Empirical and computational investigations of the relationship between intelligence and development: mental-age matching studies of cognitive variability in the normal range

Frank Daniel Baughman
School of Psychology
Birkbeck, University of London

Submitted for the degree of Doctor of Philosophy of the University of London
2009

Declaration

I hereby declare that this thesis is my own work. Where other sources of information have been used, they have been acknowledged.

Frank D. Baughman

Acknowledgements

I owe my thanks to a number of people. Firstly, I thank the Heads and teachers of the schools for allowing me to carry out this research. I am especially grateful to all the parents and guardians for giving their permission and the children who took part. I would like to thank my supervisors, Michael Thomas and Denis Mareschal, to whom I owe my sincerest thanks. To Michael, in particular, for his patience and encouragement, and for never tiring in his attempt to support my intellectual growth – thank you for inspiring me to be a better scientist. Thanks to Dagmara Annaz, for introducing me to the Developmental Neurocognition Laboratory and to all its members: Dagmara Annaz, Michael Dash, Tessa Dekker, Roberto Filippi, Neil Forrester, Themis Karaminis, Annette Karmiloff-Smith, Victoria Knowland, Fabrizio Pizzioli, Harry Purser, Martin Redington, Fiona Richardson, Angelica Ronald, Michael Thomas and Sam Wass. Thanks also to Nic Ruh and Rob Leech for their encouragement and support. To my family (on both sides of the Atlantic), thank you for all the support and encouragement over the years. I am especially grateful to my parents for teaching me the value of perseverance – in the end, I was the little blue engine that *did*. Thanks to my sister Mara and also Shiva and Nathan for reminding me that relaxation is also an important part of life and to Rosemary, Inka, Rob, Delisha and Cuba for making my visits away from home more endurable. A special thanks to my three wonderful children – Jonas, Isla and Alexander – who gave up many hours of their time to help pilot and test the computer-based tasks. Finally, to my wife Natalie, I owe my thanks more than to anyone else. Without Natalie this thesis would never have happened. It is thanks to her that I started on this journey and it is thanks to her efforts and sacrifice that I was able to complete it.

Abstract

Much research on *cognitive development* and *intelligence* proceeds independently, as if these two forms of cognitive variability were themselves independent. However, theoretical proposals of the mechanisms underlying each type of variability often coincide (e.g., both forms are explained by differences in inhibition; or explained by differences in speed of processing). Building on the work of Spitz (1982), this thesis examines the question of whether cognitive development and intelligence lie on the same dimension by employing a *mental-age matching design*. Groups of younger high ability children (YHA) and older low ability children (OLA) were recruited from the normal population in such a way that the overall mental ages (MAs) of the two groups were matched (as assessed by an intelligence test – the British Abilities Scales II [BAS]; Elliot et al., 1997) while the difference in chronological ages of the two groups was maximised. This design was repeated at two age groups, Primary School (N=40, range 6 to 10 years of age) and Secondary School (N=35, range 12 to 16 years of age). The hypothesis was as follows: if cognitive development and intelligence lie on the same dimension, matching for overall ability in groups with divergent chronological ages should nevertheless lead to identical performance in the two groups. The hypothesis was tested using two empirical methods and two analytical designs. The first empirical method investigated whether the overall mental ages of the matched YHA and OLA groups were comprised of different patterns of performance on the sub-tests of the BAS. The second empirical method investigated whether, using more sensitive experimental tasks, differences could be observed in the underlying cognitive processes of the younger and older groups. Five computer-based tasks were employed: Stroop task (selective attention), semantic priming in a lexical decision task (word recognition), two Piagetian tasks (Conservation and Balance scale), and the Tower of London reasoning task. The first analytical design took sub-groups of the YHA and OLA groups who were exactly matched for overall MA (Primary: n=14 per group, MA-matched at 8.2 years, CA difference of 4.0 years; Secondary: N=16 and 19, matched at 13.8 years, CA difference of 2.2 years). It then used Analysis of Variance to compare group performance. The second analytical design computed the disparity between mental age and chronological age (CA) for each individual (generating positive values for

YHA and negative values for OLA). This disparity was then used as a predictor in an Analysis-of-Covariance design. The ANOVA method revealed almost identical cognitive profiles for younger more able and older less able children at both Primary and Secondary level. Only faster speed of processing for the older children at Primary level generated a marked difference (1 of 24 sub-test comparisons). These results contrasted with the findings of Spitz (1982). The more sensitive ANCOVA method revealed that both MA-CA disparity and age group (YHA vs. OLA) explained variance in behaviour, in most cases explaining independent contributions. At a broad scale, the more sensitive experimental tasks yielded a similar pattern, supporting the idea that cognitive development and intelligence are largely similar types of variability. However, at a finer scale differences did emerge: (1) particularly at Primary level, the OLA group tended to respond more quickly; (2) one task (Conservation) suggested an advantageous role of greater experience in both speed and accuracy; (3) both the Balance scale and Tower of London reasoning tasks indicated advantages for greater ability, possibly linked to superior performance in combining multiple dimensions of information, and in constructing sub-goals, respectively. Development and intelligence may therefore be largely overlapping but not co-extensive forms of variation. The final section of the thesis presents a series of computational modelling studies within a dynamical systems framework (van der Maas et al., 2006). These simulations explored in greater detail the architectural conditions under which even and uneven cognitive profiles emerge across development.

Table of Contents

List of Figures

List of Tables

List of Equations

This thesis contains approximately 80,000 words.

Chapter 1 Introduction

Research question

What is the relationship between *intelligence* and *cognitive development*? Is being more intelligent like having 'a little more' development? Or, are the two things different? This question has significance within both practical (e.g., educational) and theoretical settings. But, firstly, what do intelligence and cognitive development *mean,* and how have the approaches taken to their study differed?

Background

The study of intelligence and cognitive development has been treated as separate forms of cognitive variability with different questions guiding the development of each approach. Within intelligence research, the emphasis is primarily on quantifying the abilities of large numbers of individuals of the same age (hence the term psychometrics is often used). This approach has adopted an interest in *within-age* individual differences. By contrast, for those studying cognitive development the emphasis has been on understanding the processes underlying age-related changes in ability. That is, within this approach explaining how *between-age* differences emerge is of central interest.

One recent example that reflects the importance of the research question stated earlier comes from a secondary school in Portsmouth, United Kingdom. The school in question was the very first school in the country to take the commonplace practice of streaming[1] and apply it to classes of children, regardless of their actual age (Ungoed-Thomas, 2005). That is, the curriculum was re-designed so that in each of the core subjects (i.e., English, mathematics, science and history) classes were taught to children of a variety of ages, but who shared similar overall levels of cognitive ability. This shift in teaching strategy was based on the premise that "there is no real difference in the way these children [i.e., children of different ages, but with similar

---

[1] 'Streaming' describes the practice of teaching children of similar abilities in separate groups. In Primary Schools, this is often achieved by grouping children of similar abilities at separate tables, within the same classroom. At Secondary Schools, pupils are typically divided amongst different classrooms corresponding to low, medium and high ability.

scores on measures of cognitive ability] learn" (Cheryl Heron, Head Teacher; personal communication, 2006). Thus, within the so-called 'Ability not Age' approach, and within any one of the classes teaching core subjects, one might find 12-year-olds through to 16-year-olds being taught together[2]. Proponents of the view that younger, more able and older, less able children are equivalent in the way they learn might thus be expected to argue that differences in intelligence and differences in cognitive development represent variability on the *same dimension*.

However, empirical support for this argument is lacking. In fact, it contradicts directly one crucial piece of research carried out by Merrill (1924; and later reviewed by Spitz, 1982). Merrill adopted a mental age matching design whereby younger and older children were matched on an overall ability level, using the Stanford-Binet test (1916). She showed that groups of younger, more able and older, less able children arrived at their overall similar level of ability via *different strengths and weaknesses* on the variety of subtests comprising the Stanford-Binet. Spitz (1982) argued that Merrill's data therefore indicate intelligence and cognitive development are not the same thing.

However, more recently within the literature on intelligence and cognitive development one can find a degree of overlap in the descriptions of mechanisms proposed to underlie variability. Specifically, this relates to a number of accounts positing single mechanisms. It follows, then, that if the mechanisms that are described within these separate accounts are the same, differences in intelligence and differences in cognitive development may be explained by variability on a single dimension. The role that these mechanisms play in producing variability is currently unclear and thus the precise nature of the relationship between intelligence and cognitive development remains open.

---

[2] In actuality, the widest range of ages found was generally around 12-14 years old, but 12-16 years old represented the widest possible range according to the school's new system.

The aims of the thesis

The primary aim of this thesis is to clarify the relationship between intelligence and cognitive development. Towards this aim, I identify three main steps. Firstly, I use a standardised test of ability to examine the relative strengths and weaknesses of groups of younger and older children who are matched on overall ability. In an extension of Merrill's (1924) design using more recent tests of intellectual ability, I examine whether the same differential profiles of scores on subtests emerge in these groups. This is carried out both on Primary School children (respective mean ages: 6.4 vs. 10.4 years) and Secondary School children (respective mean ages: 12.9 vs. 15.1 years). By comparing the profiles of groups of younger and older children who are of the same mental age (MA), we may see whether for example, low intelligence may be compensated for by greater chronological age, or conversely, whether lower chronological age (CA) may be compensated for by greater intelligence.

Secondly, I administer a battery of computer-based on-line cognitive tasks to the same groups of children. These tasks comprise the Stroop task, a primed lexical decision task, conservation of number and volume tasks, the balance scale task and the Tower of London task. The purpose of using on-line tasks is to provide a more refined analysis of the potential mechanisms underlying intelligence and cognitive development and therefore of the relationship between them.

Finally, this thesis argues for the importance of taking seriously the developmental process in explaining mechanisms underlying variability. Specifically, I argue it is of critical importance to specify the type of cognitive framework in which candidate mechanisms are assumed to be instantiated in. This is because the same mechanisms may have radically different outcomes depending on which architecture it operates within. Thus, in the third step I implement a series of computational models each of which characterise a different cognitive architecture. I then examine how each of the architectures contributes to development by testing the effect of damage to specific parameters relating to possible mechanisms.

It is also worth articulating what this thesis is *not* about. It is not concerned with arguments relating to the many different definitions of intelligence, or the specific number of intelligences that may exist. Nor is it about a range of other issues relating to race, or to sex differences in intelligence (see for example, Deary, 2000;

Gardner, 1999; Gottfredson, 2005; Jensen, 2002; Sternberg, Grigorenko, & Kidd, 2005; Sternberg, 2000). I make the assumption that standardised tests of intelligence are adequate in allowing children to be matched on their level of ability. That is, test scores are taken to be robust measures of current ability (and reliable predictors of academic success), but not as measures of some form of 'general intelligence' (see Chapter 2; see also Anastasi, 1992).

The structure of the thesis

The remainder of this thesis is organised into a total of 7 parts. In Part 1, Chapter 2 begins by briefly setting out some of the key definitions of 'intelligence' and how the term is used within this thesis. It then presents some of the main literature relevant to our examination of the relationship between intelligence and cognitive development.

In Part 2, Chapter 3 begins by detailing the organisation of the standardised test used in this thesis – the British Ability Scales second edition (BAS II; Elliot, Smith, & McCulloch, 1997) – and describes how scores from this test will be used to derive measures of intellectual ability. This forms the first step in replicating Merrill's (1924) findings. Chapter 3 then presents the general methodology that applies to the experimental and computer-based tasks. This avoids unnecessary repetition within each of the chapters that describe subsequent empirical studies employing the same methodology. For example, the reader should assume that participant details are identical unless stated otherwise in the relevant chapter.

In Part 3, Chapter 4 presents the results of the replication of Merrill's study (1924) using groups of younger and older children who are matched on MA. I compare their performance on the range of subtests that comprise the BAS II (Elliot, et al., 1997).

In Part 4, Chapter 5 focuses on the results of the Stroop task and Chapter 6 presents the results of the primed lexical decision task. These chapters are presented together within Part 4 due to their shared focus on measuring automaticity and spreading activation, inhibition and speed of processing.

Part 5 presents the results of the three remaining computer-based tasks that each require a greater degree of cognitive control, for instance in problem solving and working memory. This part comprises results from the conservation of number and liquid tasks, the balance scale task and the Tower of London task in Chapters 7, 8 and 9, respectively.

In Part 6, Chapter 10 presents work on a range of computational studies that further explores the question of domain-general versus domain-specific development in relation to uneven cognitive profiles. I argue that modelling is important in providing causal frameworks to explore the origins of uneven cognitive profiles in the development of multiple component systems. In this chapter I implement a series of dynamical systems models within which mechanisms may be tested. I apply these models to a particular debate within the literature in which, on one hand, development is characterised as a graded and highly interactive processes whilst, on the other hand, it is argued that markedly uneven profiles result from variability operating in independently developing modules which do not affect other components of the cognitive system. Accordingly, the models I develop portray various cognitive architectures (ranging from fully modular to fully distributed) and assess the effects of damage on cognitive profiles to one specific ‘knowledge module’ over development. We will see that this modelling has particular relevance to debates recurring in a related field of cognitive variability, namely developmental disorders (Thomas & Karmiloff-Smith, 2002; Thomas, Karmiloff-Smith, & Goswami, 2002).

Finally, in Part 7, Chapter 11 offers a final discussion, integrating the main research findings from the experimental and computational work.

Part 1

Chapter 2 Literature review

The aims of this chapter are to introduce the core concepts that relate to the main experimental design. These concepts appeal to the literature on *intelligence* (and the mechanisms that are assumed to explain intelligence) and the literature on *development* (and the mechanisms that explain developmental change). These are both vast literatures and they cannot be reviewed here. However, in the later relevant chapters, there will be more detailed reviews of the literature that pertains to the individual tasks that we will consider. Instead, within this chapter I will define these terms and focus on the cognitive level in order to consider several candidate cognitive mechanisms. This leads us to consider several cognitive accounts in which a degree of overlap exists in the descriptions of *single mechanisms* proposed to underlie variability in intelligence and cognitive development. For example, we review mechanistic accounts of *speed of processing*, *inhibition*, *capacity* and *complexity*. The similarity of descriptions given to these mechanisms raises the possibility that a single mechanism may account for differences in ability both within-ages *and* between-ages. We will then consider this possibility in the light of a key piece of research by Merrill (1924; reviewed by Spitz, 1982) and show how Merrill's findings offer the view that these are separate forms of cognitive variability. The chapter ends with a description of the empirical methodology chosen to shed light on relationship between intelligence and cognitive development. First, let us briefly review some common definitions of the term 'intelligence' and how individual differences and cognitive development approaches have traditionally assessed ability.

Conceptions of intelligence

After approximately a century of modern research on intelligence, there is still no single, agreed definition of what intelligence actually *is* (e.g., Sternberg, 2005). Indeed, Boring's early tautological statement that "*Intelligence is what is measured by intelligence tests*" (Boring, 1923, p35) appears just as valid today. Boring's statement highlighted the ambiguity of determining exactly which abilities tests of intelligence might actually tap, and thus the difficulty of using such tests to examine theory. The view that Piaget offered was that "*intelligence is assimilation to the extent that it incorporates all the given data of experience within its framework.*" (Piaget, 1950, p6). Anastasi proposed, "*The term [intelligence] denotes that combination of abilities required for survival and advancement within a particular culture*" (Anastasi, 1992, p612). And a task force commissioned by the American Psychological Association to determine what was known and unknown in intelligence research offered the following definition: "*Individuals differ from one another in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought*" (Neisser, et al., 1996, p77). With the exception of Boring, the statements above appear to share a common theme. For example, they illustrate what Sternberg has asserted to be the most general and accepted view of intelligence, that it involves "*...a person's ability to adapt to the environment and to learn from experience*" (Sternberg, 2005, p189).

More controversially, Anastasi has also stated that "*intelligence is not a single, unitary ability, but rather a composite of several functions*" (Anastasi, 1992, p614). Additionally, Gardner goes further in his proposal that intelligence is "*...a neural mechanism or computational system which is genetically programmed to be activated or 'triggered' by certain kinds of internally or externally presented information.*" (Gardner, 1983, p64). More recently, Gardner has argued that approximately nine 'module-like' intelligences can be discerned (his list includes linguistic, logical or mathematical, spatial, interpersonal, intrapersonal, musical, bodily-kinaesthetic, naturalistic and existential. See Gardner, 1999), although he admits the strength of evidence for each of these different intelligences varies (see Gardner, 1999, p47).

In contrast, Jensen has argued, "*Intelligence is a general factor that runs through all types of performance*" (Jensen, 1998, p38). Much of Jensen's efforts have focused on establishing the notion that a single general factor exists as a substantive property of the brain, and that variability in intelligence is largely determined by genetic rather than environmental factors (see e.g., Jensen, 2002).

Each of these views remains actively debated (see e.g., Sternberg, et al., 2005). The job at hand within this thesis is not to attempt to substantiate one or other definition. Indeed, as Sternberg wrote, "*...there seem to be almost as many* *definitions of intelligence as there were experts asked to define it*" (Sternberg, 1987, p376). Instead, within this thesis, intelligence is viewed according to the broader sense offered by Sternberg (2005). That is, intelligence involves a person's ability to adapt and learn. Next, I briefly outline how ability is assessed within individual differences and cognitive development approaches.

Assessing ability

The individual differences approach

Modern efforts to measure within-age individual differences are typically traced back to the work of Sir Francis Galton (1822-1911). Galton provided the basis for the individual differences approach with the view that various observable traits were correlated with intelligence. Basic sensory and motor abilities, height, respiratory power and "robust frame" (body size) were considered key indicators of intellectual ability (Galton, 1892). Later, Alfred Binet (1857-1911) and Theodore Simon (1873-1961) worked on devising a test that allowed children's abilities to be assessed relative to the performance of a normative sample of same-age peers. Thus, it provided a system for distinguishing lower ability and higher ability children from children of average ability (Binet & Simon, 1905). Charles Spearman's (1863-1945) pioneering use of mathematical models in producing *factor analysis* consolidated further the construct of general intelligence. Spearman showed that by analysing the correlations between task scores, a single variable emerged that accounted for a considerable proportion of variance in the Binet-Simon scales. These findings led Spearman to propose his two-factor theory of intelligence: that an individual's performance on tests was influenced most by a single, general factor that he called the *general factor* of intelligence (or *g*) and then by secondary 'specific factors' that he considered unique to the particular test (Spearman, 1904). Spearman interpreted the differences on tests between people as the result of different concentrations of "mental energy" – an actual property of the brain that influenced all cognitive behaviour and which explained the outcome of a single factor (for a review see e.g., Williams, et al., 2003).

Assessing children's ability relative to their peers was only one outcome of the Binet-Simon scale. Later, the test also provided the construct of *mental age* (MA) - the idea that on each test and for any given age, an age-appropriate performance score could be determined. Thus, a 10-year-old child performing at the level expected of the average 8-year-old could be said to have a MA of 8. The notion of MA subsequently led Stern (1912) to develop the intelligence quotient (IQ) – an index of ability based on an individual's MA as a function of their chronological age (CA). See Equation 1.

$$IQ = \frac{MA}{CA} \times 100$$

Equation 1. Indexing ability from MA and CA: Intelligence quotient (IQ; Stern, 1912)

In addition to the Stanford-Binet[1] (1916), numerous other intelligence tests have been developed. Examples of these include, the British Ability Scale Second Edition (BAS II; Elliot, et al., 1997), the Wechsler Intelligence Scale for Children Fourth Edition (WISC-IV; Wechsler, 2004), the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1997) and the Kaufman Assessment Battery for Children (KABC; Kaufman & Kaufman, 2004)[2]. Each of these tests shares a degree of similarity in the kinds of tasks that have been designed to tap various intellectual abilities. They also provide the means of obtaining a conceptual measure of one's overall intellectual ability. The notion that one's intellectual ability can be characterised by a unitary score has fuelled attempts aimed at isolating a single, substantive property of the brain responsible for differences in intelligence. Some of the roots to this view can be seen, for example, within the early work of Galton (1892) and Spearman (1904). However, this notion also figures prominently in many contemporary accounts (see e.g., Haier, White, & Alkire, 2003). Another key similarity across these tests is that they are comprised mostly of *off-line* tasks. The significance of this is discussed more fully towards the end of the chapter, when I contrast off-line and on-line tasks. However, briefly, one critical difference between off- and on-line tasks is that off-line tasks measure the *end stage* of performance, whereas on-line tasks aim to tap

---

[1] The Binet-Simon test formed the basis for the Stanford-Binet (1916) test.

[2] Note, in Chapter 3, I describe in some detail the intelligence test chosen for use in this research project - this is the British Ability Scale Second Edition (Elliot, et al., 1997). For a review of the other tests, see for example (Loewenthal, 2001).

processes *during* information processing. As such, off-line tasks may mask important differences in underlying processing that are revealed by on-line tasks.

The cognitive development approach

Jean Piaget (1896-1980) is well known for his theory of cognitive development in which he mapped out the growth of reasoning from infancy to adulthood (Piaget, 1954, 1972). While testing children for Simon and Binet in Paris, Piaget observed that children of different ages made different types of errors on various tests. This led him to reason that children differed in a qualitative sense in the kinds of thinking of which they were capable (1954). Piaget formulated a theory in which he argued that intellectual growth depended on units of knowledge he called 'schemas' (Piaget, 1972). For Piaget, intelligence comprised more than a process of acquiring greater numbers of schemas. It was the nature of the organisation of schemas within the cognitive system that produced changes in the quality of children's thinking (Piaget, 1972). During early development, Piaget argued that these schemas were limited to processing simple sensory and motor experiences and, that transitions to more complex stages of thinking involved a hierarchical organisation of these schemas into knowledge structures that enabled more complex representations of the world. While a number of biological and environmental factors were held to interact in providing change, Piaget viewed the child to be instrumental in bringing about change through its own contact with the world. Thus, the account of development that Piaget came to formulate was a *constructivist* one (Piaget, 1954, 1972). Piaget proposed that the fundamental mechanism underlying cognitive change was a process he called *equilibration*. This was described as a process of balancing internal mental change brought about by *adaptation* to the environment. According to Piaget, adaptation was further comprised of the processes of *assimilation* (whereby new information is dealt with by existing schemas, or knowledge structures) and *accommodation* (a process in which new schemas are developed to deal with information that existing schemas could not deal with).

Piaget developed a series of tasks to challenge children's thinking at different points in their intellectual growth, and to demonstrate the differences in their reasoning at each stage of development. Two tasks that I draw attention to here are the conservation task (Piaget, 1972) and the balance scale task (Inhelder & Piaget, 1958). Both these tasks have been studied extensively with respect to children's cognitive development and later on, in Chapters 7 and Chapter 8, I detail the results of children's performance using these tasks. In those chapters I provide a review of the relevant literature. However, briefly, I highlight the main elements of these tasks. In the conservation of liquid task, children are presented with two identical containers that hold equal volumes of water. They watch as the water from one container is poured into a third different container. If, for example, the third container is narrower and taller than the original container, the level of water appears *higher* than that in the original. Piaget found that younger, *non-conserving* children would show an over-reliance on the perceptual features that would lead them to conclude that the third container held more water. In the balance scale task, the aim is to test children's understanding of weights and distances. For example, weights are placed at different positions on a balance beam and children are asked to predict what consequence follows when two supporting blocks are removed. In both these tasks, the general idea is that information from *more than one dimension* has to be integrated. In the case of the conservation tasks, it is information relating to the perceptual and conceptual properties of objects that must be integrated. In the case of the balance scale task, it is information relating to weight and distance that must be integrated.

Piaget's theory of cognitive development has been widely acknowledged for its role in describing the qualitative changes that occur in children's thinking, as they grow older. However, it remains contentious as to precisely what mechanisms underlie cognitive change (see e.g., Johnson, Munakata, & Gilmore, 2002, for recent review). Furthermore, while on one hand there has been broad support for the kinds of developmental stages that Piaget proposed, on the other hand there have also been a number of important criticisms. These criticisms include claims that children may be competent in a given domain *earlier* than Piaget assumed (e.g., Spelke, Mehler, & Franck, 1995), that children do not progress through the discrete stages Piaget described (Fischer & Silvern, 1985), that children often do not show uniform abilities across different domains and that not all children (or even adults) appear to reach the final stage of formal operations (see e.g., Flavell, 1982). Other criticisms have focused on logical issues. For example, Fodor (1983) argued that it is not possible for a system to be capable of determining new, more complex information, if the structure for processing the more complex information is not already in place. (For a fuller review of several key criticisms of Piaget's theory see e.g., Brainerd, 1974;

Flavell, 1982). These criticisms have been used to argue support for various nativist accounts – the position that the cognitive abilities that we come to possess, as adults, are relatively independent *modules*, present from birth and a product of evolution (see e.g., Fodor, 1983). We return to the subject of modularity in Chapter 10, where I explore in more depth some of the ways modularity accounts have influenced theories of intelligence[3]. However, one key criticism of Piaget's account concerns the feasibility of learning mechanisms in bringing about the marked differences in cognitive ability seen between children of different ages. In sum, although the theory Piaget formed was comprehensive in scope, it lacked explicit descriptions of *how* exactly the mechanisms he proposed were so instrumental in producing change. As Feldman and Fowler wrote, "*Equilibration-like processes are almost certainly necessary for developmental change to occur in all regions, but they must be supplemented with additional processes at every point along the continuum*" (Feldman & Fowler, 1997, p.202).

But, what *are* the mechanisms that underlie cognitive variability? What property does one need to have *more* of to be more intelligent? And could the same property that makes one more intelligent also be responsible for the general increase in cognitive ability as one gets older?

Variability in intelligence and cognitive development: Levels of description

A number of approaches have emerged aimed at investigating the causal mechanisms underlying cognitive variability. These may be distinguished broadly according to the *level of description* they utilise in their explanations of variability. Next, I briefly distinguish between the *gene*, *brain* and *cognitive* levels of descriptions. While space precludes a full review of each, in the following sections I will highlight some of the key findings relating to the study of differences between intelligence and cognitive development.

---

[3] In Chapter 10, I attempt to contrast several differing views as to how cognition could be organised (i.e., its functional architecture) by implementing a series of dynamical systems model of development. This approach is useful in that it offers a simplified framework within which questions relating to variability in development can be tested.

The gene level

Behavioural genetics uses studies of twins and adopted children to test assumptions concerning the influence of genes versus the environment on a given behaviour. For example, the effects of genes have been claimed to account for up to 50% of the variance on tests of intelligence (see e.g., Plomin, DeFries, & McClearn, 2008). On the study of cognitive development, genes have been reported to have stronger influences at later points in development than earlier points (Plomin, et al., 1997).

Molecular genetics uses two key types of analysis (referred to as linkage analysis and association analysis) to identify the genes that contribute to a given behaviour (see e.g., Posthuma & de Geus, 2006). On the study of intelligence, a variety of genes have been reported to have significant associations to intelligence. Posthuma and de Geus (2006) summarise the following genes linked to intelligence: the ALDH5A1 gene; the APOE gene; the CTSD gene; the DRD2 gene; the CBS gene; the BDNF gene; and, the COMT gene. However, Posthuma et al. (2006) note that most of the associations involving genes and intelligence have proven difficult to replicate. Thus, it remains to be seen how the effects of genes can be interpreted into causal cognitive accounts.

The brain level

A number of magnetic resonance imaging (MRI) studies have reported correlations between measures of intelligence and total brain volume (see e.g., Andreasen, et al., 1993) and white matter volume and grey matter volume separately (e.g., Haier, et al., 2004; Posthuma, et al., 2002). The underlying logic is that if measures of brain volume correspond to differences in computational capacity, then they might provide candidate explanations for differences in intelligence. These studies show, however, that only a small proportion of the variance on tests of intelligence is accounted for. For example, total brain volume has been reported to account for around 16% of the variance on tests of intelligence (see e.g., Haier, et al., 2004). The research on the relationship between grey matter volume and white matter volume to intelligence is less conclusive, with no clear picture emerging on whether grey matter or white matter offers the best predictor of intelligence (see e.g., Gignac, et al., 2003).

Functional magnetic resonance imaging (fMRI) has been used to pinpoint the neural markers, or correlates of intelligence. These are typically achieved by investigating patterns of activation that accompany a cognitive task. While some researchers have reported that measures of intelligence are associated with increased activation levels in frontal cortical regions (see e.g., Geake & Hansen, 2005), causal accounts that relate these to cognitive change are missing.

The study of event related potentials (ERP) has demonstrated in detail the brain's response to specific types of stimuli which have enabled researchers to understand how different types of information are represented and processed throughout the brain. However, again these methods typically yield measures of correlations between neural conductivity and measures of intelligence (e.g., Deary & Caryl, 1997) and it remains to be clarified how these data should translate to cognitive theories of variability.

On the study of cognitive development, MRI studies have showed that growth of the prefrontal cortical area is positively correlated with increases on tests of intelligence (see e.g., Gray & Thompson, 2004). Similar associations have also been shown between changes in white matter over development and scores of intelligence (Mabbott, et al., 2006). Finally, studies using fMRI have showed that over the course of development the brain comes to exhibit a degree of localisation for some functions (Szameitat, et al., 2002).

The cognitive level

A number of candidate cognitive mechanisms aimed at explaining variability underlying intelligence and cognitive development can be found within the literature. Here, I focus on *speed of processing*, *inhibition*, *capacity* and *complexity* as potential mechanisms. As we shall see, the similarity of descriptions given to these mechanisms raises the possibility that a single mechanism may account for observed differences in ability – both within-ages and between-ages. This has been referred to as the *unidimensional hypothesis* (Davis & Anderson, 1999; Davis & Anderson, 2001).

Speed of processing

A key feature within a number of theories aimed at explaining the causes of individual differences and cognitive development concerns the view that cognitive systems operate under a general constraint of the speed (or rate) at which information is processed (see e.g., Anderson & Miller, 1998; Burns, Nettelbeck, & Cooper, 1999;

Jensen, 1993; Wright, et al., 2001). Additionally, if information within the system is assumed to be subject to stable decay, then it seems reasonable to assume that the system able to process information at a higher speed would have the advantage over other systems. For example, given a specific problem and a limited time-period, a faster system might be able to produce a greater number of representations and thus allow for more complex reasoning than a slower system. Under this view, slower systems (being more susceptible to the effects of decay) would fare poorly due to fewer complete representations (see e.g., Neubauer, et al., 2004). Speed of processing has been proposed to be an important factor in both within-age differences (e.g, Anderson, 1992; Nettelbeck, 1987) and between-age differences (e.g., Hale, 1990; Kail, 1996). Support for this mechanism has been claimed on the basis of studies showing correlations between measures of IQ and measures of response times on cognitive tasks and with neurophysiological data (e.g., Anderson, 1992; Jensen, 1985; Posthuma, et al., 2002). Additionally, within some theories of cognitive development, one can also find speed of processing described as a causal mechanism underlying intellectual change. For example, it has been proposed that cognitive speed changes with age and therefore gives rise to a greater capacity for thought (Wellman & Gelman, 1992). This view finds support from studies that have examined age-related changes in performance on Inspection Time tasks[4] (see e.g., Nettelbeck, 1987).

Inhibition

The ability to *inhibit* irrelevant information is widely assumed to be one of a core set of mechanisms that comprise a range of high-level cognitive behaviours referred to as *executive functions* (see e.g., Miyake, et al., 2000). Differences in the ability to inhibit has been proposed to account for within-age differences in intelligence (e.g., Dempster, 1991) and between-age differences in intelligence (Houdé, 2000).

---

[4] Inspection Time tasks provide measures of the time between the onset of a target and the onset of a subsequent masking figure. Specifically, it is the minimum time between these events that is needed for an individual to respond at a pre-specified level of accuracy.

Capacity

One cognitive process where capacity has been a particular focus is working memory. The relationship of working memory capacity to intelligence is portrayed in many different ways. Some theorists argue they are the same construct (e.g., Jensen, 1998), while others claim there is almost no relationship between them (see e.g., Deary, 2000). However, many view both within-age and between-age differences in intelligence as related in some fashion to differences in working memory capacity (Ackerman, Beier, & Boyle, 2005). However, the role of capacity as a cognitive mechanism underlying this variability is not clear.

Complexity

A number of theorists have proposed, consistent with Piaget's view, that development comprises more than just acquiring additional knowledge (see e.g., Mareschal, et al., 2007). Specifically, it is argued that changes in one's ability to reason involves a process of building more complex representations of the world. Halford (1999) has similarly argued that increases in complexity of representations provide greater processing capacity and has claimed that complexity of representations can account for both within-age and between age variability in intelligence. For example, Halford proposes that very young children (up to around 1 year-old) are limited in their ability to form complex representations and thus can only process information relating to one dimension (as evidenced by empirical data on tasks such as the A-not-B task). As children grow older and become better at forming and manipulating complex representations, they also demonstrate better performance on more complex tasks involving a greater number of dimensions. At 5 years of age, Halford (1999) estimates that children are capable of processing up to three dimensions in parallel. Thus, at this age he argues children begin to show understanding of transitive logic (e.g., if A = B and B = C, then A also equals C).

Empirical investigations into the relationship between intelligence and cognitive development

The literature shows there have been few empirical attempts to explicitly test the relationship between intelligence and cognitive development. One notable exception, to which we next turn, comes from an early study carried out by Maud Merrill

(1924). This study focuses on the question of whether intelligence and cognitive development can compensate for each other in task performance.

Merrill's (1924) study

During her work with Terman, producing the Stanford-Binet revision (1916), Merrill became interested in how children of similar abilities but of different ages compared on tests of ability. By establishing groups who were matched on their overall intellectual ability and allowing development (i.e., age) to vary somewhat, Merrill provided an analysis of the relationship between intelligence and development (1924). She applied the Stanford-Binet test (1916) to assess the abilities of 100 younger children and 350 older children with the aim of matching younger and older groups on mental age (MA). The younger group's mean chronological age (CA) was 5.5 years, the older group's mean CA was 11.9 years and the overall mean MA shared by the two groups was 8.0 years. The groups were compared on their scores across the range of subtests that make up the Stanford-Binet test (1916).

