners may have ought also take into consideration how participants themselves theorise and learn about interpretation, perhaps requiring theories of Theory of Interpretation (ToI) in analogy with theories of Theory of Mind (ToM).

Stenning and van Lambalgen (2005) provided a theory of how the suppression task is interpreted by the suppressed using a non-monotonic logic. It is an open question how this logic may be extended to model the inferences that people draw on e.g., immediate inference and syllogisms. It will be important to include some mechanism to explain both the credulous/sceptical distinction and in addition the other dimensions of individual difference that we found in our explorations. Also important will be mechanisms to allow inferences about the speaker's knowledge which aren't explicitly included in the material to be reasoned about.

More work can be done on how the various notions of model relate to each other. A key problem in reasoning credulously is how information packaging influences the selection of an intended model. In logic, the classical notion of first-order model strips away too much information about the linguistic features. An interesting line of research could involve varying the influence of various contextual features on which models are selected, e.g., how many models of a particular size are selected when the term-order of conclusions must match that of the premises? Guided by issues of computational complexity (a topic which Rooij, 2008, discusses with particular clarity, emphasising explicit consideration of the complexity of input parameters), this is likely to give a space of parameters which can feasibly characterise reasoning.

Empirically, more work could be done to investigate whether people with ASC show the same patterns of inferences as do TD individuals with strong autistic-like traits. Given the overlap between ASC and psychiatric conditions such as psychosis, it would also be interesting to extend the investigation into these clinical groups, or using measures of psychosis proneness such as OLIFE in addition to AQ or BAPQ in non-clinical groups.

I began by considering how theories of impaired theory of mind (ToM), weak central coherence (WCC), and executive dysfunction in ASC may help point to causal mechanisms explaining variation in reasoning performance. It was clear as the literature was reviewed, and more so after the data analysis was performed and the results interpreted, that there is a great deal of overlap between the phenomena explained by the theories. Now that there is evidence of connections between self-reported autistic traits and reasoning, it would be interesting to examine performance tests of theory of mind and various executive functions in tests of performance.

Given the apparent effects of stimulus-oriented processing, e.g., for syllogisms, it would be interesting to investigate if the moderating effects of autistic traits remain in the absence of stimulus, e.g., the stimulus could be presented for a limited amount of time, forcing participants to form a cognitive representation and making it impossible to refresh this representation by rereading the screen. Modifications to the syllogistic stimulus may also be interesting, e.g., comparing *all A are not B* with *no A are B* to see the effect of representing negation close to the predicate.

11.4 Postlude

In this thesis, I have argued that social, communicative, and general cognitive control processes are intertwined with reasoning processes, providing statistical models derived from logical models of correspondences. The credulous versus sceptical interpretation distinction explains some but not all of the variation in the inferences drawn. I hope that other researchers are encouraged to study qualitative features of the inferences that people draw, taking seriously deviations from competence model predictions and not just treating such deviations as residual error.

Appendix A

DSM definitions of Autistic Disorder and Asperger's Disorder

A.1 Autistic Disorder

A. A total of six (or more) items from (1), (2), and (3), with at least two from (1), and one each from (2) and (3):

(1) qualitative impairment in social interaction, as manifested by at least two of the following:

(a) marked impairment in the use of multiple nonverbal behaviors, such as eye-to-eye gaze, facial expression, body postures, and gestures to regulate social interaction

(b) failure to develop peer relationships appropriate to developmental level

(c) a lack of spontaneous seeking to share enjoyment, interests, or achievements with other people (e.g., by a lack of showing, bringing, or pointing out objects of interest)

(d) lack of social or emotional reciprocity

(2) qualitative impairments in communication as manifested by at least one of the following:

(a) delay in, or total lack of, the development of spoken language (not accompanied by an attempt to compensate through alternative modes of communication such as gesture or mime)

(b) in individuals with adequate speech, marked impairment in the ability to initiate or sustain a conversation with others

(c) stereotyped and repetitive use of language or idiosyncratic language

(d) lack of varied, spontaneous make-believe play or social imitative play appropriate to developmental level

(3) restricted, repetitive, and stereotyped patterns of behavior, interests, and activities, as manifested by at least one of the following:
   (a) encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is abnormal either in intensity or focus
   (b) apparently inflexible adherence to specific, nonfunctional routines or rituals
   (c) stereotyped and repetitive motor mannerisms (e.g., hand or finger flapping or twisting or complex whole-body movements)
   (d) persistent preoccupation with parts of objects

B. Delays or abnormal functioning in at least one of the following areas, with onset prior to age 3 years: (1) social interaction, (2) language as used in social communication, or (3) symbolic or imaginative play.

C. The disturbance is not better accounted for by Rett's disorder or childhood disintegrative disorder.

A.2 Asperger's Disorder

A. Qualitative impairment in social interaction, as manifested by at least two of the following:
   (1) marked impairments in the use of multiple nonverbal behaviors such as eye-to-eye gaze, facial expression, body postures, and gestures to regulate social interaction
   (2) failure to develop peer relationships appropriate to developmental level
   (3) a lack of spontaneous seeking to share enjoyment, interest or achievements with other people (e.g., by a lack of showing, bringing, or pointing out objects of interest to other people)
   (4) lack of social or emotional reciprocity

B. Restricted repetitive and stereotyped patterns of behavior, interests and activities, as manifested by at least one of the following:
   (1) encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is abnormal either in intensity or focus
   (2) apparently inflexible adherence to specific, nonfunctional routines or rituals
   (3) stereotyped and repetitive motor mannerisms (e.g. hand or finger flapping or twisting, or complex whole-body movements)
   (4) persistent preoccupation with parts of objects

C. The disturbance causes clinically significant impairments in social, occupational, or other important areas of functioning.

D. There is no clinically significant general delay in language (e.g., single words used by age 2 years, communicative phrases used by age 3 years).

E. There is no clinically significant delay in cognitive development or in the development of age-appropriate self-help skills, adaptive behavior (other than in social interaction), and curiosity about the environment in childhood.

F. Criteria are not met for another specific Pervasive Developmental Disorder or Schizophrenia.

Appendix B

Autistic-traits questionnaire items

B.1 Autism-spectrum Quotient

1. I prefer to do things with others rather than on my own.
2. I prefer to do things the same way over and over again.
3. If I try to imagine something, I find it very easy to create a picture in my mind.
4. I frequently get so strongly absorbed in one thing that I lose sight of other things.
5. I often notice small sounds when others do not.
6. I usually notice car number plates or similar strings of information.
7. Other people frequently tell me that what I've said is impolite, even though I think it is polite.
8. When I'm reading a story, I can easily imagine what the characters might look like.
9. I am fascinated by dates.
10. In a social group, I can easily keep track of several different people's conversations.
11. I find social situations easy.
12. I tend to notice details that others do not.
13. I would rather go to a library than a party.
14. I find making up stories easy.
15. I find myself drawn more strongly to people than to things.

16. I tend to have very strong interests which I get upset about if I can't pursue.
17. I enjoy social chit-chat.
18. When I talk, it isn't always easy for others to get a word in edgeways.
19. I am fascinated by numbers.
20. When I'm reading a story, I find it difficult to work out the characters' intentions.
21. I don't particularly enjoy reading fiction.
22. I find it hard to make new friends.
23. I notice patterns in things all the time.
24. I would rather go to the theatre than a museum.
25. It does not upset me if my daily routine is disturbed.
26. I frequently find that I don't know how to keep a conversation going.
27. I find it easy to 'read between the lines' when someone is talking to me.
28. I usually concentrate more on the whole picture, rather than the small details.
29. I am not very good at remembering phone numbers.
30. I don't usually notice small changes in a situation, or a person's appearance.
31. I know how to tell if someone listening to me is getting bored.
32. I find it easy to do more than one thing at once.
33. When I talk on the phone, I'm not sure when it's my turn to speak.
34. I enjoy doing things spontaneously.
35. I am often the last to understand the point of a joke.
36. I find it easy to work out what someone is thinking or feeling just by looking at their face.
37. If there is an interruption, I can switch back to what I was doing very quickly.
38. I am good at social chit-chat.
39. People often tell me that I keep going on and on about the same thing.
40. When I was young, I used to enjoy playing games involving pretending with other children.
41. I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.).
42. I find it difficult to imagine what it would be like to be someone else.

43. I like to plan any activities I participate in carefully.
44. I enjoy social occasions.
45. I find it difficult to work out people's intentions.
46. New situations make me anxious.
47. I enjoy meeting new people.
48. I am a good diplomat.
49. I am not very good at remembering people's date of birth.
50. I find it very easy to play games with children that involve pretending.

B.2 Broad Autism Phenotype Questionnaire

Note, for starred items (*), participants were asked to consider 'Casual interaction with acquaintances, rather than special relationships such as with close friends and family members.'

1. I like being around other people.
2. I find it hard to get my words out smoothly.
3. I am comfortable with unexpected changes in plans.
4. It's hard for me to avoid getting sidetracked in conversation.
5. I would rather talk to people to get information than to socialize.
6. People have to talk me into trying something new.
7. I am 'in-tune' with the other person during conversation.*
8. I have to warm myself up to the idea of visiting an unfamiliar place.
9. I enjoy being in social situations.
10. My voice has a flat or monotone sound to it.
11. I feel disconnected or 'out of sync' in conversations with others.*
12. People find it easy to approach me.*
13. I feel a strong need for sameness from day to day.
14. People ask me to repeat things I've said because they don't understand.
15. I am flexible about how things should be done.
16. I look forward to situations where I can meet new people.
17. I have been told that I talk too much about certain topics.
18. When I make conversation it is just to be polite.*

19. I look forward to trying new things.
20. I speak too loudly or softly.
21. I can tell when someone is not interested in what I am saying.⋆
22. I have a hard time dealing with changes in my routine.
23. I am good at making small talk.⋆
24. I act very set in my ways.
25. I feel like I am really connecting with other people.
26. People get frustrated by my unwillingness to bend.
27. Conversation bores me.⋆
28. I am warm and friendly in my interactions with others.⋆
29. I leave long pauses in conversation.
30. I alter my daily routine by trying something different.
31. I prefer to be alone rather than with others.
32. I lose track of my original point when talking to people.
33. I like to closely follow a routine while working.
34. I can tell when it is time to change topics in conversation.⋆
35. I keep doing things the way I know, even if another way might be better.
36. I enjoy chatting with people.⋆

Appendix C

Stimuli

Table C.1: The subjects and predicates used in the syllogism task, based on the selection used by Stenning and Yule (1997).

A	*B*	*C*
archaeologists	Albanians	archers
architects	Angolans	atheists
artists	Armenians	beer drinkers
bakers	Australians	boxers
builders	Austrians	calligraphers
bus drivers	Belgians	cat owners
butchers	Bolivians	chess players
chemists	Brazilians	Christians
clerks	Cambodians	coin collectors
cooks	Canadians	cricketers
counsellors	Colombians	cyclists
dentists	Croatians	dog owners
doctors	Czechs	footballers
grocers	Ecuadorians	golfers
joiners	Estonians	gourmets
judges	Germans	gymnasts

(Continued on next page)

Table C.1—continued from previous page

A	*B*	*C*
lawyers	Greeks	hang gliders
lecturers	Hungarian	high jumpers
managers	Hungarians	hill walkers
miners	Indians	horse riders
musicians	Indonesians	ice skaters
nurses	Italians	liberals
physicists	Koreans	poker players
plumbers	Latvians	rock climbers
politicians	Mauritians	rowers
sailors	Mexicans	shot putters
stockbrokers	Mongolians	skiers
students	Nigerians	squash players
teachers	Poles	swimmers
typists	Russians	tennis players
vets	Russians	vegetarians
waiters	Swedes	woodworkers

Appendix D

Enumerating first-order models

A program was developed to explore the space of first-order models for syllogisms, for instance how many models of a conclusion are true in models of the premises. To enumerate the models, they are viewed as rows in a truth table with $3n$ propositional variables, where n is the number of individuals (see Table D.1). A row in the table specifies a particular model by listing for which individuals, i_j, each predicate is true. For the first row, all the predicates are false for all individuals. For the last row, all are true for all individuals. From this arrangement it can be see that there are 2^{3n} models with n individuals since a truth table with c columns has 2^c rows.

Counting models consists then of restricting the table to appropriate subsets of models. For instance to find all models of *all A are B* and *all B are C* means restricting to rows where if $A(i_j)$ is true, then $B(i_j)$ is true, and where if $B(i_j)$ is true, then $C(i_j)$ is true. Table D.2 shows how in how many models these premises are true, for models with up to five individuals.

individuals	1	2	3	4	5
models	4	16	64	256	1024

Table D.2: Model counts for the premises *all A are B*, *all B are C*

Counter models of the conclusion in models of some set of premises may also be examined. Consider the conjecture *some A are not B, all B are C* $\models$ *all C are A*. Table D.3 shows the counts. As can be seen, for models with two individuals, countermodels of the

$A(i_1)$	$B(i_1)$	$C(i_1)$	$A(i_2)$	$B(i_2)$	$C(i_2)$	...	$A(i_n)$	$B(i_n)$	$C(i_n)$
0	0	0	0	0	0	...	0	0	0
0	0	0	0	0	0	...	0	0	1
0	0	0	0	0	0	...	0	1	0
0	0	0	0	0	0	...	0	1	1
0	0	0	0	0	0	...	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮
1	1	1	1	1	1	...	1	1	1

Table D.1: Truth table to represent syllogistic models

conclusion are in the minority (40%), whereas for models with five individuals the majority of models (85%) are countermodels of the conclusion.

individuals	1	2	3	4	5
models of premises	2	20	152	1040	6752
countermodels of conclusion	0	8	96	800	5760

Table D.3: Model counts for the conjecture *some A are not B, all B are C* $\models$ *all C are A*

Table D.5 (pp. 267 ff.) shows the counts of models and countermodels for all syllogistic conjectures, with up to five individuals.

Closed forms were found for functions generating these number sequences. (Since no proofs are provided that the expressions are correct, consider them only conjectures.) Some were straightforward; others were not so. The approach taken was to determine if unknown sequences could be defined by arithmetic in terms of the easily found sequences. The *On-Line Encyclopedia of Integer Sequences* (Sloane, 2007) was also consulted and two relevant sequences found: A051588 ($8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$), the number of $3 \times n$ binary matrices such that any two rows have a common 1, described originally by Jovović and Kilibarda (1999); and twice A016103, described only as 'Expansion of $1/(1-4x)(1-5x)(1-6x)$' with no closed form. Other functions I'd found, and the sequences close to A016103, hinted that the answer would be a function of the form

$$\alpha_1 \cdot \beta_1^n + \alpha_2 \cdot \beta_2^n + \alpha_3 \cdot \beta_3^n + \alpha_4 \cdot \beta_4^n$$

A search for the likely α's and β's eventually found the sequence $4^n+6^n-2\cdot5^n$ (dividing this by two gives a sequence matching the terms in the encyclopedia for A016103). Table D.4 gives counts of models for the syllogistic premises. Table D.6 (pp. 284 ff.) gives counts of models of the premises which are countermodels for each conclusion.

Has valid conclusion?	*Quantifiers*	*Models*
Yes	Both universals	2^{2n}
Yes	One universal, one existential	$6^n - 5^n$
No	Both universal	5^n
No	One universal, one existential	$6^n - 4^n$
No	Both existential	$8^n + 5^n - 2 \cdot 6^n$

Table D.4: Counts of models of the premises as a function of the number of individuals in the models. Existential quantifiers are *some* and *some…not*; universal quantifiers are *all* and *no*.

D.1 Haskell code for enumeration

Now the Haskell (Peyton Jones et al., 1999) code for enumerating the models will be explained. The following imports are required:

```
import Prelude hiding (showList)
import List hiding (union)
import IO
```

At the top level, formula are represented:

```
data Quant = All
           | Some
           | SomeNot
           | No
    deriving Eq

data Prop = Prop Quant String String
    deriving Eq
```

With *show* methods for ease of display:

```
instance Show Prop where
    show (Prop All      v1 v2) = "All " ++ v1 ++ " are " ++ v2
    show (Prop Some     v1 v2) = "Some " ++ v1 ++ " are " ++ v2
    show (Prop SomeNot  v1 v2) = "Some " ++ v1 ++ " are not " ++ v2
    show (Prop No       v1 v2) = "No " ++ v1 ++ " are " ++ v2
```

We want to generate a table of model and countermodel counts for all conjectures of syllogistic form, so let's enumerate all the syllogisms. First the premises:

```
quants = [All, Some, No, SomeNot]
prems = [ (Prop q1 v1a v1b, Prop q2 v2a v2b ) | q1 ← quants
                                              , q2 ← quants
                                              , v1a ← ["A", "B"]
                                              , v1b ← ["A", "B"]
                                              , v2a ← ["B", "C"]
                                              , v2b ← ["B", "C"]
                                              , v1a /= v1b
                                              , v2a /= v2b ]
```

Then the conclusions:

```
concs = [ Prop q1 v1a v1b | q1 ← quants
                          , v1a ← ["A", "C"]
                          , v1b ← ["C", "A"]
                          , v1a /= v1b ]
```

And finally the conjectures:

```
sylls = [(p1,p2,c) | ((p1,p2),c) ← apa]
   where apa = [((p1,p2),c) | (p1,p2) ← prems, c ← concs ]
```

Some more types for representing models, individuals, and classes.

```
type Model = [(Int,[Class])]
type Ind = String
type Class = String
```

Note the choice for the representation of model here. Rather than use a binary matrix, originally I considered a model to be a list of individuals, each holding a list of classes to which that individual belonged. So for instance the list

```
[(1,["A","B"]), (2,["C"])]
```

represents a model with two individuals, the first of which is an A, a B, but not a C, and the second of which is not an A, not a B, but is a C.

We define some functions which test if a proposition is true in a model:

```
test :: Prop → Model → Bool
test (Prop All      c1 c2) model = testAll c1 c2 model
test (Prop Some     c1 c2) model = testSome c1 c2 model
test (Prop SomeNot  c1 c2) model = testSomeNot c1 c2 model
test (Prop No       c1 c2) model = testNo c1 c2 model
```

The implementations of each of the test functions is a direct translation of Tarski's semantics definition, where Haskell becomes the meta-language. First a simple function to test if an element is contained in a list:

```
isIn :: Eq a ⇒ a → [a] → Bool
isIn x []     = False
isIn x (y:ys) = x == y || x ‘isIn‘ ys
```

The actual definitions are now easily expressed:

```
testAll _  _ []             = True
testAll c1 c2 ((ind,cs):ms) = if c1 ‘isIn‘ cs
                                then if c2 ‘isIn‘ cs
                                         then testAll c1 c2 ms
                                         else False
                                else testAll c1 c2 ms

testSome _  _ []              = False
testSome c1 c2 ((ind,cs):ms)  = if c1 ‘isIn‘ cs
                                  then if c2 ‘isIn‘ cs
                                           then True
                                           else testSome c1 c2 ms
                                  else testSome c1 c2 ms

testSomeNot _  _ []              = False
testSomeNot c1 c2 ((ind,cs):ms)  = if c1 ‘isIn‘ cs
                                     then if not (c2 ‘isIn‘ cs)
                                              then True
                                              else testSomeNot c1 c2 ms
                                     else testSomeNot c1 c2 ms

testNo _  _ []             = True
testNo c1 c2 ((ind,cs):ms) = if c1 ‘isIn‘ cs
                               then if not (c2 ‘isIn‘ cs)
                                        then testNo c1 c2 ms
```

```
                                                  else False
                                        else testNo c1 c2 ms
```

Now finally we can test whether a syllogistic conjecture is true in a given model:

```
testSyll (p1,p2,c) models = concsModels == premsModels
    where
    premsModels = [ m | m ← models, test p1 m, test p2 m ]
    concsModels = [ m | m ← premsModels, test c m ]
```

The main work of generating models is done using binary arithmetic. The following function converts a natural number of type *Int* to a list of binary digits:

```
rbits :: Int → [Int]
rbits 0 = [0]
rbits 1 = [1]
rbits n = n ‘mod‘ 2 : rbits (n ‘div‘ 2)
```

Enumerating models is just counting:

```
bitsList :: Int → [[Int]]
bitsList n = [ (take (width - ((length ◦ rbits) i)) (repeat 0)) -- pad with zeros
             ++ (reverse ◦ rbits) i | i ← [0..n-1] ]
  where
  width = ceiling (log (fromIntegral n) / log 2.0)
```

The following functions manipulate the lists and convert to our representation of models:

```
bitsList2models :: [[Int]] → [[(Int, [String])]]
bitsList2models []       = []
bitsList2models (is:iss) = (bits2model is 1):(bitsList2models iss)

bits2model :: [Int] → Int → [(Int, [String])]
bits2model [] _  =  []
bits2model bs n  =  (n, fixbits (take 3 bs)):(bits2model (drop 3 bs) (n+1))

fixbits :: [Int] → [String]
fixbits []      = error "Too small"
fixbits [1]     = ["C"]
fixbits [0]     = []
fixbits [1,c]   = "B":fixbits [c]
fixbits [0,c]   = fixbits [c]
fixbits [1,b,c] = "A":fixbits [b,c]
fixbits [0,b,c] = fixbits [b,c]
fixbits x       = error "Too big"
```

All 2^{3n} models with n individuals are enumerated by

```
models n = bitsList2models$bitsList (2^(3*n))
```

Finally we can find models and counter models:

```
findModels (p1,p2) models = [ m | m ← models, test p1 m, test p2 m ]

findCountermodels (p1,p2,c) models = concsModels
    where
    premsModels = [ m | m ← models, test p1 m, test p2 m ]
    concsModels = [ m | m ← premsModels, not $ test c m ]
```

The following code outputs the table as a tab-delimited text file:

```
cmct = do
    h ← openFile "modelstable5.txt" WriteMode
    hPrintCT h (cmCountTable 5)

cmCountTable :: Int → [(Prop, Prop, Prop, [(Int,Int)])]
cmCountTable n = [ (p1, p2, c, ms p1 p2 c) | (p1,p2,c) ← sylls ]
  where
  ms :: Prop → Prop → Prop → [(Int,Int)]
  ms p1 p2 c = [(length $ findCountermodels (p1,p2,c) (models i)
                ,length $ findModels (p1,p2) (models i))
                | i ← [1..n]
               ]

hPrintCT :: Handle → [(Prop, Prop, Prop, [(Int,Int)])] → IO()
hPrintCT h []                     = hClose h
hPrintCT h ((p1,p2,c,counts):rows) = do
    hPutStr h (show p1)
    hPutStr h "\t"
    hPutStr h (show p2)
    hPutStr h "\t"
    hPutStr h (show c)
    hPutStr h "\t"
    hPutStrLn h (showDamnListT (unzipMe counts))
    hPrintCT h rows

unzipMe :: [(a,a)] → [a]
unzipMe []         = []
unzipMe ((x,y):xs) = x:y:unzipMe xs

-- Sometimes plumbing can be tedious.
showDamnList [] = ""
showDamnList [x] = show x
```

```
showDamnList (x:xs) = show x ++ ", " ++ showDamnList xs

showDamnListT [] = ""
showDamnListT [x] = show x
showDamnListT (x:xs) = show x ++ "\t" ++ showDamnListT xs
```

D.2 Tables

The following pages list tables for the model counts up to 5 individuals (Table D.5) followed by conjectured closed forms for the number of countermodels (Table D.6).