Spitz (1982) redrew the key differences that Merrill found between the younger and older groups. This is duplicated in Figure 2.1 below. The key point to bear in mind while looking at these data is that if differences in intelligence and cognitive development are aspects of the same kind of variability, then we might expect there to be no differences in the performance profiles of these two groups. Age differences would be compensated for by intelligence in the younger group and differences in intelligence would be compensated for by age in the older group.

Figure 2.1 presents comparisons between the younger and older groups on 11 subtests within the Stanford-Binet test (1916). On the Y-axis, tasks are divided according to the age-level for which they were intended (i.e., the categories VII, VIII and IX contain the tests typically suited for 7, 8 and 9 year olds, respectively). On the X-axis, bars represent reliable differences in percentage points between younger and older children on a given task (i.e., the performance score for one group, subtracted from the performance score of the other group). Bars to the left of zero represent the tasks where the younger children were reliably better than the older children. Bars to the right of zero represent the tasks where the older were reliably better than the younger group. The figure shows that younger children reliably outperformed the older children on six subtests. These were: (1) *Comprehension* (showing knowledge of appropriate behaviour in various social situations); (2) *Similarities* (explaining the relationship and similarities between two objects); (3) *Superior Definitions* (providing definitions of words); (4) *Weights* (ordering objects of the same size but different weights in ascending order); (5) *Rhymes* (finding as many words as possible within one minute that rhyme with a given word); (6) and *60 words* (recalling as many words as possible in 3 minutes from a list of 60 items).

Figure 2.1 also shows five tasks in which the older children did reliably better than the younger children. These were: (1) *Fingers* (without counting, the child tells the experimenter how many fingers he/she has on one, then both hands together); (2) *Counting Backwards* (the child counts backwards from 20 to 0 in 40 seconds, making no more than one error); (3) *Change* (how much change should be expected from a given purchase); (4) *3 words* (using three words provided by the experimenter to produce a sentence); (5) and *Date* (knowing and correctly stating the date).

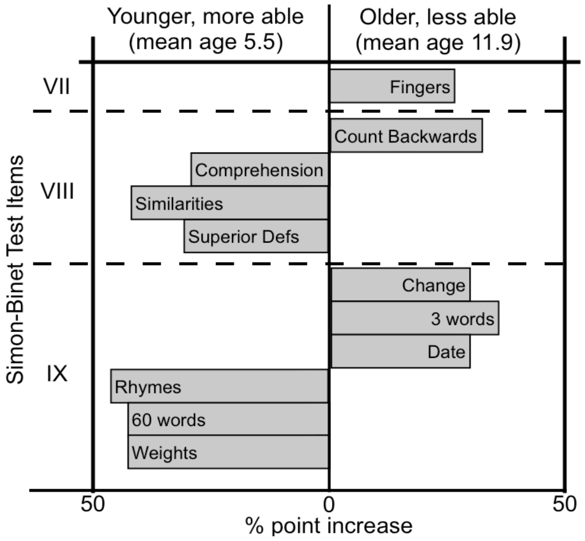


Figure 2.1. Merrill's (1924) findings of the relative strengths and weaknesses of younger more able and older less able groups of children on the Stanford-Binet (1916) test. Groups were matched on a MA of 8.0 years. Bars on the left hand-side represent the subtests on which the younger group were superior to the older group (P<.05, chi square). Bars on the right-hand side represents subtests on which the older group were superior to the younger group (P<.05, chi square).

While both groups had the same overall MA, Figure 2.1 shows that they arrived at this MA through *different relative strengths and weaknesses in abilities*. Spitz (1982) interpreted these data to imply that younger, more able children excel on tasks involving verbal reasoning and abstraction while older, less able children excel at tasks tapping experience, maturation and rote learning. On the basis of this, he argued that it is inaccurate to characterise the two groups as being of the same developmental or the same cognitive level. This point is critical as it suggests that intelligence and development contribute differentially to ability and thus are not the same thing. Key to interpreting this relationship further is the analysis of which mechanisms underlie high (or low) performance on each subtest.

A recent two-dimensional account of variability

Drawing on Spitz's review, Anderson (2001) and Davis and Anderson (1999) argued that unidimensional theories are insufficient in capturing within-age and between-age variability in intelligence. Instead, Anderson proposed that two dimensions are minimally needed to explain these forms of variability. In Anderson's 'minimal cognitive architecture' (1998) two routes to acquiring knowledge are described. The first route includes two processors specialised in verbal and spatial abilities, respectively. Within these, a set of parameters (with values that are normally distributed in the population) are assumed to determine their processing power of each. These processors are intended to explain how a person might develop to be good at verbal and poor at spatial tasks. However, overall behaviour is a result of these processes interacting with a basic processing mechanism. In Anderson's model, it is the speed at which this basic processing mechanism operates that ultimately determines the acquisition of knowledge through this route.

The second route Anderson describes to acquiring knowledge is through domain-specific information-processing *modules*[5], which take one of the three following specific forms: (1) innately pre-specified modules that allow for such abilities as obtaining 3-D representations from 2-D retinal images, and others that serve language acquisition; (2) modules that work to store and retrieve information (Anderson referred to these as "fetch and carry mechanisms in memory"); and, (3) acquired modules of the sort that develop through experience. Anderson suggested that all three types operate automatically and independently of basic processing mechanisms. According to Anderson (1998, 2001), the first route, with its constraint

---

[5] Note: these modules are different to those proposed by Fodor (1983; 2000). In Fodor's account, modules could have any number of a variety of properties (for example, he listed domain-specificity, information encapsulation, obligatory responses, fast, limited accessibility and fixed neural architecture). In Chapter 10, we return to discuss issues relating to modularity in a little more detail.

of speed of processing, influences within-age differences in intelligence. The second route, containing modules that may mature at differential rates, influences between-age changes in intellectual ability.

In Part 2 that follows, I set out the general methodology used within this thesis to replicate Merrill's design. This comprises the use of a standard test of intelligence – the British Ability Scales II (BAS II; Elliot, et al., 1997). Part 3 presents the findings of the BAS II. In Parts 4, 5 and 6 of this thesis outline new empirical and computational approaches. For example, in Part 4, new empirical work will focus on the use of "on-line" tasks to provide more sensitive measures of underlying processes (I describe the distinction between on-line and off-line tasks shortly). In Part 5, the empirical work will focus on the use of cognitive tasks. In Part 6, the computational modelling work will contrast several differing theoretical assumptions on the organisation of the cognitive system, including the extent to which either many or fewer processes contribute to performance.

Extending Merrill's (1924) study in this thesis

On the one hand, the account offered by Spitz suggests the view that intelligence and cognitive development are two *separate dimensions* of cognitive variability. On the other hand, the similarity in the descriptions of the candidate mechanisms suggests that differences in intelligence and cognitive development may be accounted for by variability on the *same dimension*. The paucity of research investigating the relationship of intelligence and cognitive development leaves open the question of what underlies these forms of cognitive variability. This thesis aims to address this shortcoming. I adopt the same design used by Merrill to examine the performance profiles of groups of younger and older children, who are within the normal range, on the subtests comprised within a standard test of intelligence. That is, in the first instance, the aim is to replicate the different patterns of strengths and weaknesses reported by Spitz (1982). Second, I use several additional computer-based tasks to probe for subtler patterns of variability between these groups. These tasks are important in their ability to offer *on-line* measures of processing. Next, I outline some of the advantages that on-line tasks have over off-line tasks.

On-line vs. off-line tasks

The distinction between these types of tasks plays an important role within this thesis. The two types of tasks differ with respect to the points in the information-processing process at which they are held to yield measurements (e.g., Shapiro, Swinney, & Borsky, 1998). For example, off-line tasks are tasks that offer measurements of the *end-point* of a process (e.g., Shapiro, et al., 1998). Examples of off-line tasks include tasks such as word definition, sentence comprehension and spatial memory tasks. In these tasks the number of correct answers (or solutions) is typically used to assess ability. However, a key problem with the sole use of this type of measure is that it offers no discriminability between individuals who achieve the same level of accuracy, but who differ in how they arrived at the same performance level. For example, on many spatial memory tasks (such as the Recall of Designs task within the British Ability Scales II; Elliot, et al., 1997; discussed in Chapter 3), there is no quantitative difference in ability attributed to a child who completes an item slowly while correcting for multiple errors, to the child who completes the same item just as accurately, but without mistakes and more quickly. In this way, off-line tasks can mask subtler patterns in underlying processing (see e.g., Shapiro, et al., 1998; Tyler, et al., 1997). By contrast, on-line tasks seek to uncover the relative contributions of different processes to a specific behaviour. In doing so, it offers the potential for a greater understanding of information processing (see Shapiro, et al., 1998). The distinction between off-line and on-line tasks may be illustrated with an example.

Let us imagine that an experiment was administered to a group of younger (Group A) and older (Group B) children and their response times were recorded. Suppose further that the task allowed discrete measurements of various sub-processing elements (for example, facilitation and interference on the Stroop task). We could thus examine the differences between groups in their performances, both at a global level (i.e., how do the groups compare overall?) and at a local level (i.e., do the groups share similar patterns in the sub-processes?). Figure 2.2 illustrates several possible outcomes in a hypothetical task comprising three sub-processes. The X-axes represent the three sub-processes that were measured on a task, and the Y-axes represent the time taken to carry out those processes. The figure portrays two scenarios in which groups take the same overall time on the task (both groups show overall RT of 3.1secs in panels 1 and 3, respectively) and two scenarios in which the groups differ in the amounts of time taken (Group A shows overall RT of 3.1secs in panels 2 and 4, while Group B shows faster RT of 2.5secs and 1.8secs, in panels 2 and 4, respectively).

These hypothetical data illustrate that the same overall levels of performance may be the consequence of different underlying factors. Specifically, Panel 1 illustrates that identical group performance is a consequence of identical patterns of processing. The two groups are identical in their information processing and ability. Panel 3 also shows identical overall group performance, but in this instance, overall performance is comprised of different strengths and weaknesses on the individual sub-processes[6].

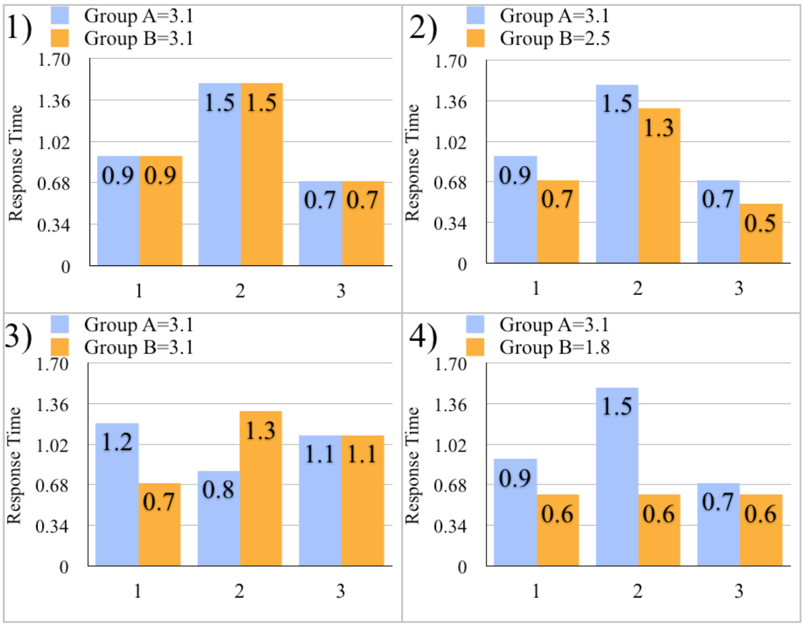


Figure 2.2 Illustrating four possible processing patterns between two groups on an on-line task.

Figure 2.2 also shows that different overall levels of performance do not necessarily imply dissimilar patterns of processing. For example, while Panel 2 portrays absolute differences between Groups A and B, the relative differences within each group are identical. That is, Group B is faster, but it mirrors the pattern of time taken to process the different elements of the task. Finally, Panel 4 portrays an instance in which

---

[6] Note: It is this sort of differential pattern of processing that Spitz (1982) reported between younger and older children of the same mental age. See Chapter 1.

different overall levels of performance are attributable to dissimilar patterns of processing. Here, Group A shows an uneven profile of RT across the individual sub-processes, whereas Group B shows an even profile.

The role of computational models in this thesis

Within this thesis, computational modelling plays two main roles. In the first instance, computational models are used to help elucidate the possible origins of variability in task performance. For example, each of the on-line tasks that were selected (i.e., the Stroop task, the lexical decision task, the conservation of number and liquid tasks, the balance scale task and the Tower of London task), were chosen because they permit an investigation of the mechanisms that drive change using formal methods. That is, for each of these tasks a computational model of normative development already exists. These models are Cohen, Dunbar and McClelland's (1990) model of the Stroop task, Plaut's (1995) model of semantic priming task, Shultz's (1998) model of conservation, McClelland's (1989), model of the balance scale task and Baughman and Cooper's (2007) model of the Tower of London task. Although these models clearly offer the additional potential of examining how mechanisms underlying differences may produce differences found in the experimental data, this is not the primary objective. Instead, within Parts 4 and 5, where the results from the battery of the computer-based cognitive tasks are presented, the corresponding existing models for each task is referred to for the purpose of interpreting the behavioural data.

In the second instance, in Part 6 Chapter 10, computational modelling is used directly to test the causal influence of small changes to the starting state of a cognitive system. I implement a series of dynamical systems models that capture a range of different possible cognitive architectures. I implement versions of the following seven types of cognitive architecture: (1) fully distributed; (2) hemispheric; (3) central processor; (4) bi-directional loop; (5) uni-directional loop; (6) hierarchical; and, (7) modular. I then examine the conditions under which uneven cognitive profiles emerge in development within these architectures.

***

In sum, the focus of this project is to replicate Merrill’s (1924) study and to extend the measures that are used to examine the relationship between intelligence and cognitive development. In the next part, I provide details of the methodology used in the empirical phase of this research. This includes details for the standard test of intelligence that was used and for the computer-based on-line cognitive tasks.

Part 2

Chapter 3 General Methodology

The British Ability Scales II (Elliot, et al., 1997)

The British Ability Scales Second Edition (BAS II; Elliot, et al., 1997) is a standardised pencil-and-paper questionnaire for assessing ability for individuals aged between 2 years and 6 months (2:6) and 17 years 11 months (17:11). The original test items were constructed from a nationwide representative sampling of children's and adolescents' abilities from around the UK between 1993 and 1996 (total sample size 1689). This sample provides the basis for standardising all subsequent data collected on the BAS II (Elliot, et al., 1997). The test comprises a number of subtests within an Early Years battery (for children between 2:6 and 5:11) and a School Age battery (for children and adolescents between 6:0 and 17:11). Each battery is administered to children individually and the results can then be used to assess children's cognitive abilities and educational achievements. Consequently, within the BAS II (Elliot, et al., 1997), test items have been arranged into item sets that are intended to be age-appropriate. For each subtest, normal starting points and expected end points are indicated for various different ages. However, when testing a child suspected to be of lower-than-average ability, the child may be administered items in the set below that usually given for their age. Similarly, a child suspected to be of higher ability may start on a set of items above the set usual for their age. In both scenarios the BAS II (Elliot, et al., 1997) procedure permits the experimenter to either go back to easier problems if the starting point appears too difficult, or skip item sets that are too easy. This offers the potential of speeding up the time taken to administer the tests and keeping the attention of the child focused. Testing ceases on a given task when the child fails a specified number of items. The key aim of this system of testing is to ascertain the level at which an individual performs relatively stably by finding the point at which they begin to fail items regularly. Next, I detail the School Age battery (6:0 and 17:11) which provides the subtest profiles of ability that is used throughout the remainder of this thesis.

British Ability Scales II (Elliot, et al., 1997) School Age battery

The British Ability Scales II (Elliot, et al., 1997) School Age battery (from here on abbreviated as BAS II-SA) is comprised of a set of 6 Core scales, 8 Diagnostic tasks and 3 Achievement tests. Within the 6 Core scales, pairs of subtest scores may be clustered into measures of Verbal Ability, Non-verbal Reasoning Ability and Spatial Ability scores. The three pairs of Core scales are as follows: (1) Word Definitions and Verbal Similarities (verbal cluster); (2) Matrices and Quantitative Reasoning (non-verbal cluster); and (3) Recall of Design and Pattern Construction (spatial cluster). The Diagnostic scales consist of: (1) Recall of Objects (Immediate / Delayed / Spatial); (2) Speed of Information Processing; (3) Recall of Digits Forward; (4) Recognition of Pictures; and (5) Recall of Digits Backward. The Achievement tests comprise: (1) Mathematics; (2) Spelling; and (3) Word Reading tests.

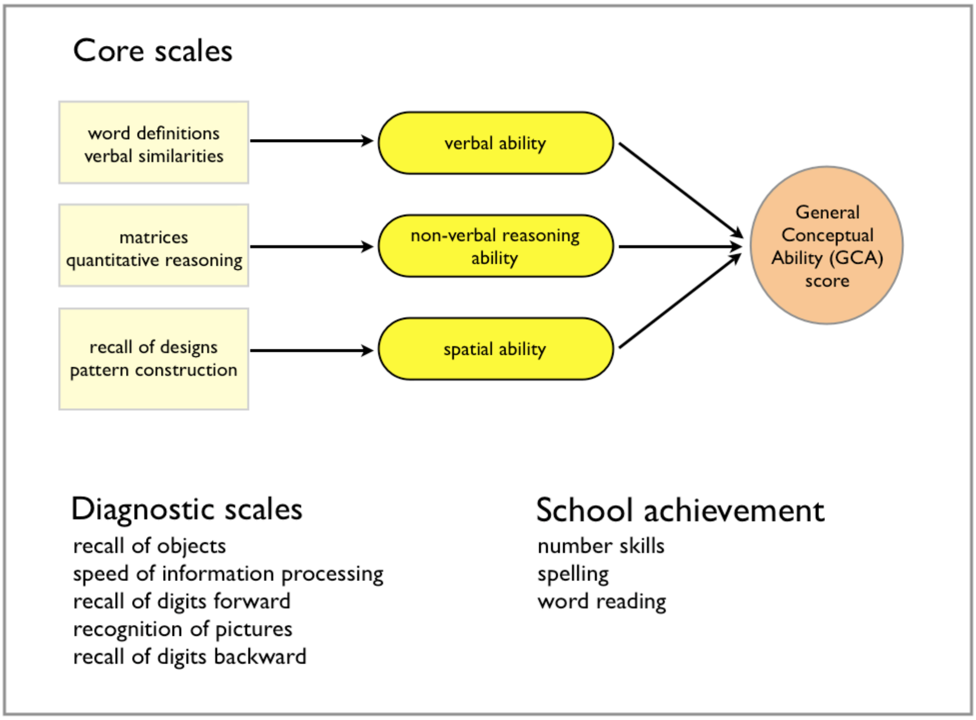


Figure 3.1. The structure of the British Ability Scales II (Elliot, et al., 1997) School Age battery (ages 6:0-17:11)

The scoring of the tests is performed manually, after testing, and involves converting raw scores into Ability scores. The Ability score is directly linked to the number of correct items achieved within a specific range of age-related items. For example, a child scoring 6 correct items within the range of items for ages 8:0-10:11 will obtain a higher Ability score than the child obtaining the same number of correct items within the 5:0-7:11 age range. (This is because the 8:0-10:11 range consist of more challenging items). Ability scores may then be converted into standardised T-scores according to age (with associated percentile ranks), which in turn may be summed to yield a General Conceptual Ability score (GCA). The GCA represents a normalised standardised ability score, analogous to an IQ score.

MA-matching

The structure of the BAS II-SA also permits the use of alternative measures of ability that do not convert participants' scores based on their chronological age. One such measure is the *proportional* ability score. This is a measure of performance that is derived by taking one's actual ability score as a proportion of the highest possible ability score for a given task. For example, the highest possible ability score on the Recall of Designs is 175. A child getting this score would have a proportional ability score of 100%, whereas the child obtaining an ability score of 123 would have a proportional ability score of approximately 70% (i.e., 123/175x100=70.2).

Additionally, the BAS II-SA provides a means to establish the mean age-level at which individuals are performing. This utilises a set of *age-equivalent* scores for each of the core scales representing the level (in terms of age in months) at which an individual is performing. For example, an individual obtaining an ability score of 88 points on Matrices has a corresponding age-equivalent score of 105 months, or 8.75 years. This approach provides another way of establishing mental age matched groups.

Figure 3.2 illustrates a typical developmental trajectory (blue dotted line) closely following an idealised MA-CA relationship of 1:1 (blue line). The normal trajectory is shown flanked by upper and lower bounds (black dashed-lines), representing higher and lower ability respectively. Two hypothetical cases corresponding to children of two different ages are shown. One child aged 5 years old (red square) and another child aged 10 years old (pink triangle) both show the same overall mental age of 7.5 years (green line).

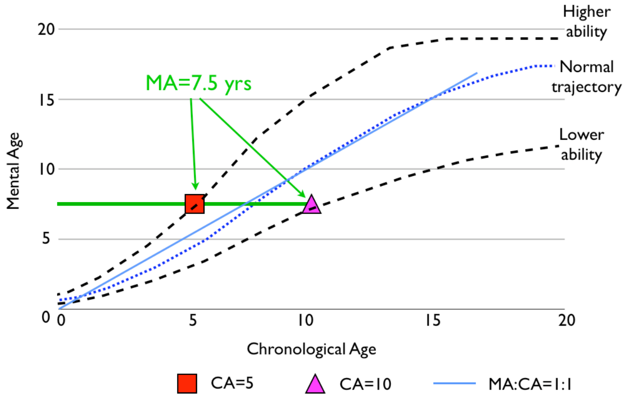


Figure 3.2. Illustrating a the trajectories of a higher ability 5-year-old (red square) and a lower ability 10-year-old (pink triangle) and matching them on a MA of 7.5 years (horizontal green line)

Figure 3.2 illustrates the differences between the mental ages of the younger child and the older child relative to their chronological ages (CA). Unless an individual's mental age is exactly average for their chronological age, contrasting the two age measures will produce a difference (or disparity) score. This is referred to as their *MA-CA disparity score*. Thus, in the example above, the 10 year old has a MA-CA disparity of -2.5 years (7.5–10=-2.5) and the 5 year-old has a MA-CA disparity of +2.5 years (7.5–5=2.5). With this background given to the BAS II-SA, I next present the basic methodology used for the standardised and computer-based tests.

Methodology of standardised tests

Schools

Three schools took part in this research: one Infant school and two Primary Schools in North London and one Secondary School in Portsmouth. The schools were (1) Martin Infant school; (2) Martin Junior school; (3) Highgate Primary School; and (4) Bridgemary Community Sports College. All schools are mixed-gender, community schools. Martin Infant and Junior Schools are located together on one site with separate infrastructures (i.e., separate heads and administrative staff) and cater separately to children aged 3-6 and 7-11. At the time of testing the two schools had 203 and 211 pupils attending, respectively. Highgate Primary School for children aged 3-11 had around 249 pupils in attendance at the time of testing. Bridgemary Community Sports College caters for children aged 11-16. At the time of testing, approximately 1080 children were attending this school. At the time of testing, this school was also unique within the UK because its style of teaching involved forming classes based on ability rather than age group. For example, in Mathematics, classes were comprised of children of different ages (e.g., ranging from 13.0-14.5) but of a similar overall level of ability.

Participants

In the Infant and Primary Schools, names were obtained of the five most able children in each Year 1 class and the names of the five least able children in each Year 5 class. This was achieved using school data of children's performance on school tests and teachers' assessments in class. Children with special educational needs (SEN) classifications were not included in any testing.

A total of 40 Primary School children were recruited and completed all testing. This sample comprised 20 children (7 males, 13 females) in the younger age group (youngest=6.0 years, oldest=6.75) and 20 children (11 males and 9 females) in the older age group (youngest=10.0, oldest=10.75).

The same selection method was applied in the Secondary School sample. However, ability assignment was based on the School's existing measures from the Cognitive Abilities Test, second edition (Thorndike, Hagen, & France, 1986). Due to the fact that classes were taught on overall ability, the ages and abilities of those recruited were not as homogeneous as in the Primary School sample. A greater overlap was obtained in the ages of the two Secondary School groups. Although the ages and the ability levels of the children in this school did not initially provide the same discrete boundaries as in the Primary School sample (i.e., younger vs. older), these groups were obtained based on a median age split. From an initial total of 40 children recruited in the Secondary School sample, 4 children from the younger age group and 1 child from the older age group withdrew from participation. The final Secondary School sample totalled 35 children: 16 children in the younger age group (youngest=11.75, oldest=13.92) and 19 children in the older age group (youngest=14.0, oldest=15.92). There were 8 males and 8 females in the younger age group and 9 males and 10 females in the older age group.

Ethics

Ethical approval for the research was obtained from the School of Psychology, Ethics Committee at Birkbeck, University of London (reference number: 2322). Prior to testing (and in accordance with the Police Act of 1997) a Police check including a Standard Disclosure was obtained for the experimenter. Parents and guardians of all children were sent an information pack about the study, which included a parental consent form. Assent was also obtained from each child.

Materials and Apparatus

The standardised test included items from the British Ability Scales Second Edition (Elliot, et al., 1997). This included all associated peripherals (i.e., test booklets, stopwatch, pencils and paper). The battery of computer-based (PC) experimental tasks required the use of one laptop (used to control task presentation), one touch screen monitor (used to view and respond on each computer-based task), a digital microphone, one pair of Seinhouser stereo headphones, a RS500 button box, one set of external speakers, a set of 12 counters and three measuring cups with water.

Design

The study adopted a quasi-experimental, mixed design. The between-groups factor was ability, yoked to age. These were younger, higher-ability children and older lower-ability children. The dependent variables were (1) performance scores within the British Ability Scales Second Edition (Elliot, et al., 1997), and (2) response time and accuracy data on the PC versions of the Stroop task, the balance scale task, the Tower of London task, a lexical decision task and conservation of number and liquid tasks.

Procedure

Several factors impinged on the time given to testing children in the Primary and Secondary Schools. These included availability of suitable space for testing, sports days and special class events and assemblies. Subsequently, the number of sessions each child took to complete standardised tests varied between 2-3 and on the PC tasks between 2-4. Each session lasted approximately 30-45 minutes. The average completion time was approximately 90 minutes for the standardised tests and 105 minutes for the PC tasks. Within these sessions, the order of the standardised tests remained fixed, with participants completing tasks in the order they appeared in the test booklet (this ensured the appropriate time interval advised for the Recall of Objects Delayed and the Recall of Objects Spatial tasks). The order of PC tasks was randomised and the experimental setup allowed testing to be postponed at the end of each task. These tasks were interlaced with the standardised tests and the majority of children completed all testing within a 1-week period.

All participants were tested individually and under supervision on the BAS II-SA (Elliot, et al., 1997). The standard procedure was applied with the administration of items beginning with sets commensurate with each child's expected ability level. This was then adjusted for, according to the child's responses to initial items (see earlier, this chapter). In the following sections, I describe the task procedures for each of the BAS II-SA components that were administered to the school children. I then describe the procedures used to administer the computer-based tasks given to all participants.

BAS II-School Age: Core Scales

In this section, I present the procedures and examples for each of the six subtests. These are Word Definitions, Verbal Similarities, Matrices, Quantitative Reasoning, Recall of Design and Pattern Construction.

Word Definition

The Word Definition is a verbal task that contributes to the Verbal ability cluster score. Assumed to be a measure of crystallised mental ability (Elliot, et al., 1997), it is intended to measure a child's acquired verbal knowledge. In this task, children are presented orally with a series of single word items and asked to give definitions for each word. The items are initially easy and become progressively more challenging. Children must provide a complete definition (e.g., rather than an example of the category) in order to obtain points on this task. No time limit applies and raw scores are converted into ability scores, based on the child's number of correct definitions within a particular item set and the set they began on. Figure 3.3 below illustrates the task.

<u>Experimenter</u>: "What does jubilant mean?"
<u>Child A</u>: "It means feeling really good, or happy"
<u>Child B</u>: "It's like when I scored a goal, I felt jubilant"

Figure 3.3. Example of Word definition item: Child A provides a correct response. Child B uses the word correctly within a context, but fails to properly define the item

Verbal Similarities

The Verbal Similarities task also contributes to the Verbal ability cluster score and is held to offer a measure of acquired verbal knowledge and reasoning (Elliot, et al., 1997). In this task, children are presented orally with three word items and asked to describe what the items have in common. Children may make a variety of responses but it is assumed that providing the name of the superordinate class for the items (e.g., like that given by Child A in Figure 3.4) indicates a greater verbal reasoning ability than giving a subordinate response (such as the one offered by Child B in Figure 3.4). No time limit is imposed on this task and ability scores are derived from the number of correct responses to items within a set.

Experimenter: “What do ‘Banana’, ‘Apple’, ‘Orange’, have in common? What could you call these things?”
Child A: “They are all fruits”
Child B: “They all have skins”

Figure 3.4. Example of Verbal similarities items: Child A shows knowledge of the super-ordinate category to which the items belong. Child B chooses a subordinate aspect of the relationship between the items.

Matrices

The Matrices task contributes to the Non-verbal reasoning cluster score and is assumed to measure an individual’s ability to infer relations and reason fluidly (Elliot, et al., 1997). The task uses a multiple-choice design in which a matrix of 4 squares (in the easier item sets) or 9 squares (in the more difficult item sets) is presented (see Figure 3.5 below). All but one of the squares contain various designs within which a logical relationship exists. To respond correctly, the child must understand the relations between abstract properties of the items in the matrix and then correctly identify the missing design from an array of alternatives. Ability is assessed by the number of correct responses within a given item set. No time limit applies on the task.

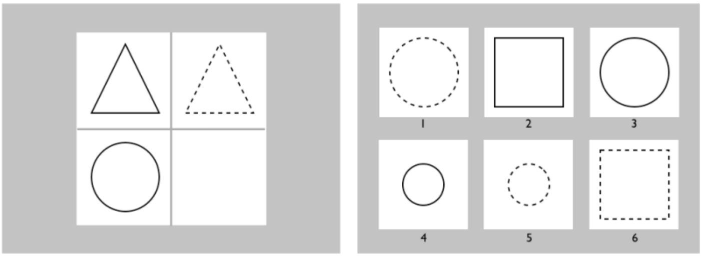


Figure 3.5. An example of a Matrices item. Question: Which figure from the 6 options on the right completes the matrix on the left? Answer: #1 (e.g., based on abstract relation ‘convert shape on left to dotted’, or ‘apply above dotted pattern to circle’).

Quantitative Reasoning

Quantitative Reasoning also contributes to the Non-verbal reasoning cluster score. The task requires the child to deduce the logical relation between pairs of numbers in order to complete one pair where a number is missing. Children are presented with 2-

dimensional pictures of a series of ‘dominoes’ showing numbers in each half. Using the first two dominoes, they must answer verbally which number completes the last domino (see Figure 3.6). No time limit is imposed on this task.

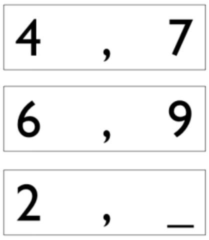


Figure 3.6. An example of a Quantitative reasoning item. Question: What number completes the last domino? Answer: 5 (e.g., based on abstract relation ‘add 3 to the number on the left’).

Recall of Design

The Recall of Designs task is assumed to tap Visuo-spatial abilities of both retention and recall (Elliot, et al., 1997). Scores on this test contribute to the Spatial ability cluster. In this test, children are presented with a series of abstract line drawings of geometric shapes. Each item is presented for 5 seconds and children are then asked to replicate the design by drawing it on a separate sheet of A6 grid paper. Erasers are permitted and no time limit applies to the task. Ability is assessed at a later stage, using acetate overlays, supplied within the BAS II School Age pack (Smith & Traynelis, unpublished) and is determined by the degree to which the child’s drawing matches the relative proportion and configuration of the original design. See Figure 3.7 below.

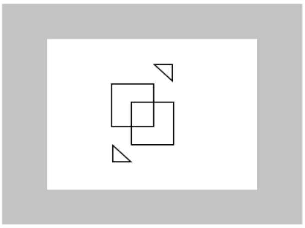


Figure 3.7. Example of Recall of Designs item: Illustrations are presented to the child for 5 seconds. It is then removed from view and the child is asked to replicate the design

Pattern Construction

The Pattern Construction task (Figure 3.8) also contributes to the Spatial ability score. In this task, participants are asked to produce a 3-dimensional match to a 2-dimensional target pattern, using a number of identical plastic cubes. Each cube has faces that are: (i) all yellow; (ii) all black; or (iii) part yellow and part black. Children begin on easier items containing fewer cubes and are given more complex patterns to construct using more cubes as they progress. On each trial, the target pattern remains in view and children are timed on their efforts to produce the match. Problem sets start out with 2 cubes in the easiest problems and increase to 9 cubes for the most difficult items. Ability on this task is accuracy yoked to time, with higher points being awarded for faster completion. Within each item set, as items get harder, the maximum time allowed for completion increases.

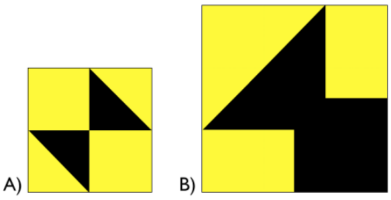


Figure 3.8. Examples of Pattern construction task: the child is asked to manipulate a set of blocks to match an illustration. Problems start out easier e.g., using 4 blocks to match the pattern in Figure 3.8A and get progressively harder, e.g., using 9 blocks in Figure 3.8B.