Table D.5: A table of counts of first-order models and counter models for syllogisms. Column CM_i denotes how many models of the premises of size i are countermodels of the conclusion; M_i gives how many models of the premises there are of size i. The rows are sorted by CM_5/M_5, so all the very classically correct conjectures are first and all the very classically wrong conjectures are last.

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
All A are B	All B are C	All A are C	0	4	0	16	0	64	0	256	0	1024
All B are A	All C are B	All C are A	0	4	0	16	0	64	0	256	0	1024
All A are B	No B are C	No A are C	0	4	0	16	0	64	0	256	0	1024
All A are B	No B are C	No C are A	0	4	0	16	0	64	0	256	0	1024
All A are B	No C are B	No A are C	0	4	0	16	0	64	0	256	0	1024
All A are B	No C are B	No C are A	0	4	0	16	0	64	0	256	0	1024
No A are B	All C are B	No A are C	0	4	0	16	0	64	0	256	0	1024
No A are B	All C are B	No C are A	0	4	0	16	0	64	0	256	0	1024
No B are A	All C are B	No A are C	0	4	0	16	0	64	0	256	0	1024
No B are A	All C are B	No C are A	0	4	0	16	0	64	0	256	0	1024
All B are A	Some B are C	Some A are C	0	1	0	11	0	91	0	671	0	4651
All B are A	Some B are C	Some C are A	0	1	0	11	0	91	0	671	0	4651
All B are A	Some C are B	Some A are C	0	1	0	11	0	91	0	671	0	4651
All B are A	Some C are B	Some C are A	0	1	0	11	0	91	0	671	0	4651
All A are B	Some C are not B	Some C are not A	0	1	0	11	0	91	0	671	0	4651
All B are A	Some B are not C	Some A are not C	0	1	0	11	0	91	0	671	0	4651
Some A are B	All B are C	Some A are C	0	1	0	11	0	91	0	671	0	4651
Some A are B	All B are C	Some C are A	0	1	0	11	0	91	0	671	0	4651
Some B are A	All B are C	Some A are C	0	1	0	11	0	91	0	671	0	4651
Some B are A	All B are C	Some C are A	0	1	0	11	0	91	0	671	0	4651
Some A are B	No B are C	Some A are not C	0	1	0	11	0	91	0	671	0	4651
Some A are B	No C are B	Some A are not C	0	1	0	11	0	91	0	671	0	4651
Some B are A	No B are C	Some A are not C	0	1	0	11	0	91	0	671	0	4651
Some B are A	No C are B	Some A are not C	0	1	0	11	0	91	0	671	0	4651
No A are B	Some B are C	Some C are not A	0	1	0	11	0	91	0	671	0	4651
No A are B	Some C are B	Some C are not A	0	1	0	11	0	91	0	671	0	4651
No B are A	Some B are C	Some C are not A	0	1	0	11	0	91	0	671	0	4651

(Continued on next page)

Table D.5—continued from previous page

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
No B are A	Some C are B	Some C are not A	0	1	0	11	0	91	0	671	0	4651
Some A are not B	All C are B	Some A are not C	0	1	0	11	0	91	0	671	0	4651
Some B are not A	All B are C	Some C are not A	0	1	0	11	0	91	0	671	0	4651
All A are B	All B are C	Some C are not A	2	4	4	16	8	64	16	256	32	1024
All B are A	All C are B	Some A are not C	2	4	4	16	8	64	16	256	32	1024
All B are A	All B are C	Some A are C	3	5	9	25	27	125	81	625	243	3125
All B are A	All B are C	Some C are A	3	5	9	25	27	125	81	625	243	3125
All B are A	No B are C	Some A are not C	3	5	9	25	27	125	81	625	243	3125
All B are A	No C are B	Some A are not C	3	5	9	25	27	125	81	625	243	3125
No A are B	All B are C	Some C are not A	3	5	9	25	27	125	81	625	243	3125
No B are A	All B are C	Some C are not A	3	5	9	25	27	125	81	625	243	3125
All A are B	Some B are C	Some C are not A	1	2	7	20	37	152	175	1040	781	6752
All A are B	Some C are B	Some C are not A	1	2	7	20	37	152	175	1040	781	6752
All B are A	Some C are not B	Some A are C	1	2	7	20	37	152	175	1040	781	6752
All B are A	Some C are not B	Some C are A	1	2	7	20	37	152	175	1040	781	6752
Some A are B	All C are B	Some A are not C	1	2	7	20	37	152	175	1040	781	6752
Some B are A	All C are B	Some A are not C	1	2	7	20	37	152	175	1040	781	6752
No A are B	Some C are not B	Some C are not A	1	2	7	20	37	152	175	1040	781	6752
No B are A	Some C are not B	Some C are not A	1	2	7	20	37	152	175	1040	781	6752
Some A are not B	All B are C	Some A are C	1	2	7	20	37	152	175	1040	781	6752
Some A are not B	All B are C	Some C are A	1	2	7	20	37	152	175	1040	781	6752
Some A are not B	No B are C	Some A are not C	1	2	7	20	37	152	175	1040	781	6752
Some A are not B	No C are B	Some A are not C	1	2	7	20	37	152	175	1040	781	6752
Some A are B	Some B are C	Some A are C	0	1	2	17	30	205	302	2129	2550	20341
Some A are B	Some B are C	Some C are A	0	1	2	17	30	205	302	2129	2550	20341
Some A are B	Some C are B	Some A are C	0	1	2	17	30	205	302	2129	2550	20341
Some A are B	Some C are B	Some C are A	0	1	2	17	30	205	302	2129	2550	20341
Some B are A	Some B are C	Some A are C	0	1	2	17	30	205	302	2129	2550	20341
Some B are A	Some B are C	Some C are A	0	1	2	17	30	205	302	2129	2550	20341
Some B are A	Some C are B	Some A are C	0	1	2	17	30	205	302	2129	2550	20341
Some B are A	Some C are B	Some C are A	0	1	2	17	30	205	302	2129	2550	20341
Some A are B	Some B are not C	Some A are not C	0	1	2	17	30	205	302	2129	2550	20341

(Continued on next page)

Table D.5—continued from previous page

Premise 1	Premise 2	Conclusion	Counts									
			CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
Some B are A	Some B are not C	Some A are not C	0	1	2	17	30	205	302	2129	2550	20341
Some B are not A	Some B are C	Some C are not A	0	1	2	17	30	205	302	2129	2550	20341
Some B are not A	Some C are B	Some C are not A	0	1	2	17	30	205	302	2129	2550	20341
Some A are not B	Some C are not B	Some A are C	0	1	2	17	30	205	302	2129	2550	20341
Some A are not B	Some C are not B	Some C are A	0	1	2	17	30	205	302	2129	2550	20341
Some A are B	Some C are not B	Some A are C	0	0	2	8	30	144	302	1760	2550	18240
Some A are B	Some C are not B	Some C are A	0	0	2	8	30	144	302	1760	2550	18240
Some B are A	Some C are not B	Some A are C	0	0	2	8	30	144	302	1760	2550	18240
Some B are A	Some C are not B	Some C are A	0	0	2	8	30	144	302	1760	2550	18240
Some A are not B	Some B are C	Some A are C	0	0	2	8	30	144	302	1760	2550	18240
Some A are not B	Some B are C	Some C are A	0	0	2	8	30	144	302	1760	2550	18240
Some A are not B	Some C are B	Some A are C	0	0	2	8	30	144	302	1760	2550	18240
Some A are not B	Some C are B	Some C are A	0	0	2	8	30	144	302	1760	2550	18240
Some A are not B	Some B are not C	Some A are not C	0	0	2	8	30	144	302	1760	2550	18240
Some B are not A	Some C are not B	Some C are not A	0	0	2	8	30	144	302	1760	2550	18240
All A are B	Some B are not C	Some C are not A	2	2	12	20	56	152	240	1040	992	6752
All B are A	Some C are not B	Some A are not C	2	2	12	20	56	152	240	1040	992	6752
Some A are not B	All B are C	Some C are not A	2	2	12	20	56	152	240	1040	992	6752
Some B are not A	All C are B	Some A are not C	2	2	12	20	56	152	240	1040	992	6752
All B are A	Some B are C	Some A are not C	1	1	7	11	37	91	175	671	781	4651
All B are A	Some C are B	Some A are not C	1	1	7	11	37	91	175	671	781	4651
All B are A	Some B are not C	Some A are C	1	1	7	11	37	91	175	671	781	4651
All B are A	Some B are not C	Some C are A	1	1	7	11	37	91	175	671	781	4651
Some A are B	All B are C	Some C are not A	1	1	7	11	37	91	175	671	781	4651
Some B are A	All B are C	Some C are not A	1	1	7	11	37	91	175	671	781	4651
No A are B	Some B are not C	Some C are not A	1	1	7	11	37	91	175	671	781	4651
No B are A	Some B are not C	Some C are not A	1	1	7	11	37	91	175	671	781	4651
Some B are not A	All B are C	Some A are C	1	1	7	11	37	91	175	671	781	4651
Some B are not A	All B are C	Some C are A	1	1	7	11	37	91	175	671	781	4651
Some B are not A	No B are C	Some A are not C	1	1	7	11	37	91	175	671	781	4651
Some B are not A	No C are B	Some A are not C	1	1	7	11	37	91	175	671	781	4651
Some A are B	Some C are not B	Some A are not C	0	0	4	8	54	144	496	1760	3870	18240

(Continued on next page)

Table D.5—continued from previous page

Premise 1	Premise 2	Conclusion	Counts									
			CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
Some A are B	Some C are not B	Some C are not A	0	0	4	8	54	144	496	1760	3870	18240
Some B are A	Some C are not B	Some A are not C	0	0	4	8	54	144	496	1760	3870	18240
Some B are A	Some C are not B	Some C are not A	0	0	4	8	54	144	496	1760	3870	18240
Some A are not B	Some B are C	Some A are not C	0	0	4	8	54	144	496	1760	3870	18240
Some A are not B	Some B are C	Some C are not A	0	0	4	8	54	144	496	1760	3870	18240
Some A are not B	Some C are B	Some A are not C	0	0	4	8	54	144	496	1760	3870	18240
Some A are not B	Some C are B	Some C are not A	0	0	4	8	54	144	496	1760	3870	18240
Some A are not B	Some B are not C	Some A are C	0	0	4	8	54	144	496	1760	3870	18240
Some A are not B	Some B are not C	Some C are A	0	0	4	8	54	144	496	1760	3870	18240
Some B are not A	Some C are not B	Some A are C	0	0	4	8	54	144	496	1760	3870	18240
Some B are not A	Some C are not B	Some C are A	0	0	4	8	54	144	496	1760	3870	18240
Some A are B	Some B are C	Some A are not C	1	1	11	17	91	205	671	2129	4651	20341
Some A are B	Some B are C	Some C are not A	1	1	11	17	91	205	671	2129	4651	20341
Some A are B	Some C are B	Some A are not C	1	1	11	17	91	205	671	2129	4651	20341
Some A are B	Some C are B	Some C are not A	1	1	11	17	91	205	671	2129	4651	20341
Some B are A	Some B are C	Some A are not C	1	1	11	17	91	205	671	2129	4651	20341
Some B are A	Some B are C	Some C are not A	1	1	11	17	91	205	671	2129	4651	20341
Some B are A	Some C are B	Some A are not C	1	1	11	17	91	205	671	2129	4651	20341
Some B are A	Some C are B	Some C are not A	1	1	11	17	91	205	671	2129	4651	20341
Some A are B	Some B are not C	Some A are C	1	1	11	17	91	205	671	2129	4651	20341
Some A are B	Some B are not C	Some C are A	1	1	11	17	91	205	671	2129	4651	20341
Some B are A	Some B are not C	Some A are C	1	1	11	17	91	205	671	2129	4651	20341
Some B are A	Some B are not C	Some C are A	1	1	11	17	91	205	671	2129	4651	20341
Some B are not A	Some B are C	Some A are C	1	1	11	17	91	205	671	2129	4651	20341
Some B are not A	Some B are C	Some C are A	1	1	11	17	91	205	671	2129	4651	20341
Some B are not A	Some C are B	Some A are C	1	1	11	17	91	205	671	2129	4651	20341
Some B are not A	Some C are B	Some C are A	1	1	11	17	91	205	671	2129	4651	20341
Some A are not B	Some C are not B	Some A are not C	1	1	11	17	91	205	671	2129	4651	20341
Some A are not B	Some C are not B	Some C are not A	1	1	11	17	91	205	671	2129	4651	20341
Some B are not A	Some B are not C	Some A are not C	1	1	11	17	91	205	671	2129	4651	20341
Some B are not A	Some B are not C	Some C are not A	1	1	11	17	91	205	671	2129	4651	20341
All A are B	All B are C	Some A are C	3	4	9	16	27	64	81	256	243	1024

(Continued on next page)

Table D.5—continued from previous page

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
All A are B	All B are C	Some C are A	3	4	9	16	27	64	81	256	243	1024
All B are A	All C are B	Some A are C	3	4	9	16	27	64	81	256	243	1024
All B are A	All C are B	Some C are A	3	4	9	16	27	64	81	256	243	1024
All A are B	No B are C	Some A are not C	3	4	9	16	27	64	81	256	243	1024
All A are B	No B are C	Some C are not A	3	4	9	16	27	64	81	256	243	1024
All A are B	No C are B	Some A are not C	3	4	9	16	27	64	81	256	243	1024
All A are B	No C are B	Some C are not A	3	4	9	16	27	64	81	256	243	1024
No A are B	All C are B	Some A are not C	3	4	9	16	27	64	81	256	243	1024
No A are B	All C are B	Some C are not A	3	4	9	16	27	64	81	256	243	1024
No B are A	All C are B	Some A are not C	3	4	9	16	27	64	81	256	243	1024
No B are A	All C are B	Some C are not A	3	4	9	16	27	64	81	256	243	1024
Some A are B	Some B are not C	Some C are not A	1	1	13	17	115	205	865	2129	5971	20341
Some B are A	Some B are not C	Some C are not A	1	1	13	17	115	205	865	2129	5971	20341
Some B are not A	Some B are C	Some A are not C	1	1	13	17	115	205	865	2129	5971	20341
Some B are not A	Some C are B	Some A are not C	1	1	13	17	115	205	865	2129	5971	20341
Some B are not A	Some B are not C	Some A are C	1	1	13	17	115	205	865	2129	5971	20341
Some B are not A	Some B are not C	Some C are A	1	1	13	17	115	205	865	2129	5971	20341
All A are B	Some B are C	Some A are C	1	2	9	20	61	152	369	1040	2101	6752
All A are B	Some B are C	Some C are A	1	2	9	20	61	152	369	1040	2101	6752
All A are B	Some C are B	Some A are C	1	2	9	20	61	152	369	1040	2101	6752
All A are B	Some C are B	Some C are A	1	2	9	20	61	152	369	1040	2101	6752
All A are B	Some B are not C	Some A are not C	1	2	9	20	61	152	369	1040	2101	6752
All B are A	Some C are not B	Some C are not A	1	2	9	20	61	152	369	1040	2101	6752
Some A are B	All C are B	Some A are C	1	2	9	20	61	152	369	1040	2101	6752
Some A are B	All C are B	Some C are A	1	2	9	20	61	152	369	1040	2101	6752
Some B are A	All C are B	Some A are C	1	2	9	20	61	152	369	1040	2101	6752
Some B are A	All C are B	Some C are A	1	2	9	20	61	152	369	1040	2101	6752
No A are B	Some C are not B	Some A are C	1	2	9	20	61	152	369	1040	2101	6752
No A are B	Some C are not B	Some C are A	1	2	9	20	61	152	369	1040	2101	6752
No B are A	Some C are not B	Some A are C	1	2	9	20	61	152	369	1040	2101	6752
No B are A	Some C are not B	Some C are A	1	2	9	20	61	152	369	1040	2101	6752
Some A are not B	All B are C	Some A are not C	1	2	9	20	61	152	369	1040	2101	6752

(Continued on next page)

Table D.5—continued from previous page

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
Some B are not A	All C are B	Some C are not A	1	2	9	20	61	152	369	1040	2101	6752
Some A are not B	No B are C	Some A are C	1	2	9	20	61	152	369	1040	2101	6752
Some A are not B	No B are C	Some C are A	1	2	9	20	61	152	369	1040	2101	6752
Some A are not B	No C are B	Some A are C	1	2	9	20	61	152	369	1040	2101	6752
Some A are not B	No C are B	Some C are A	1	2	9	20	61	152	369	1040	2101	6752
Some A are not B	Some B are not C	Some C are not A	0	0	8	8	96	144	800	1760	5760	18240
Some B are not A	Some C are not B	Some A are not C	0	0	8	8	96	144	800	1760	5760	18240
All A are B	All C are B	Some A are C	4	5	16	25	64	125	256	625	1024	3125
All A are B	All C are B	Some C are A	4	5	16	25	64	125	256	625	1024	3125
All A are B	All C are B	Some A are not C	4	5	16	25	64	125	256	625	1024	3125
All A are B	All C are B	Some C are not A	4	5	16	25	64	125	256	625	1024	3125
All B are A	All B are C	Some A are not C	4	5	16	25	64	125	256	625	1024	3125
All B are A	All B are C	Some C are not A	4	5	16	25	64	125	256	625	1024	3125
All B are A	No B are C	Some A are C	4	5	16	25	64	125	256	625	1024	3125
All B are A	No B are C	Some C are A	4	5	16	25	64	125	256	625	1024	3125
All B are A	No B are C	Some C are not A	4	5	16	25	64	125	256	625	1024	3125
All B are A	No C are B	Some A are C	4	5	16	25	64	125	256	625	1024	3125
All B are A	No C are B	Some C are A	4	5	16	25	64	125	256	625	1024	3125
All B are A	No C are B	Some C are not A	4	5	16	25	64	125	256	625	1024	3125
No A are B	All B are C	Some A are C	4	5	16	25	64	125	256	625	1024	3125
No A are B	All B are C	Some C are A	4	5	16	25	64	125	256	625	1024	3125
No A are B	All B are C	Some A are not C	4	5	16	25	64	125	256	625	1024	3125
No B are A	All B are C	Some A are C	4	5	16	25	64	125	256	625	1024	3125
No B are A	All B are C	Some C are A	4	5	16	25	64	125	256	625	1024	3125
No B are A	All B are C	Some A are not C	4	5	16	25	64	125	256	625	1024	3125
No A are B	No B are C	Some A are C	4	5	16	25	64	125	256	625	1024	3125
No A are B	No B are C	Some C are A	4	5	16	25	64	125	256	625	1024	3125
No A are B	No B are C	Some A are not C	4	5	16	25	64	125	256	625	1024	3125
No A are B	No B are C	Some C are not A	4	5	16	25	64	125	256	625	1024	3125
No A are B	No C are B	Some A are C	4	5	16	25	64	125	256	625	1024	3125
No A are B	No C are B	Some C are A	4	5	16	25	64	125	256	625	1024	3125
No A are B	No C are B	Some A are not C	4	5	16	25	64	125	256	625	1024	3125

(Continued on next page)

Table D.5—continued from previous page

Premise 1	Premise 2	Conclusion	Counts									
			CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
No A are B	No C are B	Some C are not A	4	5	16	25	64	125	256	625	1024	3125
No B are A	No B are C	Some A are C	4	5	16	25	64	125	256	625	1024	3125
No B are A	No B are C	Some C are A	4	5	16	25	64	125	256	625	1024	3125
No B are A	No B are C	Some A are not C	4	5	16	25	64	125	256	625	1024	3125
No B are A	No B are C	Some C are not A	4	5	16	25	64	125	256	625	1024	3125
No B are A	No C are B	Some A are C	4	5	16	25	64	125	256	625	1024	3125
No B are A	No C are B	Some C are A	4	5	16	25	64	125	256	625	1024	3125
No B are A	No C are B	Some A are not C	4	5	16	25	64	125	256	625	1024	3125
No B are A	No C are B	Some C are not A	4	5	16	25	64	125	256	625	1024	3125
All A are B	Some B are C	Some A are not C	2	2	16	20	98	152	544	1040	2882	6752
All A are B	Some C are B	Some A are not C	2	2	16	20	98	152	544	1040	2882	6752
All A are B	Some B are not C	Some A are C	2	2	16	20	98	152	544	1040	2882	6752
All A are B	Some B are not C	Some C are A	2	2	16	20	98	152	544	1040	2882	6752
Some A are B	All C are B	Some C are not A	2	2	16	20	98	152	544	1040	2882	6752
Some B are A	All C are B	Some C are not A	2	2	16	20	98	152	544	1040	2882	6752
No A are B	Some C are not B	Some A are not C	2	2	16	20	98	152	544	1040	2882	6752
No B are A	Some C are not B	Some A are not C	2	2	16	20	98	152	544	1040	2882	6752
Some B are not A	All C are B	Some A are C	2	2	16	20	98	152	544	1040	2882	6752
Some B are not A	All C are B	Some C are A	2	2	16	20	98	152	544	1040	2882	6752
Some A are not B	No B are C	Some C are not A	2	2	16	20	98	152	544	1040	2882	6752
Some A are not B	No C are B	Some C are not A	2	2	16	20	98	152	544	1040	2882	6752
All B are A	Some B are C	Some C are not A	1	1	9	11	61	91	369	671	2101	4651
All B are A	Some C are B	Some C are not A	1	1	9	11	61	91	369	671	2101	4651
All A are B	Some C are not B	Some A are C	1	1	9	11	61	91	369	671	2101	4651
All A are B	Some C are not B	Some C are A	1	1	9	11	61	91	369	671	2101	4651
All A are B	Some C are not B	Some A are not C	1	1	9	11	61	91	369	671	2101	4651
All B are A	Some B are not C	Some C are not A	1	1	9	11	61	91	369	671	2101	4651
Some A are B	All B are C	Some A are not C	1	1	9	11	61	91	369	671	2101	4651
Some B are A	All B are C	Some A are not C	1	1	9	11	61	91	369	671	2101	4651
Some A are B	No B are C	Some A are C	1	1	9	11	61	91	369	671	2101	4651
Some A are B	No B are C	Some C are A	1	1	9	11	61	91	369	671	2101	4651
Some A are B	No B are C	Some C are not A	1	1	9	11	61	91	369	671	2101	4651

(Continued on next page)

Table D.5—continued from previous page

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
Some A are B	No C are B	Some A are C	1	1	9	11	61	91	369	671	2101	4651
Some A are B	No C are B	Some C are A	1	1	9	11	61	91	369	671	2101	4651
Some A are B	No C are B	Some C are not A	1	1	9	11	61	91	369	671	2101	4651
Some B are A	No B are C	Some A are C	1	1	9	11	61	91	369	671	2101	4651
Some B are A	No B are C	Some C are A	1	1	9	11	61	91	369	671	2101	4651
Some B are A	No B are C	Some C are not A	1	1	9	11	61	91	369	671	2101	4651
Some B are A	No C are B	Some A are C	1	1	9	11	61	91	369	671	2101	4651
Some B are A	No C are B	Some C are A	1	1	9	11	61	91	369	671	2101	4651
Some B are A	No C are B	Some C are not A	1	1	9	11	61	91	369	671	2101	4651
No A are B	Some B are C	Some A are C	1	1	9	11	61	91	369	671	2101	4651
No A are B	Some B are C	Some C are A	1	1	9	11	61	91	369	671	2101	4651
No A are B	Some B are C	Some A are not C	1	1	9	11	61	91	369	671	2101	4651
No A are B	Some C are B	Some A are C	1	1	9	11	61	91	369	671	2101	4651
No A are B	Some C are B	Some C are A	1	1	9	11	61	91	369	671	2101	4651
No A are B	Some C are B	Some A are not C	1	1	9	11	61	91	369	671	2101	4651
No B are A	Some B are C	Some A are C	1	1	9	11	61	91	369	671	2101	4651
No B are A	Some B are C	Some C are A	1	1	9	11	61	91	369	671	2101	4651
No B are A	Some B are C	Some A are not C	1	1	9	11	61	91	369	671	2101	4651
No B are A	Some C are B	Some A are C	1	1	9	11	61	91	369	671	2101	4651
No B are A	Some C are B	Some C are A	1	1	9	11	61	91	369	671	2101	4651
No B are A	Some C are B	Some A are not C	1	1	9	11	61	91	369	671	2101	4651
No A are B	Some B are not C	Some A are C	1	1	9	11	61	91	369	671	2101	4651
No A are B	Some B are not C	Some C are A	1	1	9	11	61	91	369	671	2101	4651
No A are B	Some B are not C	Some A are not C	1	1	9	11	61	91	369	671	2101	4651
No B are A	Some B are not C	Some A are C	1	1	9	11	61	91	369	671	2101	4651
No B are A	Some B are not C	Some C are A	1	1	9	11	61	91	369	671	2101	4651
No B are A	Some B are not C	Some A are not C	1	1	9	11	61	91	369	671	2101	4651
Some A are not B	All C are B	Some A are C	1	1	9	11	61	91	369	671	2101	4651
Some A are not B	All C are B	Some C are A	1	1	9	11	61	91	369	671	2101	4651
Some A are not B	All C are B	Some C are not A	1	1	9	11	61	91	369	671	2101	4651
Some B are not A	All B are C	Some A are not C	1	1	9	11	61	91	369	671	2101	4651
Some B are not A	No B are C	Some A are C	1	1	9	11	61	91	369	671	2101	4651

(Continued on next page)

Table D.5—continued from previous page

Premise 1	Premise 2	Conclusion	Counts									
			CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
Some B are not A	No B are C	Some C are A	1	1	9	11	61	91	369	671	2101	4651
Some B are not A	No B are C	Some C are not A	1	1	9	11	61	91	369	671	2101	4651
Some B are not A	No C are B	Some A are C	1	1	9	11	61	91	369	671	2101	4651
Some B are not A	No C are B	Some C are A	1	1	9	11	61	91	369	671	2101	4651
Some B are not A	No C are B	Some C are not A	1	1	9	11	61	91	369	671	2101	4651
All B are A	Some B are C	All C are A	0	1	2	11	30	91	302	671	2550	4651
All B are A	Some C are B	All C are A	0	1	2	11	30	91	302	671	2550	4651
All A are B	Some C are not B	All A are C	0	1	2	11	30	91	302	671	2550	4651
All A are B	Some C are not B	No A are C	0	1	2	11	30	91	302	671	2550	4651
All A are B	Some C are not B	No C are A	0	1	2	11	30	91	302	671	2550	4651
All B are A	Some B are not C	All C are A	0	1	2	11	30	91	302	671	2550	4651
Some A are B	All B are C	All A are C	0	1	2	11	30	91	302	671	2550	4651
Some B are A	All B are C	All A are C	0	1	2	11	30	91	302	671	2550	4651
Some A are B	No B are C	All C are A	0	1	2	11	30	91	302	671	2550	4651
Some A are B	No B are C	No A are C	0	1	2	11	30	91	302	671	2550	4651
Some A are B	No B are C	No C are A	0	1	2	11	30	91	302	671	2550	4651
Some A are B	No C are B	All C are A	0	1	2	11	30	91	302	671	2550	4651
Some A are B	No C are B	No A are C	0	1	2	11	30	91	302	671	2550	4651
Some A are B	No C are B	No C are A	0	1	2	11	30	91	302	671	2550	4651
Some B are A	No B are C	All C are A	0	1	2	11	30	91	302	671	2550	4651
Some B are A	No B are C	No A are C	0	1	2	11	30	91	302	671	2550	4651
Some B are A	No B are C	No C are A	0	1	2	11	30	91	302	671	2550	4651
Some B are A	No C are B	All C are A	0	1	2	11	30	91	302	671	2550	4651
Some B are A	No C are B	No A are C	0	1	2	11	30	91	302	671	2550	4651
Some B are A	No C are B	No C are A	0	1	2	11	30	91	302	671	2550	4651
No A are B	Some B are C	All A are C	0	1	2	11	30	91	302	671	2550	4651
No A are B	Some B are C	No A are C	0	1	2	11	30	91	302	671	2550	4651
No A are B	Some B are C	No C are A	0	1	2	11	30	91	302	671	2550	4651
No A are B	Some C are B	All A are C	0	1	2	11	30	91	302	671	2550	4651
No A are B	Some C are B	No A are C	0	1	2	11	30	91	302	671	2550	4651
No A are B	Some C are B	No C are A	0	1	2	11	30	91	302	671	2550	4651
No B are A	Some B are C	All A are C	0	1	2	11	30	91	302	671	2550	4651

(Continued on next page)

Table D.5—continued from previous page

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
No B are A	Some B are C	No A are C	0	1	2	11	30	91	302	671	2550	4651
No B are A	Some B are C	No C are A	0	1	2	11	30	91	302	671	2550	4651
No B are A	Some C are B	All A are C	0	1	2	11	30	91	302	671	2550	4651
No B are A	Some C are B	No A are C	0	1	2	11	30	91	302	671	2550	4651
No B are A	Some C are B	No C are A	0	1	2	11	30	91	302	671	2550	4651
No A are B	Some B are not C	All A are C	0	1	2	11	30	91	302	671	2550	4651
No A are B	Some B are not C	No A are C	0	1	2	11	30	91	302	671	2550	4651
No A are B	Some B are not C	No C are A	0	1	2	11	30	91	302	671	2550	4651
No B are A	Some B are not C	All A are C	0	1	2	11	30	91	302	671	2550	4651
No B are A	Some B are not C	No A are C	0	1	2	11	30	91	302	671	2550	4651
No B are A	Some B are not C	No C are A	0	1	2	11	30	91	302	671	2550	4651
Some A are not B	All C are B	All C are A	0	1	2	11	30	91	302	671	2550	4651
Some A are not B	All C are B	No A are C	0	1	2	11	30	91	302	671	2550	4651
Some A are not B	All C are B	No C are A	0	1	2	11	30	91	302	671	2550	4651
Some B are not A	All B are C	All A are C	0	1	2	11	30	91	302	671	2550	4651
Some B are not A	No B are C	All C are A	0	1	2	11	30	91	302	671	2550	4651
Some B are not A	No B are C	No A are C	0	1	2	11	30	91	302	671	2550	4651
Some B are not A	No B are C	No C are A	0	1	2	11	30	91	302	671	2550	4651
Some B are not A	No C are B	All C are A	0	1	2	11	30	91	302	671	2550	4651
Some B are not A	No C are B	No A are C	0	1	2	11	30	91	302	671	2550	4651
Some B are not A	No C are B	No C are A	0	1	2	11	30	91	302	671	2550	4651
All A are B	Some B are C	All A are C	0	2	4	20	54	152	496	1040	3870	6752
All A are B	Some C are B	All A are C	0	2	4	20	54	152	496	1040	3870	6752
All A are B	Some B are not C	No A are C	0	2	4	20	54	152	496	1040	3870	6752
All A are B	Some B are not C	No C are A	0	2	4	20	54	152	496	1040	3870	6752
Some A are B	All C are B	All C are A	0	2	4	20	54	152	496	1040	3870	6752
Some B are A	All C are B	All C are A	0	2	4	20	54	152	496	1040	3870	6752
No A are B	Some C are not B	All A are C	0	2	4	20	54	152	496	1040	3870	6752
No B are A	Some C are not B	All A are C	0	2	4	20	54	152	496	1040	3870	6752
Some B are not A	All C are B	No A are C	0	2	4	20	54	152	496	1040	3870	6752
Some B are not A	All C are B	No C are A	0	2	4	20	54	152	496	1040	3870	6752
Some A are not B	No B are C	All C are A	0	2	4	20	54	152	496	1040	3870	6752

(Continued on next page)

Table D.5—continued from previous page

Premise 1	Premise 2	Conclusion	Counts									
			CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
Some A are not B	No C are B	All C are A	0	2	4	20	54	152	496	1040	3870	6752
All A are B	All C are B	All A are C	1	5	9	25	61	125	369	625	2101	3125
All A are B	All C are B	All C are A	1	5	9	25	61	125	369	625	2101	3125
All A are B	All C are B	No A are C	1	5	9	25	61	125	369	625	2101	3125
All A are B	All C are B	No C are A	1	5	9	25	61	125	369	625	2101	3125
All B are A	All B are C	All A are C	1	5	9	25	61	125	369	625	2101	3125
All B are A	All B are C	All C are A	1	5	9	25	61	125	369	625	2101	3125
All B are A	No B are C	All C are A	1	5	9	25	61	125	369	625	2101	3125
All B are A	No B are C	No A are C	1	5	9	25	61	125	369	625	2101	3125
All B are A	No B are C	No C are A	1	5	9	25	61	125	369	625	2101	3125
All B are A	No C are B	All C are A	1	5	9	25	61	125	369	625	2101	3125
All B are A	No C are B	No A are C	1	5	9	25	61	125	369	625	2101	3125
All B are A	No C are B	No C are A	1	5	9	25	61	125	369	625	2101	3125
No A are B	All B are C	All A are C	1	5	9	25	61	125	369	625	2101	3125
No A are B	All B are C	No A are C	1	5	9	25	61	125	369	625	2101	3125
No A are B	All B are C	No C are A	1	5	9	25	61	125	369	625	2101	3125
No B are A	All B are C	All A are C	1	5	9	25	61	125	369	625	2101	3125
No B are A	All B are C	No A are C	1	5	9	25	61	125	369	625	2101	3125
No B are A	All B are C	No C are A	1	5	9	25	61	125	369	625	2101	3125
No A are B	No B are C	All A are C	1	5	9	25	61	125	369	625	2101	3125
No A are B	No B are C	All C are A	1	5	9	25	61	125	369	625	2101	3125
No A are B	No B are C	No A are C	1	5	9	25	61	125	369	625	2101	3125
No A are B	No B are C	No C are A	1	5	9	25	61	125	369	625	2101	3125
No A are B	No C are B	All A are C	1	5	9	25	61	125	369	625	2101	3125
No A are B	No C are B	All C are A	1	5	9	25	61	125	369	625	2101	3125
No A are B	No C are B	No A are C	1	5	9	25	61	125	369	625	2101	3125
No A are B	No C are B	No C are A	1	5	9	25	61	125	369	625	2101	3125
No B are A	No B are C	All A are C	1	5	9	25	61	125	369	625	2101	3125
No B are A	No B are C	All C are A	1	5	9	25	61	125	369	625	2101	3125
No B are A	No B are C	No A are C	1	5	9	25	61	125	369	625	2101	3125
No B are A	No B are C	No C are A	1	5	9	25	61	125	369	625	2101	3125
No B are A	No C are B	All A are C	1	5	9	25	61	125	369	625	2101	3125

(Continued on next page)

Table D.5—continued from previous page

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
No B are A	No C are B	All C are A	1	5	9	25	61	125	369	625	2101	3125
No B are A	No C are B	No A are C	1	5	9	25	61	125	369	625	2101	3125
No B are A	No C are B	No C are A	1	5	9	25	61	125	369	625	2101	3125
Some A are not B	Some B are not C	All C are A	0	0	0	8	48	144	960	1760	12480	18240
Some B are not A	Some C are not B	All A are C	0	0	0	8	48	144	960	1760	12480	18240
All A are B	Some B are C	No A are C	1	2	11	20	91	152	671	1040	4651	6752
All A are B	Some B are C	No C are A	1	2	11	20	91	152	671	1040	4651	6752
All A are B	Some C are B	No A are C	1	2	11	20	91	152	671	1040	4651	6752
All A are B	Some C are B	No C are A	1	2	11	20	91	152	671	1040	4651	6752
All A are B	Some B are not C	All A are C	1	2	11	20	91	152	671	1040	4651	6752
All B are A	Some C are not B	All C are A	1	2	11	20	91	152	671	1040	4651	6752
Some A are B	All C are B	No A are C	1	2	11	20	91	152	671	1040	4651	6752
Some A are B	All C are B	No C are A	1	2	11	20	91	152	671	1040	4651	6752
Some B are A	All C are B	No A are C	1	2	11	20	91	152	671	1040	4651	6752
Some B are A	All C are B	No C are A	1	2	11	20	91	152	671	1040	4651	6752
No A are B	Some C are not B	No A are C	1	2	11	20	91	152	671	1040	4651	6752
No A are B	Some C are not B	No C are A	1	2	11	20	91	152	671	1040	4651	6752
No B are A	Some C are not B	No A are C	1	2	11	20	91	152	671	1040	4651	6752
No B are A	Some C are not B	No C are A	1	2	11	20	91	152	671	1040	4651	6752
Some A are not B	All B are C	All A are C	1	2	11	20	91	152	671	1040	4651	6752
Some B are not A	All C are B	All C are A	1	2	11	20	91	152	671	1040	4651	6752
Some A are not B	No B are C	No A are C	1	2	11	20	91	152	671	1040	4651	6752
Some A are not B	No B are C	No C are A	1	2	11	20	91	152	671	1040	4651	6752
Some A are not B	No C are B	No A are C	1	2	11	20	91	152	671	1040	4651	6752
Some A are not B	No C are B	No C are A	1	2	11	20	91	152	671	1040	4651	6752
Some A are B	Some B are not C	All C are A	0	1	4	17	90	205	1264	2129	14370	20341
Some B are A	Some B are not C	All C are A	0	1	4	17	90	205	1264	2129	14370	20341
Some B are not A	Some B are C	All A are C	0	1	4	17	90	205	1264	2129	14370	20341
Some B are not A	Some C are B	All A are C	0	1	4	17	90	205	1264	2129	14370	20341
Some B are not A	Some B are not C	No A are C	0	1	4	17	90	205	1264	2129	14370	20341
Some B are not A	Some B are not C	No C are A	0	1	4	17	90	205	1264	2129	14370	20341
All A are B	All B are C	No A are C	1	4	7	16	37	64	175	256	781	1024

(Continued on next page)

Table D.5—continued from previous page

Premise 1	Premise 2	Conclusion	Counts									
			CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
All A are B	All B are C	No C are A	1	4	7	16	37	64	175	256	781	1024
All B are A	All C are B	No A are C	1	4	7	16	37	64	175	256	781	1024
All B are A	All C are B	No C are A	1	4	7	16	37	64	175	256	781	1024
All A are B	No B are C	All A are C	1	4	7	16	37	64	175	256	781	1024
All A are B	No B are C	All C are A	1	4	7	16	37	64	175	256	781	1024
All A are B	No C are B	All A are C	1	4	7	16	37	64	175	256	781	1024
All A are B	No C are B	All C are A	1	4	7	16	37	64	175	256	781	1024
No A are B	All C are B	All A are C	1	4	7	16	37	64	175	256	781	1024
No A are B	All C are B	All C are A	1	4	7	16	37	64	175	256	781	1024
No B are A	All C are B	All A are C	1	4	7	16	37	64	175	256	781	1024
No B are A	All C are B	All C are A	1	4	7	16	37	64	175	256	781	1024
Some A are B	Some B are C	All A are C	0	1	6	17	114	205	1458	2129	15690	20341
Some A are B	Some B are C	All C are A	0	1	6	17	114	205	1458	2129	15690	20341
Some A are B	Some C are B	All A are C	0	1	6	17	114	205	1458	2129	15690	20341
Some A are B	Some C are B	All C are A	0	1	6	17	114	205	1458	2129	15690	20341
Some B are A	Some B are C	All A are C	0	1	6	17	114	205	1458	2129	15690	20341
Some B are A	Some B are C	All C are A	0	1	6	17	114	205	1458	2129	15690	20341
Some B are A	Some C are B	All A are C	0	1	6	17	114	205	1458	2129	15690	20341
Some B are A	Some C are B	All C are A	0	1	6	17	114	205	1458	2129	15690	20341
Some A are B	Some B are not C	No A are C	0	1	6	17	114	205	1458	2129	15690	20341
Some A are B	Some B are not C	No C are A	0	1	6	17	114	205	1458	2129	15690	20341
Some B are A	Some B are not C	No A are C	0	1	6	17	114	205	1458	2129	15690	20341
Some B are A	Some B are not C	No C are A	0	1	6	17	114	205	1458	2129	15690	20341
Some B are not A	Some B are C	No A are C	0	1	6	17	114	205	1458	2129	15690	20341
Some B are not A	Some B are C	No C are A	0	1	6	17	114	205	1458	2129	15690	20341
Some B are not A	Some C are B	No A are C	0	1	6	17	114	205	1458	2129	15690	20341
Some B are not A	Some C are B	No C are A	0	1	6	17	114	205	1458	2129	15690	20341
Some A are not B	Some C are not B	All A are C	0	1	6	17	114	205	1458	2129	15690	20341
Some A are not B	Some C are not B	All C are A	0	1	6	17	114	205	1458	2129	15690	20341
Some B are not A	Some B are not C	All A are C	0	1	6	17	114	205	1458	2129	15690	20341
Some B are not A	Some B are not C	All C are A	0	1	6	17	114	205	1458	2129	15690	20341
Some A are B	Some C are not B	All A are C	0	0	4	8	90	144	1264	1760	14370	18240

(Continued on next page)

Table D.5—continued from previous page

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
Some A are B	Some C are not B	All C are A	0	0	4	8	90	144	1264	1760	14370	18240
Some B are A	Some C are not B	All A are C	0	0	4	8	90	144	1264	1760	14370	18240
Some B are A	Some C are not B	All C are A	0	0	4	8	90	144	1264	1760	14370	18240
Some A are not B	Some B are C	All A are C	0	0	4	8	90	144	1264	1760	14370	18240
Some A are not B	Some B are C	All C are A	0	0	4	8	90	144	1264	1760	14370	18240
Some A are not B	Some C are B	All A are C	0	0	4	8	90	144	1264	1760	14370	18240
Some A are not B	Some C are B	All C are A	0	0	4	8	90	144	1264	1760	14370	18240
Some A are not B	Some B are not C	No A are C	0	0	4	8	90	144	1264	1760	14370	18240
Some A are not B	Some B are not C	No C are A	0	0	4	8	90	144	1264	1760	14370	18240
Some B are not A	Some C are not B	No A are C	0	0	4	8	90	144	1264	1760	14370	18240
Some B are not A	Some C are not B	No C are A	0	0	4	8	90	144	1264	1760	14370	18240
All B are A	Some B are C	All A are C	0	1	4	11	54	91	496	671	3870	4651
All B are A	Some C are B	All A are C	0	1	4	11	54	91	496	671	3870	4651
All B are A	Some B are not C	No A are C	0	1	4	11	54	91	496	671	3870	4651
All B are A	Some B are not C	No C are A	0	1	4	11	54	91	496	671	3870	4651
Some A are B	All B are C	All C are A	0	1	4	11	54	91	496	671	3870	4651
Some B are A	All B are C	All C are A	0	1	4	11	54	91	496	671	3870	4651
No A are B	Some B are not C	All C are A	0	1	4	11	54	91	496	671	3870	4651
No B are A	Some B are not C	All C are A	0	1	4	11	54	91	496	671	3870	4651
Some B are not A	All B are C	No A are C	0	1	4	11	54	91	496	671	3870	4651
Some B are not A	All B are C	No C are A	0	1	4	11	54	91	496	671	3870	4651
Some B are not A	No B are C	All A are C	0	1	4	11	54	91	496	671	3870	4651
Some B are not A	No C are B	All A are C	0	1	4	11	54	91	496	671	3870	4651
All A are B	Some B are not C	All C are A	0	2	8	20	96	152	800	1040	5760	6752
All B are A	Some C are not B	All A are C	0	2	8	20	96	152	800	1040	5760	6752
Some A are not B	All B are C	All C are A	0	2	8	20	96	152	800	1040	5760	6752
Some B are not A	All C are B	All A are C	0	2	8	20	96	152	800	1040	5760	6752
Some A are B	Some C are not B	No A are C	0	0	6	8	114	144	1458	1760	15690	18240
Some A are B	Some C are not B	No C are A	0	0	6	8	114	144	1458	1760	15690	18240
Some B are A	Some C are not B	No A are C	0	0	6	8	114	144	1458	1760	15690	18240
Some B are A	Some C are not B	No C are A	0	0	6	8	114	144	1458	1760	15690	18240
Some A are not B	Some B are C	No A are C	0	0	6	8	114	144	1458	1760	15690	18240

(Continued on next page)