BAS II-School Age: Diagnostic tasks

This section describes the Diagnostic tasks administered to children. These comprised: Recall of Objects (consisting of three types: immediate verbal recall, delayed verbal recall and delayed spatial recall), Speed of Processing, Digits Forward and Digits Backward.

Recall of Objects

Immediate Verbal Recall

This task is assumed to measure immediate verbal and visual memory (Elliot, et al., 1997). Participants are given three trials in which they are shown an A4 stimulus card with 20 pictures of coloured objects (see Figure 3.9). In trial 1, the card is shown for 40secs at which point the card is removed from view and the participant is given 60secs to verbally recall as many of the objects from the card as possible. Trials 2 and 3 follow the same format, however, the exposure time is 20secs and recall time 40secs in each.

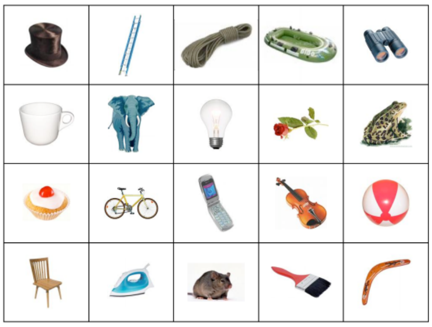


Figure 3.9. Mock example of Recall of Objects task: A4 stimulus card with 20 objects is shown for a specified time to participants and they are asked to recall as many objects as possible within a given time period.

Children's ability scores are derived directly from the sum of correctly recalled items within the time limit on the three trials (double-mentions are counted only once).

Delayed Verbal Recall

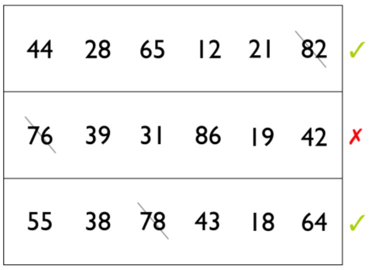
After an interval of around 15-20 minutes (see Elliot, et al., 1997), the Delayed Verbal Recall task was administered. This task measures memory retention of the previously viewed stimuli. Participants are not exposed to the actual stimulus, but are asked to recall as many of the original objects from the A4 stimulus card as possible within 60secs. In this task, raw scores (i.e., the number of correctly recalled items) do not have an associated ability score. Thus, in order to compare performance across each of the subtests, the proportion correct is used.

Spatial Recall

The delayed Spatial Recall task is assumed to rely upon the integration of verbal and spatial working memory (Elliot, et al., 1997). Following the Delayed Verbal Recall task, participants are given a spatial task that consists of a set of 20 pre-ordered, mixed picture cards. Each card has a picture identical to one of the tiles on the original A4 stimulus card. Without viewing the stimulus card again, participants are instructed to place each card in its correct position on a response grid. A maximum time of 4 minutes is given for this task. Again, as raw scores do not convert to ability scores, the proportion correct was used for comparisons between children.

Speed of Processing

The Speed of Processing task is intended to offer a measure of how quickly children process information and perform simple mental operations (Elliot, et al., 1997). Participants are presented with a series of single-sided pages containing rows of numbers and are instructed to draw a line through the highest number on each row (see Figure 3.10). Accuracy and speed are yoked to performance scores for this task with the highest scores awarded for fastest completion with the fewest errors.



44	28	65	12	21	~~82~~	✓
~~76~~	39	31	86	19	42	✗
55	38	~~78~~	43	18	64	✓

Figure 3.10. Example of Speed of processing task: Participants are presented with a series of randomised numbers and asked to draw a line through the highest number in each row. In this example, the response in row 2 was incorrect.

Digits Forward

The Digits Forward task is intended to measure short-term auditory memory (Elliot, et al., 1997). In this task, participants are read a sequence of numbers at a rate of 2 per second and asked to repeat them in the same order. The presentation followed procedures described in the BAS II (i.e., it was reproduced live, not recorded and presented in a monotone voice). Better performance on this task is assumed to be due to a child's increased ability to temporarily store verbal information. Sequences start out short (e.g., 2 digits) and get increasingly longer (maximum of 9 digits). See Figure 3.11.

Experimenter: 6-4	Child: 6-4	✓
Experimenter: 5-3-7-8	Child: 5-3-7-8	✓
Experimenter: 3-6-7-2-1-2	Child: 3-6-7-2-1-2	✓
Experimenter: 2-3-7-5-6-8-2-3	Child: 2-3-5-6-8-3-2	✗

Figure 3.11. Example of Digits Forward task: children start out repeating shorter sequences, in forward order. If correctly repeated, sequences get longer. The fourth trial contains an error.

Digits Backward

The Digits Backward task is assumed to measure a child's ability to transform items retained in short-term auditory memory into working memory (Elliot, et al., 1997). It presents the child with a similar set up to the Digits Forward. However, in this task participants are required to repeat in *reverse order* a sequence of numbers. The procedure is in all other ways identical to the Digits Forward task. That is, numbers are read out loud at 2 per second in a monotone voice and children are initially presented with shorter, easier sequences (minimum 2 digits) before progressing to more difficult, longer sequences (maximum 7 digits). See Figure 3.12.

Experimenter: 4-3	Child: 3-4	✓
Experimenter: 2-5-6-7	Child: 7-6-5-2	✓
Experimenter: 4-2-3-1-6	Child: 6-1-3-2-4	✓
Experimenter: 8-1-2-8-7-3-4	Child: 4-3-7-8-2-1-8	✓

Figure 3.12. Example of Digits Backward task: children start out repeating shorter sequences, in reverse order. If correctly repeated, sequences get longer.

Next, I briefly outline the methods that applied to the computer-based tasks. More detailed methodologies can be found in the chapters that deal with the results of the individual tasks.

Methodology of computer-based cognitive tasks

Computer-based (PC) versions of six cognitive tasks were administered. These consisted of PC versions of (1) the Stroop task, (2) a primed lexical decision task, (3) conservation of number and liquid tasks, (4) the balance scale task and (5) the Tower of London task. Brief literature reviews and detailed methods for each task are included in the relevant chapters. The following details were common to all computer-based tasks.

Participants sat directly in front of a touchscreen monitor at a comfortable arm's length. The experimenter sat to one side of, or adjacent to, the participant and controlled the presentation of tasks from a laptop connected to the touchscreen (see Figure 3.13). Each task began with an instruction video. In the case of the conservation of number and liquid tasks, there was also a short physical demonstration outlining the task requirements. Participants then performed a test trial to demonstrate their understanding of the task and were given the opportunity to hear the instructions again, and/or repeat the test trial. This was then followed immediately by the actual trials.

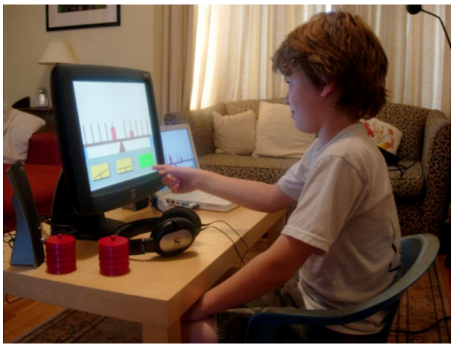


Figure 3.13. Piloting the tasks: a child responding on the Balance Scale task

With the exception of the Lexical Decision task (written in SuperLab™ and donated for testing by Rob Leech) all PC tasks were implemented as programs in MATLAB™ by Frank Baughman. The visual presentation toolbox Psychtoolbox 3.0 was used to improve onscreen performance (Kleiner, Brainard, & Pelli, 2007). Each program was designed to output data capturing the problem being administered, participants' response times (with millisecond level accuracy), overall solution times, errors made (e.g., if three buttons were presented and one wrong option was chosen, what that option corresponded to) and proportion of problems correct.

***

With the details of the methodologies used for the standardised tests and computer-based tasks complete, we can now turn to the results of BAS II-SA and the creation of ability matched groups at Primary and Secondary level.

Part 3

Chapter 4 Results of standardised tests

Results on the Core scales and Diagnostic tests from the BAS II School Age (BAS II-SA;Elliot, et al., 1997) are presented here in three sections that comprise (1) the data from the Primary School; (2) the data from the Secondary School; and (3) combined data from both school levels. The Primary School and Secondary School sections are each divided into parts that offer categorical and continuous comparisons of performance on subtests. In the categorical analysis, multivariate analysis of variance (MANOVA) is used to compare the performances of mental age (MA) matched groups on the sub-tests within the Core scales and Diagnostic tests. The MANOVA addresses whether the sub-test profiles differ between groups.

In the Primary School data, these MA-matched groups are achieved by filtering out a number of children from younger and older groups to leave a reduced dataset. In the Secondary School data, filtering is not performed since initial groups share overall MA scores. In the continuous comparisons, multivariate analysis of covariance (MANCOVA) is applied separately to full samples of Primary and Secondary School data. This analysis follows a developmental trajectories approach[1] using MA-CA disparity as the covariate and Group as the between-subjects factor. This allows us to assess the main effects of Group, main effects of MA-CA disparity and interactions between Group and MA-CA.

Finally, in the last section, data from the Primary and Secondary School levels are combined and MANCOVA are once again applied. MA-CA disparity remains the covariate. However, in this section I test age effects in two ways using both Group (younger vs. older) and School Level (Primary vs. Secondary) as the between-subjects factors. This chapter ends with a discussion of how these findings compare to those Spitz (1982) presented.

---

[1] The developmental trajectory approach allows the performances of groups to be compared in terms of both their gradients and their intercepts of lines-of-best-fit (for details of this approach see Thomas, Annaz, et al., 2009, and accompanying worksheet).

Primary School results

The Core scales and Diagnostic tests from the BAS II-SA battery were administered to a total of 40 Primary School children of two age groups. This comprised 20 children (7 males, 13 females) in the younger age group (youngest=6.0, oldest=6.75) and 20 children (11 males and 9 females) in the older age group (youngest=10.0, oldest=10.75). See Chapter 3, General Methodology for details.

Data for the Primary School children were complete on all BAS II-SA measures, with the exception of two cases in the older children group. For one child, data for Verbal Similarities, Speed of Processing, Recall of Digits Forward and Recall of Digits Backward were missing. Verbal Similarities was replaced with the mean of their scores on the other five Core scales. The other missing data were not replaced. For a second child data for all Diagnostic tests were missing. These data were also not replaced. Where data were not replaced, these children were omitted from comparisons.

The overall CAs and MAs for the younger and older groups are presented in Figure 4.1. For the CA data, a one-way ANOVA revealed the differences of 47.4 months between younger and older age groups to be reliable (F(1,39)=2850.65, p<.001, η^2=.993). Similarly, the difference of 20.1 months between the MAs of younger and older groups seen in Figure 4.1 was also found to be significant (F(1,39)=14.77, p<.001, η^2=.280). The computed IQ scores[2] for these groups were thus: Younger IQ=119; and Older IQ=90.

---

[2] Younger IQ=91.7/77.2 x 100=118.7. Older IQ=111.8/124.6 x 100=89.7

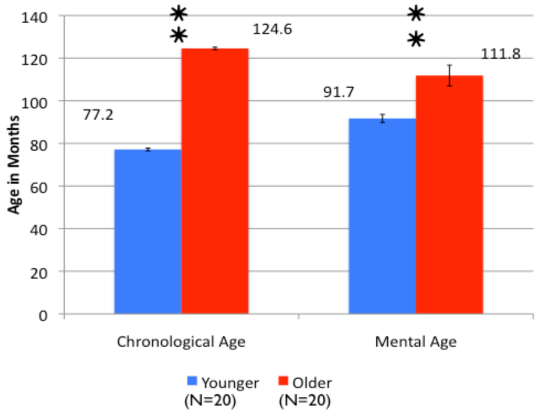


Figure 4.1. Mean chronological ages (bars on left) and mental ages (bars on right) in full sample of younger and older Primary School groups. (Error bars show standard errors of the mean. Double stars represent significant differences at the .001 level.)

Figure 4.1 shows that initial recruitment via school assessments was not successful in obtaining groups with maximum CA differences but exactly matched on MA. This is more clearly demonstrated in Figure 4.2 by looking at the disparity between each child's MA and their CA in the full sample of Primary School children.

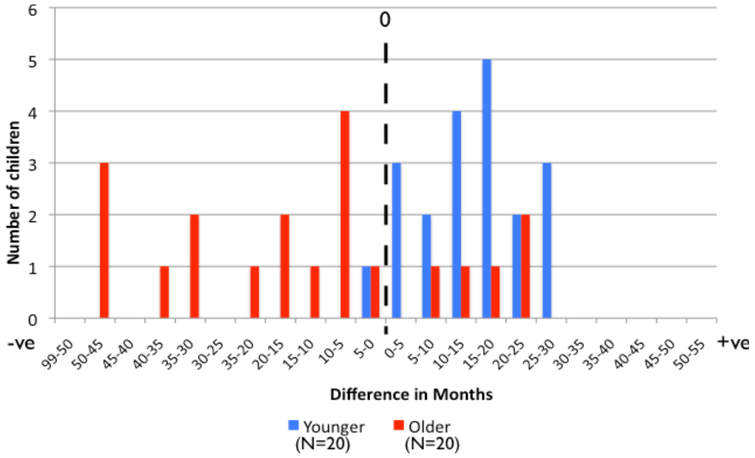


Figure 4.2. The actual MA-CA disparities in younger and older Primary School groups. Data to the left of the dashed vertical line (at point 0) represent children with MAs lower than their CAs. Data to the right represent children with MAs higher than their CAs.

If the sampling procedure had been successful Figure 4.2 would show no overlap between the MA-CA disparities of the groups. That is, the difference between each younger child's MA and their CA would result in a positive MA-CA disparity and for each older, less able child, a negative MA-CA disparity. However, Figure 4.2 shows an overlap suggesting greater accuracy in the school's assessments of younger children's ability and less accuracy in the assessment of older children's ability. Five of the children in the older children age group obtained MAs well above the level typically associated with their age and thus were not lower ability. This discrepancy between the school's assessments and the child's ability as measured by the BAS II-SA is noteworthy and relevant to educational issues, such as streaming classes based on ability. Minimally, the implication is that some of the older children were performing poorly on school assessments for reasons other than their cognitive ability. These issues and the possible causes of such a discrepancy are considered further in the discussion.

I proceed here in two ways: in the first, I continue with the aim of obtaining two groups with maximum CA and minimum MA differences by selecting a reduced data set from the full sample. For this, the younger children of average ability and the older children of above average ability are removed to achieve closer matching. In the second, I use the full Primary sample but add each child's MA-CA disparity as a covariate in the analysis. Respectively, these provide a categorical and a continuous view of the effect of MA-CA disparity on the subtest profiles of the BAS II-SA.

Primary School MA-matched group comparisons

Children in the younger age group with MA-CA differences greater than 7.5 months and older children with MA-CA differences more negative than -7.5 were selected for the reduced dataset. This resulted in a reduced dataset of 14 younger and 14 older children. Figure 4.3A below illustrates the MA-CA disparities for the reduced dataset of younger and older children. This figure shows a larger degree of variability in the MA-CA disparities within the older age group compared to the younger age group. Indeed the largest difference shown in the older age group equates to an MA 4 years below their actual CA. This difference is surprising given that the selection of participants was aimed to purposefully exclude children with any special educational needs. However, this raises another salient issue relating to how special educational needs of children are handled, within the education system. This topic is returned to later in the discussion.

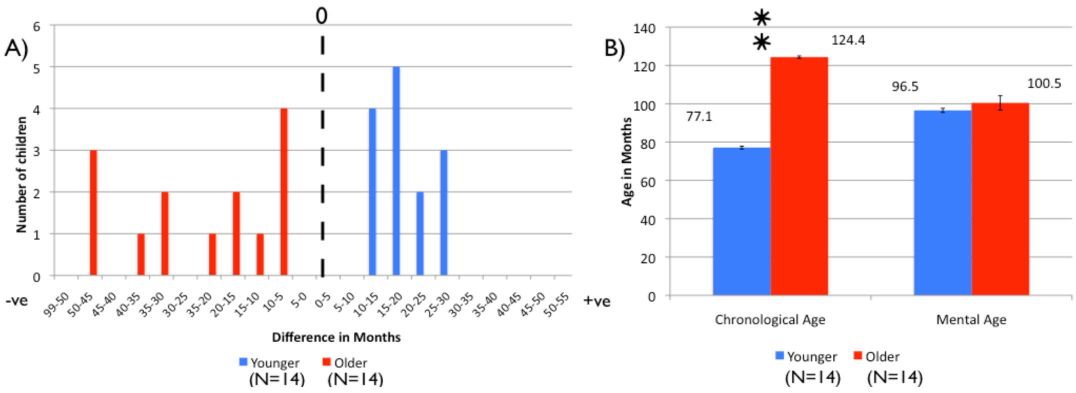


Figure 4.3A Mean MA-CA disparities of younger and older groups in reduced set of Primary school children. Data to the left of the dashed vertical line (at point 0) represent children with MAs lower than their CA. Data to the right represent children with MAs higher than their CA. Figure 4.3B Mean CAs (bars on left) and MAs (bars on right) for reduced set of Primary school children. (Error bars show standard errors of the mean. Double stars represent significant differences at the .001 level).

Figure 4.3B compares the mean CAs and MAs of the younger and older groups of Primary School children in the reduced set. A univariate ANOVA showed the mean difference of 47.3 months between the groups' CAs was highly reliable (F(1,27)=2319.30, p< .001, η^2=.99), while the mean difference of 4 months in their MA was not (F(1,27)=1.02, p=.321, η^2=.04). Taking a reduced dataset that showed no overlap in their MA-CA disparities we can now more clearly define these groups as *younger higher ability* (YHA) and *older lower ability* (OLA) groups. The computed mean IQ of these groups were: Younger IQ=125; Older IQ=81.

The mean proportion correct on each of the six Core scales is presented in Figure 4.4. A MANOVA performed on these data showed only one significant difference between groups. This was on the Pattern Construction sub-test (see ‘Pattern cons’, Figure 4.4). However, this difference did not survive a bonferonni correction (uncorrected: F(1,26)=6.98, p=0.14,η^2=.212) and no other differences reached statistically reliable levels in the respective profiles. These data do not replicate the findings of Spitz (1982), who reported reliable differences in the respective profiles of groups matched on an MA of 8.0 years.

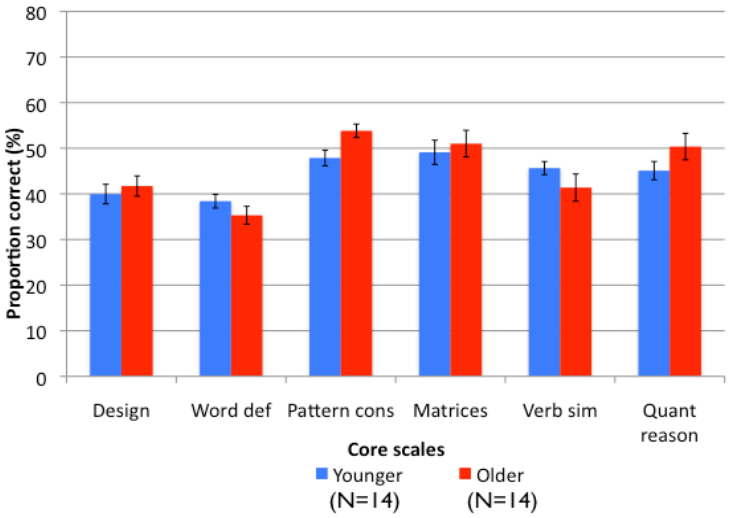


Figure 4.4. Mean proportion correct on each of the six BAS II-SA Core scales for MA-matched YHA and OLA Primary School groups. (Error bars show standard errors of the mean.)

Figure 4.5 shows the mean proportion correct across each of the six Diagnostic tests for YHA and OLA groups. A MANOVA performed on these data revealed an overall main effect of Group in the respective profiles (F(1,25)=7.40, p=.012, η^2=.228). However, Figure 4.5 shows this effect was primarily due to the large between-group differences on the Speed of Processing sub-test. On the Diagnostic sub-tests, the univariate results of the MANOVA showed that the YHA and OLA did differ reliably on two tasks. These were the Object Immediate and the Speed of Processing sub-tests. However, following a bonferonni correction the differences between groups on the Object Immediate sub-test were revealed to be not reliably different [uncorrected: (F(1,25)=4.07, p=.05, η^2=.140)]. The differences observed on the Speed of Processing sub-test did survive the correction and showed the OLA performed reliably better than the YHA (F(1,25)=38.64, p<.001, η^2=.607). Once again, no test indicated superior performance of the YHA group.

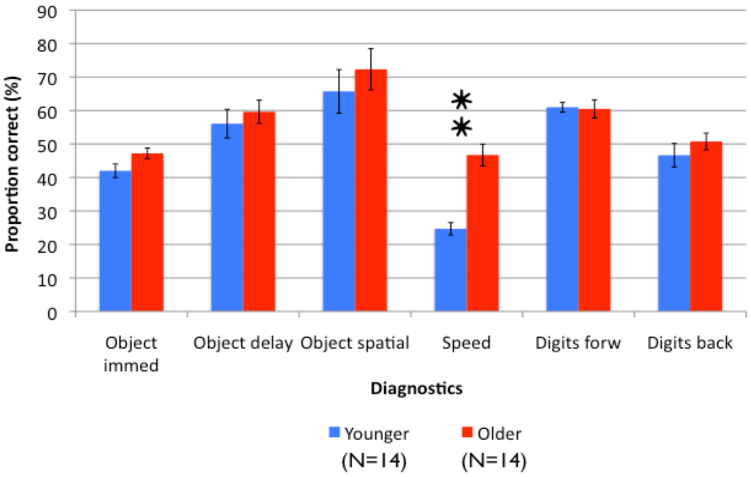


Figure 4.5. Mean proportion correct on each of the six BAS II-SA Diagnostic tests for MA-matched YHA and OLA Primary School groups. (Error bars show standard errors of the mean. Double stars represent significant differences at the .01 level.)

Primary School full sample analysis

The fact that variability exists in the MAs of the younger and older groups provides the opportunity to examine the relationship of MA-CA differences to performance. The greater the magnitude of MA-CA difference, the further the individual falls within the tails of the normal distribution of cognitive ability. The more positive this difference, the greater the advantage (in terms of cognitive ability) the child has. This allows us to establish whether cognitive profiles will be modulated by the extent to which better performance is achieved via greater cognitive ability, rather than greater chronological age.

Figure 4.6a-e distinguishes five possible outcomes of the covariate analysis. If one defines, for the purposes of this figure, MA-CA disparity to be “ability”, then these simplified illustrations demonstrate: (a) pure effects of ability, (b) effects of ability and age in which effects *do not* interact, (c) effects of ability and age in which effects *do* interact, (d) pure effect of age, and (e) no effects of either age, or ability.

In the interpretations that follow, I assume: (1) a main effect of Group to be an indicator that one’s chronological age reliably modulates performance; (2) a main effect of MA-CA disparity to be an indicator that greater ability, or one’s “advantage” of mental age over chronological age, predicts performance; (3) an interaction of Group and MA-CA disparity to indicate that these effects are combining more or less than additively and thus advantage contributes differently to performance at different age levels; and (4) that conversely, in the absence of any interaction (but where main effects of Group and MA-CA are present) that a given level of performance may be obtained either through greater age and lower ability, or greater ability and lower age.

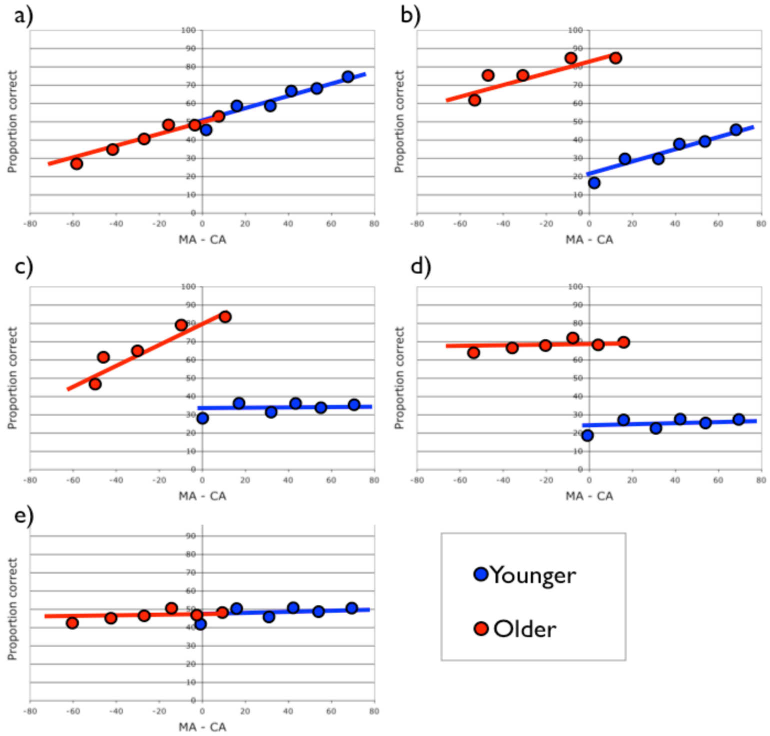


Figure 4.6. Five possible outcomes of the covariate analysis: (a) pure effect of ability, (b) effects of ability and age – with no interaction, (c) effects of ability and age – with interaction, (d) pure effect of age, and (e) no effects of either age, or ability.

Of the outcomes depicted in Figure 4.6, Tile E is the only one that appears to be consistent with the results of the categorical analysis in which the YHA and OLA groups were shown to be statistically indistinguishable from each other. That is, this pattern shows no differences in the mean scores of the two groups.

For the Speed of Processing subtest (the only subtest in which the categorical analysis did reveal reliable group differences), Figure 4.6 provides four possible scenarios that may be representative of those data. These are tiles (a), (b), (c) and (d). The patterns depicted in these tiles in Figure 4.6 each portray different group means performance levels, and differential effects of age and advantage.

To examine whether any of these patterns are indeed representative of the actual pattern of data underlying performance in the full Primary School sample, we turn to the results of multiple analysis of covariance (MANCOVA). This analysis allows us to evaluate whether age (Group) and/or advantage (MA-CA disparity) modulates the pattern found across the six Core scales and the six Diagnostic tests and whether their effects also interact. I begin by assessing the pattern of results across the Core scales in the full sample that comprises 40 Primary School children.

The results of this analysis showed that performance on the Core scales was reliably modulated by Group (F(1,38)=13.85, p=.001, η^2=.267) and MA-CA disparity (F(1,38)=4.05, p=.05, η^2=.096) and that overall, Group and MA-CA disparity interacted in a reliable way (F(2,37)=11.08, p<.001, η^2=.375). That is, the groups showed reliably different subtest profiles, but this depended on the level of MA-CA disparity. The analysis also showed that on the individual Core scales sub-tests, Group and MA-CA disparity accounted for significant proportions of variance. For example, Group predicted 22.9% of the variance on Verbal Similarities (F(1,36)=10.68, p=.002, η^2=.229), 32.5% on Matrices (F(1,36)=17.35, p=.001, η^2=.325), 33.0% on Recall of Designs (F(1,36)=17.73, p<.001, η^2=.330), 46.1% on Quantitative Reasoning (F(1,36)=30.80, p<.001, η^2=.461) and 60.8% on Pattern Construction (F(1,36)=55.74, p<.001, η^2=.608), but failed to reliably predict variance on Word Definitions. MA-CA disparity was reliable in modulating performance in 5 out of 6 of the Core scales (the exception was the Matrices subtest). MA-CA disparity was found to account for 22.9% of the variance on Recall of Designs (F(1,36)=10.70, p=.002, η^2=.229), 24.8% on Verbal Similarities (F(1,36)=11.85, p=.001, η^2=.248), 27.5% on Word Definitions (F(1,36)=13.67, p=.001, η^2=.275), 28.3% on Quantitative Reasoning (F(1,36)=14.21, p=.001, η^2=.283) and 34.5% on Pattern Construction (F(1,36)=18.96, p<.001, η^2=.345). Finally, the analysis showed no reliable Group x MA-CA disparity interaction on the individual core scales.

The finding that age and advantage modulate performance in reliable ways, but that their effects do not interact on the individual core scales suggests that the data may be more accurately characterised by Figure 4.6b – showing both age and advantage effects, with no interaction.

On the next page, in Figure 4.7 we may compare the illustrative outcomes to the actual patterns on the Core scales. In this figure 6 tiles are given, each representing the performance of younger and older Primary school groups. Each data point shows the proportion correct a single child obtained on a given sub-test (proportion correct appears on the Y-axis) and the MA-CA disparity for that child (disparity scores appear on the X-axis). In these tiles, blue lines and red lines distinguish the younger group data and older group data, respectively. Each tile also includes a marker that indicates whether age (Group) and/or advantage (MA-CA disparity) reliably predict performance at the 0.05 level. These markers appear within the ellipses in the top left-hand corner of each tile.

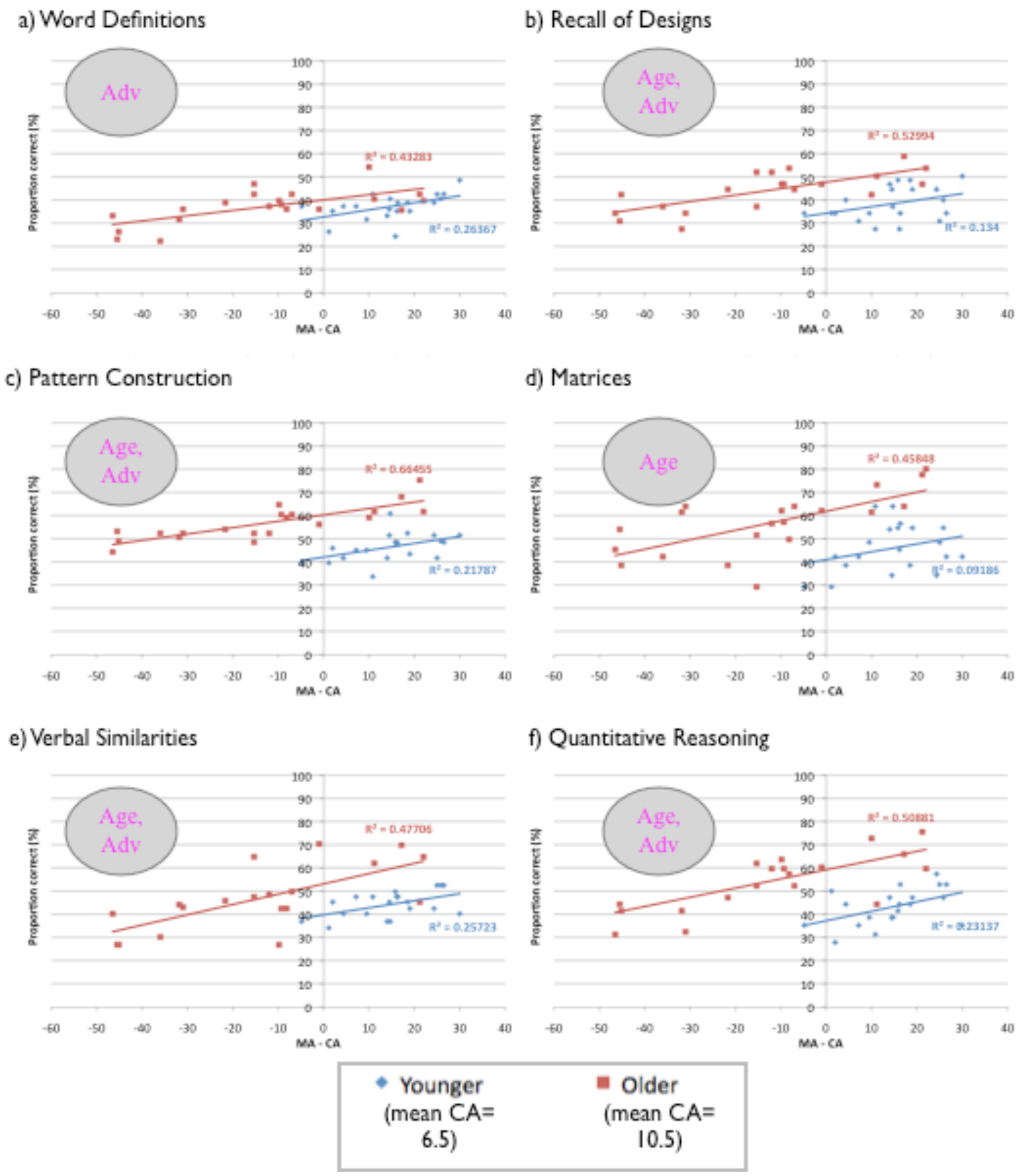


Figure 4.7 Plotting MA-CA disparity and performance on each of the Core scales for the full Primary school sample. Ellipses indicate reliable predictors of performance. No 2-way interactions were found. R^2 values indicate the proportion of variance explained by each trajectory.

The results of the MANCOVA on the Diagnostic tests revealed Group was reliable in modulating overall performance and accounted for 40.7% of the total variance ($F(1,36)=24.75$, $p<.001$, $\eta^2=.407$). On the individual subtests, Group reliably predicted performance on 4 out of 6 tests. Group accounted for 22.8% on Recall of Objects-Spatial ($F(1,34)=10.05$, $p=.003$, $\eta^2=.228$), 31.3% on Recall of Objects-Immediate ($F(1,34)=15.50$, $p<.001$, $\eta^2=.313$), 34.0% on Digits Backward ($F(1,34)=17.50$, $p<.001$, $\eta^2=.340$) and 39.3% on Speed of Processing ($F(1,34)=22.04$, $p<.001$, $\eta^2=.393$), with all surviving a bonferroni correction for multiple comparisons.