Table D.5—continued from previous page

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
Some A are not B	Some B are C	No C are A	0	0	6	8	114	144	1458	1760	15690	18240
Some A are not B	Some C are B	No A are C	0	0	6	8	114	144	1458	1760	15690	18240
Some A are not B	Some C are B	No C are A	0	0	6	8	114	144	1458	1760	15690	18240
Some A are not B	Some B are not C	All A are C	0	0	6	8	114	144	1458	1760	15690	18240
Some B are not A	Some C are not B	All C are A	0	0	6	8	114	144	1458	1760	15690	18240
Some A are B	Some B are C	No A are C	1	1	15	17	175	205	1827	2129	17791	20341
Some A are B	Some B are C	No C are A	1	1	15	17	175	205	1827	2129	17791	20341
Some A are B	Some C are B	No A are C	1	1	15	17	175	205	1827	2129	17791	20341
Some A are B	Some C are B	No C are A	1	1	15	17	175	205	1827	2129	17791	20341
Some B are A	Some B are C	No A are C	1	1	15	17	175	205	1827	2129	17791	20341
Some B are A	Some B are C	No C are A	1	1	15	17	175	205	1827	2129	17791	20341
Some B are A	Some C are B	No A are C	1	1	15	17	175	205	1827	2129	17791	20341
Some B are A	Some C are B	No C are A	1	1	15	17	175	205	1827	2129	17791	20341
Some A are B	Some B are not C	All A are C	1	1	15	17	175	205	1827	2129	17791	20341
Some B are A	Some B are not C	All A are C	1	1	15	17	175	205	1827	2129	17791	20341
Some B are not A	Some B are C	All C are A	1	1	15	17	175	205	1827	2129	17791	20341
Some B are not A	Some C are B	All C are A	1	1	15	17	175	205	1827	2129	17791	20341
Some A are not B	Some C are not B	No A are C	1	1	15	17	175	205	1827	2129	17791	20341
Some A are not B	Some C are not B	No C are A	1	1	15	17	175	205	1827	2129	17791	20341
All A are B	Some B are C	All C are A	1	2	13	20	115	152	865	1040	5971	6752
All A are B	Some C are B	All C are A	1	2	13	20	115	152	865	1040	5971	6752
All B are A	Some C are not B	No A are C	1	2	13	20	115	152	865	1040	5971	6752
All B are A	Some C are not B	No C are A	1	2	13	20	115	152	865	1040	5971	6752
Some A are B	All C are B	All A are C	1	2	13	20	115	152	865	1040	5971	6752
Some B are A	All C are B	All A are C	1	2	13	20	115	152	865	1040	5971	6752
No A are B	Some C are not B	All C are A	1	2	13	20	115	152	865	1040	5971	6752
No B are A	Some C are not B	All C are A	1	2	13	20	115	152	865	1040	5971	6752
Some A are not B	All B are C	No A are C	1	2	13	20	115	152	865	1040	5971	6752
Some A are not B	All B are C	No C are A	1	2	13	20	115	152	865	1040	5971	6752
Some A are not B	No B are C	All A are C	1	2	13	20	115	152	865	1040	5971	6752
Some A are not B	No C are B	All A are C	1	2	13	20	115	152	865	1040	5971	6752
All B are A	All B are C	No A are C	2	5	16	25	98	125	544	625	2882	3125

(Continued on next page)

Table D.5—continued from previous page

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
All B are A	All B are C	No C are A	2	5	16	25	98	125	544	625	2882	3125
All B are A	No B are C	All A are C	2	5	16	25	98	125	544	625	2882	3125
All B are A	No C are B	All A are C	2	5	16	25	98	125	544	625	2882	3125
No A are B	All B are C	All C are A	2	5	16	25	98	125	544	625	2882	3125
No B are A	All B are C	All C are A	2	5	16	25	98	125	544	625	2882	3125
All A are B	All B are C	All C are A	2	4	12	16	56	64	240	256	992	1024
All B are A	All C are B	All A are C	2	4	12	16	56	64	240	256	992	1024
All A are B	All B are C	Some A are not C	4	4	16	16	64	64	256	256	1024	1024
All B are A	All C are B	Some C are not A	4	4	16	16	64	64	256	256	1024	1024
All A are B	No B are C	Some A are C	4	4	16	16	64	64	256	256	1024	1024
All A are B	No B are C	Some C are A	4	4	16	16	64	64	256	256	1024	1024
All A are B	No C are B	Some A are C	4	4	16	16	64	64	256	256	1024	1024
All A are B	No C are B	Some C are A	4	4	16	16	64	64	256	256	1024	1024
No A are B	All C are B	Some A are C	4	4	16	16	64	64	256	256	1024	1024
No A are B	All C are B	Some C are A	4	4	16	16	64	64	256	256	1024	1024
No B are A	All C are B	Some A are C	4	4	16	16	64	64	256	256	1024	1024
No B are A	All C are B	Some C are A	4	4	16	16	64	64	256	256	1024	1024
All B are A	Some B are C	No A are C	1	1	11	11	91	91	671	671	4651	4651
All B are A	Some B are C	No C are A	1	1	11	11	91	91	671	671	4651	4651
All B are A	Some C are B	No A are C	1	1	11	11	91	91	671	671	4651	4651
All B are A	Some C are B	No C are A	1	1	11	11	91	91	671	671	4651	4651
All A are B	Some C are not B	All C are A	1	1	11	11	91	91	671	671	4651	4651
All B are A	Some B are not C	All A are C	1	1	11	11	91	91	671	671	4651	4651
Some A are B	All B are C	No A are C	1	1	11	11	91	91	671	671	4651	4651
Some A are B	All B are C	No C are A	1	1	11	11	91	91	671	671	4651	4651
Some B are A	All B are C	No A are C	1	1	11	11	91	91	671	671	4651	4651
Some B are A	All B are C	No C are A	1	1	11	11	91	91	671	671	4651	4651
Some A are B	No B are C	All A are C	1	1	11	11	91	91	671	671	4651	4651
Some A are B	No C are B	All A are C	1	1	11	11	91	91	671	671	4651	4651
Some B are A	No B are C	All A are C	1	1	11	11	91	91	671	671	4651	4651
Some B are A	No C are B	All A are C	1	1	11	11	91	91	671	671	4651	4651
No A are B	Some B are C	All C are A	1	1	11	11	91	91	671	671	4651	4651

(Continued on next page)

Table D.5—continued from previous page

			Counts									
Premise 1	Premise 2	Conclusion	CM_1	M_1	CM_2	M_2	CM_3	M_3	CM_4	M_4	CM_5	M_5
No A are B	Some C are B	All C are A	1	1	11	11	91	91	671	671	4651	4651
No B are A	Some B are C	All C are A	1	1	11	11	91	91	671	671	4651	4651
No B are A	Some C are B	All C are A	1	1	11	11	91	91	671	671	4651	4651
Some A are not B	All C are B	All A are C	1	1	11	11	91	91	671	671	4651	4651
Some B are not A	All B are C	All C are A	1	1	11	11	91	91	671	671	4651	4651

Table D.6: Conjectured closed forms for the number of models of the premises which are countermodels of the conclusions (with n individuals). Conjectures for the premise model counts are also included.

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
All A are B	All B are C	All A are C	2^{2n}	0
All B are A	All C are B	All C are A	2^{2n}	0
All A are B	No B are C	No A are C	2^{2n}	0
All A are B	No B are C	No C are A	2^{2n}	0
All A are B	No C are B	No A are C	2^{2n}	0
All A are B	No C are B	No C are A	2^{2n}	0
No A are B	All C are B	No A are C	2^{2n}	0
No A are B	All C are B	No C are A	2^{2n}	0
No B are A	All C are B	No A are C	2^{2n}	0
No B are A	All C are B	No C are A	2^{2n}	0
All B are A	Some B are C	Some A are C	$6^n - 5^n$	0
All B are A	Some B are C	Some C are A	$6^n - 5^n$	0
All B are A	Some C are B	Some A are C	$6^n - 5^n$	0
All B are A	Some C are B	Some C are A	$6^n - 5^n$	0
All A are B	Some C are not B	Some C are not A	$6^n - 5^n$	0
All B are A	Some B are not C	Some A are not C	$6^n - 5^n$	0
Some A are B	All B are C	Some A are C	$6^n - 5^n$	0
Some A are B	All B are C	Some C are A	$6^n - 5^n$	0
Some B are A	All B are C	Some A are C	$6^n - 5^n$	0
Some B are A	All B are C	Some C are A	$6^n - 5^n$	0
Some A are B	No B are C	Some A are not C	$6^n - 5^n$	0
Some A are B	No C are B	Some A are not C	$6^n - 5^n$	0
Some B are A	No B are C	Some A are not C	$6^n - 5^n$	0
Some B are A	No C are B	Some A are not C	$6^n - 5^n$	0
No A are B	Some B are C	Some C are not A	$6^n - 5^n$	0
No A are B	Some C are B	Some C are not A	$6^n - 5^n$	0
No B are A	Some B are C	Some C are not A	$6^n - 5^n$	0
No B are A	Some C are B	Some C are not A	$6^n - 5^n$	0
Some A are not B	All C are B	Some A are not C	$6^n - 5^n$	0
Some B are not A	All B are C	Some C are not A	$6^n - 5^n$	0
All A are B	All B are C	Some C are not A	2^{2n}	2^n
All B are A	All C are B	Some A are not C	2^{2n}	2^n
All A are B	All B are C	Some A are C	2^{2n}	3^n
All A are B	All B are C	Some C are A	2^{2n}	3^n
All B are A	All C are B	Some A are C	2^{2n}	3^n
All B are A	All C are B	Some C are A	2^{2n}	3^n
All A are B	No B are C	Some A are not C	2^{2n}	3^n
All A are B	No B are C	Some C are not A	2^{2n}	3^n
All A are B	No C are B	Some A are not C	2^{2n}	3^n
All A are B	No C are B	Some C are not A	2^{2n}	3^n
No A are B	All C are B	Some A are not C	2^{2n}	3^n

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
No A are B	All C are B	Some C are not A	2^{2n}	3^n
No B are A	All C are B	Some A are not C	2^{2n}	3^n
No B are A	All C are B	Some C are not A	2^{2n}	3^n
All B are A	All B are C	Some A are C	5^n	3^n
All B are A	All B are C	Some C are A	5^n	3^n
All B are A	No B are C	Some A are not C	5^n	3^n
All B are A	No C are B	Some A are not C	5^n	3^n
No A are B	All B are C	Some C are not A	5^n	3^n
No B are A	All B are C	Some C are not A	5^n	3^n
All A are B	All B are C	No A are C	2^{2n}	$4^n - 3^n$
All A are B	All B are C	No C are A	2^{2n}	$4^n - 3^n$
All B are A	All C are B	No A are C	2^{2n}	$4^n - 3^n$
All B are A	All C are B	No C are A	2^{2n}	$4^n - 3^n$
All A are B	No B are C	All A are C	2^{2n}	$4^n - 3^n$
All A are B	No B are C	All C are A	2^{2n}	$4^n - 3^n$
All A are B	No C are B	All A are C	2^{2n}	$4^n - 3^n$
All A are B	No C are B	All C are A	2^{2n}	$4^n - 3^n$
No A are B	All C are B	All A are C	2^{2n}	$4^n - 3^n$
No A are B	All C are B	All C are A	2^{2n}	$4^n - 3^n$
No B are A	All C are B	All A are C	2^{2n}	$4^n - 3^n$
No B are A	All C are B	All C are A	2^{2n}	$4^n - 3^n$
All B are A	Some B are C	Some A are not C	$6^n - 5^n$	$4^n - 3^n$
All B are A	Some C are B	Some A are not C	$6^n - 5^n$	$4^n - 3^n$
All B are A	Some B are not C	Some A are C	$6^n - 5^n$	$4^n - 3^n$
All B are A	Some B are not C	Some C are A	$6^n - 5^n$	$4^n - 3^n$
Some A are B	All B are C	Some C are not A	$6^n - 5^n$	$4^n - 3^n$
Some B are A	All B are C	Some C are not A	$6^n - 5^n$	$4^n - 3^n$
No A are B	Some B are not C	Some C are not A	$6^n - 5^n$	$4^n - 3^n$
No B are A	Some B are not C	Some C are not A	$6^n - 5^n$	$4^n - 3^n$
Some B are not A	All B are C	Some A are C	$6^n - 5^n$	$4^n - 3^n$
Some B are not A	All B are C	Some C are A	$6^n - 5^n$	$4^n - 3^n$
Some B are not A	No B are C	Some A are not C	$6^n - 5^n$	$4^n - 3^n$
Some B are not A	No C are B	Some A are not C	$6^n - 5^n$	$4^n - 3^n$
All A are B	Some B are C	Some C are not A	$6^n - 4^n$	$4^n - 3^n$
All A are B	Some C are B	Some C are not A	$6^n - 4^n$	$4^n - 3^n$
All B are A	Some C are not B	Some A are C	$6^n - 4^n$	$4^n - 3^n$
All B are A	Some C are not B	Some C are A	$6^n - 4^n$	$4^n - 3^n$
Some A are B	All C are B	Some A are not C	$6^n - 4^n$	$4^n - 3^n$
Some B are A	All C are B	Some A are not C	$6^n - 4^n$	$4^n - 3^n$
No A are B	Some C are not B	Some C are not A	$6^n - 4^n$	$4^n - 3^n$
No B are A	Some C are not B	Some C are not A	$6^n - 4^n$	$4^n - 3^n$
Some A are not B	All B are C	Some A are C	$6^n - 4^n$	$4^n - 3^n$
Some A are not B	All B are C	Some C are A	$6^n - 4^n$	$4^n - 3^n$
Some A are not B	No B are C	Some A are not C	$6^n - 4^n$	$4^n - 3^n$
Some A are not B	No C are B	Some A are not C	$6^n - 4^n$	$4^n - 3^n$

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
All A are B	All B are C	All C are A	2^{2n}	$4^n - 2^n$
All B are A	All C are B	All A are C	2^{2n}	$4^n - 2^n$
All A are B	Some B are not C	Some C are not A	$6^n - 4^n$	$4^n - 2^n$
All B are A	Some C are not B	Some A are not C	$6^n - 4^n$	$4^n - 2^n$
Some A are not B	All B are C	Some C are not A	$6^n - 4^n$	$4^n - 2^n$
Some B are not A	All C are B	Some A are not C	$6^n - 4^n$	$4^n - 2^n$
All A are B	All B are C	Some A are not C	2^{2n}	4^n
All B are A	All C are B	Some C are not A	2^{2n}	4^n
All A are B	No B are C	Some A are C	2^{2n}	4^n
All A are B	No B are C	Some C are A	2^{2n}	4^n
All A are B	No C are B	Some A are C	2^{2n}	4^n
All A are B	No C are B	Some C are A	2^{2n}	4^n
No A are B	All C are B	Some A are C	2^{2n}	4^n
No A are B	All C are B	Some C are A	2^{2n}	4^n
No B are A	All C are B	Some A are C	2^{2n}	4^n
No B are A	All C are B	Some C are A	2^{2n}	4^n
All A are B	All C are B	Some A are C	5^n	4^n
All A are B	All C are B	Some C are A	5^n	4^n
All A are B	All C are B	Some A are not C	5^n	4^n
All A are B	All C are B	Some C are not A	5^n	4^n
All B are A	All B are C	Some A are not C	5^n	4^n
All B are A	All B are C	Some C are not A	5^n	4^n
All B are A	No B are C	Some A are C	5^n	4^n
All B are A	No B are C	Some C are A	5^n	4^n
All B are A	No B are C	Some C are not A	5^n	4^n
All B are A	No C are B	Some A are C	5^n	4^n
All B are A	No C are B	Some C are A	5^n	4^n
All B are A	No C are B	Some C are not A	5^n	4^n
No A are B	All B are C	Some A are C	5^n	4^n
No A are B	All B are C	Some C are A	5^n	4^n
No A are B	All B are C	Some A are not C	5^n	4^n
No B are A	All B are C	Some A are C	5^n	4^n
No B are A	All B are C	Some C are A	5^n	4^n
No B are A	All B are C	Some A are not C	5^n	4^n
No A are B	No B are C	Some A are C	5^n	4^n
No A are B	No B are C	Some C are A	5^n	4^n
No A are B	No B are C	Some A are not C	5^n	4^n
No A are B	No B are C	Some C are not A	5^n	4^n
No A are B	No C are B	Some A are C	5^n	4^n
No A are B	No C are B	Some C are A	5^n	4^n
No A are B	No C are B	Some A are not C	5^n	4^n
No A are B	No C are B	Some C are not A	5^n	4^n
No B are A	No B are C	Some A are C	5^n	4^n
No B are A	No B are C	Some C are A	5^n	4^n
No B are A	No B are C	Some A are not C	5^n	4^n

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
No B are A	No B are C	Some C are not A	5^n	4^n
No B are A	No C are B	Some A are C	5^n	4^n
No B are A	No C are B	Some C are A	5^n	4^n
No B are A	No C are B	Some A are not C	5^n	4^n
No B are A	No C are B	Some C are not A	5^n	4^n
All A are B	All C are B	All A are C	5^n	$5^n - 4^n$
All A are B	All C are B	All C are A	5^n	$5^n - 4^n$
All A are B	All C are B	No A are C	5^n	$5^n - 4^n$
All A are B	All C are B	No C are A	5^n	$5^n - 4^n$
All B are A	All B are C	All A are C	5^n	$5^n - 4^n$
All B are A	All B are C	All C are A	5^n	$5^n - 4^n$
All B are A	No B are C	All C are A	5^n	$5^n - 4^n$
All B are A	No B are C	No A are C	5^n	$5^n - 4^n$
All B are A	No B are C	No C are A	5^n	$5^n - 4^n$
All B are A	No C are B	All C are A	5^n	$5^n - 4^n$
All B are A	No C are B	No A are C	5^n	$5^n - 4^n$
All B are A	No C are B	No C are A	5^n	$5^n - 4^n$
No A are B	All B are C	All A are C	5^n	$5^n - 4^n$
No A are B	All B are C	No A are C	5^n	$5^n - 4^n$
No A are B	All B are C	No C are A	5^n	$5^n - 4^n$
No B are A	All B are C	All A are C	5^n	$5^n - 4^n$
No B are A	All B are C	No A are C	5^n	$5^n - 4^n$
No B are A	All B are C	No C are A	5^n	$5^n - 4^n$
No A are B	No B are C	All A are C	5^n	$5^n - 4^n$
No A are B	No B are C	All C are A	5^n	$5^n - 4^n$
No A are B	No B are C	No A are C	5^n	$5^n - 4^n$
No A are B	No B are C	No C are A	5^n	$5^n - 4^n$
No A are B	No C are B	All A are C	5^n	$5^n - 4^n$
No A are B	No C are B	All C are A	5^n	$5^n - 4^n$
No A are B	No C are B	No A are C	5^n	$5^n - 4^n$
No A are B	No C are B	No C are A	5^n	$5^n - 4^n$
No B are A	No B are C	All A are C	5^n	$5^n - 4^n$
No B are A	No B are C	All C are A	5^n	$5^n - 4^n$
No B are A	No B are C	No A are C	5^n	$5^n - 4^n$
No B are A	No B are C	No C are A	5^n	$5^n - 4^n$
No B are A	No C are B	All A are C	5^n	$5^n - 4^n$
No B are A	No C are B	All C are A	5^n	$5^n - 4^n$
No B are A	No C are B	No A are C	5^n	$5^n - 4^n$
No B are A	No C are B	No C are A	5^n	$5^n - 4^n$
All B are A	Some B are C	Some C are not A	$6^n - 5^n$	$5^n - 4^n$
All B are A	Some C are B	Some C are not A	$6^n - 5^n$	$5^n - 4^n$
All A are B	Some C are not B	Some A are C	$6^n - 5^n$	$5^n - 4^n$
All A are B	Some C are not B	Some C are A	$6^n - 5^n$	$5^n - 4^n$
All A are B	Some C are not B	Some A are not C	$6^n - 5^n$	$5^n - 4^n$
All B are A	Some B are not C	Some C are not A	$6^n - 5^n$	$5^n - 4^n$

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
Some A are B	All B are C	Some A are not C	$6^n - 5^n$	$5^n - 4^n$
Some B are A	All B are C	Some A are not C	$6^n - 5^n$	$5^n - 4^n$
Some A are B	No B are C	Some A are C	$6^n - 5^n$	$5^n - 4^n$
Some A are B	No B are C	Some C are A	$6^n - 5^n$	$5^n - 4^n$
Some A are B	No B are C	Some C are not A	$6^n - 5^n$	$5^n - 4^n$
Some A are B	No C are B	Some A are C	$6^n - 5^n$	$5^n - 4^n$
Some A are B	No C are B	Some C are A	$6^n - 5^n$	$5^n - 4^n$
Some A are B	No C are B	Some C are not A	$6^n - 5^n$	$5^n - 4^n$
Some B are A	No B are C	Some A are C	$6^n - 5^n$	$5^n - 4^n$
Some B are A	No B are C	Some C are A	$6^n - 5^n$	$5^n - 4^n$
Some B are A	No B are C	Some C are not A	$6^n - 5^n$	$5^n - 4^n$
Some B are A	No C are B	Some A are C	$6^n - 5^n$	$5^n - 4^n$
Some B are A	No C are B	Some C are A	$6^n - 5^n$	$5^n - 4^n$
Some B are A	No C are B	Some C are not A	$6^n - 5^n$	$5^n - 4^n$
No A are B	Some B are C	Some A are C	$6^n - 5^n$	$5^n - 4^n$
No A are B	Some B are C	Some C are A	$6^n - 5^n$	$5^n - 4^n$
No A are B	Some B are C	Some A are not C	$6^n - 5^n$	$5^n - 4^n$
No A are B	Some C are B	Some A are C	$6^n - 5^n$	$5^n - 4^n$
No A are B	Some C are B	Some C are A	$6^n - 5^n$	$5^n - 4^n$
No A are B	Some C are B	Some A are not C	$6^n - 5^n$	$5^n - 4^n$
No B are A	Some B are C	Some A are C	$6^n - 5^n$	$5^n - 4^n$
No B are A	Some B are C	Some C are A	$6^n - 5^n$	$5^n - 4^n$
No B are A	Some B are C	Some A are not C	$6^n - 5^n$	$5^n - 4^n$
No B are A	Some C are B	Some A are C	$6^n - 5^n$	$5^n - 4^n$
No B are A	Some C are B	Some C are A	$6^n - 5^n$	$5^n - 4^n$
No B are A	Some C are B	Some A are not C	$6^n - 5^n$	$5^n - 4^n$
No A are B	Some B are not C	Some A are C	$6^n - 5^n$	$5^n - 4^n$
No A are B	Some B are not C	Some C are A	$6^n - 5^n$	$5^n - 4^n$
No A are B	Some B are not C	Some A are not C	$6^n - 5^n$	$5^n - 4^n$
No B are A	Some B are not C	Some A are C	$6^n - 5^n$	$5^n - 4^n$
No B are A	Some B are not C	Some C are A	$6^n - 5^n$	$5^n - 4^n$
No B are A	Some B are not C	Some A are not C	$6^n - 5^n$	$5^n - 4^n$
Some A are not B	All C are B	Some A are C	$6^n - 5^n$	$5^n - 4^n$
Some A are not B	All C are B	Some C are A	$6^n - 5^n$	$5^n - 4^n$
Some A are not B	All C are B	Some C are not A	$6^n - 5^n$	$5^n - 4^n$
Some B are not A	All B are C	Some A are not C	$6^n - 5^n$	$5^n - 4^n$
Some B are not A	No B are C	Some A are C	$6^n - 5^n$	$5^n - 4^n$
Some B are not A	No B are C	Some C are A	$6^n - 5^n$	$5^n - 4^n$
Some B are not A	No B are C	Some C are not A	$6^n - 5^n$	$5^n - 4^n$
Some B are not A	No C are B	Some A are C	$6^n - 5^n$	$5^n - 4^n$
Some B are not A	No C are B	Some C are A	$6^n - 5^n$	$5^n - 4^n$
Some B are not A	No C are B	Some C are not A	$6^n - 5^n$	$5^n - 4^n$
All A are B	Some B are C	Some A are C	$6^n - 4^n$	$5^n - 4^n$
All A are B	Some B are C	Some C are A	$6^n - 4^n$	$5^n - 4^n$
All A are B	Some C are B	Some A are C	$6^n - 4^n$	$5^n - 4^n$