By contrast, MA-CA disparity scores failed to reliably modulate overall performance on the Diagnostic tests and was reliable in modulating the performance of only one of the subtests: Digits Backward ($F(1,36)=11.38$, $p=.002$, $\eta^2=.251$). This analysis also showed no reliable interaction between Group and MA-CA disparity. With regard to the illustrative outcomes identified earlier for the reliable MA-matched group differences in Speed of Processing, the results here support the pattern depicted in Figure 4.6d. This figure depicts an effect of age and no effect of advantage on performance.

The actual performance of younger and older Primary School children on each of the Diagnostic tests is shown on the following page in Figure 4.8. This figure shows 6 tiles corresponding to the individual sub-tests. Each data point shows the proportion correct a single child obtains on a sub-test (proportion correct appears on the Y-axis) and the MA-CA disparity for that child (disparity scores appear on the X-axis). Again, tiles display blue and red lines to distinguish the younger group data and older group data, respectively. Ellipses indicate whether Group (age) or MA-CA disparity (advantage) reliably predict performance.

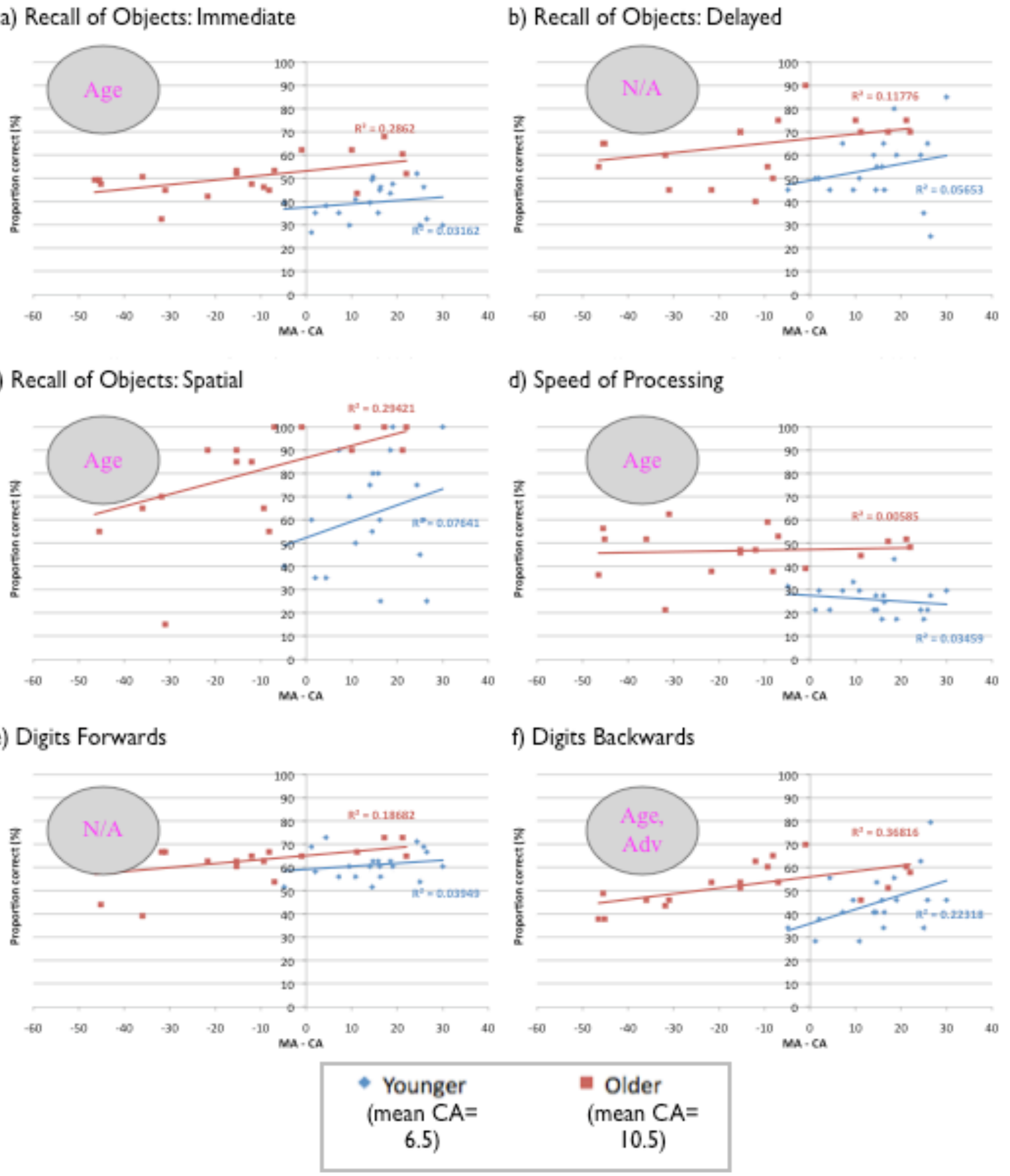


Figure 4.8 Plotting MA-CA disparity and performance on each of the Diagnostic tests for the full Primary school sample. Ellipses indicate reliable predictors of (N/A indicates neither). No 2-way interactions were found. R^2 values indicate the proportion of variance explained by each trajectory.

Summary

Following the creation of two MA-matched groups (mean MA=8.2 years) with a disparity in age of 3.9 years, analyses of the Core scales yielded no overall difference in performance profiles, and no reliable differences on any of the individual sub-tests. The analysis of these groups' respective profiles on the Diagnostic tests revealed an advantage for the older less able group in only one task: Speed of Processing. These findings contrast with Spitz (1982) who reported advantages for YHA over OLA on verbal reasoning tasks (e.g., superior definitions, comprehension and word similarities) and abstract reasoning tasks (e.g., weights) and advantages for OLA over YHA on tasks tapping experience and rote learning (e.g., counting backwards, change and date).

In the continuous analyses on the full dataset for the Core scales, age effects and advantage effects were revealed. Group was found to account for significant proportions of variance on 5 out of 6 subtests: Recall of Designs, Pattern Construction, Matrices, Verbal Similarities and Quantitative Reasoning.

MA-CA disparity scores were also found to predict performance on 5 out of 6 of these Core scale subtests: Recall of Designs, Word Definitions Pattern Construction, Verbal Similarities, Quantitative Reasoning. In all cases, reliable age effects correspond to better performance for the older group. Yet, no interactions were found between Group and MA-CA disparities on the individual sub-tests, thus suggesting largely separate influences of age and advantage. Table 4-1 summarises these findings.

Table 4-1. Summary table of Primary school results on Core scales BAS II-SA.

Subtest	*Reliable predictor*	*Interpretation*
Recall of Designs	Group, MA-CA	Age and Advantage effects
Word Definitions	MA-CA	Advantage effects
Pattern Construction	Group, MA-CA	Age and Advantage effects
Matrices	Group	Age effects
Verbal Similarities	Group, MA-CA	Age and Advantage effects
Quantitative Reasoning	Group, MA-CA	Age and Advantage effects

Within the Diagnostic tests, MANCOVA revealed advantage was only reliable in influencing performance on the Digits Backward task. Instead, differences in performance were mostly attributable to age effects. Group accounted for significant proportions of variance on 4 out of 6 subtests. These findings are summarised in Table 4-2.

Table 4-2. Summary table of Primary school results on Diagnostic tests BAS II-SA.

Subtest	*Reliable predictor*	*Interpretation*
Recall of Objects-Immediate	Group	Age effects
Recall of Objects-Delayed	N/A	
Recall of Objects-Spatial	Group	Age effects
Speed of Processing	Group	Age effects
Digits Forwards	N/A	
Digits Backward	Group, MA-CA	Age and Advantage effects

Secondary School results

In this section, results are presented for 35 Secondary School children who were administered the same set of Core scales and Diagnostic tests from the BAS II-SA given to the Primary School children. The sampling procedure differed from that followed with the Primary School children in that the Secondary School children were selected from pre-streamed classes with similar notional common ability levels but with a wide chronological age range (for details, see Chapter 3, 'General Methodology'). Rather than two discrete age groups, the children's ages ranged from 11.75 through to 15.92. However, groups of younger and older children were formed, based on a median age split of 13.98 years. These groups are comprised of 16 children with chronological ages between 11.75 and 13.92 years (mean=12.90, sd=0.72) and 19 children aged between 14.00 and 15.92 years (mean=15.08, sd=0.57). Gender was split equally with 8 males and 8 females in the younger age group and 9 males and 10 females in the older age group.

Figure 4.9 below shows the mean CAs and MAs of younger and older Secondary School children. A univariate analysis of variance of these data showed the mean difference of 26.1 months in CA between younger and older groups was reliable ($F(1,34)=100.84$, $p< .001$, $\eta^2=.753$). It also found no reliable differences between the mean MAs of the younger and older groups ($F(1,34)=0.30$, $p=.587$, $\eta^2=.009$). The computed mean IQ of these groups were: Younger IQ=110; and Older IQ=89.

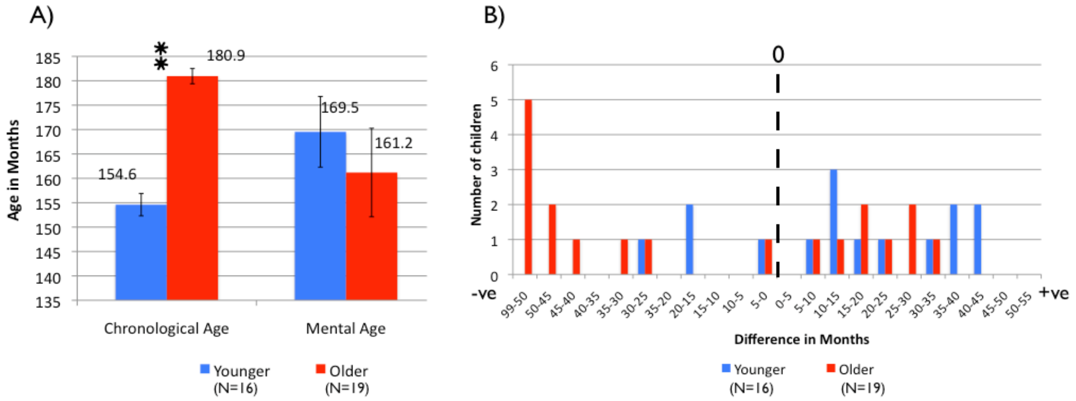


Figure 4.9A. Mean chronological ages (bars on left) and mental ages (bars on right) in full sample of younger and older Secondary School groups. (Error bars show standard errors of the mean. Single stars represent significant differences at the .05 level. Double stars represent significant differences at the .001 level.). Figure 4.9B. MA-CA disparities in younger and older Secondary School groups. Data to the left of the dashed vertical line (at point 0) represent children with MAs lower than their CA. Data to the right represent children with MAs higher than their CA.

Figure 4.9B shows the MA-CA disparities for each child in the Secondary School sample, split by younger and older groups. The figure shows that a large degree of variability exists within both groups. This is accounted for by the fact that children were selected from basic and more advanced school streams in which wide ranges of chronological ages were present (see Chapter 3, ‘General Methodology’).

While the univariate analysis of variance indicated that the two different age groups were matched on MA, Figure 4.9B shows a wide degree of overlap within each group. Although taking a reduced dataset was effective in matching the two age groups on MA in the Primary School sample, this was not the case in the Secondary School sample: removing the younger less able children and the older more able children from the dataset yielded respective mean MAs of 15:1 and 11:0 for the two groups. Consequently, I continue here by using the whole sample as initially recruited, first in the categorical, then in the continuous design. If null categorical effects are artefacts of overlap, this should become apparent in the continuous analyses.

Secondary School MA-matched group comparisons

The results from two MANOVAs found no reliable effect of Group on the Core scales or on the Diagnostic tests. These analyses further revealed that Group also did not reliably modulate performance of any of the sub-tests within the Core scales or Diagnostic tests. Figure 4.10 shows the mean proportion correct on each of the Core scales for YHA and OLA groups.

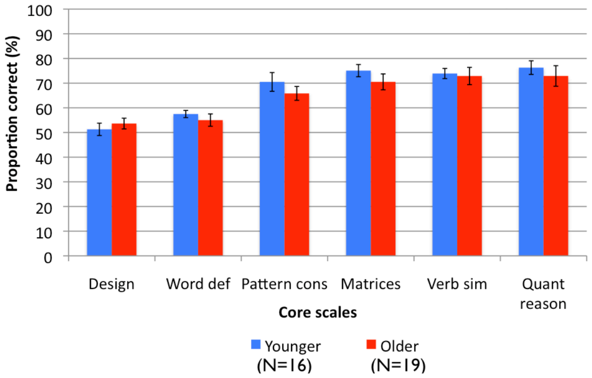


Figure 4.10. Mean proportion correct on each of the six BAS II-SA Core scales for MA-matched YHA and OLA Secondary School groups. (Error bars show standard errors of the mean.)

Figure 4.11 below shows the mean proportion correct on each of the Diagnostic tests for YHA and OLA groups. The absence of a Speed of Processing difference contrasts with the Primary School results.

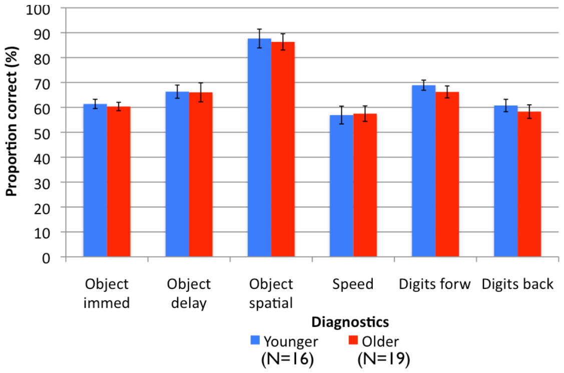


Figure 4.11. Mean proportion correct on each of the six BAS II-SA Diagnostic tests for MA-matched YHA and OLA Secondary School groups. (Error bars show standard errors of the mean.)

Secondary School full sample analysis

As was the case in the Primary School data, the Secondary School group comparisons showed profiles for YHA and OLA that were statistically indistinguishable from each other on the Core scales and Diagnostic tests. Taking the Secondary School sample as a whole, a MANCOVA was performed on the proportion correct across the BAS II-SA Core scales and Diagnostic tests, using MA-CA disparity (in months) as the covariate.

On the Core scales analysis, results showed no main effect of Group, but did show that Group modulated performance on 4 out of 6 of the individual sub-tests. Group was reliable in accounting for 21.5% of the variance on Verbal Similarities (F(1,31)=8.48, p=.007, η^2=.215), 22.5% on Pattern Construction (F(1,31)=8.98, p=.005, η^2=.225), 31.8% on Quantitative Reasoning (F(1,31)=14.43, p=.001, η^2=.318) and 46.0% on Recall of Designs (F(1,31)=26.46, p<.001, η^2=.460), but failed to modulate performance on Word Definitions and Matrices. MA-CA disparities were reliable in predicting overall performance on Core scales (F(1,33)=144.78, p<.001, η^2=.814) and accounted for significant proportions of the variance on all 6 sub-tests. MA-CA disparities accounted for 48.0% of the variance in Word Definitions (F(1,31)=28.62, p<.001, η^2=.480), 52.7% on Verbal Similarities (F(1,31)=34.52, p<.001, η^2=.527), 64.3% on Recall of Designs (F(1,31)=55.88, p<.001, η^2=.643), 65.4% on Matrices (F(1,31)=55.59, p<.001, η^2=.654), 72.1% on Quantitative Reasoning (F(1,31)=80.30, p<.001, η^2=.721) and 73.3% on Pattern Construction (F(1,31)=85.24, p<.001, η^2=.733). In Figure 4.12, six tiled scatterplots represent the performance of younger and older Secondary School groups on the Core scales. Blue lines and red lines distinguish younger and older children's data respectively. As before, each tile includes a marker that indicates whether Group (age) or MA-CA disparity (advantage) reliably predicts performance. These markers appear within the ellipses in the bottom left-hand corner of each tile. A single asterisk indicates where findings are *different* to the pattern of findings in the Primary School data.

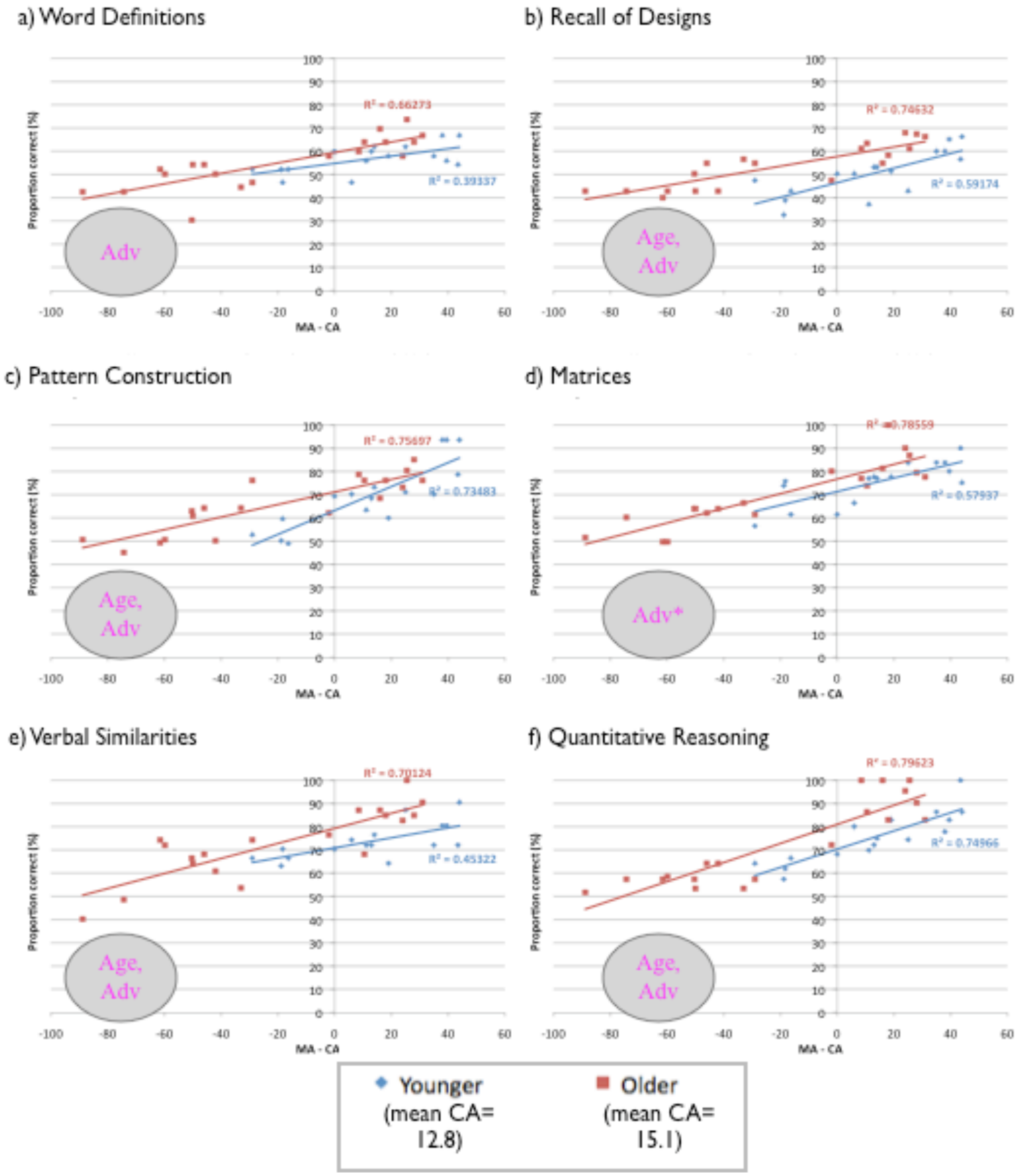
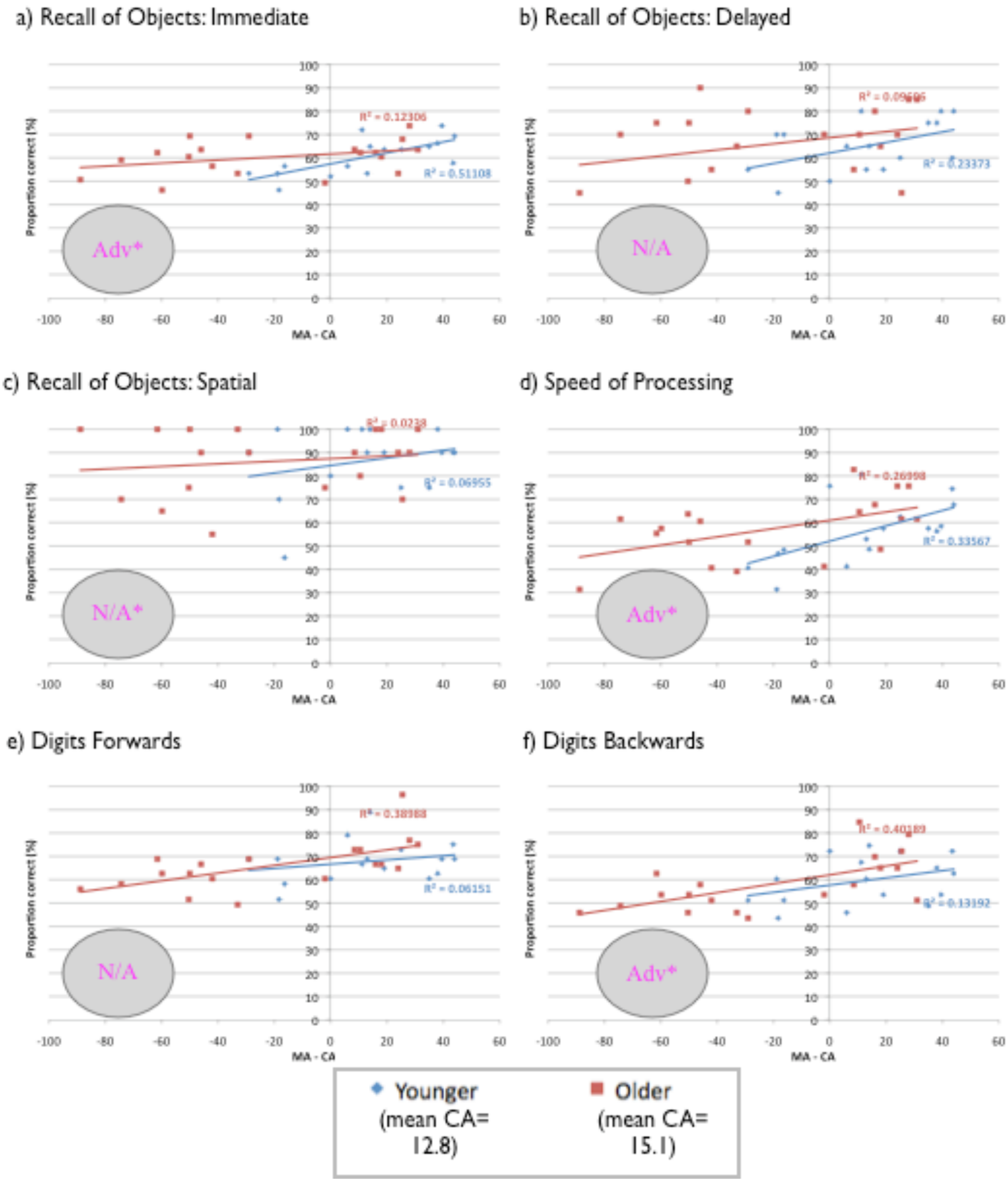


Figure 4.12. Plotting MA-CA disparity and performance on each of the Core scales for the full Secondary School sample. Ellipses indicate reliable predictors of performance. No 2-way interactions were found. Asterisks within ellipses indicate findings *different* to Primary School data.

In comparing the Secondary School and Primary School results on the Core scales (see Figure 4.7), we can see that the pattern of findings is identical, with the exception of one sub-test. In the Secondary school data, advantage offers the only predictor of performance on the Matrices sub-test. By contrast, in the Primary School data, it was age that offered the only predictor. The data represented in Figure 4.12 indicate a greater amount of heterogeneity within YHA and OLA groups in the Secondary School sample. This is different to the results obtained in the Primary

School dataset. Referring back to the simplified outcomes illustrated in Figure 4.6, a mixture of age and advantage effects appear to underlie the data represented above. For example, the data for Word Definitions and Matrices resemble Figure 4.6a showing advantage effects, whereas Recall of Designs, Pattern Construction, Verbal Similarities and Quantitative Reasoning more closely resemble a mixture of age and advantage effects, as depicted in Figure 4.6b.

The results of the MANCOVA on the Diagnostic tests showed Group was not reliable at modulating the overall profile of performance and did not modulate performance on any of the sub-tests. On the other hand, MA-CA disparity was reliable at predicting overall performance ($F(1,33)=16.83$, $p<.001$, $\eta^2=.338$) and the analysis revealed a reliable overall interaction between Group and MA-CA disparity ($F(2,32)=8.40$, $p=.001$, $\eta^2=.344$). MA-CA disparity further accounted for significant proportions of variance on 3 Diagnostic sub-tests. MA-CA disparity accounted for 21.6% on Digits Backward ($F(1,31)=8.52$, $p=.006$, $\eta^2=.216$), 28.6% on Speed of Processing ($F(1,31)=12.42$, $p=.001$, $\eta^2=.286$) and 31.4% on Recall of Objects-Immediate ($F(1,31)=14.20$, $p=.001$, $\eta^2=.314$). Plots of proportion correct on the Diagnostic sub-tests are given in Figure 4.13 below. Blue lines and red lines represent the data for younger and older children, respectively. Within each tile, ellipses display an asterisk where patterns of results are different to those found in the Primary School analysis. These ellipses also indicate whether age and or ability were reliable in predicting performance.



a) Recall of Objects: Immediate

b) Recall of Objects: Delayed

c) Recall of Objects: Spatial

d) Speed of Processing

e) Digits Forwards

f) Digits Backwards

Figure 4.13. Plotting MA-CA disparity and performance on each of the Diagnostic sub-tests in the full Secondary School sample. Ellipses indicate reliable predictors of performance (N/A indicates neither). No 2-way interactions were found. Asterisks within ellipses indicate findings *different* to Primary School data.

Summary

For Secondary School groups matched on a mean mental age of 13.8 years with mean CA disparity of 2.3 years, no reliable group differences were found between YHA and OLA on the Core scales. Although the YHA showed marginally better performance on 5 out of 6 of the individual sub-tests none of the subsequent analyses showed these differences to be reliable. On the Diagnostic tests, comparisons of group's profiles also showed no reliable differences overall, or on the sub-tests.

In the continuous analysis on the Core scales, the results of the MANCOVA showed advantage to be a strong predictor of overall performance and of performance on each of the individual Core scale subtests. Age also reliably modulated performance, but did so on fewer sub-tests (see summary

Table 4-3).

Table 4-3. Summary table of Secondary school results on Core scales BAS II-SA (**indicates finding different to Primary school data*)

Subtest	*Reliable predictor*	*Interpretation*
Recall of Designs	Group, MA-CA	Age and Advantage effects
Word Definitions	MA-CA	Advantage effects
Pattern Construction	Group, MA-CA	Age and Advantage effects
Matrices	MA-CA*	Advantage effects
Verbal Similarities	Group, MA-CA	Age and Advantage effects
Quantitative Reasoning	Group, MA-CA	Age and Advantage effects

The results from the second MANCOVA on the Diagnostic tests revealed that age did not reliably predict performance, whereas advantage modulated performance on 4 out of 6 sub-tests tests (see summary Table 4-4).

Table 4-4. Summary table of Secondary school results on Diagnostic tests BAS II-SA (**indicates finding different to Primary school data*)

Subtest	*Reliable predictor*	*Interpretation*
Recall of Objects-Immediate	MA-CA*	Advantage effects
Recall of Objects-Delayed	N/A	
Recall of Objects-Spatial	N/A*	
Speed of Processing	MA-CA*	Advantage effects
Digits Forwards	N/A	
Digits Backward	MA-CA*	Advantage effects

The results from the Secondary school analysis also contrast with Spitz (1982) who reported reliable advantages for YHA over OLA on verbal reasoning tasks and advantages for OLA over YHA on tasks tapping experience.

Combined Primary and Secondary School data

The data from Primary and Secondary School levels were combined in two final MANCOVAs and applied to the Core scales and Diagnostic tests. In these analyses, the effects of age and advantage were assessed over the entire experimental sample. MA-CA disparity remained the covariate and Group (younger vs. older) and Level (Primary vs. Secondary) formed the between-subjects factors[3]. The total sample in this analysis was 75 children and comprised the 40 Primary School children and 35 Secondary School children described in previous sections.

Figures plotting each child's MA-CA disparity by performance follow shortly. But first, let us consider some of the possible outcomes that the covariate analysis may reveal to aid in the interpretation of the data. I apply the following interpretations to the results of these analyses: (1) main effects of Group or Level to indicate an effect of age; (2) a main effect of MA-CA disparity indicates that one's "advantage" of MA given CA expectations modulates performance in reliable ways; (3) a Group by Level interaction indicates that differences in the performance between Groups are dissimilar at the different school levels (i.e., the effect of an age disparity between two groups depends on what age we are studying); (4) a Group/Level by MA-CA interaction indicates that the effects of age and advantage combine more or less than additively and thus that the effect of advantage on performance differs across the age range; (5) that the absence of a Group/Level by MA-CA interaction, but where main effects of Group/Level and MA-CA are present, indicates that a given level of performance can be reached independently either through lower age and greater ability, or through greater age and lower ability; and finally, (6) that a 3-way interaction between Group, Level and MA-CA indicates performance is modulated by a mix of factors including age and advantage and this changes over school level depending on one's ability.

Figure 4.14a-e depicts five possible outcomes that illustrate: (a) pure effects of advantage, (b) main effects of age (both Group and Level) and advantage, where there is no interaction, (c) a 3-way interaction of Group, Level and advantage, (d) a pure effect of age (both Group and Level); and (e) no effect of age or advantage.

---

[3] Note, that when we come to analyse the experiments we will find it more useful to combine the Group by Level information into a single Group variable consisting of four levels (Primary Younger, Primary Older, Secondary Younger, Secondary Older). The reader will be reminded of this where it is first used, in Chapter 5.

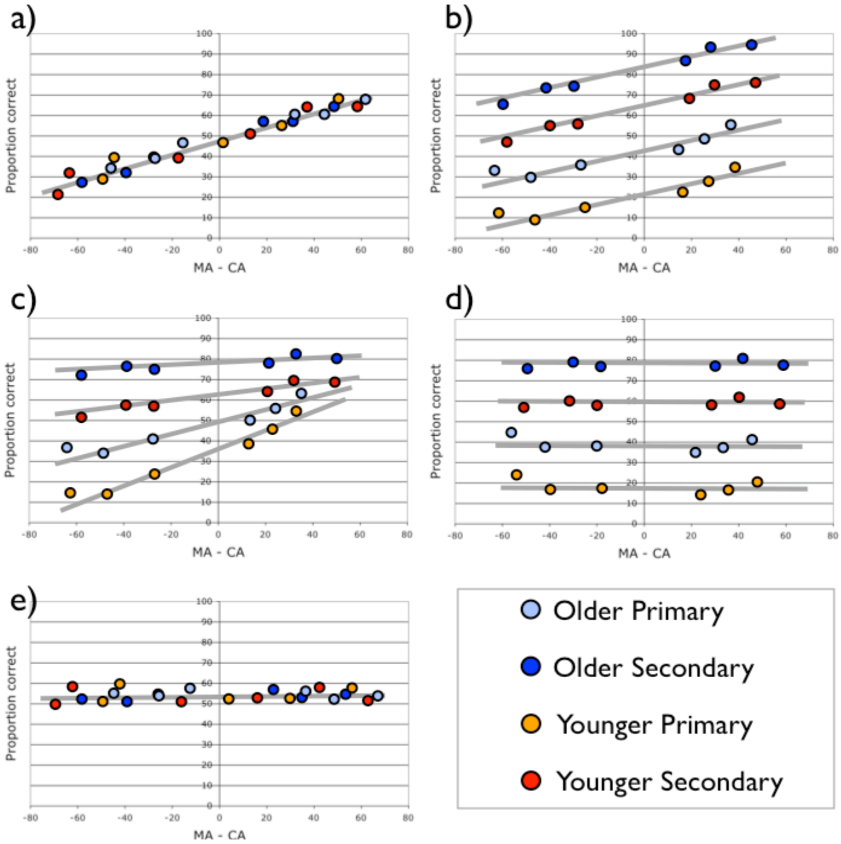


Figure 4.14 Illustrating five possible outcomes of the combined covariate analysis: (a) pure effect of advantage, (b) advantage and age effects-no interaction, (c) age and advantage effects-with interaction, (d) pure age effect of age; and (e) no effects of age or advantage.

Full sample analysis

Beginning with the results of the analysis on the Core scales, the first MANCOVA found main effects of Group (F(3,71)=39.73, p<.001, η^2=.627) and of Level (F(1,73)=104.03, p<.001, η^2=.588), indicating that age reliably modulates overall performance on the sub-tests. This analysis also showed MA-CA disparity was reliable in predicting overall performance (F(1,73)=12.63, p=.001, η^2=.147). Furthermore, each of the 2-way overall interactions were reliable: Group x MA-CA (F(4,70)=22.68, p<.001, η^2=.564), Level x MA-CA (F(2,72) =7.11, p=.002, η^2=.165) and Level x Group (F(3,71)=39.73,p<.001, η^2=.627). The reliable interaction between Level and Group indicates differential effects of age on performance in younger and older children at the two school levels. This might lead us to expect patterns of MA-CA disparities and performance resembling Figure 4.14c, in the illustrative outcomes above.