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
All A are B	Some C are B	Some C are A	$6^n - 4^n$	$5^n - 4^n$
All A are B	Some B are not C	Some A are not C	$6^n - 4^n$	$5^n - 4^n$
All B are A	Some C are not B	Some C are not A	$6^n - 4^n$	$5^n - 4^n$
Some A are B	All C are B	Some A are C	$6^n - 4^n$	$5^n - 4^n$
Some A are B	All C are B	Some C are A	$6^n - 4^n$	$5^n - 4^n$
Some B are A	All C are B	Some A are C	$6^n - 4^n$	$5^n - 4^n$
Some B are A	All C are B	Some C are A	$6^n - 4^n$	$5^n - 4^n$
No A are B	Some C are not B	Some A are C	$6^n - 4^n$	$5^n - 4^n$
No A are B	Some C are not B	Some C are A	$6^n - 4^n$	$5^n - 4^n$
No B are A	Some C are not B	Some A are C	$6^n - 4^n$	$5^n - 4^n$
No B are A	Some C are not B	Some C are A	$6^n - 4^n$	$5^n - 4^n$
Some A are not B	All B are C	Some A are not C	$6^n - 4^n$	$5^n - 4^n$
Some B are not A	All C are B	Some C are not A	$6^n - 4^n$	$5^n - 4^n$
Some A are not B	No B are C	Some A are C	$6^n - 4^n$	$5^n - 4^n$
Some A are not B	No B are C	Some C are A	$6^n - 4^n$	$5^n - 4^n$
Some A are not B	No C are B	Some A are C	$6^n - 4^n$	$5^n - 4^n$
Some A are not B	No C are B	Some C are A	$6^n - 4^n$	$5^n - 4^n$
All B are A	Some B are C	All C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
All B are A	Some C are B	All C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
All A are B	Some C are not B	All A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
All A are B	Some C are not B	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
All A are B	Some C are not B	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
All B are A	Some B are not C	All C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	All B are C	All A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	All B are C	All A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	No B are C	All C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	No B are C	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	No B are C	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	No C are B	All C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	No C are B	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	No C are B	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	No B are C	All C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	No B are C	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	No B are C	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	No C are B	All C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	No C are B	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	No C are B	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No A are B	Some B are C	All A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No A are B	Some B are C	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No A are B	Some B are C	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No A are B	Some C are B	All A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No A are B	Some C are B	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No A are B	Some C are B	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No B are A	Some B are C	All A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No B are A	Some B are C	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
No B are A	Some B are C	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No B are A	Some C are B	All A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No B are A	Some C are B	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No B are A	Some C are B	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No A are B	Some B are not C	All A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No A are B	Some B are not C	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No A are B	Some B are not C	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No B are A	Some B are not C	All A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No B are A	Some B are not C	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
No B are A	Some B are not C	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are not B	All C are B	All C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are not B	All C are B	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are not B	All C are B	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are not A	All B are C	All A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are not A	No B are C	All C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are not A	No B are C	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are not A	No B are C	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are not A	No C are B	All C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are not A	No C are B	No A are C	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are not A	No C are B	No C are A	$6^n - 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	Some C are not B	Some A are C	$8^n - 2 \cdot 6^n + 4^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	Some C are not B	Some C are A	$8^n - 2 \cdot 6^n + 4^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	Some C are not B	Some A are C	$8^n - 2 \cdot 6^n + 4^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	Some C are not B	Some C are A	$8^n - 2 \cdot 6^n + 4^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are not B	Some B are C	Some A are C	$8^n - 2 \cdot 6^n + 4^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are not B	Some B are C	Some C are A	$8^n - 2 \cdot 6^n + 4^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are not B	Some C are B	Some A are C	$8^n - 2 \cdot 6^n + 4^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are not B	Some C are B	Some C are A	$8^n - 2 \cdot 6^n + 4^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are not B	Some B are not C	Some A are not C	$8^n - 2 \cdot 6^n + 4^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are not A	Some C are not B	Some C are not A	$8^n - 2 \cdot 6^n + 4^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	Some B are C	Some A are C	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	Some B are C	Some C are A	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	Some C are B	Some A are C	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	Some C are B	Some C are A	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	Some B are C	Some A are C	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	Some B are C	Some C are A	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	Some C are B	Some A are C	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	Some C are B	Some C are A	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are B	Some B are not C	Some A are not C	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are A	Some B are not C	Some A are not C	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are not A	Some B are C	Some C are not A	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some B are not A	Some C are B	Some C are not A	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are not B	Some C are not B	Some A are C	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
Some A are not B	Some C are not B	Some C are A	$8^n - 2 \cdot 6^n + 5^n$	$4^n + 6^n - 2 \cdot 5^n$
All B are A	All B are C	No A are C	5^n	$5^n - 3^n$

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
All B are A	All B are C	No C are A	5^n	$5^n - 3^n$
All B are A	No B are C	All A are C	5^n	$5^n - 3^n$
All B are A	No C are B	All A are C	5^n	$5^n - 3^n$
No A are B	All B are C	All C are A	5^n	$5^n - 3^n$
No B are A	All B are C	All C are A	5^n	$5^n - 3^n$
All A are B	Some B are C	Some A are not C	$6^n - 4^n$	$5^n - 3^n$
All A are B	Some C are B	Some A are not C	$6^n - 4^n$	$5^n - 3^n$
All A are B	Some B are not C	Some A are C	$6^n - 4^n$	$5^n - 3^n$
All A are B	Some B are not C	Some C are A	$6^n - 4^n$	$5^n - 3^n$
Some A are B	All C are B	Some C are not A	$6^n - 4^n$	$5^n - 3^n$
Some B are A	All C are B	Some C are not A	$6^n - 4^n$	$5^n - 3^n$
No A are B	Some C are not B	Some A are not C	$6^n - 4^n$	$5^n - 3^n$
No B are A	Some C are not B	Some A are not C	$6^n - 4^n$	$5^n - 3^n$
Some B are not A	All C are B	Some A are C	$6^n - 4^n$	$5^n - 3^n$
Some B are not A	All C are B	Some C are A	$6^n - 4^n$	$5^n - 3^n$
Some A are not B	No B are C	Some C are not A	$6^n - 4^n$	$5^n - 3^n$
Some A are not B	No C are B	Some C are not A	$6^n - 4^n$	$5^n - 3^n$
All B are A	Some B are C	All A are C	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
All B are A	Some C are B	All A are C	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
All B are A	Some B are not C	No A are C	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
All B are A	Some B are not C	No C are A	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
Some A are B	All B are C	All C are A	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
Some B are A	All B are C	All C are A	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
No A are B	Some B are not C	All C are A	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
No B are A	Some B are not C	All C are A	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
Some B are not A	All B are C	No A are C	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
Some B are not A	All B are C	No C are A	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
Some B are not A	No B are C	All A are C	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
Some B are not A	No C are B	All A are C	$6^n - 5^n$	$3^n + 6^n - 4^n - 5^n$
All A are B	Some B are C	All A are C	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
All A are B	Some C are B	All A are C	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
All A are B	Some B are not C	No A are C	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
All A are B	Some B are not C	No C are A	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
Some A are B	All C are B	All C are A	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
Some B are A	All C are B	All C are A	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
No A are B	Some C are not B	All A are C	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
No B are A	Some C are not B	All A are C	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
Some B are not A	All C are B	No A are C	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
Some B are not A	All C are B	No C are A	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
Some A are not B	No B are C	All C are A	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
Some A are not B	No C are B	All C are A	$6^n - 4^n$	$3^n + 6^n - 4^n - 5^n$
Some A are B	Some C are not B	Some A are not C	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
Some A are B	Some C are not B	Some C are not A	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
Some B are A	Some C are not B	Some A are not C	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
Some B are A	Some C are not B	Some C are not A	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
Some A are not B	Some B are C	Some A are not C	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
Some A are not B	Some B are C	Some C are not A	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
Some A are not B	Some C are B	Some A are not C	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
Some A are not B	Some C are B	Some C are not A	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
Some A are not B	Some B are not C	Some A are C	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
Some A are not B	Some B are not C	Some C are A	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
Some B are not A	Some C are not B	Some A are C	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
Some B are not A	Some C are not B	Some C are A	$8^n - 2 \cdot 6^n + 4^n$	$3^n + 6^n - 4^n - 5^n$
All B are A	Some B are C	No A are C	$6^n - 5^n$	$6^n - 5^n$
All B are A	Some B are C	No C are A	$6^n - 5^n$	$6^n - 5^n$
All B are A	Some C are B	No A are C	$6^n - 5^n$	$6^n - 5^n$
All B are A	Some C are B	No C are A	$6^n - 5^n$	$6^n - 5^n$
All A are B	Some C are not B	All C are A	$6^n - 5^n$	$6^n - 5^n$
All B are A	Some B are not C	All A are C	$6^n - 5^n$	$6^n - 5^n$
Some A are B	All B are C	No A are C	$6^n - 5^n$	$6^n - 5^n$
Some A are B	All B are C	No C are A	$6^n - 5^n$	$6^n - 5^n$
Some B are A	All B are C	No A are C	$6^n - 5^n$	$6^n - 5^n$
Some B are A	All B are C	No C are A	$6^n - 5^n$	$6^n - 5^n$
Some A are B	No B are C	All A are C	$6^n - 5^n$	$6^n - 5^n$
Some A are B	No C are B	All A are C	$6^n - 5^n$	$6^n - 5^n$
Some B are A	No B are C	All A are C	$6^n - 5^n$	$6^n - 5^n$
Some B are A	No C are B	All A are C	$6^n - 5^n$	$6^n - 5^n$
No A are B	Some B are C	All C are A	$6^n - 5^n$	$6^n - 5^n$
No A are B	Some C are B	All C are A	$6^n - 5^n$	$6^n - 5^n$
No B are A	Some B are C	All C are A	$6^n - 5^n$	$6^n - 5^n$
No B are A	Some C are B	All C are A	$6^n - 5^n$	$6^n - 5^n$
Some A are not B	All C are B	All A are C	$6^n - 5^n$	$6^n - 5^n$
Some B are not A	All B are C	All C are A	$6^n - 5^n$	$6^n - 5^n$
All A are B	Some B are C	No A are C	$6^n - 4^n$	$6^n - 5^n$
All A are B	Some B are C	No C are A	$6^n - 4^n$	$6^n - 5^n$
All A are B	Some C are B	No A are C	$6^n - 4^n$	$6^n - 5^n$
All A are B	Some C are B	No C are A	$6^n - 4^n$	$6^n - 5^n$
All A are B	Some B are not C	All A are C	$6^n - 4^n$	$6^n - 5^n$
All B are A	Some C are not B	All C are A	$6^n - 4^n$	$6^n - 5^n$
Some A are B	All C are B	No A are C	$6^n - 4^n$	$6^n - 5^n$
Some A are B	All C are B	No C are A	$6^n - 4^n$	$6^n - 5^n$
Some B are A	All C are B	No A are C	$6^n - 4^n$	$6^n - 5^n$
Some B are A	All C are B	No C are A	$6^n - 4^n$	$6^n - 5^n$
No A are B	Some C are not B	No A are C	$6^n - 4^n$	$6^n - 5^n$
No A are B	Some C are not B	No C are A	$6^n - 4^n$	$6^n - 5^n$
No B are A	Some C are not B	No A are C	$6^n - 4^n$	$6^n - 5^n$
No B are A	Some C are not B	No C are A	$6^n - 4^n$	$6^n - 5^n$
Some A are not B	All B are C	All A are C	$6^n - 4^n$	$6^n - 5^n$
Some B are not A	All C are B	All C are A	$6^n - 4^n$	$6^n - 5^n$
Some A are not B	No B are C	No A are C	$6^n - 4^n$	$6^n - 5^n$

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
Some A are not B	No B are C	No C are A	$6^n - 4^n$	$6^n - 5^n$
Some A are not B	No C are B	No A are C	$6^n - 4^n$	$6^n - 5^n$
Some A are not B	No C are B	No C are A	$6^n - 4^n$	$6^n - 5^n$
Some A are B	Some B are C	Some A are not C	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some A are B	Some B are C	Some C are not A	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some A are B	Some C are B	Some A are not C	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some A are B	Some C are B	Some C are not A	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are A	Some B are C	Some A are not C	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are A	Some B are C	Some C are not A	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are A	Some C are B	Some A are not C	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are A	Some C are B	Some C are not A	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some A are B	Some B are not C	Some A are C	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some A are B	Some B are not C	Some C are A	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are A	Some B are not C	Some A are C	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are A	Some B are not C	Some C are A	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are not A	Some B are C	Some A are C	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are not A	Some B are C	Some C are A	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are not A	Some C are B	Some A are C	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are not A	Some C are B	Some C are A	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some A are not B	Some C are not B	Some A are not C	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some A are not B	Some C are not B	Some C are not A	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are not A	Some B are not C	Some A are not C	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
Some B are not A	Some B are not C	Some C are not A	$8^n - 2 \cdot 6^n + 5^n$	$6^n - 5^n$
All A are B	Some B are not C	All C are A	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
All B are A	Some C are not B	All A are C	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some A are not B	All B are C	All C are A	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some B are not A	All C are B	All A are C	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some A are not B	Some B are not C	Some C are not A	$8^n - 2 \cdot 6^n + 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some B are not A	Some C are not B	Some A are not C	$8^n - 2 \cdot 6^n + 4^n$	$2^n + 6^n - 2 \cdot 4^n$
All A are B	Some B are C	All C are A	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
All A are B	Some C are B	All C are A	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
All B are A	Some C are not B	No A are C	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
All B are A	Some C are not B	No C are A	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some A are B	All C are B	All A are C	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some B are A	All C are B	All A are C	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
No A are B	Some C are not B	All C are A	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
No B are A	Some C are not B	All C are A	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some A are not B	All B are C	No A are C	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some A are not B	All B are C	No C are A	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some A are not B	No B are C	All A are C	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some A are not B	No C are B	All A are C	$6^n - 4^n$	$2^n + 6^n - 2 \cdot 4^n$
Some A are B	Some B are not C	Some C are not A	$8^n - 2 \cdot 6^n + 5^n$	$2^n + 6^n - 2 \cdot 4^n$
Some B are A	Some B are not C	Some C are not A	$8^n - 2 \cdot 6^n + 5^n$	$2^n + 6^n - 2 \cdot 4^n$
Some B are not A	Some B are C	Some A are not C	$8^n - 2 \cdot 6^n + 5^n$	$2^n + 6^n - 2 \cdot 4^n$
Some B are not A	Some C are B	Some A are not C	$8^n - 2 \cdot 6^n + 5^n$	$2^n + 6^n - 2 \cdot 4^n$

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
Some B are not A	Some B are not C	Some A are C	$8^n - 2 \cdot 6^n + 5^n$	$2^n + 6^n - 2 \cdot 4^n$
Some B are not A	Some B are not C	Some C are A	$8^n - 2 \cdot 6^n + 5^n$	$2^n + 6^n - 2 \cdot 4^n$
Some A are not B	Some B are not C	All C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 3 \cdot 4^n - 2^n$
Some B are not A	Some C are not B	All A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 3 \cdot 4^n - 2^n$
Some A are B	Some C are not B	All A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some A are B	Some C are not B	All C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some B are A	Some C are not B	All A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some B are A	Some C are not B	All C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some A are not B	Some B are C	All A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some A are not B	Some B are C	All C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some A are not B	Some C are B	All A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some A are not B	Some C are B	All C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some A are not B	Some B are not C	No A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some A are not B	Some B are not C	No C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some B are not A	Some C are not B	No A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some B are not A	Some C are not B	No C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some A are B	Some B are not C	All C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some B are A	Some B are not C	All C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some B are not A	Some B are C	All A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some B are not A	Some C are B	All A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some B are not A	Some B are not C	No A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some B are not A	Some B are not C	No C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 4^n - 3^n + 5^n$
Some A are B	Some C are not B	No A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are B	Some C are not B	No C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are A	Some C are not B	No A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are A	Some C are not B	No C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are not B	Some B are C	No A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are not B	Some B are C	No C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are not B	Some C are B	No A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are not B	Some C are B	No C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are not B	Some B are not C	All A are C	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are not A	Some C are not B	All C are A	$8^n - 2 \cdot 6^n + 4^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are B	Some B are C	All A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are B	Some B are C	All C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are B	Some C are B	All A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are B	Some C are B	All C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are A	Some B are C	All A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are A	Some B are C	All C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are A	Some C are B	All A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are A	Some C are B	All C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are B	Some B are not C	No A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are B	Some B are not C	No C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are A	Some B are not C	No A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are A	Some B are not C	No C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are not A	Some B are C	No A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$

(Continued on next page)

Table D.6—continued from previous page

Premise 1	*Premise 2*	*Conclusion*	*Models*	*Countermodels*
Some B are not A	Some B are C	No C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are not A	Some C are B	No A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are not A	Some C are B	No C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are not B	Some C are not B	All A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are not B	Some C are not B	All C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are not A	Some B are not C	All A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some B are not A	Some B are not C	All C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 2 \cdot 5^n$
Some A are B	Some B are C	No A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some A are B	Some B are C	No C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some A are B	Some C are B	No A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some A are B	Some C are B	No C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some B are A	Some B are C	No A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some B are A	Some B are C	No C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some B are A	Some C are B	No A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some B are A	Some C are B	No C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some A are B	Some B are not C	All A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some B are A	Some B are not C	All A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some B are not A	Some B are C	All C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some B are not A	Some C are B	All C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some A are not B	Some C are not B	No A are C	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$
Some A are not B	Some C are not B	No C are A	$8^n - 2 \cdot 6^n + 5^n$	$8^n - 3 \cdot 6^n + 3 \cdot 5^n - 4^n$

Appendix E

Responses in syllogistic reasoning

Table E.1: Percentages of inferences drawn of each type for the syllogistic problems (rounded to the nearest natural number). (*Some′* denotes the quantifier *some ... not.*)

			Inference drawn (%)							
		NVC	All		Some		No		Some′	
Quantifiers	*Figure*		AC	CA	AC	CA	AC	CA	AC	CA
All	ABBC	3	69	10	10	5	0	0	1	1
All	ABCB	27	37	11	18	5	0	0	1	1
	BABC	6	57	3	24	7	0	0	3	0
	BACB	4	37	34	15	6	1	0	2	1
All	ABBC	2	0	0	2	0	72	12	9	2
No	ABCB	5	1	0	1	1	50	28	11	2
	BABC	7	1	0	2	3	63	7	15	2
	BACB	6	2	0	5	3	39	22	17	6
All	ABBC	2	1	1	62	20	0	1	6	7
Some	ABCB	13	0	2	40	36	1	0	3	5
	BABC	0	2	9	55	24	1	0	6	3
	BACB	2	1	2	41	47	1	0	5	0
All	ABBC	6	0	1	27	8	3	0	47	8
Some′	ABCB	10	1	0	27	11	0	0	24	26
	BABC	1	1	3	13	15	0	0	60	7
	BACB	6	1	0	23	23	0	1	15	33
Some	ABBC	0	1	2	45	41	1	0	2	7
All	ABCB	6	3	0	43	34	2	0	3	8
	BABC	1	1	4	26	59	1	2	1	5
	BACB	5	1	0	28	57	0	2	1	6

(Continued on next page)

Table E.1—continued from previous page

			Inference drawn (%)							
		NVC	All		Some		No		Some′	
Quantifiers	*Figure*		AC	CA	AC	CA	AC	CA	AC	CA
Some No	ABBC	7	0	0	15	5	12	12	46	5
	ABCB	10	1	0	13	1	10	9	53	2
	BABC	11	0	0	9	6	17	9	45	3
	BACB	9	0	0	11	2	15	13	42	8
Some Some	ABBC	16	0	0	57	9	4	1	8	5
	ABCB	26	0	1	47	7	2	1	11	5
	BABC	26	1	1	55	5	0	0	8	5
	BACB	21	0	1	43	20	1	0	11	2
Some Some′	ABBC	19	0	0	26	5	0	2	41	8
	ABCB	22	0	0	28	5	1	1	24	19
	BABC	18	0	0	23	5	1	0	48	6
	BACB	25	1	0	23	9	0	1	32	9
No All	ABBC	7	0	2	2	5	27	39	5	14
	ABCB	8	0	0	1	3	37	43	3	5
	BABC	13	1	2	3	3	17	41	3	16
	BACB	3	0	2	5	3	25	51	2	9
No No	ABBC	35	3	0	18	5	26	5	5	3
	ABCB	49	6	1	11	2	18	7	6	0
	BABC	37	5	0	17	0	27	3	9	2
	BACB	39	5	0	10	9	22	11	5	0
No Some	ABBC	14	0	1	12	5	21	8	6	35
	ABCB	12	0	0	2	6	18	15	3	44
	BABC	7	0	0	7	6	20	20	3	36
	BACB	5	0	0	8	8	13	11	8	48
No Some′	ABBC	21	1	1	12	5	13	8	8	32
	ABCB	16	0	0	8	22	6	7	8	34
	BABC	17	0	0	10	7	10	12	8	36
	BACB	12	0	0	9	31	9	6	4	29
Some′ All	ABBC	3	1	0	29	17	0	0	32	18
	ABCB	10	1	0	14	28	1	1	21	24
	BABC	2	2	0	18	15	0	2	11	49
	BACB	3	1	1	22	19	0	0	7	47
Some′ No	ABBC	15	1	1	25	1	7	7	37	7
	ABCB	9	0	1	24	9	6	8	36	7
	BABC	12	0	0	8	7	12	9	44	9
	BACB	20	1	0	8	7	10	9	38	7
Some′ Some	ABBC	16	0	0	32	17	0	0	14	21
	ABCB	26	1	0	19	14	0	2	18	19
	BABC	20	0	0	23	10	2	1	9	35
	BACB	14	1	0	23	12	0	3	7	39

(Continued on next page)

Table E.1—continued from previous page

			Inference drawn (%)							
		NVC	All		Some		No		Some′	
Quantifiers	*Figure*		AC	CA	AC	CA	AC	CA	AC	CA
Some′	ABBC	21	2	0	27	6	1	0	38	5
Some′	ABCB	38	2	0	31	6	1	3	10	9
	BABC	37	1	1	21	6	0	1	29	4
	BACB	26	1	0	20	15	0	0	19	19

Table E.2: A comparison of the percentages of each type of response obtained in my sample with that obtained by Chater and Oaksford (1999) in their meta analysis (rounded to the nearest natural number). (*Some′* denotes the quantifier *some … not*.)