On the individual sub-tests, Group, Level and MA-CA disparities were each reliable in accounting for significant proportions of variance on all 6 Core scales sub-tests. Table 4-5 summarises these results. Only one 2-way interaction was found reliable in these analyses. This was a Level x Group interaction on the Matrices sub-test (F(1,67)=8.18, p=.006, η^2=.109). Collapsing over MA-CA disparity, examination of each group's performance on the Matrices sub-test showed that at the Primary School level, the YHA group was less accurate (mean=45.9%, se=3.8) compared to the OLA group (mean=56.6%, se=4.8). However, at the Secondary School level the YHA were *more* accurate (mean=75.1%, se=3.7) compared to the OLA (mean=70.5%, se=5.2). Thus, at the different school levels the effect of Group on performance was not identical. In terms of the trajectories, the age difference encoded by Group was more important at Primary (trajectories further apart) than at Secondary (trajectories closer together). While this might be expected (the CA difference between the younger and older groups *was* larger for Primary than Secondary), what is notable is that Matrices was alone in shaping this interaction. It suggests that for Matrices in particular, age differences become less important for predicting performance as age increases, when differences in intelligence controlled for.

Table 4-5. Summary table of significant results on sub-tests of BAS II School Age Core scales: Primary and Secondary School data combined (df=1, df_{error}=67).

	BAS II School Age: Core scales	F	Sig.	Partial Eta Squared
Level	Recall of Designs	33.72	<.001	0.335
	Word Definitions	139.28	<.001	0.675
	Pattern Construction	74.00	<.001	0.525
	Matrices	69.84	<.001	0.510
	Verbal Similarities	139.66	<.001	0.676
	Quantitative Reasoning	129.56	<.001	0.659
Group	Recall of Designs	41.60	<.001	0.383
	Word Definitions	11.73	.001	0.149
	Pattern Construction	49.61	<.001	0.425
	Matrices	23.57	<.001	0.260
	Verbal Similarities	19.19	<.001	0.223
	Quantitative Reasoning	45.94	<.001	0.407
MA-CA	Recall of Designs	36.58	<.001	0.353
	Word Definitions	30.97	<.001	0.316
	Pattern Construction	60.67	<.001	0.475
	Matrices	28.63	<.001	0.299
	Verbal Similarities	31.24	<.001	0.318
	Quantitative Reasoning	50.75	<.001	0.431

In Figure 4.15 below, six tiled scatterplots are presented for each of the Core scales. Each tile shows the performance reached for each child in younger Primary School group (Younger-Pri), older Primary School group (Older-Pri), younger Secondary School group (Younger-Sec) and older Secondary School group (Older-Sec).

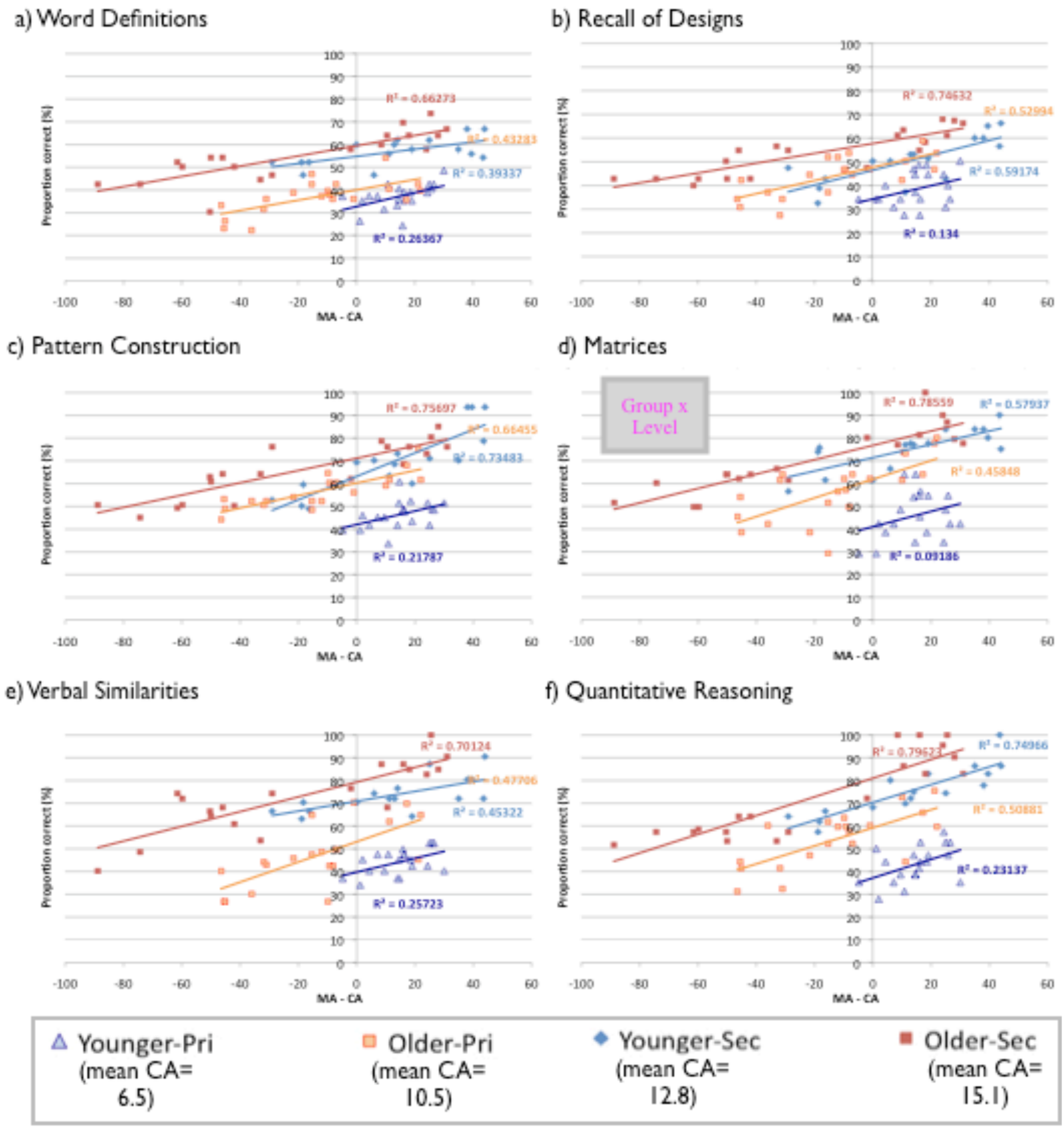


Figure 4.15 Plotting MA-CA disparity and performance on each of the Core scales for the full Primary and Secondary School samples. Main effects of Group, Level and MA-CA were reliable for all subtests. Tile (d) shows the sole reliable 2-way interaction for Matrices.

The results of the second MANCOVA on the Diagnostic tests revealed a main effect of Group (F(3,69)=25.68, p<.001, η^2=.528) and Level (F(1,71)=42.18, p<.001, η^2=.373), but no main effect of MA-CA on the overall subtest profile. These results indicate that age is reliable in modulating overall performance. The analysis further revealed a reliable 2-way interaction of Group x MA-CA (F(4,68)=10.36, p<.001, η^2=.379) and Group x Level interaction (F(3,69)=25.68, p<.001, η^2=.528). The former would suggest that the effect of advantage on performance is modulated by age and the latter again indicates that at the different school levels the effect of

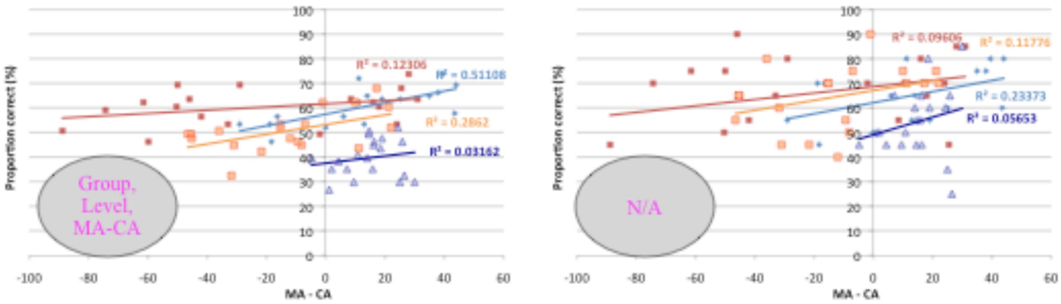
Group is not equal (as in part expected by the different age differences at Primary and Secondary levels).

On the individual Diagnostic sub-tests, age was a reliable predictor of performance – a significant proportion of the variance was accounted for in 4 out of 6 tasks by Level and in 5 out of 6 tasks by Group. MA-CA disparity modulated performance reliably in only 2 out of 6 of the Diagnostic sub-tests. Lastly, the analysis showed no reliable 2-way or 3-way interactions on the individual sub-tests. These results are summarised in Table 4-6.

Table 4-6. Summary table of significant results on sub-tests of BAS II School Age Diagnostic tests: Primary and Secondary School data combined (df=1, df_{error}=65)

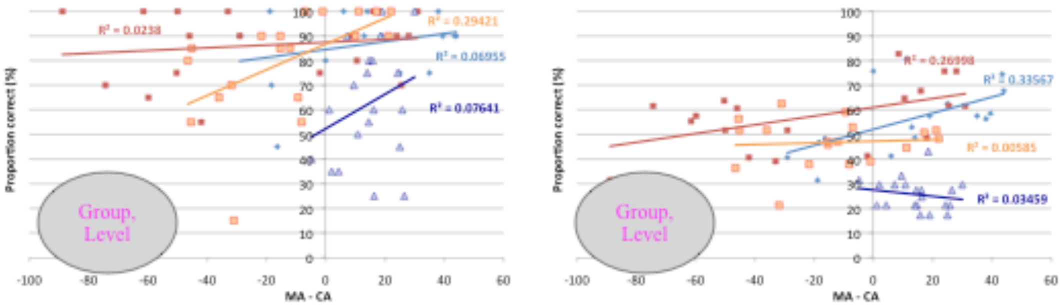
	BAS II School Age: Diagnostic Tests	F	Sig.	Partial Eta Squared
Level	Recall of Objects:Immediate	41.91	<.001	0.392
	Recall of Objects:Spatial	7.66	.007	0.105
	Speed of Processing	35.94	<.001	0.356
	Digits Backwards	21.28	<.001	0.247
Group	Recall of Objects:Immediate	18.40	<.001	0.221
	Recall of Objects:Delayed	7.57	.008	0.104
	Recall of Objects:Spatial	10.10	.002	0.135
	Speed of Processing	20.12	<.001	0.236
	Digits Backwards	16.24	<.001	0.200
MA-CA	Recall of Objects:Immediate	9.01	.004	0.122
	Digits Backwards	18.26	<.001	0.219

In Figure 4.16 below, six tiled scatterplots are presented for each of the Diagnostic tests. Each tile shows the performance reached for each child in younger Primary School group (light blue: Younger-Pri), older Primary School group (orange: Older-Pri), younger Secondary School group (dark blue: Younger-Sec) and older Secondary School group (red: Older-Sec). Each tile also includes a marker that indicates whether Group, Level or MA-CA disparity (advantage) reliably predicts performance. These markers appear within the ellipses in the bottom left-hand corner of each tile



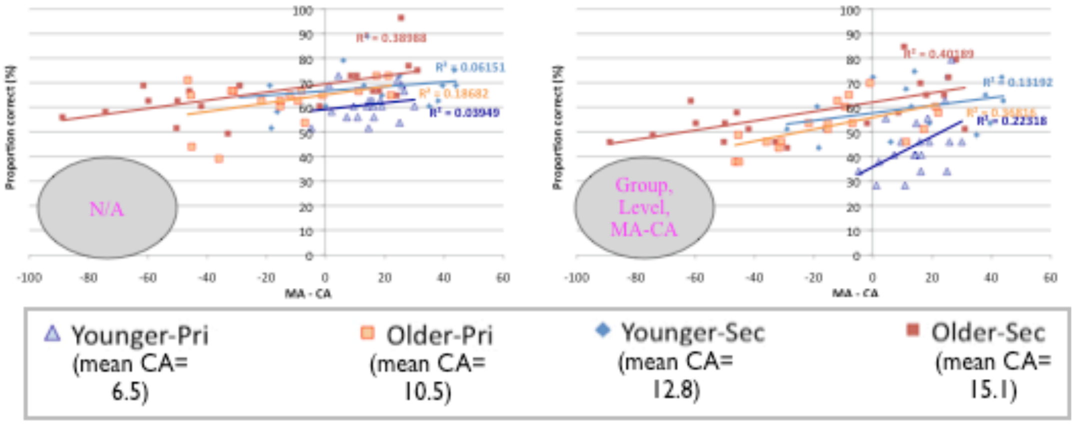
a) Recall of Objects: Immediate

b) Recall of Objects: Delayed



c) Recall of Objects: Spatial

d) Speed of Processing



e) Digits Forwards

f) Digits Backwards

Figure 4.16 Plotting MA-CA disparity and performance on each of the Diagnostic tests for the full Primary and Secondary school samples. Ellipses indicate reliable predictors of performance (N/A indicates neither). No 2-way or 3-way interactions were found.

Summary

In combining the Primary School and Secondary School data the results of the continuous analysis on the Core scales showed a mix of Group, Level and MA-CA effects. However, these effects appeared largely separate and did not interact in influencing performance on the majority of sub-tests. These results tie in with the results found separately at the Primary and Secondary levels, suggesting once again a given level of performance may be reached either through higher ability and lower age, or lower ability and higher age. One exception was the single interaction that was found in the Core scales on the Matrices sub-test. For this subtest in particular, age differences were less predictive of performance differences at Secondary School level than at Primary School level (with differences in MA-CA disparity controlled for).

Discussion

Drawing on Merrill's (1924) earlier work, Spitz (1982) showed that on the Stanford-Binet test (1916) the YHA group had the advantage on tasks tapping verbal abilities (e.g., Comprehension, Similarities, Superior Definitions, Rhymes and 60 words) and abstract reasoning (e.g., Weights), and the OLA group had the advantage on tasks tapping maturation and experience (e.g., Fingers, Counting backwards, Change, 3 Words and Date). On the basis of these findings Spitz argued that MA-matched children of different actual ages *did not* possess similar skills or abilities – he argued they are at different cognitive and intellectual levels.

In replicating Merrill's original design, we might therefore have expected to find similar differences in abilities between YHA and OLA groups on tasks within the Core scales of the BAS II-SA. For instance, we might have expected the YHA to outperform the OLA on Word Definitions, or on Verbal Similarities and the OLA to perhaps have outperformed the YHA on tasks such as Recall of Designs, or on Pattern Construction. This was not found to be the case. The kinds of differential patterns of strengths and weaknesses that Spitz showed in YHA and OLA groups were not revealed here, either in Primary School or Secondary School samples of MA-matched children. Indeed, at both school levels the YHA and OLA were statistically indistinguishable in their performances on the Core scales, and there was little difference between the YHA and the OLA on the Diagnostic tests. Only one reliable difference emerged between groups. This was in the Primary School groups on the Speed of Processing task, where the OLA showed superior performance to the YHA. The results of both Primary and Secondary level analyses showed that both age and advantage reliably modulated performance on the Core scales. However, in the Diagnostic tests there appeared to be a shift from a stronger influence of age at the Primary level to an influence of advantage at the Secondary level. That is, while at the Primary level, Group was the only reliable predictor of performance, at the Secondary level, it was one's MA-CA disparity.

This raises several questions. Firstly, what might account for the presence of an OLA advantage in Speed of Processing at the Primary School level, but an absence of any such advantage at Secondary School level? Might, say, changes brought about by later maturing frontal brain areas explain these findings? The cognitive systems of the older Primary School children have had more time to mature compared to the cognitive systems of children in the Primary YHA group. Additionally, by Secondary School level, any age-related changes that may have occurred in neural development may have reached their peak, thus explaining the absence of Speed of Processing differences at this level. On the other hand, one might imagine that motor differences are involved in explaining the differences in these data. Recall that in the Speed of Processing task, children were tested for how quickly they read and crossed out the highest number on each line of a page containing several lines of numbers. Children's performance was assessed on their approximate time taken to complete a page (e.g., children received 5 points for finishing between 0-10 seconds, 4 points for 11-15 seconds and so on). Thus, it is possible that differences in fine motor control might exaggerate the apparent abilities of younger and older groups at the Primary School level, but show negligible differences at Secondary School level (when control over fine motor skills has already been acquired).

One challenge in understanding the source of the differences in Speed of Processing is that there does not appear to be any equivalent tasks we can compare on the measures Spitz used from the Stanford-Binet test (1916). Although three of those tasks included timed components (Rhymes, 60 words and Counting backwards), they each only imposed a deadline for completing the task and thus do not offer any continuous or discrete measures of ability that is yoked to speed.

Secondly, on the Diagnostic tests, what might explain the apparent shift between Primary and Secondary samples from an influence of age to an influence of advantage in one's performance? One possibility is that the benefit of younger children's advantage *decreases*, as they get older. But as children get older, age may also bring with it experience which can be a factor in the success of older less able children. Additionally, experience may offer older children alternative routes to achieving the same outcome. Research strategies for addressing these questions will be returned to shortly.

Overall, the results presented here point to a very different argument than the one put forward by Spitz. The data here suggest that the cognitive and intellectual abilities of MA-matched groups of different ages are on the whole very *similar*. Whatever disadvantages the OLA group experiences as a result of their lower abilities, these are compensated for by their greater age; and conversely, whatever the YHA group lacks by virtue of their lower age is compensated for by their greater ability. These findings are critical to the question at hand within this thesis, as they suggest *no difference between intelligence and cognitive development*. That is, one's level of performance may be reached equally through either greater age and lower ability, or lower age and greater ability. This interpretation could not be more different to Spitz's and thus leads us to consider how the two sets of findings can be reconciled to account for the differences in results.

In comparing some of the characteristics of the samples used here and those of Merrill's, we can see that Merrill achieved a much larger sample size – consisting of 450 children who were recruited separately from "special classes" and gifted groups. Her original samples comprised 350 lower-ability children and 100 higher ability children. From this sample, Spitz compared the abilities of a total of 69 children who were matched on a mental age of 8.0 years. In this final sample of 69 children (reflecting around 15% of the original sample size), 54 were of lower ability (mean CA 11.9) and 15 were of higher ability (mean CA=5.5). Thus, calculating the mean IQs of the two groups would give: OLA=67.2; YHA=145.5.

By comparison, within this research project the numbers of children in the Primary School YHA and OLA groups were considerably smaller, once children were matched for mental age of 8.2 years. In this thesis, Primary School MA-matched groups comprised of 14 children each (OLA mean CA=10.4 years, YHA mean CA=6.4 years) with mean IQs of 78.8 and 128.1, respectively. Although the samples achieved here were also YHA and OLA, the disparity between these groups was less extreme than Spitz reported. It is possible then that the group of OLA children Spitz used in his comparisons included children with learning disorders, or children whose development was atypical. Given that the aim of this thesis is to examine the relationship between intelligence and cognitive development in *typically developing* children, children outside the normal range (or those with known learning disorders, e.g., classified as having special educational needs; SEN) were purposefully excluded from the recruitment. The inclusion of children with possible learning disorders in Spitz's samples and the exclusion of these children in this thesis may thus account for the contrasting findings.

Another possibility is that the tests themselves could account for the differences in results seen here and those of Spitz. While the Stanford-Binet (1916) was the first revision to Binet's early test items (1906), the development of the BAS II (Elliot, et al., 1997) has profited from over 80 years of adjustments and refinements. Though over time all intelligence tests may have drawn criticisms for their power to accurately assess abilities, particularly relevant to the discussion here is the suggestion that the early versions of the Stanford-Binet test were *verbally loaded* (see e.g., Becker, 2003). In addition to problems with the administration and coding of the Stanford-Binet, Becker claims that the verbal instructions that accompanied several of the sub-tests would have unfairly disadvantaged children with a poorer grasp of language. Indeed, from Becker's analysis of the history of the Stanford-Binet test (1916; 1937; 1960; 1973; 1986; 2003), the test remained verbally loaded until its 6th revision in 1986 (Becker, 2003). Such a limitation may have favoured the younger, more articulate and well-read children in Merrill's data and disadvantaged the older ones. However, the test's verbal loading would not account for the advantages that Spitz showed in the OLA. Herein lies another potentially important difference between the Stanford-Binet (1916) and the BAS II (1997). Whereas the Stanford-Binet test (1916) included items to tap age, or maturational effects (test items such as tying shoes, counting change and correctly stating the date), there do not seem to be equivalent tasks in the BAS II (Elliot, et al., 1997). Though the BAS II includes Recall of Designs sub-test (a drawing task in which motor control might be assumed to be more advanced in older children), this task may likely also involve some component of frontal, executive control in planning the copy of that drawing from memory. Taken together, the presence of a verbally loaded Stanford-Binet (which may have advantaged Merrill's YHA group) and lack of pure age/maturational tests in the BAS II (which might have disadvantaged the OLA in this thesis), may explain the differences between the results presented here and those of Spitz.

During the analysis of this data, I flagged two other pertinent issues that deserve discussion here. To be clear, these each comprise complex and sensitive matters that cannot be fully considered here. Minimally, their discussion is aimed at highlighting some of the difficulties that could be faced recruiting and testing samples in schools. The first of these issues concerns the finding that within the Primary School OLA group there were a number of children who obtained scores *above* the scores typically associated for their age. Out of 20 children recruited to the OLA group in the Primary School data, 5 children were found to be high ability according to the BAS II measures. This was an unexpected finding given that schools were asked to provide only samples of older children who had been assessed (by teachers and/or by school tests) as *below average in their ability* and who did not have any known learning problems. This finding also appeared to surprise the Heads of the schools and the teachers of those children.

What reasons might there be for these older children performing poorly at school and yet doing so well on the BAS II? Within the samples tested in this thesis, some children experienced significant difficulties at home that may have accounted for their apparent lack of interest at school. But examples of these 'home difficulties' were varied. For example, at one school, I met two 11-year-old boys who had witnessed first-hand the violence of war. These children, who lived with guardians (who were themselves first generation asylum-seekers), may have had a range of concerns that made the demands of school life less relevant. Another 11-year-old boy who was assessed as lower ability, from an average middle class background, seemed simply disenchanted with school. However, outside school he was apparently an exceptional and talented musician. Each of these children did better on measures of the BAS II-SA than the level expected for their age. However, at school each child was performing poorly and each was assessed as below-average ability.

It is now commonly accepted that a range of factors, including social, motivational and emotional issues are important in influencing children's learning. But how exactly these factors interact to influence children's behaviour is not understood. What combination explains the difference between one under-privileged child who works hard and seizes each opportunity and another similarly under-privileged child who never tries? What combination is at play in influencing the middle class child from a stable home to become indifferent to learning and wasteful of their opportunities? For educationalists, these questions surely represent great challenges in understanding how to be effective in helping children achieve their full potential.

A second issue that was flagged in the analysis concerns the magnitude of negative MA-CA differences that were found in the Primary School and Secondary School OLA groups. Here, the data showed in the Primary OLA group there were three children with a maximum disparity of an MA approximately 4 years below their actual CA. In the Secondary School sample, the largest MA-CA disparity was an MA 7.3 years below their actual CA. These disparities are staggering given that schools were asked to supply only samples of OLA children who were lower ability but *who did not have any special educational needs (SENs)*. It is difficult to understand how children with such large MA-CA disparities could not be classified as having SEN. However, *part* of the answer may lie in an intricate balance that seems to exist for schools in: (1) receiving the funding they need for extra resources required for children with SEN; and (2) maintaining high overall standards of achievement, and thus remaining an attractive choice to prospective parents[4]. It is thus possible that, had neither of these two previous influences existed, these children showing the largest negative MA-CA disparities would have been classified as SEN. The results further indicate that excluding children with SEN will not guarantee a sample of children for whom performance is in the normal range.

Lastly, it is possible that the similarity of profiles between the YHA and OLA groups may appear different if tested at a second time point – a possible consequence of influences such as regression to the mean (e.g., Bland & Altman, 1994). Briefly, this phenomenon refers to the empirical observation that when tested on more than one occasion, children's scores tend to regress towards the mean of their group. While it may be reasonable to assume some fluctuation in children's scores at two different testing sessions, the direction of such fluctuation cannot be easily predicted. That is, would the scores of the YHA and OLA groups regress towards the mean score of peers of the same chronological age, or the same mental age? It was not possible to include a second round of testing of children within the current research project. However, future planned work is aimed at addressing this issue.

Within this chapter I have highlighted a number of questions that may be addressed to help more fully understand the relationship between intelligence and cognitive development. In the remaining chapters of this thesis some of these questions will be addressed.

Part 4 aims to assess whether the Speed of Processing difference seen in younger versus older Primary School groups indicates true differences in children's ability to process information. In this part of the thesis, MA-matched groups are

---

[4] There was a suggestion that prospective average/middle-class parents might feel less inclined to send their children to schools with 'high' proportions of children with SEN, presumably because classrooms with larger numbers of children needing additional help in their learning would detract from the attention or resources given to their own children.

compared on two cognitive tasks that tap *automatic* and *implicit information processing*. These are the Stroop task (Chapter 5) and a Word Priming task (Chapter 6).

Part 5 addresses the questions of whether the influence of one's advantage decreases with age and whether age offers alternative routes to success. Part 5 compares the performances of MA-matched groups on three separate cognitive tasks tapping higher-level reasoning and problem solving. These are the conservation of number and liquid tasks (Chapter 7), the balance scale task (Chapter 8) and the Tower of London task (Chapter 9). These chapters offer another approach to gauging the similarity of MA-matched groups by examining the relative contributions of various underlying processes to behaviour. Thus, we may be able to determine whether MA-matched groups of different ages are similar in more than just their end behaviour, but whether they are also similar in how they represent and process information.

A number of other questions fall outside of the scope of the current project but could be addressed in future work. For example, although the results presented here indicate that MA-matched children of different ages are largely equal in their abilities, there is no empirical evidence yet to support the notion that they *continue* at similar levels throughout their school careers. Thus, future studies that include a longitudinal dimension, providing periodic, but regular assessments of children's abilities would be useful in clarifying this question.

To examine the possibility that extreme sample differences may have accounted for the difference between Spitz's findings and those here, future work might also aim to compare groups of MA-matched children who display larger disparities in their mean CAs.

Finally, issues to do with schools' assessments of children's abilities and the balance between funding and servicing the needs of pupils will likely continue to present significant challenges. In spite of these challenges, the results reported in this chapter suggest that if schools are accurate in their assessments of children's abilities then the novel practice of teaching children in classes based on ability and not age, as carried out by our Secondary School, may be a viable one.

Part 4

In Part 3, the results of the BAS II, demonstrated that younger more able and older less able children matched on mental age were broadly equivalent in their overall abilities. Those results differed to what might have been predicted given Spitz's work (1982) and were more consistent with unidimensional, mechanistic accounts of within-age and between-age differences (see Chapter 2).

In Part 4, Chapters 5 and 6 continue to explore the relationship between intelligence and cognitive development by comparing the performances of MA-matched groups on two well-known on-line tasks. In Part 4, Chapter 5 presents the results of the Stroop task in which I probe for differences between younger highly able children and older less able children in their selective attention, and specifically a) their ability to *inhibit* irrelevant information and b) their *speed of response*. Chapter 6 then presents the results from a word-priming study. In that chapter, I examine the profiles of younger and older MA-matched groups in terms of the priming and *spreading activation* of words.

Each chapter begins with a brief literature review, explaining the nature of the task, its origins, the populations to whom it has been applied and the major findings that stem from it. In particular, its application to research on intelligence, or variability between children is assessed. For each chapter, the structure will be similar, presenting in turn the data from the Primary School children, Secondary School children and then Primary and Secondary Schools' results combined, on the target measures. In each case the key interest will be in determining whether there are main effects of Group, main effects of MA-CA and whether there is any interaction between Group and MA-CA, in addition to determining the effects of the experimental manipulation (e.g., task and condition). Therefore, in Primary and Secondary School results sections, analyses begin with categorical comparisons of MA-matched groups and are then followed by continuous comparisons on each dataset. In the full dataset analysis, all data across age group are collapsed into a single analysis. The disparity between each child's MA and their CA in months is used as a covariate to explore whether this continuous dimension modulated patterns on the dependent variables. For this covariate, the older children were characterised predominantly by negative values (because their CAs exceeded their MAs) and the younger children were characterised by positive values (because their MAs exceeded their CAs). However, this analysis also reflects the variation present within each group.

Chapter 5 The Stroop task

Introduction

In 1935, a dissertation paper by John Ridley Stroop was published examining the strength of associations between reading words and naming colours. While previous studies had shown that naming colours typically took longer than reading words (e.g., Ligon, 1932; Telford, 1930), Stroop was the first to demonstrate *interference* between these behaviours. He showed how interference could be manipulated on colour-naming tasks when a word's meaning was *incongruent* with the colour of the ink in which it was printed (Stroop, 1935). For example, the time taken to correctly name the colour of the incongruent word item **RED** printed in green ink, is typically longer than the time taken to name a patch of green. Stroop interpreted these results as interference between two learned behaviours. Naming colours and reading words are products of training, each comprised of different strengths of associations. Stroop argued that word reading is fast, automatic and highly practised, while colour naming requires more attention in choosing one of several alternative names.

This effect, referred to as the Stroop effect, is now viewed more generally in that any time that a stimulus comprises more than one dimension and the experimental task requires attending to just one dimension, responses may show a separation. Patches of colour are described as *neutral* because in naming their colour, only one dimension is presented and thus only one response need be generated.

Over the years a range of modifications have been made to Stroop's original design. One revision commonly found within studies examining the Stroop effect involves turning the presentation of stimulus into a sequential design, rather than presenting items all at once. Combined with more precise response time measurements, this method offers greater accuracy in evaluating potential differences in processing the different word items. A second revision commonly used in experiments using the Stroop task concerns the addition of an extra condition. This condition combines word items with colours, such that the colour of ink is *congruent* with the word meaning (e.g., the word item **BLUE**, in blue ink). Thus, in studies using the Stroop task, single presentations of neutral, incongruent and congruent items are frequently found.

Compared to modern behavioural techniques (capable of measuring response times at millisecond accuracy) Stroop's techniques were relatively rudimentary. However, similarly robust differences in response times have been found in many subsequent replications using more precise response time measurements and looking only at correct performance (for a review, see MacLeod, 1992). For instance, the time taken to respond to neutral items remains the benchmark for each child's colour naming performance. Typically, responses are relatively fast for these items. Responses for incongruent items remain the slowest relative to the other items, and the difference between the time taken to name colours of neutral and incongruent items offers a measure of interference. Lastly, in those studies that include congruent word items, response times are generally found to be the fastest, compared to neutral and incongruent items (see e.g., Kane & Engle, 2003). Word meaning appears to aid faster naming for items when their meaning is congruent with the colour of ink it appears in, even when the task is only to name colours. This effect has been termed *facilitation* and it is measured as the difference between neutral and congruent response times. Figure 5.1 illustrates a pattern of RT data consistent with these effects on the Stroop task. The red line represents the Naming Colour task on each of the three conditions (Neutral, Incongruent and Congruent). This figure shows that naming times for Congruent words are faster than for the Neutral baseline and slowest for Incongruent. Response times for reading words, here represented by the blue line, are only marginally affected in the Incongruent colour condition.

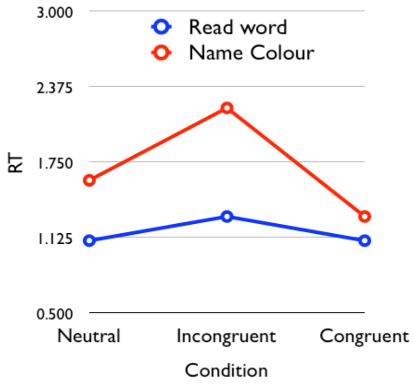


Figure 5.1. Illustrating a typical pattern of response times for Reading words and Naming colours on the Stroop task.

Several theories have emerged over the years to account for the Stroop effect (for a review, see MacLeod & Dunbar, 1988). For example, according to the *speed of* *processing* theory (see Posner & Snyder, 1975) the effect is explained by (1) a difference in the relative speed at which information is processed in parallel for words vs. colours; and (2) a single channel for output that the processing streams compete towards. The analogy that has been applied to this theory is one of a horse race, where horses race in parallel and without any interference, but where the track narrows to a single lane in which only one horse can enter and therefore win.

The speed-of-processing hypothesis is closely linked to the *automaticity hypothesis* in which it is argued that because we have greater practice of reading words, reading becomes an automatic behaviour while naming colour requires effortful control (MacLeod, 1991). The processing of information on one dimension requires more attentional control than attending to the information on the other. In contrasting these theories on their ability to account for a range of empirical phenomena, MacLeod (1991) concluded they were both lacking. He proposed that a parallel processing account, in which information was processed from both the relevant and the irrelevant dimensions, was more accurate in describing the Stroop patterns (MacLeod, 1991). Each of these theories were subsequently formalised in a range of computational implementations (Altmann & Davidson, 2001; Cohen, et al., 1990; Phaf, et al., 1998).