		Inference drawn (%)									
		NVC		All		Some		No		Some′	
Quantifiers	*Figure*	Meta	Sample	Meta	Sample	Meta	Sample	Meta	Sample	Meta	Sample
All	ABBC	4	3	75	79	16	15	1	0	1	2
All	ABCB	28	27	58	48	8	23	1	0	1	2
	BABC	14	6	57	60	29	31	0	0	0	3
	BACB	2	4	90	71	5	21	0	1	0	4
All	ABBC	3	2	0	0	3	2	87	84	2	12
No	ABCB	6	5	0	1	0	2	88	78	1	14
	BABC	23	7	0	1	1	6	61	69	13	17
	BACB	32	6	0	2	3	8	59	61	6	23
All	ABBC	27	2	0	2	71	82	0	1	1	12
Some	ABCB	29	13	0	2	57	76	3	1	11	8
	BABC	4	0	1	11	89	79	1	1	3	9
	BACB	3	2	0	3	92	89	3	1	3	5
All	ABBC	20	6	0	1	5	35	3	3	72	55
Some′	ABCB	22	10	0	1	6	39	3	0	67	50
	BABC	19	1	0	5	10	28	0	0	66	67
	BACB	31	6	1	1	6	45	1	1	57	47
No	ABBC	24	7	1	2	3	7	61	66	8	18
All	ABCB	4	8	0	0	0	5	89	79	3	8
	BABC	12	13	0	3	0	7	64	58	22	19
	BACB	4	3	0	2	1	8	87	75	3	11
No	ABBC	63	35	0	3	3	23	31	31	1	8
No	ABCB	77	49	3	7	3	14	14	25	3	6
	BABC	78	37	0	5	0	17	18	30	3	11
	BACB	61	39	0	5	1	19	34	33	1	5

(Continued on next page)

Table E.2—continued from previous page

		Inference drawn (%)									
		NVC		All		Some		No		Some′	
Quantifiers	*Figure*	Meta	Sample	Meta	Sample	Meta	Sample	Meta	Sample	Meta	Sample
No Some	ABBC	37	14	0	1	2	16	32	29	27	40
	ABCB	22	12	1	0	1	8	21	33	52	47
	BABC	27	7	0	0	6	13	15	41	48	40
	BACB	14	5	0	0	5	16	15	24	66	56
No Some′	ABBC	70	21	0	2	5	16	8	21	12	40
	ABCB	65	16	0	0	13	30	7	12	11	42
	BABC	58	17	0	0	0	17	9	23	28	44
	BACB	59	12	1	0	8	40	8	15	23	33
Some All	ABBC	4	0	0	3	91	86	1	1	1	9
	ABCB	31	6	13	3	49	77	3	2	12	11
	BABC	6	1	2	5	85	85	1	4	4	6
	BACB	22	5	0	1	72	85	0	2	6	7
Some No	ABBC	24	7	0	0	1	20	28	23	44	51
	ABCB	27	10	0	1	0	14	39	19	30	56
	BABC	35	11	0	0	1	15	30	26	33	48
	BACB	57	9	1	0	1	14	22	27	16	50
Some Some	ABBC	57	16	0	0	42	66	0	5	1	13
	ABCB	49	26	1	1	42	53	3	3	3	16
	BABC	72	26	0	2	24	59	3	0	1	13
	BACB	51	21	0	1	41	63	3	1	4	14
Some Some′	ABBC	47	19	0	0	5	30	1	2	44	49
	ABCB	49	22	1	0	5	33	4	2	37	43
	BABC	59	18	0	0	9	27	1	1	29	53
	BACB	61	25	3	1	4	32	1	1	30	41
Some′ All	ABBC	33	3	1	1	3	46	6	0	27	49
	ABCB	23	10	0	1	11	42	5	2	56	45
	BABC	8	2	0	2	15	33	3	2	69	61
	BACB	23	3	0	2	3	41	3	0	68	53

(Continued on next page)

Table E.2—continued from previous page

		Inference drawn (%)									
		NVC		All		Some		No		Some′	
Quantifiers	*Figure*	Meta	Sample	Meta	Sample	Meta	Sample	Meta	Sample	Meta	Sample
Some′	ABBC	56	15	0	2	19	26	9	14	14	44
No	ABCB	63	9	0	1	8	33	11	14	16	43
	BABC	61	12	0	0	5	15	12	21	18	53
	BACB	77	20	1	1	0	15	14	20	5	45
Some′	ABBC	55	16	3	0	8	49	2	0	29	36
Some	ABCB	50	26	0	1	8	33	3	2	35	38
	BABC	56	20	1	0	9	33	1	3	31	44
	BACB	54	14	4	1	6	36	0	3	35	46
Some′	ABBC	65	21	1	2	4	33	1	1	25	42
Some′	ABCB	68	38	0	2	16	36	5	5	10	19
	BABC	77	37	1	2	6	27	0	1	15	33
	BACB	66	26	1	1	8	35	1	0	22	38

Appendix F

Distribution of response times in immediate inference

Figures F.1–F.4 on the following pages show the distribution of reaction times (in seconds) for immediate inference problems, ordered from those showing most variation in times to those showing the least.

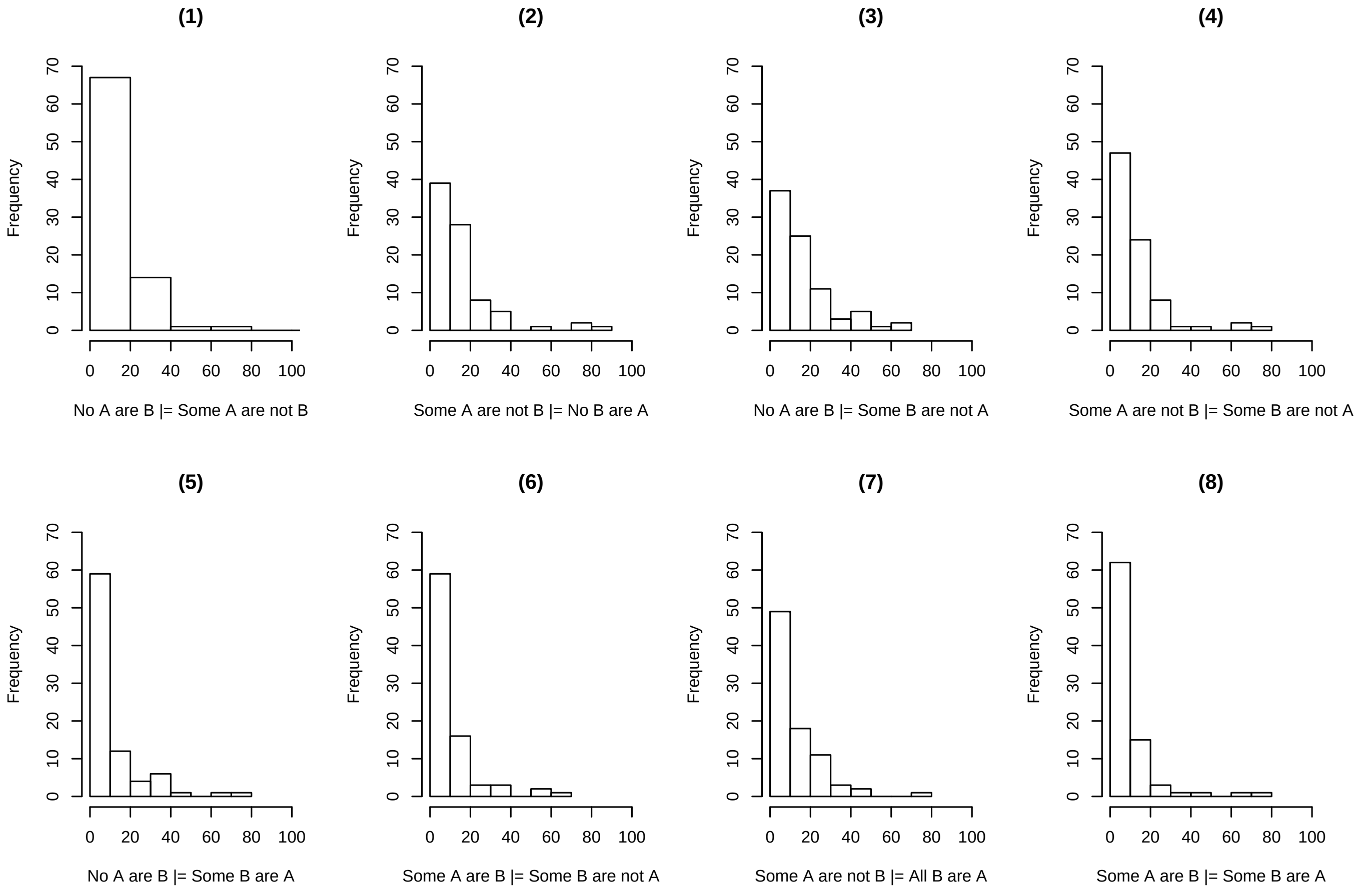

Figure F.1: Histogram of response times for problems 1–8.

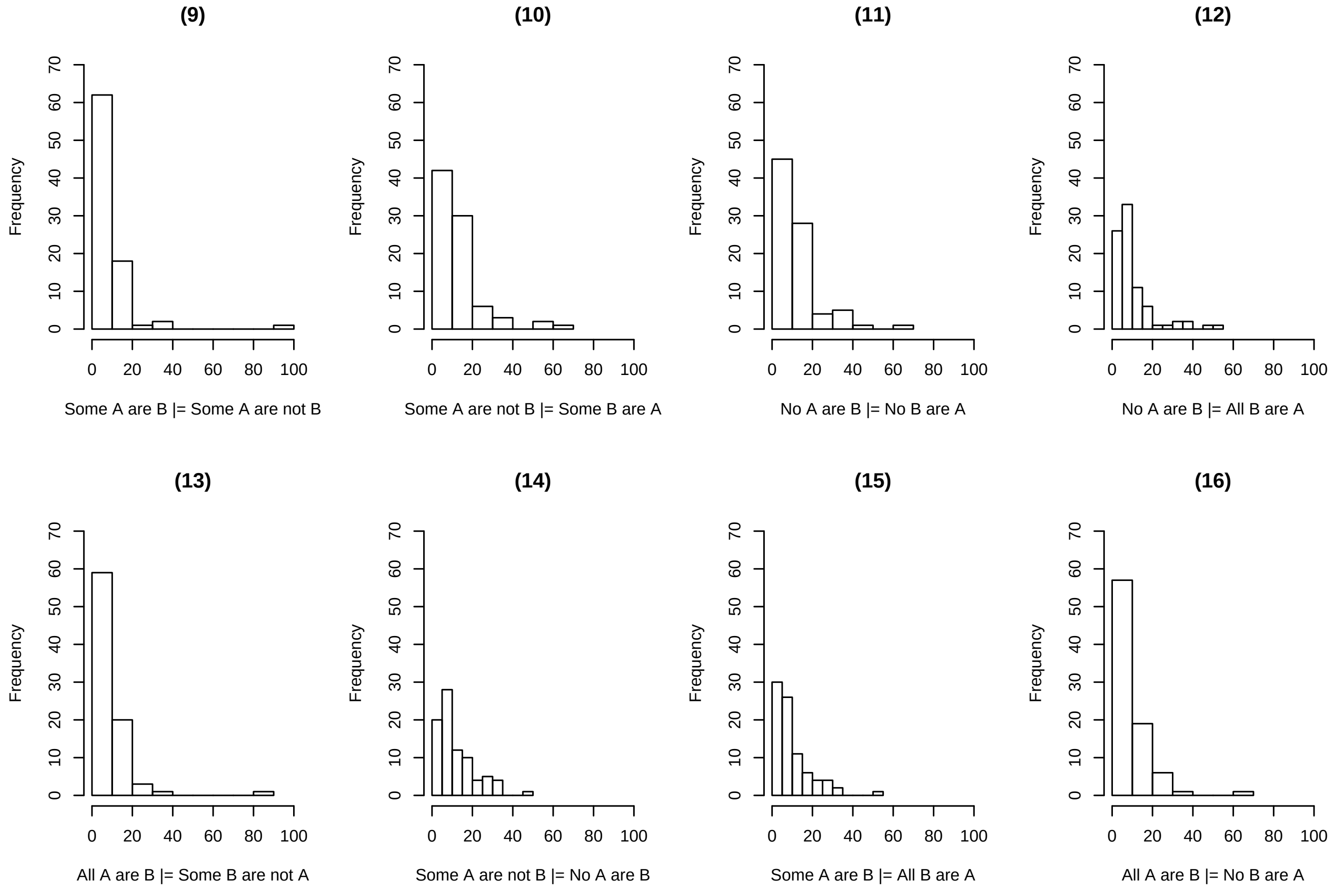

Figure F.2: Histogram of response times for problems 9–16.

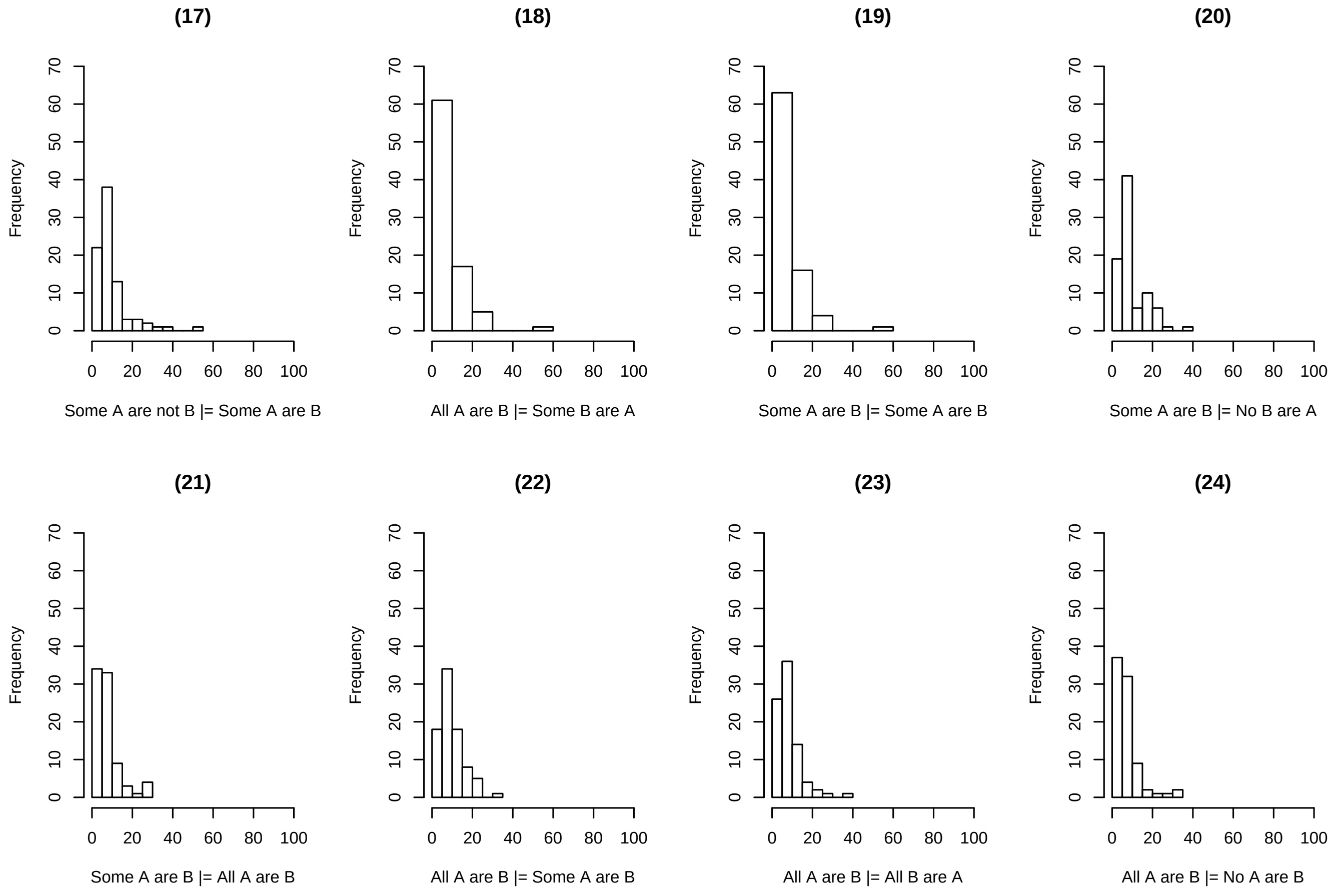

Figure F.3: Histogram of response times for problems 17–24.

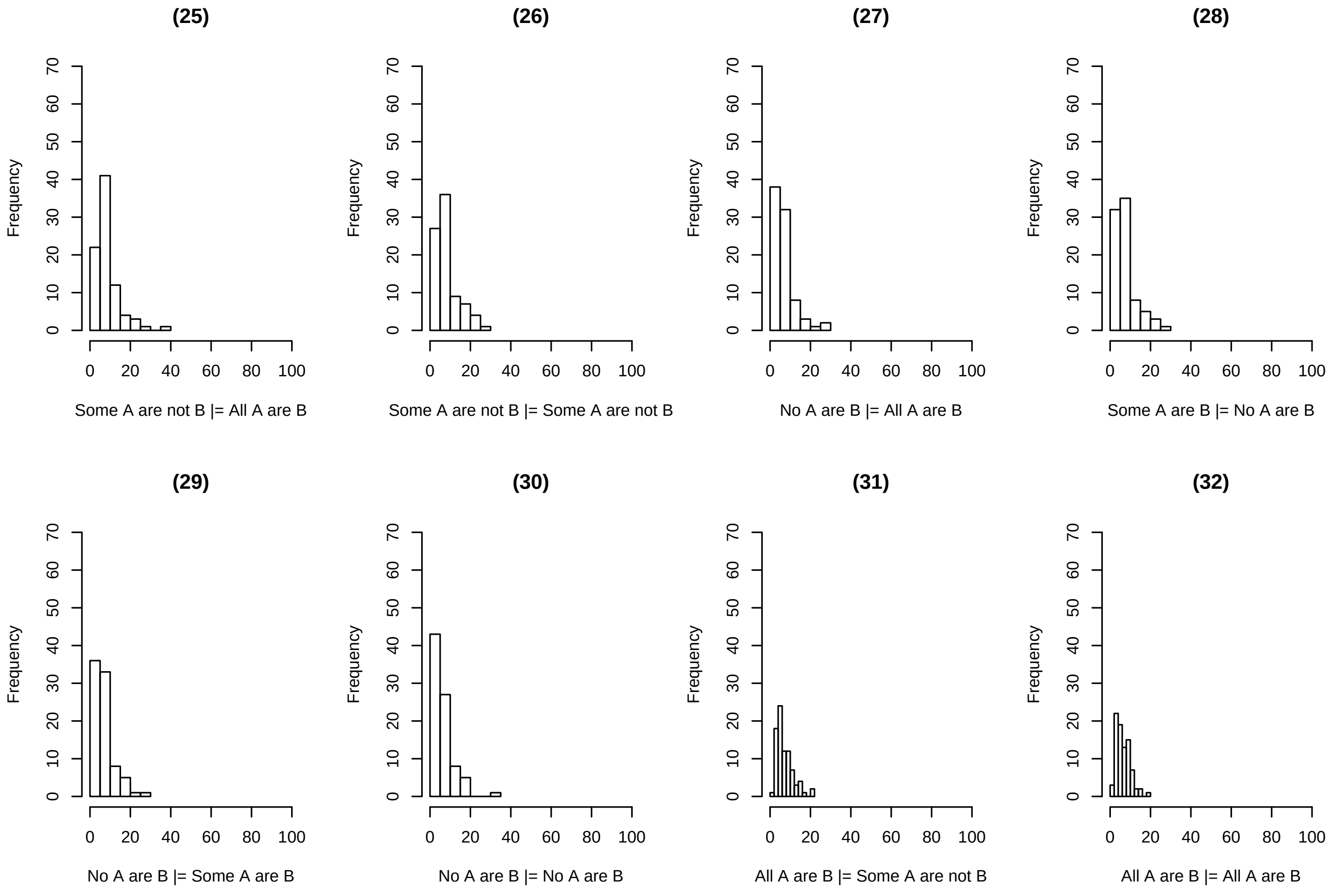

Figure F.4: Histogram of response times for problems 25–32.

Appendix G

Supplementary Analyses

G.1 The Linda Problem

Recall again that the possible responses were as follows.

A. Linda is a bank teller
B. Linda is active in the feminist movement
C. Linda is a bank teller and is active in the feminist movement

A. Bill plays in a rock band for a hobby
B. Bill is an accountant
C. Bill is an accountant and plays in a rock band for a hobby

Participants were asked to give the most probable statement first. According to probability theory, $P(\alpha) \geq P(\alpha \wedge \beta)$ and $P(\beta) \geq P(\alpha \wedge \beta)$, so, according to this criterion, participants should respond with both A and B (in either order) before C in both cases. See Table G.1 for the responses given.

	Proportion of participants (%)	
Response	Bill	Linda
A	0	2
B	18	17
C	4	1
BA	0	1
BC	1	0
ABC	0	2
ACA	1	0
ACB	2	2
BAC	18	17
BBA	1	0
BCA	52	49
CAB	0	1
CBA	2	7

Table G.1: Percentage of people who gave each responses for the two problems, rounded to the nearest natural number.

For both problems, around 18% of participants gave only one answer, that of the second conjunct. Around the same number (18%) do indicate that the individual conjuncts are more probable than the conjunction. However the modal answer, with around half the participants, is to respond with $P(A \wedge B) \geq P(A)$. Also around 8% of participants responded that the conjunction was more probable than both of the individual conjuncts.

Bibliography

Agresti, A. (2002). *Categorical data analysis.* NJ, USA: Wiley.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders DSM-IV-TR (text revision)* (4th ed.). Washington DC: American Psychiatric Association.

Anderson, B. (1993). Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: intelligence? *Neuroscience Letters*, *153*(1), 98–102. Available from http://dx.doi.org/10.1016/0304-3940(93)90086-Z

Ando, J., Shikishima, C., Hiraishi, K., Sugimoto, Y., Takemura, R., & Okada, M. (2006). At the crossroad of logic, psychology, and behavioral genetics: Development of the deductive reasoning test in the Keio Twin-BAROCO project. In D. Andler, O. Mitsuhiro, O. Yoshinori, & W. Shigeru (Eds.), *Reasoning and cognition* (Vol. 2, pp. 19–36). Tokyo, Japan: Keio University Press.