One thing that different accounts of the Stroop effect have in common is that the ability to selectively attend to the colour dimension on the Stroop task (i.e., naming the colours of incongruent word items, when the task is to name colours) is assumed to require the participant to successfully *inhibit* the irrelevant dimension (e.g., information relating to word meaning must be inhibited on the colour naming task. See MacLeod, 1991). Inhibition has been argued elsewhere to be one of several key 'executive functions' used to "*modulate the operation of various cognitive subprocesses and thereby regulate the dynamics of cognition*" (Miyake, et al., 2000). Consequently, the use of the Stroop task has become popular within a range of studies exploring the relationship of inhibition to typical and atypical development. For example, within typically developing groups it has been used to gather evidence of the general changes in the ability to inhibit as one grows older (see e.g., Bub, Masson, & Lalonde, 2006). Within atypical groups the task has been applied to the study of addiction (Cox, Fadardi, & Pothos, 2006), children and adolescents with ADHD (MacLeod & Prior, 1996; Martel, Nikolas, & Nigg, 2007; Rubia, Smith, & Taylor, 2007), anxiety disorders (Benoit, et al., 2007), conduct disorder (Herba, et al., 2006), learning disorders (Nichelli, et al., 2005), and personality and psychiatric disorders (see e.g., Chen, et al., 2001; MacLeod & Prior, 1996). Facilitation, on the other hand has been linked to problems of *goal neglect*[1]. For example, schizophrenics who often exhibit goal neglect on tasks like the Stroop task have shown greater levels of facilitation compared to normal controls (see e.g., Taylor, Kornblum, & Tandon, 1996). Additionally, in many of the above examples, evidence from neuroscience strongly links the maturation of the prefrontal cortex with the development of executive functions (see e.g., Braver, et al., 2006). For example, evidence has been reported linking the abnormal development of the prefrontal cortex with inhibitory difficulties assessed by perseveration tasks in some disorders, such as Down's syndrome (see e.g., Rowe, Lavender, & Turk, 2006)[2].

Current aims

Using the Stroop task with groups of younger and older MA-matched children is relevant to this thesis for the following reasons. First, the task offers millisecond level measurements of information processing, which may help shed further light on the nature of the speed of processing differences highlighted in Chapter 4. Do MA-matched groups differ in their accuracy or in their speed of response on the Stroop task? Second, the task offers an on-line measurement[3] of facilitation and interference – measures that may tap components of executive functions, such as goal neglect and inhibition. Examining the pattern of processes underlying selective attention may reveal subtle differences in the ways information is processed where overall speed and absolute values such as accuracy might not. Thus, the use of the Stroop task offers the possibility of determining whether MA-matched children of different ages show evidence of differences in inhibition or goal neglect. Third, the Stroop task offers a means of examining in parallel two separate accounts that argue causal roles for inhibition in producing variation in cognitive ability. Specifically, if differences

---

[1] Goal neglect refers to a problem whereby some aspects of a task are ignored, even though participants show awareness of them.

[2] The task that Rowe et al. (2006) used was based on Luria's preservation task (1980) and required participants to tap a table once if the experimenter tapped twice and tap twice if the experimenter tapped once.

[3] The advantages of on-line tasks were discussed previously in Chapter 2, for their ability to reveal the relative contributions of underlying local processes to overt behaviour.

in the ability to inhibit account for within-age differences in intelligence, then we might expect MA-CA disparity to reliably predict measures of interference, and groups matched on overall MA to demonstrate identical interference effects. On the other hand, if differences in inhibition account for between-age differences in cognitive ability, then we might expect Group to reliably predict interference. Fourth, because inhibition has been linked to the later development of the prefrontal cortex, this might lead us to expect that participant's age offers the most reliable indicator of performance. However, it is not clear how greater ability (i.e., each participants MA-CA advantage) may modulate the ability to inhibit irrelevant information. For these reasons the Stroop task is ideally suited to provide an insight into MA-matched groups who have reached this overall level of ability either through greater advantage or greater age.

Method

Participants

See ‘Participants’ section, Chapter 3 General Methodology. All participants, bar one, completed this task. One child from the YHA group (participant number 31) was excluded from the analyses. The experimenter abandoned the task after two blocks when this child showed no interest in responding. This was the only time a participant appeared unengaged with the task.

Design

Primary School and Secondary School samples were analysed separately using mixed-design ANOVAs and then combined in a mixed ANCOVA. In the separate analyses, Group was the between-participants factor (YHA vs. OLA) and Condition (Neutral, Incongruent, Congruent) and Task (Read Word vs. Name Colour) were the within-participants factors. Groups were compared on two dependent variables (Accuracy and RT) and two specific contrasts: Facilitation ($RT_{neutral}$ - $RT_{congruent}$) and Inhibition ($RT_{neutral}$ - $RT_{incongruent}$). In the combined analysis, Group was transformed to consist of four levels (Primary-Younger, Primary-Older, Secondary-Younger, Secondary-Older) and added as a between-groups factor to the mixed ANCOVAs.

Procedure

Items were displayed individually and in sequence every 5000 milliseconds on a computerised display, and digital recording of verbal responses were used for later coding of accuracy and response time[4]. Participants were seated directly in front of the display, at a distance of approximately 30-50cm. Each participant was administered 4 blocks of 15 items. The order of blocks was rotated, with half the participants completing the task in the order of (1) Read-Word; (2) Name-Colour; (3) Read-Word; (4) Name-Colour, and the other half in the order of (1) Name-Colour; (2) Read-Word; (3) Name-Colour; (4) Read-Word. Within each block, items were presented in random order, but with an equal total frequency over the entire task.

---

[4] Audio recording and off-line coding were chosen in preference to a voice-activated microphone because it was believed to be less prone to data-loss, caused by background noises in schools.

Stimuli were presented on screen in 48-point Ariel bold font. The colour-words used were RED, GREEN, BLUE, PINK and YELLOW and the colours that words appeared in were red, green, blue, pink and yellow. There were a total of 60 items: 30 items in the Read Word task: 10 x Neutral (black ink), 10 x Congruent, 10 x Incongruent; and 30 items in the Name Colour task: 10 x Neutral (colour patches), 10 x Congruent, 10 x Incongruent.

Unfortunately, a technical error led to just two of the neutral colour naming items being presented - these were for red, and blue patches of colour (the possible consequences of this are discussed later). While standard comparisons between neutral and congruent and neutral and incongruent conditions were retained, analytically, this led to a focus on the difference between Congruent and Incongruent conditions as a direct measure of inhibitory control.

In all conditions, participants were instructed to respond verbally and as quickly and as clearly as possible. Responses were scored for accuracy by the experimenter and response times were recorded digitally via the computer that presented the stimuli. A timeline illustrating one trial is given in Figure 5.2. This figure shows that: (1) an auditory signal was presented for 1000ms (this served as an alert for the participant and was used as the marker for later coding RT data); (2) the stimulus was presented on screen 1000ms after the auditory signal stopped (in the figure, this point is represented by the vertical, black dashed-line); (3) the participant responded by either naming the colour or reading the word.

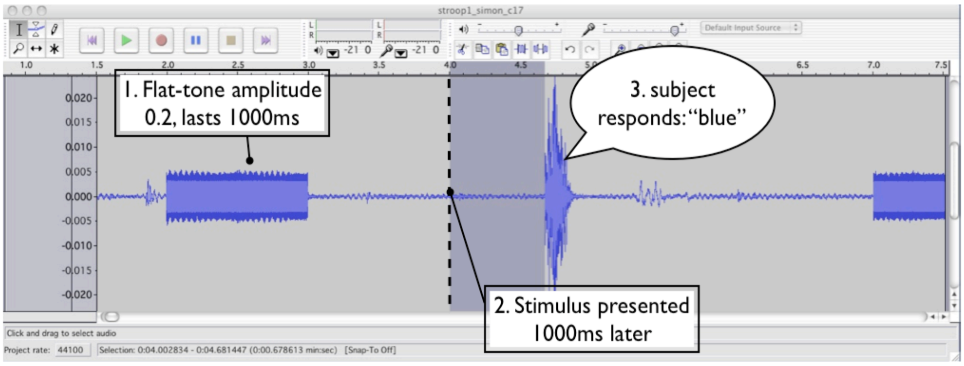


Figure 5.2. Screenshot showing the timeline of the Stroop presentation and data recording in Audacity®. (1) The trial begins and a tone is presented for 1000ms; (2) after a delay of a further 1000ms a stimulus is presented; (3) the participant responds.

As Figure 5.2 illustrates, the time it took an individual to respond to the stimulus was the total time between the end of the auditory marker and their actual response, minus the 1000ms interval between the end of the marker and the stimulus presentation (represented by the dark-grey area). Each item remained on screen for a fixed 4000ms.

Materials

The experimental tasks were coded in MATLAB^{TM} by Frank Baughman and required the use of a touchscreen display, a digital microphone and external speakers attached to an Apple Mac G4 iBook laptop computer (1.33 GHz, 1GB RAM).

Primary School results

Results are first presented for the MA-matched YHA (n=13) and OLA (n=14) groups, then the full sample (total=39; younger n=19, older n=20).

Primary School MA-matched group comparisons

Accuracy: Overall accuracy was at ceiling for both the YHA (mean accuracy=97.9%, se=.70) and the OLA groups (mean accuracy=99.5%, se=.60). Figure 5.3 shows accuracy for YHA and OLA groups across the three conditions (Neutral, Incongruent and Congruent) for both tasks (Read Words vs. Name Colours). While some variability can be seen in the YHA group's performance on the Name Colours task for Neutral and Incongruent word items, mean accuracy remained above 90%. Ceiling scores of this nature render the data inappropriate for statistical analysis, thus only RT data were analysed.

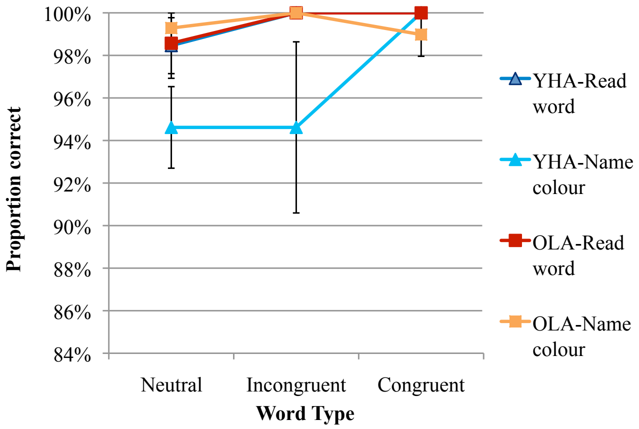


Figure 5.3. Mean accuracy on Stroop for MA-matched YHA and OLA Primary School groups. Error bars show standard errors of the mean.

Response times: A 2x2x3 ANOVA was performed on the RT data with Group (YHA vs. OLA) as the between-participants factor and Task (Read Word vs. Name Colour) and Condition (Neutral, Incongruent, Congruent) as the within-participants factors[5]. This analysis revealed a main effect of Group (F(1,25)=12.83, p=.001, η^2=.339), which stemmed from the OLA responding comparatively faster (mean RT=1.09s,

---

[5] The outcome of the second analysis on log-transformed RT data did not differ from the outcome of results on un-transformed data.

se=.06) than the YHA (mean RT=1.39s, se=.06). A significant interaction was found between Group and Task (F(1,25)=4.41, p=.046, η^2=.150). This was due to the YHA showing relatively larger changes in RT over Task (Name Colour mean RT=1.47, se=.07; Read Word mean RT=1.31, se=.07), with the OLA showing relatively little change (Name Colour mean RT=1.09, se=.06; Read Word mean RT=1.08, se=.07). However, no reliable interaction was found for Group x Condition, or Group x Task x Condition, suggesting similar relative changes between YHA and OLA for the within-participant variables. This analysis did reveal a reliable Task x Condition interaction (F(2,50)=12.44, p<.001, η^2=.332), indicating that the experimental manipulation of Task and Condition had differential effects on RT. Separately, a main effect of Task (F(1,25)=5.23, p=.031, η^2=.173), reflected that the overall RTs for Read Words (mean RT=1.20, se=.05) and Name Colours (mean RT=1.28, se=.05) were reliably different. A main effect of Condition was also found (F(2,50)=57.15, p<.001, η^2=.696). This stemmed from overall faster RT for Neutral (mean RT=1.01, se=.04) compared to Incongruent (mean RT=1.40, se=.05) and Congruent (mean RT=1.31, se=.05) items. The mean RT data for YHA and OLA on each of the Stroop conditions are presented in Figure 5.4. This figure shows that in YHA and OLA groups, performance in both Read Word and Name Colour tasks was fastest in the Neutral condition.

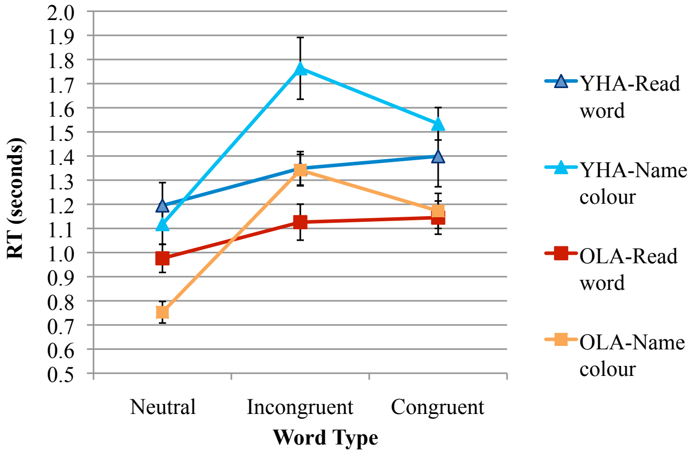


Figure 5.4. Mean RT on Stroop for MA-matched YHA and OLA Primary School groups. Error bars show standard error of the mean.

The pattern of RT found here differs from the pattern illustrated earlier in Figure 5.1, in which fastest response times are typically shown for Congruent items. This disparity is most likely due to the technical error that resulted in a limited number of items in the experimental set. Because facilitation is measured against the baseline of neutral responses, this led to the slightly odd situation of negative facilitatory effects. However, the important contrast is the size of the difference between congruent and neutral and incongruent and neutral for each group. Since the neutral condition emerged as faster in both groups, the key contrast should be unaffected by the divergence from the normal pattern of the Stroop response times. However, by taking the difference of RTs between Congruent and Incongruent items, we can obtain a measure of inhibitory control that is not dependent on RT in the Neutral condition.

Facilitation and interference: Taking the differences in RT between Neutral and Congruent (Facilitation) and Neutral and Incongruent (Interference), a further one-way ANOVA found no evidence of reliable between-group differences. Figure 5.5 below depicts these measures for YHA (blue bars) and OLA (red bars) for Interference (upper two bars) and Facilitation (lower two bars). Thus, although the older group responded more quickly than the younger group, the relative sizes of the interference and facilitatory effects were not different.

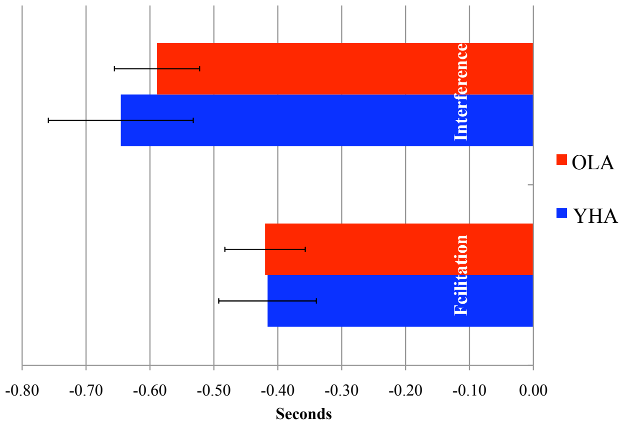


Figure 5.5. Mean interference and facilitation effects for MA-matched YHA and OLA Primary School groups. Error bars show standard error of the mean.

Primary School full sample analysis

Response times: In the previous ANOVA on the MA-matched data, Group was found to account for approximately 34% of the variance in RT scores. Using the full Primary School sample in a repeated measures ANCOVA, with MA-CA disparity as the covariate, Group as the between-participants factor, and Task and Condition as the within-participant factors, a reliable main effect was found of Group on RT

(F(1,35)=6.41, p=.016, η^2=.155). However, the proportion of total variance accounted for in this model was less (around 15%). This analysis revealed no main effect of MA-CA disparity and no Group x MA-CA disparity interaction, indicating that children's response times when taken together and also within each group, were not differentially modulated by their advantage. These results are illustrated in Figure 5.6 (Tiles A-D). Here, effects of age can be observed for Neutral (Tile A), Incongruent (Tile B) and Congruent (Tile C) word items. Tile D displays each child's MA-CA disparity by the difference in their response times for Incongruent versus Congruent items. In this tile, no clear effects of Group or MA-CA are apparent.

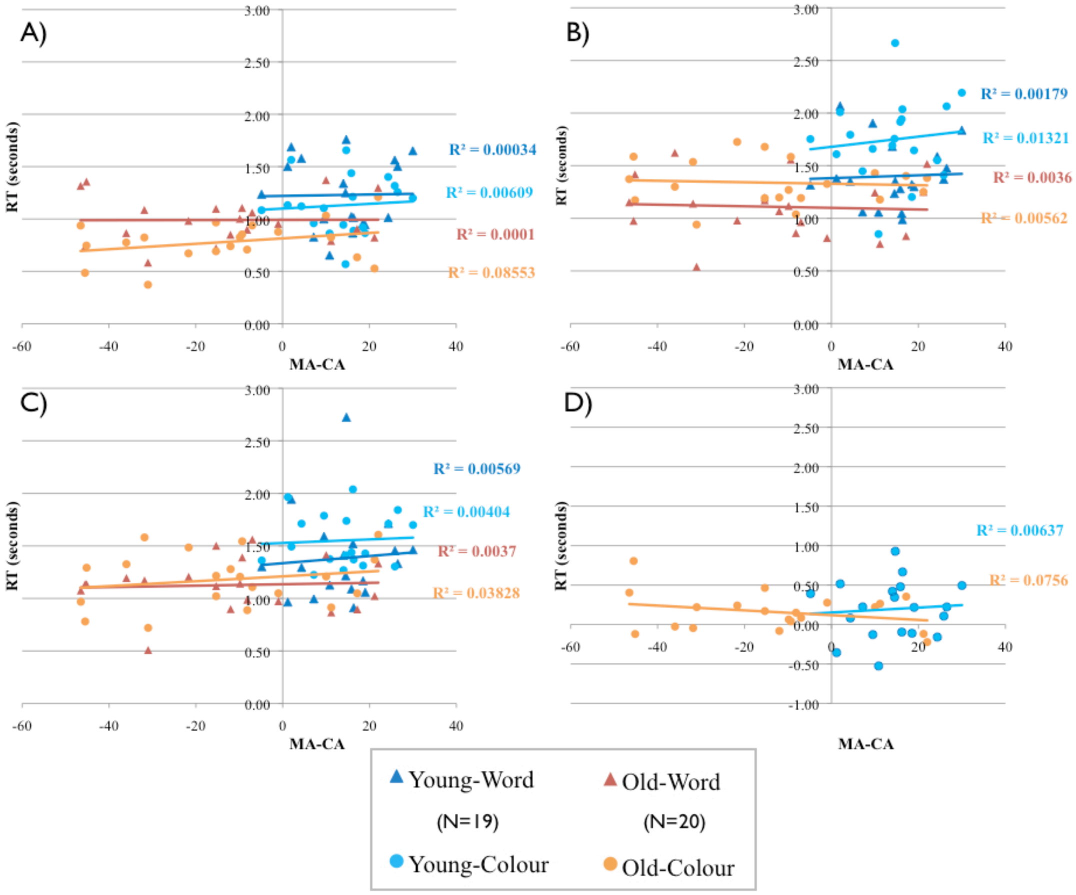


Figure 5.6. Full Primary School sample MA-CA disparities by RT for Neutral (Tile A), Incongruent (Tile B), Congruent (Tile C) word items, and (Tile D) Difference scores for Incongruent – Congruent RTs.

<u>Facilitation and interference</u>: The results of a one-way ANCOVA, with MA-CA disparities as the covariate and Group as the between-participant factor revealed that neither Group nor MA-CA disparity were reliable predictors of Facilitation and

Interference scores. Figure 5.7 depicts trendlines that are not significantly different to zero for the younger and older groups on Facilitation (Tile A) and Interference (Tile B).

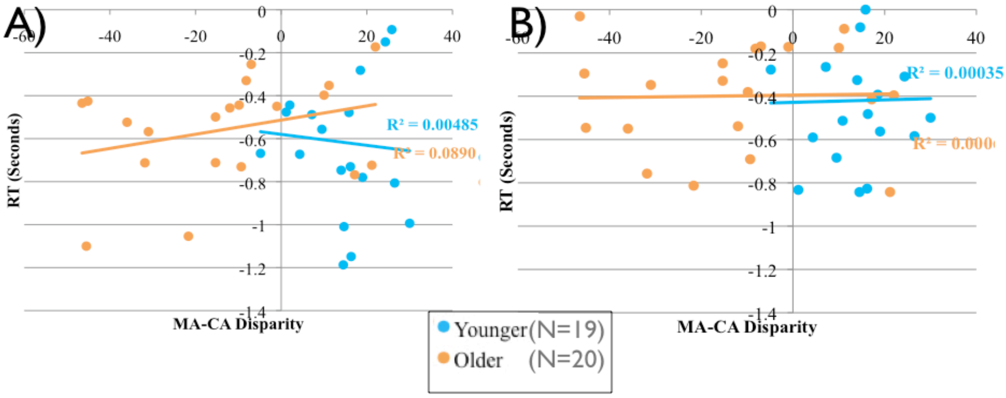


Figure 5.7. MA-CA disparity by Interference (Tile A) and Facilitation (Tile B) for younger and older Primary School children.

Summary

For the MA-matched groups, accuracy was comparable and close to ceiling. On response times, the OLA group was faster than the YHA group. However, facilitation effects and interference effects of each group were of the same relative size. In the full sample analysis, MA-CA disparity scores did not modulate children's response times. To the extent that interference reflects reduced levels of inhibition and facilitation reflects reduced goal neglect, these processes appeared at the same level of efficiency once MA was matched. However, chronological age seemed to predict faster naming responses.

Secondary School results

Secondary School MA-matched group comparisons

Accuracy: As was the case in the Primary School data, the accuracy of YHA (mean accuracy=99.1%, se=.07) and the OLA (mean accuracy=98.9%, se=.06) in the Secondary School sample was at ceiling (see Figure 5.8). Overall performance was observed to fall slightly in both groups on the Incongruent Name Colour condition. This was due to a total of 10 errors in the younger group and 12 errors in the older group. Again, these data were not suitable for analysis, therefore only RT data were used.

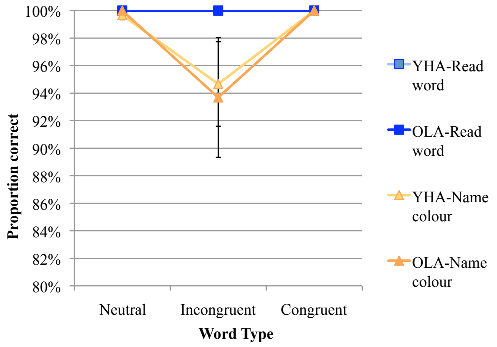


Figure 5.8 Mean accuracy on Stroop for MA-matched YHA and OLA Secondary School groups. Error bars show standard error of the mean.

Response times: Overall the OLA were marginally faster at responding (mean RT=0.94s, se=.04) compared to the YHA (mean RT=0.98s, se=.04). However, in contrast to the Primary School results, no reliable main effect of Group was found using a 2x2x3 ANOVA on the RT data with Group (YHA vs. OLA) as the between-participants factor and Task (Name Colour vs. Read Word) and Condition (Neutral, Incongruent, Congruent) as the within-participants variables.

Similarly to the Primary School results, the analysis did reveal a main effect of Task (F(1,33)=9.74, p=.004, η^2=.228) that stemmed from faster overall RT for Read Words (mean RT=0.93, se=.03) compared to Name Colours (mean RT=0.99, se=.03) and a main effect of Condition was also found (F(2,66)=180.50, p<.001, η^2=.845). This was due to faster overall RT for Neutral (mean RT=0.75s, se=.03) compared to Incongruent (mean RT=1.12s, se=.04) and Congruent (mean RT=1.01s, se=.03) items. A significant interaction was observed for Task x Condition (F(2,66)=64.39, p<.001, η^2=.661). However, the absence of interactions involving Group (i.e., Group x Task, Group x Condition, or Group x Task x Condition) indicated that Task and Condition had similar effects in both YHA and OLA groups. Figure 5.9 below presents the mean RT data for Secondary School YHA and OLA groups on each of the Stroop conditions.

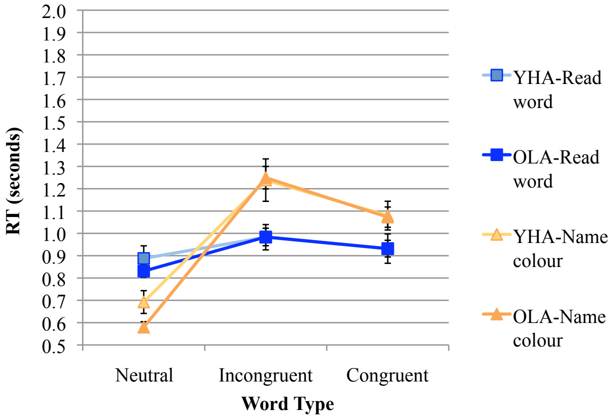


Figure 5.9. Mean RT on Stroop for MA-matched YHA and OLA Secondary School groups. Error bars show standard error of the mean.

Facilitation and interference: Figure 5.10 contrasts overall measures of Facilitation (lower two bars) and Interference (upper two bars) for YHA (blue bars) and OLA (red bars). Although this figure suggests that the OLA had both higher interference and facilitation, a one-way ANOVA found these differences were not reliable. Thus, MA-matched comparisons of response time data in the Secondary School groups revealed the younger and older groups to be statistically indistinguishable.

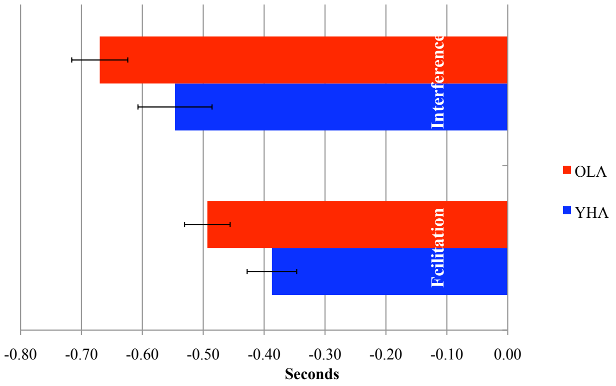


Figure 5.10. Mean interference and facilitation effects for Naming Colours on the Stroop for MA-matched YHA and OLA Secondary School groups. Error bars show standard error of the mean.

Secondary School full sample analysis

Response times: Using MA-CA disparities as the covariate, the results of a 2 x 2 x 3 repeated measures ANCOVA on RT revealed no main effects of Group or MA-CA disparity, and no Group x MA-CA interaction. This differed from the Primary School results, where Group was found to reliably modulate the RT of younger versus older children. Figure 5.11 shows each child's MA-CA disparity scores and their response time on each condition. The figure shows separate near-zero trendlines representing the younger and older groups on Neutral (Tile A), Incongruent (Tile B) and Congruent (Tile C) word items. Tile D displays the difference between children's Incongruent versus Congruent items and MA-CA difference. These tiles show no effect of Group or MA-CA disparity.

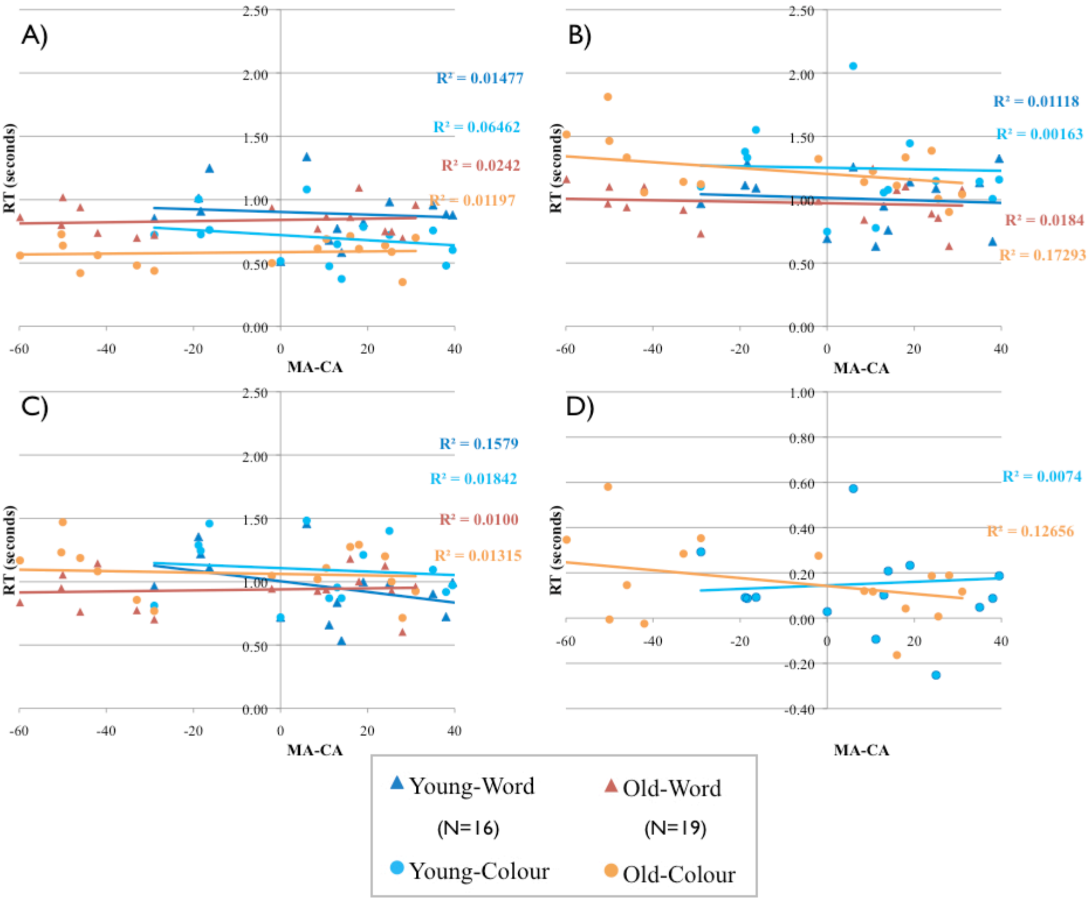


Figure 5.11. Full Secondary School sample RT by MA-CA disparities for Neutral (A), Incongruent (B) and Congruent (C) word items. Tile D depicts Incongruent – Congruent RT.

Facilitation and interference: Figure 5.12 plots each child's MA-CA disparity by measures of Interference and Facilitation (Tile B). Though the figure suggests a possible relationship between MA-CA and RT (see Older group in Tile A), the results of a one-way ANCOVA also showed that neither Group, nor MA-CA disparity, nor their interaction reliably modulated the sizes of these effects.

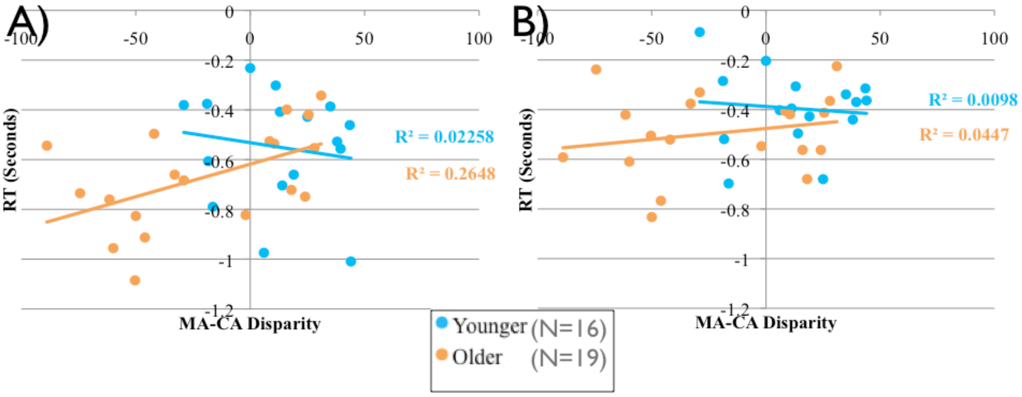


Figure 5.12. MA-CA disparity by Interference (Tile A) and Facilitation (Tile B) for younger and older Secondary School children.

Summary

For MA-matched groups, the performances of Secondary School younger and older children were indistinguishable in terms of their speed and accuracy. Where the full sample was used, neither Group nor MA-CA disparity scores predicted response times on any of the conditions for Read Word, or Name Colour.

Combined Primary and Secondary School results

In this section (and within all subsequent chapters where the results for the combined Primary School and Secondary School data are presented), Group is comprised of four levels: Primary-Younger; Primary-Older; Secondary-Younger; and Secondary-Older.

Response times: Using all data from the Primary School and Secondary School samples with MA-CA disparity as the covariate, a 4x2x3 repeated measures ANCOVA was performed on RT data. The between-participants factor was Group (Primary-Younger, Primary-Older, Secondary-Younger, Secondary-Older) and the within-participant factors were Task (Read Words vs. Name Colours) and Condition (Neutral, Incongruent, Congruent). This analysis revealed a reliable main effect of Group ($F(3,66)=6.92$, $p<.001$, $\eta^2=.239$). Figure 5.13 shows this effect stemmed from slowest overall RT in the youngest age group (Primary-Younger mean RT=1.37s, se=.10), followed by the next oldest group (Primary-Older mean RT=1.10s, se=.05), then by Secondary-Younger (mean RT=1.00s, se=.06), with the fastest overall RT in the oldest group (Secondary-Older mean RT=0.94s, se=.05).