Asperger, H. (1944/1991). 'Autistic psychopathy' in childhood. In U. Frith (Ed.), *Autism and Asperger's syndrome* (p. 37-92). Cambridge, UK: Cambridge University Press. (Translated and annotated by Uta Frith.)

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Differences*, *38*(2), 451–460.

Bach, P., & Tipper, S. P. (2007). Implicit action encoding influences personal-trait judgments. *Cognition*, *102*(2), 151–178. Available from http://dx.doi.org/10.1016/j.cognition.2005.11.003

Bacon, A. M., Handley, S. J., & McDonald, E. L. (2007). Reasoning and dyslexia: A spatial strategy may impede reasoning with visually rich information. *British Journal of Psychology*, *98*, 79–92.

Bacon, A. M., Handley, S. J., & Newstead, S. E. (2003). Individual differences in strategies for syllogistic reasoning. *Thinking and Reasoning*, *9*(2), 133–168.

Baddeley, A. (1986). *Working memory.* Oxford, UK: Oxford Science Publications.

Bailey, A., Couteur, A. L., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., et al. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological Medicine*, *25*(1), 63–77.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., et al. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet*, *368*(9531), 210–215. Available from http://dx.doi.org/10.1016/S0140-6736(06)69041-7

Barbeau, E., Mendrek, A., & Mottron, L. (2008). Are autistic traits autistic? *British Journal of Psychology*. Available from http://dx.doi.org/10.1348/000712608X337788 (Fast track article)

Barnes, R. (2008, January 23). Court rules inmates can't sue for property loss. *The Washington Post*, A02.

Baron-Cohen, S. (1997). *Mindblindness.* The MIT Press.

Baron-Cohen, S., & Hammer, J. (1997). Is autism an extreme form of the 'male brain'? *Advances in Infancy Research*, *11*, 193–217.

Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition*, *21*, 37–46.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society of London: Series B—Biological Sciences*, *358*(1430), 361–374.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism Developmental Disorders*, *34*(2), 163–175.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'reading the mind in the eyes' test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17.

Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed-effects models using s4 classes [Computer software manual]. Available from http://lme4.r-forge.r-project.org/ (R package version 0.999375-25)

Bayliss, A. P., Pellegrino, G. di, & Tipper, S. P. (2005). Sex differences in eye gaze and symbolic cueing of attention. *The Quarterly Journal of Experimental Psychology A*, *58*(4), 631–650.

Bayliss, A. P., & Tipper, S. P. (2005). Gaze and arrow cueing of attention reveals individual differences along the autism spectrum as a function of target context. *British Journal of Psychology*, *96*(Pt 1), 95–114. Available from http://dx.doi.org/10.1348/000712604X15626

Beckert, B., & Posegga, J. (1995). leanT^AP: Lean tableau-based deduction. *Journal of Automated Reasoning*, *15*(3), 339–358.

Begg, I., & Harris, G. (1982). On the interpretation of syllogisms. *Journal of Verbal Learning and Verbal Behavior*, *21*(5), 595–620.

Beglinger, L. J., & Smith, T. H. (2001). A review of subtyping in autism and proposed dimensional classification model. *Journal of Autism and Developmental Disorders*, *31*(4), 411–422.

Bentall, R. (2006). Madness explained: why we must reject the Kraepelinian paradigm and replace it with a 'complaint-orientated' approach to understanding mental illness. *Medical Hypotheses*, *66*(2), 220–233. Available from http://dx.doi.org/10.1016/j.mehy.2005.09.026

Benthem, J. van. (2008). Logic and reasoning: Do the facts matter? *Studia Logica*, *88*(1), 67–84.

Bishop, D. V. M., Maybery, M., Maley, A., Wong, D., Hill, W., & Hallmayer, J. (2004).

Using self-report to identify the broad phenotype in parents of children with autistic spectrum disorders: a study using the Autism-Spectrum Quotient. *Journal of Child Psychology and Psychiatry*, *45*(8), 1431–1436. Available from http://dx.doi.org/10.1111/j.1469-7610.2004.00849.x

Bobzien, S. (2000). Why the order of the figures of the hypothetical syllogisms was changed. *The Classical Quarterly*, *50*(1), 247–251. Available from http://www.jstor.org/stable/1558949

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Reviews of Psychology*, *53*, 605–634.

Bölte, S., Rudolf, L., & Poustka, F. (2002). The cognitive structure of higher functioning autism and schizophrenia: a comparative study. *Comprehensive Psychiatry*, *43*(4), 325–330. Available from http://dx.doi.org/10.1053/comp.2002.33490

Bonnefon, J.-F., Eid, M., Vautier, S., & Jmel, S. (2007). A mixed Rasch model of dual-process conditional reasoning. *Quarterly Journal of Experimental Psychology*, *61*, 809–824. Available from http://dx.doi.org/10.1080/17470210701434573

Bonnefon, J.-F., Vautier, S., & Eid, M. (2007). Modeling individual differences in contrapositive reasoning with continuous latent state and trait variables. *Personality and Individual Differences*, *42*(7), 1403–1413. Available from http://dx.doi.org/10.1016/j.paid.2006.10.017

Boolos, G. S. (1984). On 'syllogistic inference'. *Cognition*, *17*(2), 181–184.

Boolos, G. S., & Jeffrey, R. C. (1980). *Computability and logic* (2nd ed.). Cambridge University Press.

Booth, R., Charlton, R., Hughes, C., & Happé, F. (2003). Disentangling weak coherence and executive dysfunction: planning drawing in autism and attention-deficit/hyperactivity disorder. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *358*(1430), 387–392. Available from http://dx.doi.org/10.1098/rstb.2002.1204

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425–440.

Borsboom, D., Mellenbergh, G. J., & Heerden, J. van. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457. Available from http://dx.doi.org/10.1016/j.jml.2004.05.006

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, *27*(3), 281–291. Available from http://jpubhealth.oxfordjournals.org/cgi/content/full/27/3/281

Burgess, J. P. (2008). Tarski's tort. In *Mathematics, models, and modality: Selected philosophical essays* (pp. 149–168). Cambridge, UK: Cambridge University Press.

Burgess, P. W., Dumontheil, I., & Gilbert, S. J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in Cognitive Sciences*, *11*(7), 290–298. Available from http://dx.doi.org/10.1016/j.tics.2007.05.004

Burgess, P. W., Gilbert, S. J., & Dumontheil, I. (2007). Function and localization within rostral prefrontal cortex (area 10). *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *362*(1481), 887–899. Available from http://dx.doi.org/10.1098/rstb.2007.2095

Burgess, P. W., Simons, J. S., Dumontheil, I., & Gilbert, S. J. (2005). The gateway hypothesis of rostral prefrontal cortex (area 10) function. In J. Duncan, L. Phillips, & P. McLeod (Eds.), *Measuring the mind: Speed, control, and age* (pp. 217–248). Oxford, UK: Oxford University Press.

Burnette, C. P., Mundy, P. C., Meyer, J. A., Sutton, S. K., Vaughan, A. E., & Charak, D. (2005, Feb). Weak central coherence and its relations to theory of mind and anxiety in autism. *Journal of Autism and Developmental Disorders*, *35*(1), 63–73.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*(1), 61–83.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431.

Chabris, C. F. (2007). Cognitive and neurobiological mechanisms of the law of general intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus* *domain specific processes in higher cognition* (chap. 19). Hove, UK: Psychology Press.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, *38*(2), 191–258.

Cheng, P. W., & Hollyoak, K. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391–416.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, *76*(4), 387–404.

Colom, R., Abad, F. J., García, L. F., & Juan-Espinosa, M. (2002). Education, Wechsler's Full Scale IQ, and *g*. *Intelligence*, *30*(5), 449–462.

Couteur, A. S. L. (2003). *National Autism Plan for Children (NAPC).* The National Autistic Society for NIASA in collaboration with The Royal College of Psychiatrists (RCPsych), The Royal College of Paediatrics and Child Health (RCPCH) and the All Party Parliamentary Group on Autism (APPGA). Available from http://www.nas.org.uk/nas/jsp/polopoly.jsp?a=2178&d=368

Crespi, B., & Badcock, C. (2008). Psychosis and autism as diametrical disorders of the social brain. *Behavioral and Brain Sciences*, *31*, 241–261.

Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, *23*(5), 646–658.

Da Silva Neves, R., Bonnefon, J.-F., & Raufaste, E. (2002). An empirical test of patterns for nonmonotonic inference. *Annals of Mathematics and Artificial Intelligence*, *34*(1–3), 107–130.

Dalen, D. van. (1997). *Logic and structure* (Third ed.). Berlin: Springer.

Davidoff, J., Fonteneau, E., & Fagot, J. (2008). Local and global processing: Observations from a remote culture. *Cognition*, *108*(3), 702–709.

Dawson, M., Soulières, I., Gernsbacher, M. A., & Mottron, L. (2007). The level and nature of autistic intelligence. *Psychological Science*, *18*(8), 657–662. Available from http://dx.doi.org/10.1111/j.1467-9280.2007.01954.x

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, *54*(2), 128–133. Available from http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=zea-54-2-128&site=ehost-live

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* Oxford University Press.

Dehaene, S., & Changeux, J. P. (1991). The Wisconsin Card Sorting Test: theoretical analysis and modeling in a neuronal network. *Cerebral Cortex*, *1*(1), 62–79.

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. *Intelligence*, *21*(2), 135–155.

Dias, M., & Harris, P. L. (1988). The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology*, *6*, 207–221.

Dick, P. K. (1977/2006). *A scanner darkly.* London, UK: Gollancz.

Dinno, A. (2008). paran: Horn's test of principal components/factors [Computer software manual]. (R package version 1.3.2)

Dodgson, C. L. (1958). *Mathematical recreations of Lewis Carroll: Symbolic Logic and the Game of Logic.* New York, USA: Dover.

Duncan, J., Johnson, R., Swales, M., & Freer, C. (1997). Frontal lobe deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology*, *14*(5), 713–741.

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

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory and Cognition*, *35*(7), 1772–1784.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, *112*(4), 1040–1052.

Eysenck, H. J., & Eysenck, S. B. G. (1976). *Psychoticism as a dimension of personality.* London, UK: Hodder and Stroughton.

Falmagne, R. J., & Gonsalves, J. (1995). Deductive inference. *Annual Review of Psychol-*

ogy, *46*, 525–559.

Faraway, J. J. (2006). *Extending the linear model with R.* Chapman & Hall/CRC.

Ferrando, P. J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicológica*, *21*, 301–323.

Flint, J., & Munafò, M. R. (2007). The endophenotype concept in psychiatric genetics. *Psychological Medicine*, *37*(2), 163–180. Available from http://dx.doi.org/10.1017/S0033291706008750

Folstein, S., & Rutter, M. (1977). Genetic influences and infantile autism. *Nature*, *265*(5596), 726–728.

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition*, *54*(1), 1–71.

Fox, J. (2007). sem: Structural Equation Models [Computer software manual]. (R package version 0.9-11)

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225.

Frith, C. (2004). Is autism a disconnection disorder? *The Lancet Neurology*, *3*(10), 577. Available from http://dx.doi.org/10.1016/S1474-4422(04)00875-0

Frith, U. (1989). *Autism: Explaining the enigma.* Oxford, UK: Basil Blackwell.

Frith, U., & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *British Journal of Developmental Psychology*, *1*, 329–342.

Geis, M. L., & Zwicky, A. M. (1971). On invited inferences. *Linguistic Inquiry*, *2*, 560-566.

Gelman, A. (To appear). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*. Available from http://www.stat.columbia.edu/~gelman/research/published/standardizing7.pdf

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge University Press.

Gentzen, G. (1935). Untersuchungen über das logische Schließen. *Mathematische Zeitschrift*, *39*, 176–210, 405–431. (1934–5)

Geurts, B. (2003). Reasoning with quantifiers. *Cognition*, *86*, 223–251.

Geurts, B., & Slik, F. van der. (2005). Monotonicity and processing load. *Journal of Semantics*, *22*, 97–117.

Ghaziuddin, M., Ghaziuddin, N., & Greden, J. (2002). Depression in persons with autism: implications for research and clinical care. *Journal of Autism and Developmental Disorders*, *32*(4), 299–306.

Gilbert, S. J., Bird, G., Brindley, R., Frith, C. D., & Burgess, P. W. (2008). Atypical recruitment of medial prefrontal cortex in autism spectrum disorders: An fMRI study of two executive function tasks. *Neuropsychologia*, *46*(9), 2281–2291. Available from http://dx.doi.org/10.1016/j.neuropsychologia.2008.03.025

Gilbert, S. J., & Burgess, P. W. (2008). Executive function. *Current Biology*, *18*, R110–R114.

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory and Cognition*, *21*(1), 115–124.

Gillberg, C. (1983). Are autism and anorexia nervosa related? *British Journal of Psychiatry*, *142*, 428.

Girard, J.-Y. (1999). On the meaning of logical rules I: syntax vs. semantics. In U. Berger & H. Schwichtenberg (Eds.), *Computational logic* (Vol. 165, pp. 215–272). Springer Verlag.

Girard, J.-Y. (2003). From foundations to ludics. *The Bulletin of Symbolic Logic*, *9*(2), 131–168. Available from http://www.jstor.org/stable/3094787

Girard, J.-Y., Taylor, P., & Lafont, Y. (1989/2003). *Proofs and types* (No. 7). Cambridge University Press. (Reprinted for the Web 2003)

Girotto, V., Mazzocco, A., & Tasso, A. (1997). The effect of premise order in conditional reasoning: a test of the mental model theory. *Cognition*, *63*(1), 1–28.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, *55*(3), 377–393.

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The 'reading the mind in the voice' test-revised: a study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental*

Disorders, *37*(6), 1096–1106. Available from http://dx.doi.org/10.1007/s10803-006-0252-5

Gorden, R. L. (1953). The effect of attitude toward Russia on logical reasoning. *Journal of Social Psychology; Political, Racial and Differential Psychology*, *37*(1), 103-111.

Gottesman, I. I., & Shields, J. (1973). Genetic theorizing and schizophrenia. *British Journal of Psychiatry*, *122*(566), 15–30.

Grice, H. P. (1975). Logic and conversation. In C. Peter & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts.* Academic Press.

Grimmett, G., & Welsh, D. (1986). *Probability: An introduction.* Oxford Science Publications. (Reprinted with corrections, 2003)

Hahn, B., Ross, T. J., & Stein, E. A. (2006). Neuroanatomical dissociation between bottomup and topdown processes of visuospatial selective attention. *NeuroImage*, *32*(2), 842–853.

Hambrook, D., Tchanturia, K., Schmidt, U., Russell, T., & Treasure, J. (2008). Empathy, systemizing, and autistic traits in anorexia nervosa: A pilot study. *British Journal of Clinical Psychology*. Available from http://dx.doi.org/10.1348/014466507X272475 (In Press)

Hamel, R., & Schmittmann, V. D. (2006). The 20-minute version as a predictor of the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement*, *66*(6), 1039–1046.

Happé, F. (1994). Wechsler IQ profile and theory of mind in autism: a research note. *Journal of Child Psychology and Psychiatry*, *35*(8), 1461–1471.

Happé, F., & Booth, R. D. L. (2008, Jan). The power of the positive: revisiting weak coherence in autism spectrum disorders. *The Quarterly Journal of Experimental Psychology*, *61*(1), 50–63. Available from http://dx.doi.org/10.1080/17470210701508731

Happé, F., & Frith, U. (2006, Jan). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. Available from http://dx.doi.org/10.1007/s10803-005-0039-0

Happé, F., & Ronald, A. (In press). The 'fractionable autism triad': A review of evidence from behavioural, genetic, cognitive and neural research. *Neuropsychology Review*. Available from http://dx.doi.org/10.1007/s11065-008-9076-8

Happé, F. G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154.

Harvey, A. G., Watkins, E., Mansell, W., & Shafran, R. (2004). *Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment*. Oxford University Press.

Hauck, J., Walter W., & Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association*, *72*(360), 851–853. Available from http://links.jstor.org/sici?sici=0162-1459%28197712%2972%3A360%3C851%3AWTAATH%3E2.0.CO%3B2-K

Heeger, D., & Ress, D. (2002). What does fMRI tell us about neuronal activity? *Nature Reviews Neuroscience*, *3*(2), 142–151.

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review*, *69*(4), 366–378.

Hill, E. L., & Bird, C. M. (2006). Executive processes in Asperger syndrome: Patterns of performance in a multiple case series. *Neuropsychologia*, *44*(14), 2822–2835.

Hoare, C. A. R. (1969). An axiomatic basis for computer programming. *Communications of the ACM*, *12*(10), 576–580, 583.

Hodges, W. (2005). Detecting the logical content: Burley's 'Purity of Logic'. In S. N. Artëmov, H. Barringer, A. S. d'Avila Garcez, L. C. Lamb, & J. Woods (Eds.), *We Will Show Them! Essays in Honour of Dov Gabbay, Volume Two* (pp. 69–116). London: King's College Publications.

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the autism-spectrum quotient (aq): A study in dutch population and patient groups. *Journal of Autism and Developmental Disorders*. Available from http://dx.doi.org/10.1007/s10803-008-0538-x (Online first.)

Hoekstra, R. A., Bartels, M., Verweij, C. J. H., & Boomsma, D. I. (2007). Heritability of autistic traits in the general population. *Archives of Pediatrics and Adolescent Medicine*, *161*(4), 372–377. Available from http://dx.doi.org/10.1001/archpedi.161.4.372

Horn, J. L. (1965). A rationale and a test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185.

Hughes, C., Leboyer, M., & Bouvard, M. (1997, Jan). Executive function in parents of children with autism. *Psychological Medicine*, *27*(1), 209–220.

Hughes, C., Russell, J., & Robbins, T. (1994). Evidence for executive dysfunction in autism. *Neuropsychologia*, *32*, 477–492. Available from http://dx.doi.org/10.1016/0028-3932(94)90092-2

Hunt, E. B. (1974). Quote the Raven? Nevermore! In L. W. Gregg (Ed.), *Knowlege and cognition* (pp. 129–157). Potomac, Maryland, US: Lawrence Erlbaum Associates.

Hurley, R. S., Losh, M., Parlier, M., Reznick, J. S., & Piven, J. (2007). The broad autism phenotype questionnaire. *Journal of Autism and Developmental Disorders*, *37*, 1679–1690.

Hurst, R. M., Mitchell, J. T., Kimbrel, N. A., Kwapil, T. K., & Nelson-Gray, R. O. (2007). Examination of the reliability and factor structure of the Autism Spectrum Quotient (AQ) in a non-clinical sample. *Personality and Individual Differences*, *43*(7), 1938–1949. Available from http://dx.doi.org/10.1016/j.paid.2007.06.012

Hurst, R. M., Nelson-Gray, R. O., Mitchell, J. T., & Kwapil, T. R. (2007). The relationship of Asperger's characteristics and schizotypal personality traits in a non-clinical adult sample. *Journal of Autism and Developmental Disorders*, *37*(9), 1711–1720. Available from http://www.springerlink.com/content/h832442007x71848/

James, I. (2003). Singular scientists. *Journal of the Royal Society of Medicine*, *96*, 36–39. Available from http://www.jrsm.org/cgi/content/full/96/1/36

Jarrold, C., Butler, D. W., Cottington, E. M., & Jimenez, F. (2000). Linking theory of mind and central coherence bias in autism and in the general population. *Developmental Psychology*, *36*(1), 126–138.

Jobe, L. E., & White, S. W. (2007). Loneliness, social relationships, and a broader autism phenotype in college students. *Personality and Individual Differences*, *42*(8), 1479–1489. Available from http://dx.doi.org/10.1016/j.paid.2006.10.021

John, B., & Lewis, K. R. (1966, May). Chromosome variability and geographic distribution in insects. *Science*, *152*(3723), 711–721. Available from http://dx.doi.org/10.1126/science.152.3723.711

Johnson, W., Bouchard, J., T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *184*, 95–107. Available from http://dx.doi.org/10.1016/S0160-2896(03)00062-X

Johnson, W., & Bouchard, T. J., Jr. (2007). Sex differences in mental abilities: *g* masks the dimensions on which they lie. *Intelligence*, *35*, 23–39.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge University Press.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Reviews of Psychology*, *50*, 109–135.

Johnson-Laird, P. N., & Bara, B. G. (1984). Logical expertise as a cause of error: A reply to boolos. *Cognition*, *17*(2), 183–184.

Johnson-Laird, P. N., Byrne, R. M., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, *99*(3), 418–439.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Lawrence Erlbaum Associates.

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, *10*, 64–99.

Jolliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory: linguistic processing in high-functioning adults with autism or asperger syndrome: is local coherence impaired? *Cognition*, *71*(2), 149–185.

Jonge, M. V. de, Kemner, C., & Engeland, H. van. (2006). Superior disembedding performance of high-functioning individuals with autism spectrum disorders and their parents: the need for subtle measures. *Journal of Autism and Developmental Disorders*, *36*(5), 677–683. Available from http://dx.doi.org/10.1007/s10803-006-0113-2

Jovović, V., & Kilibarda, G. (1999). On the number of Boolean functions in the Post classes f_8^μ. *Diskretnaya Matematika*, *11*, 127–138.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–50.

Kant, S. G., Wit, J. M., & Breuning, M. H. (2005). Genetic analysis of tall stature. *Hormone Research*, *64*(3), 149–156. Available from http://dx.doi.org/10.1159/000088589

Ketelaars, C., Horwitz, E., Sytema, S., Bos, J., Wiersma, D., Minderaa, R., et al. (2008).

Brief report: Adults with mild autism spectrum disorders (ASD): Scores on the Autism Spectrum Quotient (AQ) and comorbid psychopathology. *Journal of Autism and Developmental Disorders*, *38*(1), 176–180. Available from http://dx.doi.org/10.1007/s10803-007-0358-4

Klaczynski, P. A., & Daniel, D. B. (2005). Individual differences in conditional reasoning: A dual-process account. *Thinking & Reasoning*, *11*(4), 305–325.

Knauff, M. (2007). How our brains reason logically. *Topoi*, *26*, 19–36.

Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: toward an integrated framework of cognitive style. *Psychological Bulletin*, *133*(3), 464–481. Available from http://dx.doi.org/10.1037/0033-2909.133.3.464

Kunihira, Y., Senju, A., Dairoku, H., Wakabayashi, A., & Hasegawa, T. (2006). Autistic traits in non-autistic Japanese populations: Relationships with personality traits and cognitive ability. *Journal of Autism and Developmental Disorders*, *36*(4), 553–566. Available from http://dx.doi.org/10.1007/s10803-006-0094-1

Kurita, H., & Koyama, T. (2006). Autism-spectrum quotient Japanese version measures mental health problems other than autistic traits. *Psychiatry and Clinical Neurosciences*, *60*(3), 373–378. Available from http://dx.doi.org/10.1111/j.1440-1819.2006.01516.x

Lawley, D. N., & Maxwell, A. E. (1962). Factor analysis as a statistical method. *The Statistician*, *12*(3), 209–229.