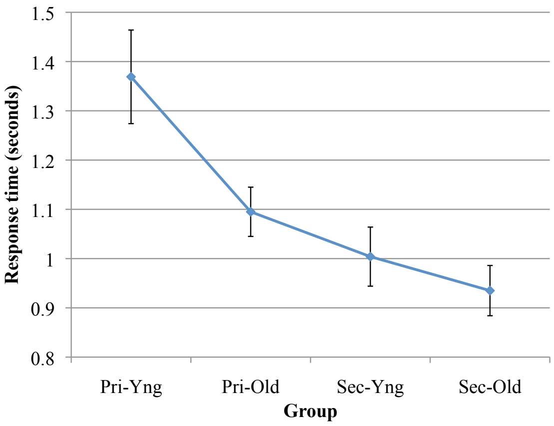
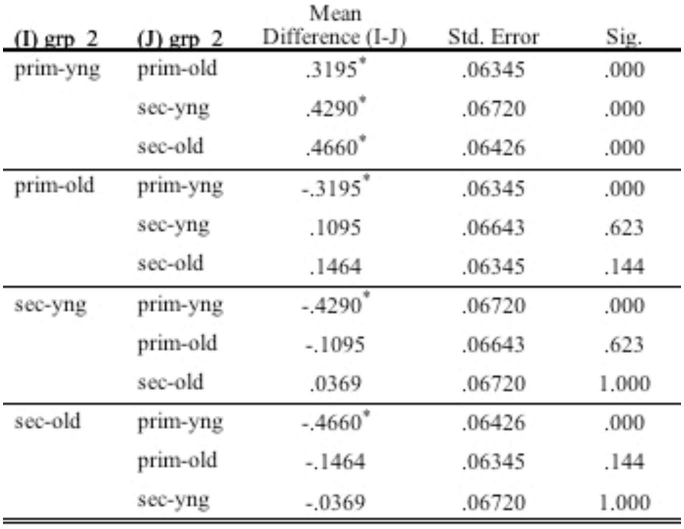


Figure 5.13. Overall mean response times for combined Primary and Secondary School groups (Pri-Yng=Primary-Younger, Pri-Old=Primary-Older, Sec-Yng= Secondary-Younger, Sec-Old=Secondary-Older).

Post-hoc tests (provided in Table 5-1) showed that the reliable main effect of Group was entirely due to the overall slower RT in the Primary-Younger group.

Table 5-1. Post-hoc tests of mean differences in RT between Group (Primary-Younger, Primary-Older, Secondary-Younger, Secondary-Older).



(I) grp_2	(J) grp_2	Mean Difference (I-J)	Std. Error	Sig.
prim-yng	prim-old	.3195*	.06345	.000
	sec-yng	.4290*	.06720	.000
	sec-old	.4660*	.06426	.000
prim-old	prim-yng	-.3195*	.06345	.000
	sec-yng	.1095	.06643	.623
	sec-old	.1464	.06345	.144
sec-yng	prim-yng	-.4290*	.06720	.000
	prim-old	-.1095	.06643	.623
	sec-old	.0369	.06720	1.000
sec-old	prim-yng	-.4660*	.06426	.000
	prim-old	-.1464	.06345	.144
	sec-yng	-.0369	.06720	1.000

Again, MA-CA disparities failed to reliably predict RT, indicating that age most reliably modulated performance on the Stroop task. In the full data set, the analysis showed main effects for Condition (F(2,132)=108.20, p<.001, η^2=.621) and Task (F(1,66)=7.23, p=.009, η^2=.099) and a reliable Condition x Task (F(2,132)=26.51, p<.001, η^2=.287) interaction, demonstrating that the experimental manipulations were effective within the entire sample of children spanning an age-range of approximately 11 years (i.e., 5-16 years-old). These were the only reliable within-participant effects found.

Facilitation and interference: Neither Group, nor MA-CA disparity were found to be reliable predictors of Facilitation or Interference. Figure 5.14 plots each child's MA-CA disparity by the time taken to respond on the Name Colour task for Neutral (Tile A), Incongruent (Tile B) and Congruent (Tile C) word items. This figure illustrates the effect of age found in the analysis. Irrespective of children's advantage, RT appears fastest overall for the oldest children (Sec-Old), then the younger Secondary School children (Sec-Yng), followed by the older Primary School children (Pri-Old). The youngest group (Pri-Yng) is the slowest to correctly name colours of word items.

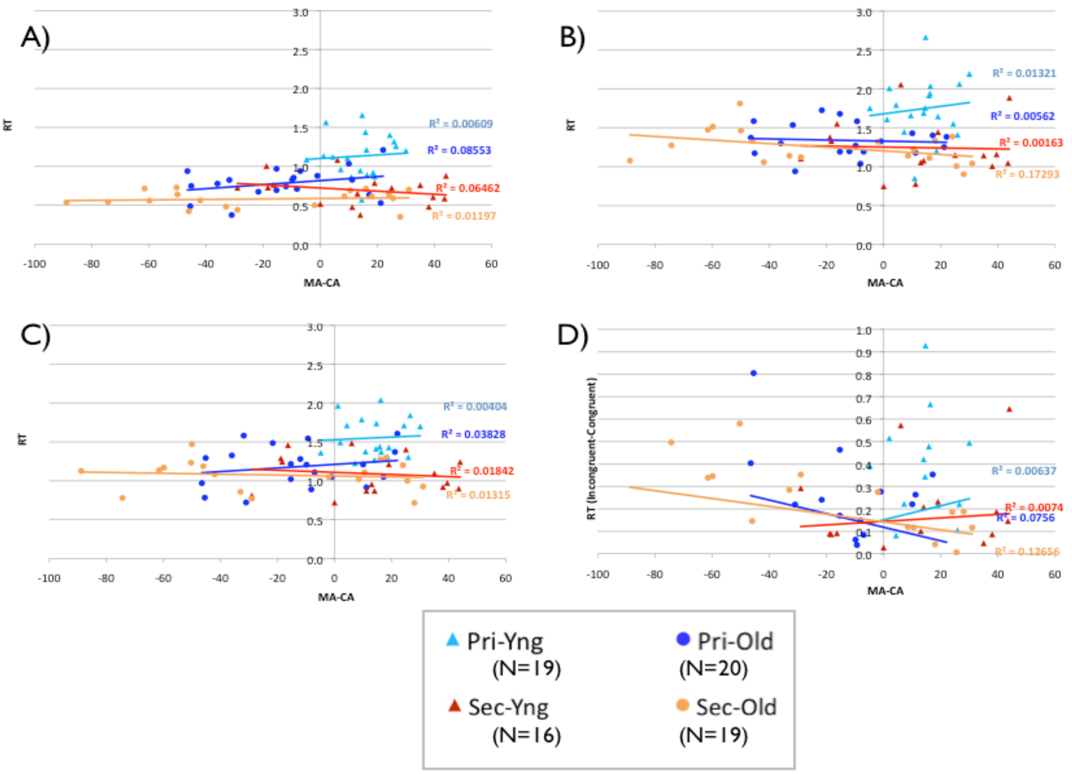


Figure 5.14. Primary and Secondary School samples RT by MA-CA disparities on Name Colour task for Neutral (A), Incongruent (B) and Congruent (C) word items. Tile D depicts Incongruent-Congruent RT differences.

Summary

The results of the analysis using the combined Primary School and Secondary School data revealed that participants' age, and not their advantage was important in influencing performance on the Stroop task. That is, Group reliably modulated performance, but MA-CA disparity did not. Furthermore, the results showed no evidence of any interaction between MA-CA disparities and Group.

Discussion

The Stroop task was employed to examine possible information processing differences between younger and older groups of MA-matched children. The following questions were set out for study: (1) do these groups differ in their overall *accuracy* or *speed of response*? (2) What evidence is there of group differences in *goal neglect* or *inhibition*? (3) What evidence is there that differences in inhibition might explain *within-age* differences in intelligence? And, finally, (4) what evidence is there that differences in inhibition might explain *between-age* differences in ability?

To address these questions, groups of younger higher ability children and older lower ability children at Primary and Secondary School levels were compared in their accuracy, response times and on measures of interference and facilitation for naming colour words. While accuracy was found to be at ceiling in each group and at both school levels (and thus did not offer the means of discriminating ability between groups), a clear advantage in speed of response was shown by the older children, both in the MA-matched analysis and the whole sample Primary School analysis. These results mirrored earlier results in Chapter 4, in which the OLA were shown to be reliably faster than the YHA on the Speed of Processing sub-test (BAS II-SA; Elliot, et al., 1997). On the Stroop task, MA-CA disparity failed to reliably modulate response times, suggesting that speed of response varies independently of cognitive ability, but not of age. However, as was also seen in the Speed of Processing sub-test (Chapter 4), on the Stroop task, the reliable response time differences between Primary younger and older children was not present at Secondary School level. At this school level, neither age nor advantage reliably modulated performance and groups were statistically indistinguishable from each other.

On measures of interference and facilitation, the results showed no reliable group differences (at either school level) in the relative effects of neutral, incongruent and congruent word items on the speed of naming colours. That is, irrespective of the speed at which children in each group were responding they showed the same levels of interference and facilitation. The interpretation of these results is made more challenging given the presence of the speed of response differences between groups at the Primary School level.

In examining the relationship between MA-CA disparity and interference, the results of the separate and combined school analyses revealed no reliable effect of advantage. Accordingly, these data offer no support for the claim that differences in the ability to inhibit may account for within-age differences in intelligence. Likewise, however, the results also offer no support for the alternative claim that differences in the ability to inhibit may account for differences in ability between-ages. Results also showed that age group did not modulate children's interference, either in separate, or combined school analyses.

The findings presented in this chapter raise a number of questions. Firstly, the faster performance of the OLA over the YHA in the Primary School data suggests that age is influential in predicting performance on the Stroop task. But, what accounts for the absence of any such effect at Secondary School? The power-law function depicted in Figure 5.15 offers a suitable characterisation of the data presented in this study. For example, it shows larger gains in speed are obtained during earlier ages and only small gains in speed at later ages. However, a number of different causes might underlie this generalised pattern.

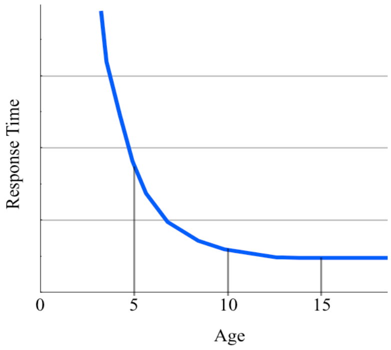


Figure 5.15. A power-law function showing non-linear reduction in RT gain over age.

One possible explanation is that, say by age 10, the information processes underlying colour naming have reached their peak and thus the speed of information processing is identical between younger and older groups. Alternatively, it is possible that differences in speed of processing *do* exist between the Secondary School younger and older children, but the task loses sensitivity at these ages. That is, both the younger and older children at Secondary School level may be demonstrating that faster verbal responses cannot be accurately made on these items and thus both groups are at ceiling on the task. Another possibility is that the relationship between intelligence and cognitive development does genuinely change across age such that development and intelligence are different at Time1 and the same at Time2. Yet another possibility is that the disappearance of group differences between Primary School level and Secondary School level reflect differences in verbal motor speed and not any form of informational processing.

If we assume first that verbal motor speed differences do not account for the faster response times of the older Primary School children, what does the same pattern of interference indicate vis-à-vis the relationship of development to individual differences? Had MA-CA disparity been found to predict interference, then this would suggest that cognitive ability influences the development of inhibition. Yet, neither Group nor MA-CA was predictive of interference. Instead, the results suggest that the ability to inhibit develops independently of age and ability, thus leading to the possibility that other factors are responsible for the differences found. Indeed, the literature shows that a number of studies have found speed advantages for normal controls compared to children with ADHD, but no differences in interference (for a review see e.g., van Mourik, Oosterlaan, & Sergeant, 2005). Thus, if the Stroop task targets inhibition in later-maturing frontal cortex, the current findings appear to present a challenge for both individual difference theories and cognitive development theories that claim a role of inhibition in accounting for variability in intelligence.

Several issues that emerged during the analysis of this task may need to be addressed in order to clarify the findings presented here. Firstly, it is possible that the pattern of results found were influenced by the absence of data for 3 items in the colour naming neutral task. Recall that these patches of colour provide the basis for calculations of interference and facilitation scores. During the testing, an error in the program controlling the task resulted in random interleaving of only 2 of the 5 colour patches (green, pink and yellow patches were not shown). Because participants had to choose from a narrower set of responses on this task, it seems feasible that this is the cause of the faster response times for neutral items. While response times did show variability in this condition, clearly data for the full item set are needed in order to rule out this error as a potential confound. Secondly, it seems critical that future work should establish the extent to which differences in verbal motor speed account for the differences in response times seen between Primary younger and older children, and the absence of differences at Secondary School level. Some work is already underway using an inspection time (IT) task (Baughman, et al., in prep). IT tasks involve measuring the minimum time needed to accurately make judgements of the lengths of two lines and performance on this task has been studied extensively within the context of individual differences and cognitive development (see e.g., Anderson, 1988; Anderson, Reid, & Nelson, 2001; Burns, et al., 1999; Nettelbeck, 2001). IT tasks do not require motor movements during the information processing stage, but require participants to detect changes in increasingly brief exposures. As such, this type of task may provide a more accurate picture of the information processing abilities of MA-matched groups of younger and older children. If this work reveals speed of processing differences remain (i.e., once possible effects of differences in motor control are removed), then other factors would need to be considered in order to understand how inhibition may vary independently of age and ability. Here, factors such as working memory capacity might be examined for their influence on interference and facilitation.

In attempting to identify the relative contributions of different candidate mechanisms in tasks involving information processing, computational methodologies may offer an important service (see e.g., Thomas, McClelland, et al., 2009). For example, using existing computational models of the Stroop task, it is possible to explore the role of various candidate mechanisms such as inhibition or working memory capacity. According to Cohen, Dunbar and McClelland's (1990) model, speed of processing and interference effects are related to a common underlying variable that the authors referred to as 'strength of processing'. Thus, the question arises, how might their model produce faster response times, without changing the patterns of interference, or facilitation?

Finally, if the Stroop task does tap inhibition and goal neglect then simple linear relationships between interference and facilitation may not be adequate for assessing children's abilities. It is possible that the causes of distinct behavioural profiles of ability will only be understood if differential combinations of inhibition, development and intelligence are considered together. Indeed, it has already been argued that combinations of low vs. high inhibition with low vs. high intelligence may account for distinct behavioural outcomes. For example, Carlson, Peterson and Higgins (2003) have suggested low-inhibition, low-intelligence patterns in schizophrenic individuals and low-inhibition, high-intelligence patterns in geniuses and highly creative individuals.

In the next chapter, I report on the findings of a word-priming test, aimed once more at exploring possible differences in information processing between MA-matched children of different ages. Because this task also requires that children make motor responses (pressing one of two buttons), a point of interest is to determine whether the only difference between younger and older children is their speed of response. If results show different underlying patterns of processing, and in particular in size of priming, then this is evidence that factors other than age, and thus more-than-better control of motor movements, are at play in influencing the performance of younger and older MA-matched children.

Chapter 6 Primed lexical decision task

Introduction

A number of off-line and on-line tasks have been developed to tackle questions regarding the relationship between language and cognition. These tasks include, for example, word definition, word finding, word recall, word priming and lexical decision tasks. These tasks differ with respect to the level at which they are assumed to tap language processes (see e.g., Shapiro, et al., 1998; Tyler, et al., 1997). For instance, off-line tasks such as the word definition task, assess lexical knowledge on the basis of whether or not a correct definition is given for a word item. That is, this task measures the *end stage* of processing. Because individuals may monitor their responses, or engage in other forms of metalinguistic processing, response time measurements will be ambiguous vis-à-vis the component processes they reflect (see e.g., Karmiloff-Smith, et al., 1998; Shapiro, et al., 1998; Tyler, et al., 1997). By contrast, on-line tasks such as lexical decision tasks are aimed at measuring the behaviour of component processes *during* information processing. Consequently, lexical decision tasks are assumed to be more sensitive to the timecourse of operations of the underlying processes (see e.g., Shapiro, et al., 1998; Thomas, et al., 2008)[1].

Primed lexical decision tasks (of the sort used within this chapter) present participants with a set of stimulus-pairs in sequence. The first stimulus is a word (referred to as the *prime*) and its pair is either a real word, or a non-word[2] (referred to as the *target*). Participants indicate (usually by pressing one of two buttons) whether the target item is a real word or a non-word, thereby providing response time (RT) measurements of the time taken to decide the target's lexical status. The key empirical phenomenon that emerges from this task is the finding that the speeds at which lexical decisions are made are directly influenced by the semantic relationship of the target to the prime. For example, when adults are presented with a prime and a target that are *related* in meaning (e.g., BOAT and SHIP), SHIP is typically

---

[1] The distinction between off-line and on-line tasks was discussed earlier, in Chapter 2.

[2] A non-word is a string of phonemes that adheres to language-specific constraints on the possible combinations of phonemes but is not a real word. For example, "flin" in English.

responded to faster than when preceded by the semantically *unrelated* item such as GRILL (e.g., Meyer & Schvaneveldt, 1971). The degree to which the prime facilitates faster recognition of the target can thus be measured by subtracting their respective RTs. That is, the *priming effect* is the difference in time to respond to unrelated primes (e.g., SHIP preceded by GRILL), minus the time taken to respond to related primes (e.g., SHIP preceded by BOAT).

Several theories have been put forward to account for this effect (Thomas, 1997). Two broad classes of theory that may be distinguished are *spreading activation theories* (e.g., Anderson, 1983; McNamara, 1992) and *parallel distributed theories* (e.g., Cree, McRae, & McNorgan, 1999; Plaut, 1995).

Within spreading activation theories, the emphasis is placed on the *fast* and *automatic* spread of activation through a network of interconnected units that represent semantic information. When presented with a word, activation spreads from one concept to other related concepts, thereby raising the activation levels for all related word meanings. Words that are closer in their proximity to the prime are activated more strongly. Greater semantic activation provides top down facilitation of lexical entries, thereby accounting for the priming effects of related words.

Within parallel-distributed accounts, the key idea is that semantic information is represented in the cognitive system in distributed patterns of activity. Words that are very similar in their meaning elicit similar, or *overlapping* patterns. Activation induced by the prime is taken to persist during the recognition of the target. The closer the meaning of a prime to a target word, the greater the similarity between the overlapping patterns. Related primes mean the recognition system is already some way to reaching the activation state required to recognise the target word. In this situation, the system will be faster to generate the appropriate response.

Each of these frameworks has subsequently provided the basis for various computational approaches aimed at the study of mechanisms underlying language processes (e.g., Cree, et al., 1999; McNamara, 2005; Plaut, 1995; Plaut & Booth, 2000). While there is agreement that differences in the magnitude of priming effects in lexical decision tasks reveal something about how lexical and semantic knowledge is organised within the cognitive system, there is disagreement over which account offers the best explanation (see e.g., McNamara, 2005; Plaut & Booth, 2000).

Additional challenges for accounts of priming are presented by evidence that a range of other factors modulate the size of the effect. The full list includes, for example, the mean frequency with which words occur both in printed text and speech (referred to as word frequency), the latency between prime and target (referred to as the stimulus onset asynchrony; SOA), the age at which words are typically acquired (age-of-acquisition), the type of semantic relationship between the prime and target (semantic association) and how strongly one word is associated with the other (association strength). For a review of each factor see Neely (1991). In the current study I focus on the influences of *semantic association* and *association strength* on priming effects in younger and older MA-matched groups.

Semantic association refers to the distinction that words may share either a *categorical* or a *functional* (sometimes called 'thematic') relationship. For example, BROTHER and SISTER are category coordinates within the superordinate category of 'family', whereas UMBRELLA and RAIN are related by their function[3]. Studies of young children's word naming preferences (see e.g., Blewitt & Toppino, 1991) and adults' response times have shown an advantage for functionally-related versus categorically-related words (Moss, et al., 1995), thus suggesting differential processing patterns for these two types of relations. One explanation for this is that functional relations are learned earlier during development and that knowledge about the different categories that objects belong to follows later (Mandler, 1994).

Association strength refers to the frequency with which words are freely associated with each other in normative tests of word association (e.g., Nelson, McEvoy, & Schreiber, 1994). For example, when given the word CAT, a highly typical response is DOG. Consequently, CAT-DOG is rated as a high-association strength word-pair. Low-association strength word-pairs have been determined in the same way. For example, given the word PINK, a highly unusual response is GREEN. Studies examining association strength have claimed evidence of an 'associative boost' for high-association versus low-association word-pairs (see e.g., Moss, et al., 1995). That is, the size of priming effect has been shown to increase for words that appear frequently together within sentences. Moss et al. (1995) demonstrated that this associative boost occurred for both functionally and categorically related word-pairs. However, Nation and Snowling (1999) have claimed differential patterns of processing exist between children with normal and poorer reading comprehension.

---

[3] This is an example of *script-based* thematic relation. The prime is intended to evoke a script or schema from which typically related items might be activated.

For example, compared to children with normal reading comprehension, they found poorer comprehenders did not exhibit an associative boost for high-associated categorical word-pairs. They attributed these findings to a lack of sensitivity towards abstract relations in the poorer comprehenders.

Each of the theoretical approaches described earlier have sought to account for the variety of effects found in priming studies[4]. For example, the associative boost has been modelled using simple recurrent networks with overlapping patterns of semantic and phonological relations (Moss, et al., 1994). Based on this work, evidence claims have been put forward for a separate associative priming mechanism (Moss, et al., 1995). Plaut and Booth (2000) however, have argued for a single mechanism account that they claim explains the associative boost and a variety of other priming-related phenomena. Using a parallel-distributed model they further assert that both within-age and between-age differences in priming may be accounted for by the same model. However, their argument appears to be based largely on two assumptions. The first is that the only difference between children and adults is the amount of experience they have with reading words and not, say, in the network architectures of younger and older systems. Secondly, it seems that the ‘single mechanism’ they refer to is in fact the entire, distributed system. According to the authors, the network constitutes a single mechanism because the independent contributions of any single factor cannot easily be teased out of a system where there is a set of computational principles displaying complex and interactive effects (Plaut & Booth, 2000).

Due to its assumed ability to tap fast and automatic processing, the lexical decision task has been used within a variety of research contexts with respect to individual variability. For example, studies have shown greater priming effects in older children compared to younger children in the visual modality (e.g., Newman & German, 2002). This finding suggests a shift away from slow and controlled processes of reading in younger years, towards automatic processing in older children and adults (Raduege & Schwantes, 1987). In other studies focusing on the relationship between intelligence (as measured by Wechsler IQ scale) and perceptual

---

[4] Plaut and Booth (2000) contrast a number of approaches for differing levels of success in accounting for the range of modulatory effects.

abilities (as measured by a 'match-to-sample' task[5]), evidence has been claimed of different patterns of processing in lower-IQ versus higher-IQ children (Detterman & Daniel, 1989)[6]. It has been suggested that such differences account for the poorer comprehension of some children (e.g., Nation & Snowling, 1999). Differences in priming effects have also been examined within a number of disorders. For example, studies have examined differences between typically developing individuals and children diagnosed with specific language impairments (e.g., Edwards & Lahey, 1996), individuals with schizophrenia (for review see e.g., Minzenberg, Ober, & Vinogradov, 2002) and Williams Syndrome (see Karmiloff-Smith, et al., 1998; Thomas, et al., 2008).

Current aims

Because the lexical decision task offers a window onto the possible cognitive processes underlying language, it is an ideal task to test the relationship between intelligence and cognitive development. However, to date it appears there have been no studies that have investigated this by comparing groups of younger and older children who share the same overall mental age. First, while there is some evidence that older children process semantic information faster than younger children (e.g., Raduege & Schwantes, 1987), it is not clear how greater or poorer mental ability will influence this. Second, on the basis of previous work that has suggested high-IQ and low-IQ children may differ in their low-level perceptual processes (Plaut & Booth, 2000), it seems reasonable to expect MA-CA differences will modulate priming effects. Thus, the primary research questions I set out to address are: (1) do YHA and OLA groups differ in their speed of response on lexical decisions? (2) Do YHA and OLA groups differ in the overall sizes of their respective priming effects? (3) Does MA-CA disparity (as a continuous variable) modulate the speed of response, and/or the priming effect?

---

[5] Match-to-sample tasks, often used with infants (or monkeys), present participants with a target object and requires that they select a match for that object from a pair of alternatives (for fuller description, see e.g., Goswami, 1998). Performance is measured in terms of accuracy.

[6] Detterman and Daniel (1989) found moderately strong correlations between scores on a perceptual efficiency task and IQ in a low-IQ group (r=.60), but weak correlations between perceptual efficiency scores and IQ in a high-IQ group (r=.26) (1989).

We have seen that the patterns of influence of semantic association and association strength have been linked to the quality of the knowledge encoded by children. Of secondary interest, then, are the following two questions: (1) How do YHA and OLA groups compare in their response times for semantically associated (categorical versus functional) items? And (2) how do YHA and OLA compare in their respective RT for high-association vs. low-association strength items?

To test these questions, I administered a computer-based version of the lexical decision task developed by Leech (2006), in the auditory modality. I applied this computerised task to the groups of younger and older children at Primary and Secondary School level, described in Chapter 3. The computerised version of the task was designed specifically to replicate Nation and Snowling's study (1999). It includes the same stimulus set, which has not been modified here. The task provides the same RT measurements, allowing for an analysis of: (1) overall speed of response and hence differences in priming; (2) differences in priming for category vs. function relations (association type); and (3) differences in priming for low-association vs. high-association strength items (association strength), between groups.

Method

Participants

All participants completed this task (see ‘Participants’ section, Chapter 3 General Methodology). However, one child, Participant 27 was later removed from the Primary School OLA group data set because her overall accuracy was equivalent to chance levels, suggesting she was likely guessing. Participant 27 scored 50% correct on the task, whereas all other children in this group scored above 90%.

Design

A 2x2x2x2 mixed design was used, comprising 3 within-participant factors and 1 between-participant factor. The between-participants factor was Group (YHA vs. OLA) and the 3 within-participant factors were Prime (Unrelated vs. Related), Strength (High-Associated vs. Low-Associated) and Association (Category vs. Functional).

Materials

The experimental tasks were coded in SuperLab 1.5™ and was donated by Rob Leech (2006). The list of stimuli used in this study and which are included in full in Appendix F, originated from the stimuli used by Nation and Snowling (1999). Their stimuli were a subset of 48 highly familiar[7] word-pairs taken from Moss et al. (1995). Nation et al. constructed a set of word-pairs in which half were related through category membership, the other half through their functional membership. The association strength of target words to prime words was manipulated using normative lists of word association provided by Carroll, Davies and Richman (1971). This list details the frequency with which one word was given freely as an associate word to another. Word-pairs with high association strength had a mean frequency of 37.65% and word-pairs of low association strength had a mean frequency occurrence of 0.49%. The complete set of related word-pair items consisted of 12 pairs of: (1) categorically-related high-associated words; (2) categorically-related low-associated words; (3) functionally-related high-associated words; and (4) functionally-related

---

[7] Nation and Snowling (1999) used words that were rated for their familiarity for 10-year-old children by adults experienced in working with children.

low-associated words. A further 48 word pairs in which the target was preceded by an unrelated item served as controls for each of the above conditions. Finally, the entire set was balanced using 96 prime-target pairs where the target was a non-word. The related and unrelated groups are represented in Table 6-1. This table shows, for instance, category related high–associated word-pair item ‘bat-ball’ and its control word-pair ‘belt-ball’. There were an equal number of word and non-word targets for a total of 196 trials, split into two blocks that counter-balanced the relatedness of prime to target (i.e., in Block A, ‘bat’ was paired with ‘ball’ and in Block B, ‘belt’ was paired with ‘ball’).

Table 6-1. Examples of high-associated and low-associated related and unrelated word-pairs for Category vs. Function conditions.

Semantic type	Association strength	Condition	Prime	Target
Category related	High-associated	Related	bat	ball
		Unrelated	belt	ball
	Low-associated	Related	cow	goat
		Unrelated	hospital	goat
Function related	High-associated	Related	grill	toast
		Unrelated	coat	toast
	Low-associated	Related	war	army
		Unrelated	salt	army

All word items had been previously recorded in a soundproofed recording booth and were produced by British English female (primes) and British English male (targets) speakers. A female and male speaker combination was used to disambiguate task responses (see Procedure). The audio files were presented and controlled through a laptop running SuperLab 1.5™, also used to record response data. The experiment used a monitor to display instructions and which the experimenter also read aloud (verbatim instructions are provided in Appendix E). Finally, a response button box and a pair of Sennheiser stereo headphones were used.

Procedure

Children sat in front of a touchscreen monitor and were read a set of instructions that also appeared onscreen, explaining the nature of the task. The touchscreen was not used for any other element of the task. The instructions informed the children that they would hear two words. Children were told that the first word (spoken by a woman) did not require any response, but that the second word (spoken by a man) did. They were informed that the second word would be either a real word like DOG, or a made up word like TOAG. Children sat with their arm positioned on a table with their dominant hand resting upon the button box, also situated on the table in front of them. Participants were told they should respond by pressing one of two buttons to indicate whether this second word was a real word or a non-word. Children were instructed to make their decision as quickly and as accurately as possible, using one finger to press either a red button (to indicate non-words) or a green button (to indicate real words). Once the instructions were finished, a practice trial consisting of 8 examples was given (these items were not repeated in the experimental trial). The practice trial examples were played through open speakers, so that in the event of confusion, the experimenter could hear and demonstrate where responses should be made. For the experimental trials, the procedure differed only in that children wore closed headphones. The order of the blocks they received was counter balanced and these ran back-to-back, with half the participants completing first Block A then Block B and the other half completing Block B then Block A. The full task took approximately 4 minutes to complete. The stimulus onset asynchrony between prime and target was fixed at 500ms and the interval between the end of one word-pair and the beginning of another was 1520ms, following the response of pressing a button.

Primary School results

Results are first presented for the MA-matched YHA (n=14) and OLA (n=13) groups, then the full sample of Younger (n=20) and Older (n=19) children. In the MA-matched comparisons, two 2x2x2x2 repeated measures ANOVAs were performed on Accuracy and then on RT data. In each case, the between-participants factor was Group (YHA vs. OLA) and the within-participant factors were Prime (Related vs. Unrelated), Strength (High-Associated vs. Low-Associated) and Association Type (Categorical vs. Functional). The same analytical design was applied to the full sample, with MA-CA Disparity added as a covariate. To avoid undue influence of outliers, individual RT data points were cropped where individuals had RTs ±2 standard deviations from the group mean. These data were replaced with the mean RT for their group (Ratcliff, 1993). In the full Primary sample, RT data was cropped for 5 individuals in a total of 7 instances in the YHA, and for 6 individuals in a total 13 instances, in the OLA group. One participant was excluded from the analysis (see ‘Participants’ section, for details).

Primary School MA-matched group comparisons

With the loss of one participant from the OLA group, the YHA and OLA groups were still nevertheless matched overall for MA (i.e., the results of an independent 2-tailed t-test yielded: t(25)=.165, ns). A summary of Accuracy and RT is given in Table 6-2 for each of the experimental conditions (Prime, Strength and Association Type) for YHA and OLA groups. The table also shows measures of priming effects, derived by subtracting the RT on related conditions from the RT on unrelated conditions.

Table 6-2 Summary table of data for Primary School MA-matched groups. Table shows mean accuracy, mean cropped RT (ms) and sizes of priming effect.

				Accuracy		**RT (ms)**		**Priming**
Group	*Association Type*	*Strength*	*Prime*	*Mean*	*Std. Error*	*Mean*	*Std. Error*	*Difference**
YHA	Categorical	High-Associated	Related	94.0%	2.90%	1199	54	18
			Unrelated	96.2%	2.40%	1217	43	
		Low-Associated	Related	97.4%	1.70%	1257	42	94
			Unrelated	92.4%	2.90%	1351	34	
	Functional	High-Associated	Related	95.2%	2.30%	1180	48	195
			Unrelated	97.6%	1.90%	1375	40	
		Low-Associated	Related	95.2%	2.50%	1250	37	43
			Unrelated	92.1%	3.50%	1293	41	
OLA	Categorical	High-Associated	Related	97.4%	3.10%	911	56	19
			Unrelated	94.9%	2.50%	930	44	
		Low-Associated	Related	97.4%	1.80%	1010	43	57
			Unrelated	98.7%	3.00%	1068	35	
	Functional	High-Associated	Related	100.0%	0.00%	1001	49	69
			Unrelated	95.9%	1.90%	1070	42	
		Low-Associated	Related	96.8%	2.60%	1035	38	21
			Unrelated	98.7%	3.70%	1055	42	

*Unrelated RT - Related RT

<u>Overall accuracy</u>: Initial comparisons showed that, overall, the OLA were more accurate (mean accuracy=97.5%, se=1.9%) compared to the YHA (mean accuracy=95.0%, se=1.8%). However, these scores were at ceiling and the analysis showed that the difference was not statistically reliable.

<u>Overall response times</u>: The ANOVA on RT data revealed a reliable main effect of Group ($F(1,25)=68.07$, $p<.001$, $\eta^2=.731$). This was due to the fact that, overall, the OLA was faster at responding (mean RT=1009ms, se=22ms) compared to the YHA (mean RT=1265ms, se=21ms).