Lechler, A. (2004). *Interpretation of conditionals in the suppression task.* MSc Thesis, Human Communication Research Centre, The University of Edinburgh, Scotland, UK.

Le Couteur, A., Lord, C., & Rutter, M. (2003). *The Autism Diagnostic Interview—Revised (ADI-R).* Los Angeles, CA: Western Psychological Services.

LeDoux, J. E. (1995). Emotion: Clues from the brain. *Annual Reviews of Psychology*, *46*, 209–235.

Leemis, L. M., & McQueston, J. T. (2008). Univariate distribution relationships. *The American Statistician*, *62*(1), 45–53.

Leevers, H. J., & Harris, P. L. (2000). Counterfactual syllogistic reasoning in normal 4-year-olds, children with learning disabilities, and children with autism. *Journal* *of Experimental Child Psychology*, *76*(1), 64–87. Available from http://dx.doi.org/10.1006/jecp.1999.2542

Loehlin, J. C. (1990). Component analysis versus common factor analysis: A case of disputed authorship. *Multivariate Behavioral Research*, *25*(1), 29–31.

Lombardo, M. V., Barnes, J. L., Wheelwright, S. J., & Baron-Cohen, S. (2007). Self-referential cognition and empathy in autism. *PLoS ONE*, *2*(9), e883. Available from http://dx.doi.org/10.1371/journal.pone.0000883

Lopez, B. R., Lincoln, A. J., Ozonoff, S., & Lai, Z. (2005, Aug). Examining the relationship between executive functions and restricted, repetitive symptoms of autistic disorder. *Journal of Autism and Developmental Disorders*, *35*(4), 445–460. Available from http://dx.doi.org/10.1007/s10803-005-5035-x

Lopez, C., Tchanturia, K., Stahl, D., Booth, R., Holliday, J., & Treasure, J. (2008). An examination of the concept of central coherence in women with anorexia nervosa. *International Journal of Eating Disorders*, *41*(2), 143–152. Available from http://dx.doi.org/10.1002/eat.20478

Lord, C., Rutter, M., DiLavore, P. D., & Risi, S. (2001). *Autism Diagnostic Observation Schedule.* Los Angeles, CA: Western Psychological Services.

Losh, M., & Piven, J. (2007). Social-cognition and the broad autism phenotype: identifying genetically meaningful phenotypes. *Journal of Child Psychology and Psychiatry*, *48*(1), 105–112. Available from http://dx.doi.org/10.1111/j.1469-7610.2006.01594.x

Lovett, A., Forbus, K., & Usher, J. (2007). Analogy with qualitative spatial representations can simulate solving Raven's Progressive Matrices. In *Proceedings of the 29th annual conference of the cognitive science society* (pp. 449–454). Nashville, TN, USA: Cognitive Science Society.

Lund, N. W. (1930). The presence of chiasmus in the New Testament. *The Journal of Religion*, *10*(1), 74–93. Available from http://www.jstor.org/stable/1196953

Luquet, G.-H. (1927/2001). *Le dessin enfantin (children's drawings).* Free Association Books Ltd. (Translated and with an introduction by Alan Costall)

Luria, A. R. (1962/1966). *Higher cortical functions in man.* London, UK: Tavistock Publications. (Authorized translation from the Russian by Basil Haigh. The original Russian was published by Moscow University Press.)

Makinson, D. (2005). *Bridges from classical to nonomonotonic logic.* London: King's College Publications.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco, USA: W. H. Freeman and Company.

Martin-Löf, P. (1984). *Intuitionistic type theory* (Vol. 1). Naples: Bibliopolis. (Notes by Giovanni Sambin of a series of lectures given in Padua, June 1980)

Mason, O., Claridge, G., & Jackson, M. (1995). New scales for the assessment of schizotypy. *Personality and Individual Differences*, *18*(1), 7–13. Available from http://dx.doi.org/10.1016/0191-8869(94)00132-C

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215.

McKenna, F. P. (1984). Measures of field dependence: Cognitive style or cognitive ability. *Journal of Personality and Social Psychology*, *47*(3), 593–603.

Medvedev, G. (1991). *The truth about Chernobyl.* London: I. B. Tauris & Co. Ltd. (Translated from the Russian by Evelyn Rossiter)

Meo, M., Roberts, M. J., & Marucci, F. S. (2007). Element salience as a predictor of item difficulty for Raven's Progressive Matrices. *Intelligence*, *35*(4), 359-368.

Messick, S. (1976). Personality consistencies in cognition and creativity. In S. Messick & Associates (Eds.), *Individuality in learning* (pp. 4–22). Jossey-Bass Publishers.

Miller, J. N., & Ozonoff, S. (2000). The external validity of Asperger disorder: lack of evidence from the domain of neuropsychology. *Journal of Abnormal Psychology*, *109*(2), 227–238.

Miyake, A., Friedman, N. P., Emerson, M., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. Available from http://dx.doi.org/10.1007/s10803-005-0040-7

Moutier, S., Angeard, N., & Houdé, O. (2002). Deductive reasoning and matching-bias inhibition training: Evidence from a debiasing paradigm. *Thinking and Reasoning*, *8*(3), 205–224.

National Security Agency. (1968). Mokusatsu: One word, two lessons. *National Security Agency Technical Journal*, *XIII*(4), 95–100. Available from http://www.nsa.gov/public/pdf/mokusatsu.pdf

Nettle, D. (2007). Empathizing and systemizing: What are they, and what do they contribute to our understanding of psychological sex differences? *British Journal of Psychology*, *98*, 237–255.

Newstead, S. E. (1995). Gricean implicatures and syllogistic reasoning. *Journal of Memory and Language*, *34*, 644–664.

Newstead, S. E. (2003). Can natural language semantics explain syllogistic reasoning? *Cognition*, *90*, 193–199.

Newstead, S. E., & Griggs, R. A. (1983). Drawing inferences from quantified statements: a study of the square of opposition. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 535–546.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *The Quarterly Journal of Experimental Psychology*, *57A*(1), 33–60.

Norman, D. A., & Shallice, T. (1980). *Attention to action: Willed and automatic control of behavior* (Technical report No. 99). La Jolla, USA: Center for Human Information Processing, University of California.

Noveck, I. A., Guelminger, R., Georgieff, N., & Labruyere, N. (2007). What autism can reveal about *Every . . . not* sentences. *Journal of Semantics*, *24*(1), 73–90. Available from http://jos.oxfordjournals.org/cgi/content/abstract/24/1/73

Noveck, I. A., & Reboul, A. (2008). Experimental pragmatics: a gricean turn in the study of language. *Trends in Cognitive Sciences*. Available from http://dx.doi.org/10.1016/j.tics.2008.07.009

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: the probabilistic approach to human reasoning.* Oxford University Press.

Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J., & Sloman, S. A.

(2007). The probability of causal conditionals. *Cognitive Psychology*, *54*(1), 62–97.

Ozonoff, S., Goodlin-Jones, B. L., & Solomon, M. (2005). Evidence-based assessment of autism spectrum disorders in children and adolescents. *Joumal of Clinical Child and Adolescent Psychology*, *34*(3), 523–540. Available from http://dx.doi.org/10.1207/s15374424jccp3403_8

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991, Nov). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *Journal of Child Psychology and Psychiatry*, *32*(7), 1081–1105.

Ozonoff, S., Rogers, S. J., & Pennington, B. F. (1991). Asperger's syndrome: evidence of an empirical distinction from high-functioning autism. *Journal of Child Psychology and Psychiatry*, *32*(7), 1107–1122.

Ozonoff, S., South, M., & Miller, J. N. (2000). DSM-IV-defined Asperger Syndrome: Cognitive, behavioral and early history differentiation from High-Functioning Autism. *Autism*, *4*(1), 29–46.

Paris, J. B. (1995). *The uncertain reasoner's companion: A mathematical perspective.* Cambridge, England: Cambridge University Press.

Peirce, C. S. (1880). On the algebra of logic. *American Journal of Mathematics*, *3*(1), 15–57. Available from http://links.jstor.org/sici?sici=0002-9327%28188003%293%3A1%3C15%3AOTAOL%3E2.0.CO%3B2-6

Peyton Jones, S., Augustsson, L., Barton, D., Boutel, B., Burton, W., Fraser, S., et al. (1999). *Haskell 98—A non-strict, purely functional language.* Available from http://www.haskell.org/onlinereport.

Pfeifer, N., & Kleiter, G. D. (In press). Framing human inference by coherence based probability logic. *Journal of Applied Logic*.

Pijnacker, J., Geurts, B., Lambalgen, M. van, Buitelaar, J., Kan, C., & Hagoort, P. (In press). Conditional reasoning in high-functioning adults with autism. *Neuropsychologia*.

Piven, J. (2001). The broad autism phenotype: a complementary strategy for molecular genetic studies of autism. *American Journal of Medical Genetics (Neuropsychiatric Genetics)*, *105*(1), 34–35.

Piven, J., Palmer, P., Jacobi, D., Childress, D., & Arndt, S. (1997). Broader autism phenotype: Evidence from a family history study of multiple-incidence autism families. *The American Journal of Psychiatry*, *154*(2), 185–190.

Politzer, G. (1986). Laws of language use and formal logic. *Journal of Psycholinguistic Research*, *15*(1), 47–92. Available from http://dx.doi.org/10.1007/BF01067391

Politzer, G. (2004). Some precursors of current theories of syllogistic reasoning. In *Psychology of reasoning: theoretical and historical perspectives* (pp. 214–240). Psychology Press.

Politzer, G., Henst, J.-B. V. der, Luche, C. D., & Noveck, I. A. (2006). The interpretation of classically quantified sentences: A set-theoretic approach. *Cognitive Science*, *30*(4), 691–723.

Politzer, G., & Mercier, H. (In press). Solving categorical syllogisms with singular premises. *Thinking & Reasoning*.

Prado, J., Kaliuzhna, M., Cheylus, A., & Noveck, I. A. (2008). Overcoming perceptual features in logical reasoning: An event-related potentials study. *Neuropsychologia*. Available from http://dx.doi.org/10.1016/j.neuropsychologia.2008.04.017

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.

Quine, W. v. O. (1966). *Methods of logic* (2nd ed.). Routledge & Kegan Paul. (Reprinted with corrections)

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from http://www.R-project.org

Ranta, A. (1994). *Type-theoretical grammar.* Oxford University Press.

Raven, J. C. (1936). *Mental tests used in genetic studies: The performance of related individuals on tests mainly eductive and mainly reproductive.* Unpublished MSc Thesis, University of London.

Raven, J. C., Raven, J. E., & Court, J. H. (1998). Section 4: Advanced progressive matrices. In *Manual for Raven's progressive matrices and vocabulary scales.* Oxford, UK: Oxford Psychologists Press.

Rawlings, D., & Locarnini, A. (2008). Dimensional schizotypy, autism, and unusual word associations in artists and scientists. *Journal of Research in Personality*, *42*(2), 465–471.

Rijmen, F., & Boeck, P. D. (2003). A latent class model for individual differences in the interpretation of conditionals. *Psychological Research*, *67*(3), 219–231. Available from http://www.springerlink.com/content/npxjwxhd0fma34cj/

Ring, H., Woodbury-Smith, M., Watson, P., Wheelwright, S., & Baron-Cohen, S. (2008). Clinical heterogeneity among people with high functioning autism spectrum conditions: evidence favouring a continuous severity gradient. *Behavioral and Brain Functions*, *4*(1), 11. Available from http://dx.doi.org/10.1186/1744-9081-4-11

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking.* Cambridge, MA, USA: The MIT Press.

Rips, L. J., & Conrad, F. G. (1989). Folk psychology of mental activities. *Psychological Review*, *96*(2), 187–207.

Roberts, M. J. (2005). Expanding the universe of categorical syllogisms: A challenge for reasoning researchers. *Behavior Research Methods*, *37*(4), 560–580.

Robinson, R. (1929). The contrast between inference and perception. *The Philosophical Review*, *38*(3), 246–257.

Ronald, A., Happé, F., Bolton, P., Butcher, L. M., Price, T. S., Wheelwright, S., et al. (2006). Genetic heterogeneity between the three components of the autism spectrum: A twin study. *The Journal of the American Academy of Child and Adolescent Psychiatry*, *45*(6), 691–699.

Rooij, I. van. (2008). The tractable cognition thesis. *Cognitive Science*, *32*(6), 939–984.

Rumain, B., Connell, J., & Braine, M. D. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. *Developmental Psychology*, *19*(4), 471–481.

Rumsey, J. M. (1985). Conceptual problem-solving in highly verbal, nonretarded autistic men. *Journal of Autism and Developmental Disorders*, *15*(1), 23–36.

Russell, T. A., Tchanturia, K., Rahman, Q., & Schmidt, U. (2007). Sex differences in theory of mind: A male advantage on happé's 'cartoon' task. *Cognition & Emotion*, *21*(7), 1554–1564.

Ryan, R. S., & Schooler, J. W. (1999). Whom do words hurt? Individual differences in susceptibility to verbal overshadowing. *Applied Cognitive Psychology*, *12*(7), S105–

S125.

Schaeken, W., & d'Ydewalle, G. (1996). Must reasoning. *Acta Psychologica*, *92*(2), 209–220. Available from http://dx.doi.org/10.1016/0001-6918(95)00008-9

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382. Available from http://www.jstor.org/stable/413107

Schmidt, G. L., Kimel, L. K., Winterrowd, E., Pennington, B. F., Hepburn, S. L., & Rojas, D. C. (2008). Impairments in phonological processing and nonverbal intellectual function in parents of children with autism. *Journal of Clinical and Experimental Neuropsychology*, *30*(5), 557–567. Available from http://dx.doi.org/10.1080/13803390701551225

Schmidt, J. R., & Thompson, V. A. (2008). 'At least one' problem with 'some' formal reasoning paradigms. *Memory & Cognition*, *36*(1), 217–229.

Schnall, S. (1999). Life as the problem: Karl Duncker's context. In *From past to future: Papers on the history of psychology* (Vol. 1, pp. 13–28).

Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). E-Prime user's guide. [Computer software manual]. Pittsburgh, PA, USA.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36–71.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. Available from http://www.jstor.org/stable/2958889

Scott, F. J., & Baron-Cohen, S. (1996). Logical, analogical, and psychological reasoning in autism: A test of the Cosmides theory. *Development and Psychopathology*, *8*(1), 235–245.

Scott, F. J., Baron-Cohen, S., & Leslie, A. (1999). 'if pigs could fly': A test of counterfactual reasoning and pretence in children with autism. *British Journal of Developmental Psychology*, *17*, 349–362.

Sellen, J. L., Oaksford, M., & Gray, N. S. (2005). Schizotypy and conditional reasoning. *Schizophrenia Bulletin*, *31*(1), 105–116. Available from http://dx.doi.org/10.1093/schbul/sbi012

Semple, D., Smyth, R., Burns, J., Darjee, R., & McIntosh, A. (2005). *Oxford handbook of* *psychiatry.* Oxford, UK: Oxford University Press.

Shah, A., & Frith, U. (1983). An islet of ability in autistic children: A research note. *Journal of Child Psychology and Psychiatry*, *24*(4), 613–620. Available from http://dx.doi.org/10.1111/j.1469-7610.1983.tb00137.x

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *298*(1089), 199–209.

Sieg, W., & Byrnes, J. (1998). Normal natural deduction proofs (in classical logic). *Studia Logica*, *60*, 67–106.

Sloane, N. J. A. (2007). *The on-line encyclopedia of integer sequences.* Available from www.research.att.com/~njas/sequences/

Slugoski, B. R., & Wilson, A. E. (1998). Contribution of conversation skills to the production of judgmental errors. *European Journal of Social Psychology*, *28*, 575–601.

Smedslund, J. (1970). Circular relation between understanding and logic. *Scandinavian Journal of Psychology*, *11*, 217–219.

Southgate, L., Tchanturia, K., & Treasure, J. (In press). Information processing bias in anorexia nervosa. *Psychiatry Research*. Available from http://dx.doi.org/10.1016/j.psychres.2007.07.017

Spearman, C. (1904). 'General Intelligence,' objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292. Available from http://links.jstor.org/sici?sici=0002-9556%28190404%2915%3A2%3C201%3A%22IODAM%3E2.0.CO%3B2-9

Sperber, D., & Girotto, V. (2002). Use or misuse of the selection task? Rejoinder to Fiddick, Cosmides and Tooby. *Cognition*, *85*(3), 277-290.

Spiker, D., Lotspeich, L. J., Dimiceli, S., Myers, R. M., & Risch, N. (2002). Behavioral phenotypic variation in autism multiplex families: evidence for a continuous severity gradient. *American Journal of Medical Genetics*, *114*(2), 129–136.

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

Stapleton, G., Masthoff, J., Flower, J., Fish, A., & Southern, J. (2007). Automated theorem proving in Euler diagram systems. *Journal of Automated Reasoning*, *39*(4), 431–470.

Stenning, K. (2002). *Seeing reason.* Oxford University Press.

Stenning, K., & Cox, R. (2006). Reconnecting interpretation to reasoning through individual differences. *Quarterly Journal of Experimental Psychology*, *59*(8), 1454–1483.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, *19*(1), 97–140.

Stenning, K., & van Lambalgen, M. (2004). A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning. *Cognitive Science*, *28*(4), 481–529.

Stenning, K., & van Lambalgen, M. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science*, *29*, 919–960.

Stenning, K., & van Lambalgen, M. (2007). Logic in the study of psychiatric disorders: executive function and rule-following. *Topoi*, *26*(1), 97–114.

Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science.* Cambridge, Massachusetts, USA: MIT Press.

Stenning, K., & Yule, P. (1997). Image and language in human reasoning: a syllogistic illustration. *Cognitive Psychology*, *34*(2), 109–159.

Sternberg, R. J. (2000). The ability is not general, and neither are the conclusions. *Behavioral and Brain Sciences*, *23*, 697698.

Störring, G. (1908). Experimentelle Untersuchungen über einfache Schlussprozesse. *Archiv für die Gesamte Psychologie*, *11*, 1–127.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

Tärnlund, S.-Å. (1977). Horn clause computability. *BIT Numerical Mathematics*, *17*(2), 215–226.

Tarski, A., & Vaught, R. L. (1956). Arithmetical extensions of relational systems. *Compositio Mathematica*, *13*, 81–102.

The Independent. (2008, October 23). Watchdog rejects yacht probe request. *The Independent*. Available from http://www.independent.co.uk/news/uk/politics/cameron-slams-pm-over-osborne-row-970933.html

The Observer. (2007, July 22). The Observer and autism: a clarification. *The Observer*. Available from http://www.guardian.co.uk/society/2007/jul/22/health.medicineandhealth

Thompson, D. G. (1878). Intuition and inference. *Mind*, *3*(12), 468–479. Available from http://www.jstor.org/stable/2246752

Tsakanikos, E., & Reed, P. (2003). Visuo-spatial processing and dimensions of schizotypy: figure-ground segregation as a function of psychotic-like features. *Personality and Individual Differences*, *35*, 703–712. Available from http://dx.doi.org/10.1016/S0191-8869(02)00263-5

Tversky, A., & Kahneman, D. (1982). Judgement under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), (pp. 84–100). Cambridge University Press.

van Lambalgen, M., & Hamm, F. (2005). *The proper treatment of events.* Blackwell Publishing.

van Lambalgen, M., & Smid, H. (2004). Reasoning patterns in autism: rules and exceptions. In L. Perez Miranda & J. Larrazabal (Eds.), *Proceedings eighth international colloquium on cognitive science.* Kluwer.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*(1), 1–28.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th, Ed.). New York, USA: Springer.

Voracek, M., & Dressler, S. G. (2006). Lack of correlation between digit ratio (2D:4D) and Baron-Cohen's 'Reading the Mind in the Eyes' test, empathy, systemising, and autismspectrum quotients in a general population sample. *Personality and Individual Differences*, *41*(8), 1481–1491.

Wakabayashi, A., Baron-Cohen, S., Uchiyama, T., Yoshida, Y., Kuroda, M., & Wheelwright, S. (2007). Empathizing and systemizing in adults with and without autism spectrum conditions: Cross-cultural stability. *Journal of Autism and Developmental Disorders*, *37*(10), 1823–1832. Available from http://dx.doi.org/10.1007/s10803-006-0316-6

Wakabayashi, A., Tojo, Y., Baron-Cohen, S., & Wheelwright, S. (2004, Apr). The Autism-Spectrum Quotient (AQ) Japanese version: evidence from high-functioning clinical group and normal adults. *Shinrigaku Kenkyu*, *75*(1), 78–84.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Penguin Books.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273–281.

Westerståhl, D. (1989). Aristotelian syllogisms and generalized quantifiers. *Studia Logica*, *48*(4), 577–585.

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., et al. (2006, Mar). Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Research*, *1079*(1), 47–56. Available from http://dx.doi.org/10.1016/j.brainres.2006.01.012

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Wing, L. (1981). Asperger's syndrome: a clinical account. *Psychological Medicine*, *11*(1), 115–129.

Wing, L. (1988). The continuum of autistic characteristics. In E. Schopler & G. B. Mesibov (Eds.), *Diagnosis and assessment in autism* (pp. 91–111). New York, USA: Plenum Press.

Witkin, H. A. (1950). Individual differences in ease of perception of embedded figures. *Journal of Personality*, *19*(1), 1–15.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. (1962). *Psychological differentiation: Studies of development.* John Wiley and Sons.

Wittgenstein, L. (1953/2001). *Philosophical investigations.* Oxford, UK: Blackwell Publishing. (Translated by G. E. M. Anscombe)

Wong, D., Maybery, M., Bishop, D. V. M., Maley, A., & Hallmayer, J. (2006). Profiles of executive function in parents and siblings of individuals with autism spectrum disorders. *Genes, Brain and Behavior*, *5*(8), 561–576. Available from http://dx.doi.org/10.1111/j.1601-183X.2005.00199.x

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger Syndrome using the AQ: a preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, *35*(3), 331–335.

World Health Organization. (2007). *ICD-10: International statistical classification of diseases and related health problems* (10th Revision ed.). Geneva, Switzerland: World Health Organization.

Zelazo, P. D., Jacques, S., Burack, J. A., & Frye, D. (2002). The relation between theory of mind and rule use: evidence from persons with autism-spectrum disorders. *Infant and Child Development*, *11*(2), 171–195.

Zhang, L. fang, & Sternberg, R. J. (2006). *The nature of intellectual styles.* Lawrence Erlbaum Associates.