<u>Priming effect</u>: The analysis showed a main effect of Prime ($F(1,25)=9.50$, $p=.005$, $\eta^2=.275$), indicating that the experimental manipulation of Related and Unrelated word-pairs was effective in modulating response times. However, the analysis revealed no Group x Prime interaction. Thus, the mean sizes of priming effects in the YHA group (87ms, se=25ms) and the OLA group (41ms, se=26ms) were not reliably different (i.e., Group x Prime: $F(1,25)=1.20$, $p=.283$, $\eta^2=.046$). Figure 6.1 shows the mean priming effects for YHA and OLA groups across each level of the experimental conditions. The figure indicates a greater priming effect for High-Associated Functional items. However, neither Strength nor Association Type, nor their combination was found to reliably modulate the size of the priming effect and Group did not reliably modulate these either.

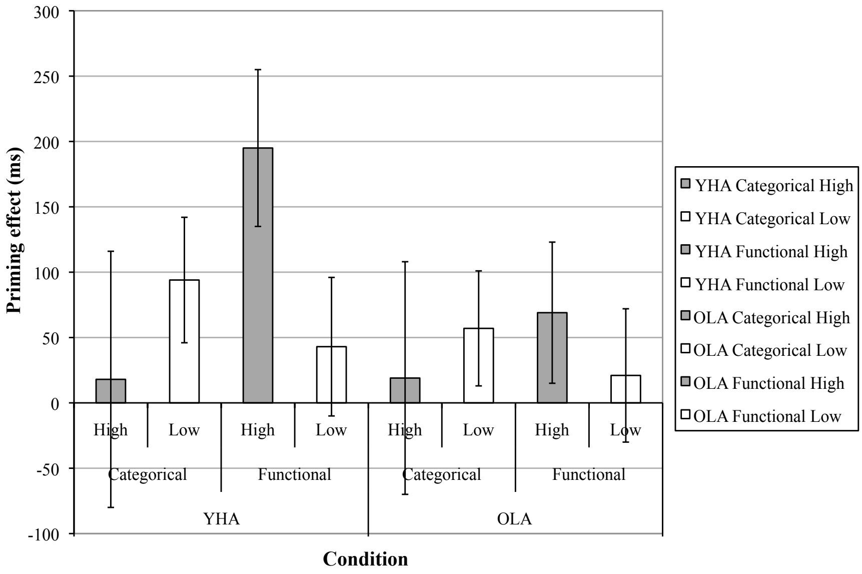


Figure 6.1. Mean priming effects across Semantic Association (Categorical vs. Functional) and Association Strength (High vs. Low) for YHA and OLA Primary School groups. Error bars show standard error of the mean.

Primary School full sample analysis

Overall accuracy: The results of the ANCOVA on Accuracy data showed MA-CA disparity was not reliable in modulating performance and no Group x MA-CA disparity interaction was found. However, using the full sample, a main effect of Group was revealed (F(1,35)=4.92, p=.033,η^2=.123). This stemmed from the higher overall accuracy in the OLA (mean accuracy=97.7%, se=2.60) compared to the YHA (mean accuracy=88.2%, se=3.60).

Overall response times: On the RT data, the results of the second ANCOVA showed MA-CA disparity did not reliably modulate overall speed of response and no interaction between Group and MA-CA disparity was revealed. The overall RTs for each group remained reliably different (F(1,35)=21.73, p<.001, η^2=.383). This stemmed from faster RT in the OLA (mean RT=1018ms, se=24ms) compared to the YHA (mean RT=1226ms, se=33ms). While this mirrored earlier findings in the MA-matched analysis, the proportion of variance accounted for in the full sample was substantially lower (MA-matched η^2=.731 vs. Full sample η^2=.383). Data are shown in Figure 6.2.

Priming effect: Contrasting the RT for Related and Unrelated conditions revealed reliable interactions for MA-CA disparity x Prime (F(1,35)=5.74, p=.022, η^2=.141) and Group x Prime (F(1,35)=4.80, p=.035, η^2=.121). Figure 6.2 shows the source of the MA-CA x Prime interaction with the dashed (Unrelated) and un-dashed (Related) black trendlines, both showing positive, yet different gradients (R^2=0.36 vs. 0.21, respectively). This figure suggests that the more MA exceeds CA, the greater the priming effect. The Group x Prime interaction is depicted by divergences within each group between Related and Unrelated conditions (YHA: light-blue and dark blue trendlines, respectively; OLA: orange and red trendlines, respectively). Finally, the results of this analysis further showed that neither Strength nor Association Type, nor their combination reliably modulate the size of the priming effect in the full sample and neither MA-CA disparity nor Group reliably modulated these either.

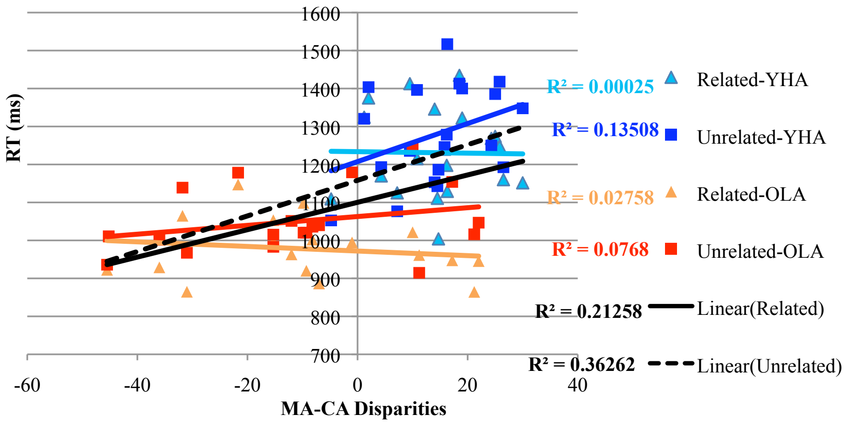


Figure 6.2. Plotting MA-CA disparity by RT on Related and Unrelated conditions in full Primary School sample. Black lines show the functions collapsed across group.

Summary

Younger and older groups of MA-matched Primary School children were found to differ reliably only in their speed of response. While the OLA group was faster than the YHA group, the relative size of priming in the two groups was not different. Neither Strength nor Association Type was found to be reliable in modulating the size of the priming effect. When the full sample was used and MA-CA disparity was added as a covariate, both Group and MA-CA disparity were found to interact reliably with RT on Related and Unrelated conditions. A greater, positive MA-CA disparity predicted larger priming effects in both groups.

Secondary School results

The same analytical design that was used in the Primary School results was applied to Secondary School data. However, in this sample, accuracy was found to be at ceiling and thus not suitable for analysis. Results are first presented for the MA-matched comparisons between YHA (n=16) and OLA (n=19) groups. These are then followed by the results of the continuous analysis, where MA-CA disparity was used as a covariate. To avoid undue influence of outliers, RT data were cropped for participants who had original RTs of ±2 standard deviations from the group mean. Data were cropped for 5 participants in a total of 6 instances in YHA and for 5 participants in a total of 6 instances in OLA group. No participants were excluded.

Secondary School MA-matched group comparisons

A summary of the mean cropped response times, accuracy (percentage correct in brackets) and priming effects (Unrelated RT-Related RT) for YHA and OLA groups on each of the experimental factors is given in Table 6-3.

Table 6-3 Summary table of data for Secondary School MA-matched groups. Table shows mean accuracy, mean cropped RT (ms) and sizes of priming effect.

				Accuracy		**RT (ms)**		**Priming**
Group	*Association Type*	*Strength*	*Prime*	*Mean*	*Std. Error*	*Mean*	*Std. Error*	*Difference**
YHA	Categorical	High-Associated	Related	100.0%	0.00%	842	33	115
			Unrelated	99.0%	1.30%	958	30	
		Low-Associated	Related	100.0%	0.00%	953	42	119
			Unrelated	100.0%	1.00%	1071	35	
	Functional	High-Associated	Related	100.0%	0.00%	964	37	55
			Unrelated	94.8%	1.80%	1019	30	
		Low-Associated	Related	100.0%	0.00%	963	24	85
			Unrelated	99.0%	0.70%	1048	38	
OLA	Categorical	High-Associated	Related	100.0%	0.00%	786	31	31
			Unrelated	97.4%	1.20%	817	28	
		Low-Associated	Related	98.2%	1.30%	899	39	-7
			Unrelated	98.2%	0.90%	892	32	
	Functional	High-Associated	Related	100.0%	0.00%	879	34	85
			Unrelated	99.1%	1.70%	964	28	
		Low-Associated	Related	100.0%	0.00%	934	22	-8
			Unrelated	100.0%	0.00%	926	34	

Overall accuracy: The overall accuracy of the YHA and OLA was identical (YHA mean accuracy=99.1%, se=.30; OLA mean accuracy= 99.1%, se=.30) and at ceiling (see Table 6-3).

Overall response time: As was the case in the Primary School data, the overall RT of the two groups differed reliably (F(1,33)=6.11, p=.019, η^2=.156), with the OLA showing faster overall RTs (mean RT=887ms, se=24) compared to the YHA (mean RT=977ms, se=26).

Priming effect: Similar to the Primary School MA-matched results, the analysis of Secondary RT data also showed a main effect of Prime (F(1,33)=29.09, p<.001, η^2=.469). However, in contrast to the Primary School results, the Secondary School results showed a reliable Group x Prime interaction (F(1,33)=9.59, p=.004, η^2=.225). Figure 6.3 illustrates these effects and shows the source of the interaction is due largely to the greater priming effect in the YHA. Figure 6.3 shows that the differences between Related and Unrelated conditions were smaller in the OLA (red line) and larger in the YHA (blue line). The overall size of priming effects was 94ms in the YHA (i.e., Unrelated mean RT=1024ms, se=27.96 vs. Related mean RT=930, se=28.13; ∴ Priming=94ms) and 26ms in the OLA (i.e., Unrelated mean RT=900ms, se=25.65 vs. Related mean RT=874, se=25.81; ∴ Priming=26ms). Figure 6.3 also shows the faster overall RT of the OLA compared to the YHA. Once again, the results showed that neither Strength nor Association Type were reliable in modulating the size of priming effects and this was the same in both groups.

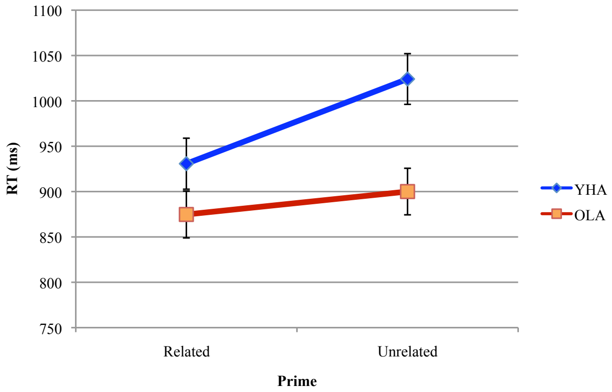


Figure 6.3. Mean RT on Related and Unrelated conditions for YHA and OLA Secondary School groups. Error bars show standard error of the mean.

Figure 6.4 shows the mean priming effects for YHA and OLA groups across each level of the experimental conditions. The figure depicts relatively equal sizes in priming effects for the majority of items.

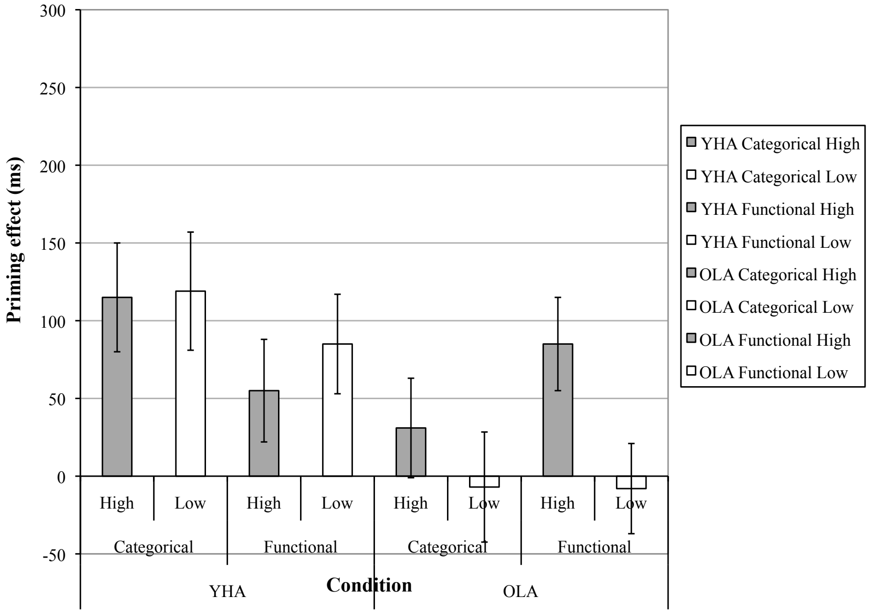


Figure 6.4. Mean priming effects across Semantic Association (Categorical vs. Functional) and Association Strength (High vs. Low) for YHA and OLA Secondary School groups. Error bars show standard error of the mean.

Secondary School full sample analysis

Overall response time: The ANCOVA on RT showed MA-CA disparity reliably modulated overall response times (F(1,31)=4.32, p=.046, η^2=.122). It also showed a main effect of Group (F(1,31)=10.23, p=.003, η^2=.248) that stemmed from faster overall responses from the OLA (mean RT=879ms, se=25) compared to the YHA (mean RT=1010ms, se=32). However, no Group x MA-CA interaction was present, indicating that MA-CA disparity had similar effects on overall RT in YHA and OLA groups.

Priming effect: Three reliable interactions were found on the analysis of priming effects. These were: Group x MA-CA x Prime (F(1,31)=25.12, p<.001, η^2=.448), Group x Prime (F(1,31)=19.82, p<.001, η^2=.390) and MA-CA x Prime (F(1,31)= 7.68, p=.009, η2=.199). These results were broadly in line with those found in the Primary School data. However, as Figure 6.5 illustrates, MA-CA appears to have a different influence on RT in the Secondary School data. Whereas in the full Primary School data, increases in one's advantage (MA-CA disparity) were related to larger priming effects (see earlier Figure 6.2), here it is the reverse. Figure 6.5 shows negative MA-CA disparity x RT relationships for the Related condition in YHA (light-blue trendline) and OLA (orange trendline) groups. Finally, in the full sample, the results showed again that neither Strength nor Association Type were reliable in modulating priming effects and neither MA-CA disparity nor Group were found to interact with these.

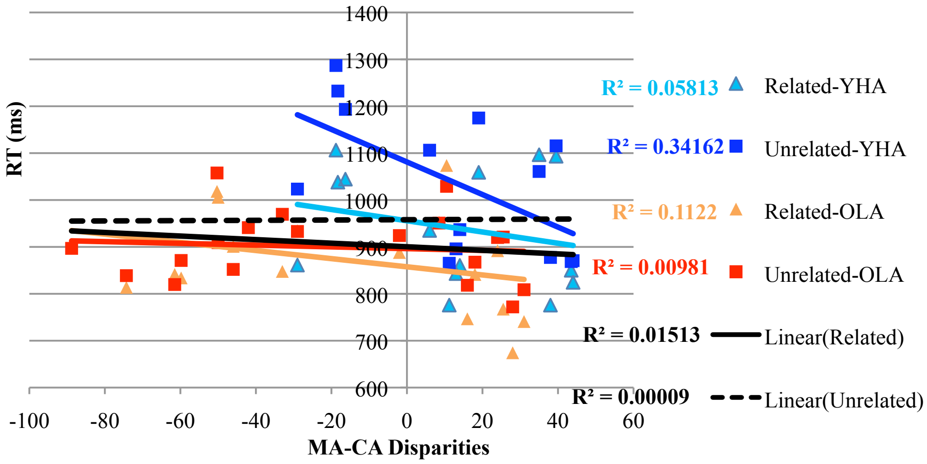


Figure 6.5. Plotting MA-CA disparity by RT on Related and Unrelated conditions in full Secondary School sample. Black lines show the functions collapsed across group.

Summary

The results of the MA-matched Secondary School comparisons revealed that the YHA were slower at deciding the lexical status of words than the OLA and that this difference was reliable. Furthermore, the groups differed in their overall size of priming effects, with the YHA showing greater priming effects than the OLA. In the analysis of the full Secondary School data, MA-CA was found to reliably modulate overall response times and size of priming effect.

Combined Primary and Secondary School Results

A final 4x2x2x2 ANCOVA on RT data was performed on combined Primary and Secondary School samples. The covariate was MA-CA disparity, the between-participant factor was Group (Primary-Younger, Primary-Older, Secondary-Younger, Secondary-Older) and the within-participant factors were Prime (Related vs. Unrelated), Strength (High-Associated vs. Low-Associated) and Association Type (Category vs. Function).

Overall response times: Group was found to reliably modulate overall response times (F(3,67)=18.60, p<.001, η^2=.454). Figure 6.6 shows that overall RT was slowest in the youngest group (Primary Younger) and fastest in the oldest (Secondary Older), while the performances of the Primary Older and Secondary Younger were more closely matched. The analysis revealed no main effect of MA-CA and no MA-CA by Group interaction.

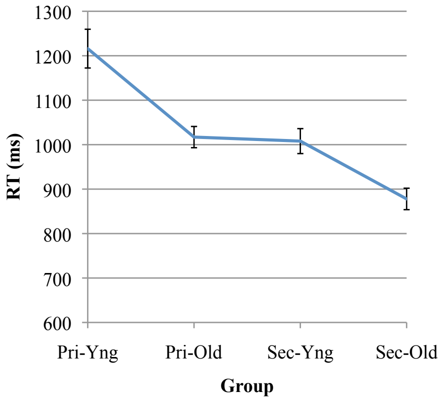


Figure 6.6. Overall response times for Primary and Secondary School combined data. The mean chronological ages of these groups were: Pr-Yng=6.4yrs; Pri-Old=10.4yrs; Sec-Yng=12.9; and Sec-Old=15.1yrs.

Priming effect: Focusing on the effects of MA-CA disparity and Group, the full sample analysis showed reliable Group x Prime (F(3,67)=4.65, p=.005, η^2=.172) and Group x MA-CA x Prime (F(3,67)=4.69, p=.005, η^2=.173) interactions. Figure 6.7 attempts to depict the source of these effects. For instance, although the Secondary School OLA group are fastest in their RT, MA-CA disparity does not appear to modulate their response times (mauve and purple trendlines). This differs from the trendlines depicting YHA Secondary group which showed that MA-CA reliably modulates RT on Unrelated (dark-green trendline), but not Related (light-green trendline) word-pairs. The greater the MA-CA advantage, the faster the response.

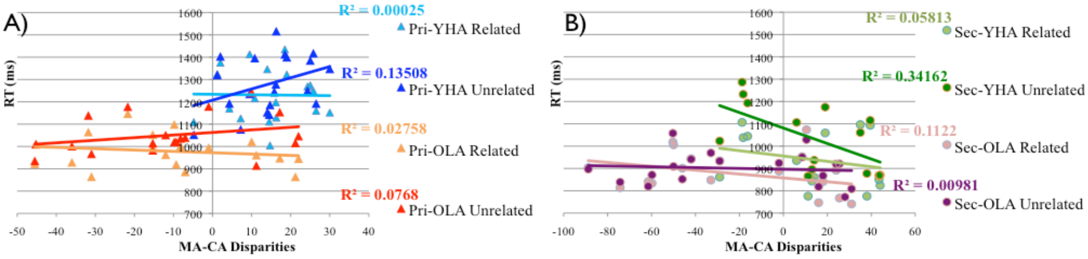


Figure 6.7. Combined Primary and Secondary School MA-CA disparities by Priming effects.

Summary

Taking Primary and Secondary School data together, the results of the continuous analysis revealed that Group reliably predicted response times and size of priming. MA-CA disparity did not reliably modulate performance on its own, but was found to interact with Group in accounting for variability in priming. One key difference noted from the combined analysis is that RT was modulated differentially by one’s advantage, depending on development, as indexed by chronological age.

Discussion

The semantic priming task described in this chapter was used to (a) further examine the speed of processing differences that were found in Chapter 4 on the BAS II results and the response time differences that were found in Chapter 5 on the Stroop task; and (b) to continue to explore possible differences in information processing between MA-matched groups of younger and older children.

Consistent with findings in the previous two chapters, the results in this chapter showed reliable group differences in overall speed of response at Primary School level. For example, in the Primary MA-matched groups, the OLA were consistently faster than the YHA on deciding the lexical status of words. Furthermore, the relative sizes of priming effects in the two groups were not different. These findings mirror those reported in the chapter on the Stroop task (Chapter 5), where it was RT and not measures of interference or facilitation which differentiated the two Primary School groups. Additionally, on examining the effects of association type and strength (two factors each previously found to modulate size of priming effects, e.g., Mandler, 1994; Nation & Snowling, 1999), the results showed no evidence of an associative boost for category-related high-associated words. Given that the Primary YHA and OLA groups did not differ in either their accuracy or the relative sizes of their priming effects, these groups appear equivalent in their underlying semantic representations, be it spreading activation, or overlap of persisting activation patterns. Once again these results suggest no differences between intelligence and cognitive development, at these ages.

Turning to the results of the Secondary School MA-matched groups, the pattern of findings appears to contrast with those from the Primary School level. Although again the OLA again demonstrated faster overall response times compared to the YHA, at this school level a reliable difference was observed in the relative sizes of priming effects. The results showed this was due to a larger overall priming effect in the younger group compared to the older group. While decreases in priming effects between younger and older children have been reported, it is not clear what accounts for this change (see e.g., Kang & Simpson, 1996; Plaut & Booth, 2000). In the previous chapter, we considered the possibility that differences in motor control might account for the faster responses of the older children. Specifically, the possibility discussed was that at some ages (particularly in younger children), greater age offers an advantage on speed of response due to a greater ability to control motor movements, but that this advantage decreases over time as children reach a ceiling in their ability to make movements any faster (a power-law function was used to illustrate this argument). Applied to the current data, this account appears to fit the Primary School data, where the older children were faster than the younger but where there was no interaction between groups in their size of priming. However, if differences in motor control were the only (albeit, non-linear) influence then presumably the difference between unrelated and related RT would hold constant over development and the pattern would resemble Tile A of Figure 6.8. Instead, the data resembled the pattern in Tile B. This suggests that the causes of changes in priming effects between younger and older children are due to more than differences in motor control.

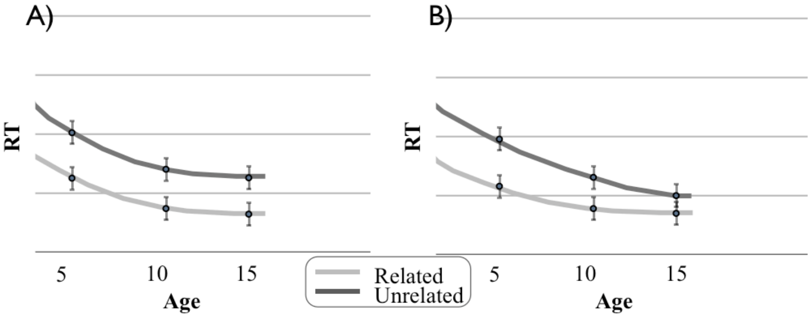


Figure 6.8 Tile A illustrates a general decrease in response times that might be attributable to age-effects, with priming contributing an independent, additive component. Tile B illustrates age-related decrease in response times and a decrease in the size of priming effect over age, in the form of a non-additive component.

One way to examine the contribution of priming effects would be to use existing models of lexical decision. For example, spreading activation accounts might propose simple strength of connection differences within the network of connected units. Parallel-distributed accounts might suggest that the differences are accountable by greater distances between overlapping patterns. However, any model seeking to simulate these differences would also have to demonstrate no differences in the overall size of priming effects for semantic information.

In a different field of research, the mirror opposite of the pattern depicted in Tile B has been reported in priming differences between younger and older adults.

That is, the difference between the times taken to respond to unrelated word-pairs versus related word-pairs grows larger in old age. Presently, there is no consensus on the causes of these differences. For instance, the literature shows these changes have been attributed to both process-specific changes in speed of processing (e.g., Laver & Burke, 1993) and more general, system-wide cognitive slowing (e.g., Myerson, et al., 1992).

From the results of the full samples analyses, another notable finding to emerge from this research was the MA-CA x Prime interaction. In the Primary School data, the results showed that as children's MA exceeded their CA, response times increased. That is, slower response times appeared associated with more positive MA-CA disparities. However, the direction of this MA-CA disparity x Prime interaction contrasted with the Secondary School data where faster response times were associated with more positive MA-CA disparities. If this is a genuine (i.e., replicable) effect, one explanation for this might be at older ages children simply have greater familiarity with words.

In the following section we turn to the results of the computer-based cognitive tasks. These tasks, now targeting higher-level components of processing, offer us the chance to examine whether groups of MA-matched children differ on more than just their speed of response or whether they also differ in their reasoning processes.

Part 5

Part 4 examined the performances of groups of younger and older MA-matched children on the Stroop task and a primed lexical decision task – on-line tasks aimed at measuring fast and automatic information processes. The results of those chapters generally supported the results offered previously by the BAS II in Chapter 4 and provided further support for the notion that differences in intelligence are overall very similar to differences in cognitive development. In this part of the thesis we move on to explore possible group differences on a range of computer-based tasks assumed to tap abilities relating to *reasoning* and *planning*. Chapter 7 presents the results of the conservation of number and liquid tasks (Piaget, 1954). Chapter 8 presents the results of the balance scale task (Piaget, 1954). Chapter 9 presents the results of the Tower of London task (Shallice, 1988). These chapters form a logical progression of the research aims because they are theoretically relevant to aspects of perceptual and conceptual ability that are taken to characterise different stages of cognitive development. For example, conservation tasks require the ability to integrate both *perceptual* and *conceptual* information about quantities of objects and their invariance under transformations, the balance scale task requires the combination of information about *distances* and *weights* and the Tower of London task requires the ability to *sub-goal* efficiently in solving problems that may permit multiple solutions. While the tasks used in this part of the thesis are traditionally used as *off-line* measures of ability (i.e., they typically measure the end stage in information processing), a further aim is to examine whether the use of computer-based versions of these tasks, in which response times are gathered, may allow these tasks to be transformed into on-line tasks that are sensitive to the timecourse of operations. The structures of the chapters are similar to those in Part 4. Chapters present, in turn, the findings of the categorical and then continuous analysis for the Primary School children, then the Secondary School children and then Primary and Secondary School results combined on the target measures.

Chapter 7 Conservation of Number and Liquid

Introduction

Piaget described cognitive development as a progression through a series of stages that enabled ever more complex forms of reasoning (1954). During children's early years he believed that the underlying factors driving this change were primarily biological, but that later as children grew older, it was experience (Piaget, 1972).

Based on his observations of children, Piaget argued that at around the ages of 6 to 7-years children begin to acquire the ability to reason more effectively, but they are limited to reasoning about concrete objects (1954, 1972). He argued that a hallmark feature characterising children's development at this age is their acquisition of *operations* – the ability to mentally manipulate complex information about the world. These representations are more elaborate than those typically assumed in children of younger ages (1954). In this view, children with concrete operations are able to represent transformations that occur to objects in their environment and then mentally reverse these operations – thereby conserving knowledge of the original properties of objects, as their features change. For example, Piaget noted that younger children would typically understand that a physical set had changed following a transformation in which items were either added to or subtracted from the set. However, Piaget also found that younger children believed the quantity of a set was different when only the physical appearance had been altered (e.g., by elongating or compressing the set). In other words, he found younger children less able to integrate perceptual information with existing conceptual information, or knowledge.

The variety of tasks that emerged to assess the development of this new form of reasoning are called *conservation* tasks. They include, for instance, conservation of the following forms: number, liquid, length, area, mass, weight, and volume. Figure 7.1 provides an illustration of the conservation of number task. In this task, a child might be presented with two rows each with six counters. In the pre-transformation state, children would typically state that both rows A and B had an equal number of counters. However, transforming the length of one of the rows by altering the spacing (in this example, row $B \Rightarrow B^I$) typically leaves younger, preoperational children with the belief that the number had also changed (Piaget, 1954). These children would state that row A now had more counters than row B^{I}.

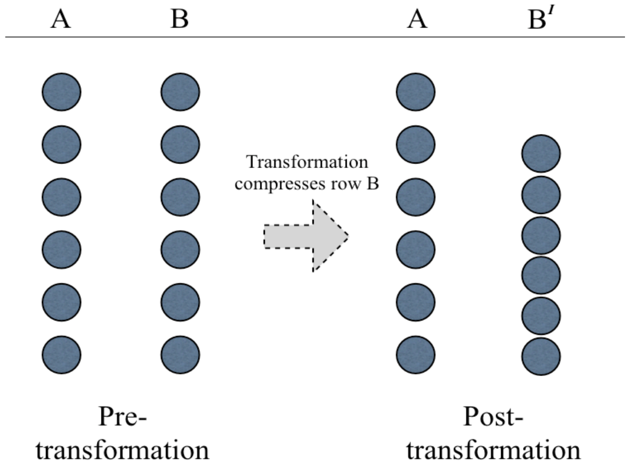


Figure 7.1. Example of a conservation of number problem. Two rows of six counters are presented (A & B). The length of one row (B) is then compressed (resulting in B^{I}), leaving the number the same. Non-conservers would state that Row A had *more* counters after the transformation. Concrete operational children would understand that the transformation had not altered the number in row B^{I} and thus they remain equal to Row A.

According to Piagetian theory, the ability to conserve information about quantity in the face of perceptual differences is a product of a progressive elaboration of the structures underlying mental representations. Children in the concrete operations stage who are reliable at conserving have acquired the operation of reversibility (Piaget, 1954). Since this theory was proposed, questions have remained as to how exactly children reach this stage. For example, what mechanisms underlie the transitions to different stages of reasoning and thereby explain variability in performance across development? One avenue of research has aimed to investigate in greater detail children's behaviour on conservation tasks, in the hope of finding clues to the underlying mechanisms (see e.g., Case, 1998; Siegler, 2006; Siegler, 1995; Thomas, et al., 2002).

Theories that attempt to address this issue have also aimed at providing an account of a number of behavioural phenomenon. These begin with the *acquisition* of conservation, which is often described as a sudden, or abrupt shift in the ability to conserve. Other effects include the *length bias* effect, whereby non-conservers are overly influenced by perceptual features that indicate greater quantity (for example, a plasticine ball rolled into a long ‘sausage’, a row of coins where the row is elongated, or a beaker of water poured into a taller, narrower beaker); the *problem size* effect, whereby conservation is demonstrated earlier for sets of smaller size, then for sets of larger size (e.g., Siegler, 1981); and the *screening* effect, whereby younger children conserve better on problems in which they do not observe the result of the transformation (Piaget, Inhelder, & Chilton, 1997).

Among the theoretical frameworks used to explain variability on conservation tasks are the staircase model (Case, 1992), the overlapping waves model (Siegler, Granott, & Parziale, 2002), dynamical systems theory (e.g., van der Maas, 1995) and connectionist and generative neural network models (Richardson, Forrester, et al., 2006; Shultz, 1998). Behavioural research has also extended to consider the area of atypical development where, for example, empirical studies have contrasted conservation performance between populations of children with specific language impairments (Mainela-Arnold, Evans, & Alibali, 2006), children with Downs syndrome (e.g., Pasnak, Willson-Quayle, & Whitten, 1998) and children with dyscalculia (e.g., Butterworth, 2005). Studies of variability in the ability to conserve has been examined in groups of gifted children (e.g., Brekke, 1976) and groups of bright, average and lower ability children (see e.g., DeVries, 1971; Goodnow & Bethon, 1966; Hood, 1962). However, these studies have led to a mixture of conclusions. For example, Hood (1962) reported reliable differences between lower-ability individuals and children of average ability, yet Goodnow et al. (1966) pointed out that these differences were likely a result of unusual sampling differences (the CAs of Hood’s average group ranged from 3-8 years, whereas the lower ability CA range was 9-41 years). By contrast, Goodnow et al. found no evidence of differences in groups more closely matched on CA.

No previous study appears to have compared the conservation performance of younger and older children in the normal range who are matched on mental age. It is an open question, then, whether children with different CAs but equal overall MAs are similar or different in their ability to conserve. As conservation requires the integration of information from both perceptual and conceptual dimensions we can therefore examine possible group differences on these dimensions. If CA is a marker for cognitive development (maturation) and intelligence is equivalent to having more or less development (unidimensional account), then MA-matched groups of different ages should not be different.

Overview

This chapter presents the findings from two tasks of conservation that were administered to groups of Primary School and Secondary School children. These were the conservation of Number and Liquid tasks. Results are presented for conservation Accuracy, Response Time and Length Bias. These measures are defined as follows: *Accuracy* – the proportion of correct responses on conservation problems (given a transformation that altered only the appearance of one of two identical rows of coins / containers of liquid); *Response Time* – the amount of time taken to correctly respond to conservation problems; *Length Bias* – the proportion of errors in which the longer of two lines of coins was chosen, but where both lines had the same number; and *Height Bias* – the proportion of errors where the taller, narrower of two containers was chosen, but where their volumes were equal.

Additionally, the method used in this study is slightly more speculative. This is because traditional tasks of conservation are completed off-line, usually using physical apparatus and recording performance simply as either ‘pass’ or ‘fail’. Given that around three-quarters of the total sample of children tested in this research project are above the age typically tested on the conservation tasks, we might expect a significant proportion will be at ceiling. This chapter assesses whether response time measurements can transform the conservation task into an on-line task that is potentially sensitive to differences between YHA and OLA. This is attempted in the following three ways: (1) the task is computerised, with random presentations of a large number of animated events; (2) RT is recorded in addition to accuracy; and (3) markers indicating reasoning differences are assessed, in this case, length bias and height bias.

Method

Participants

See ‘Participants’ section, Chapter 3 General Methodology. All participants took part in this experiment.

Design

On both conservation of Number and Liquid tasks the between-participants variable was Age Group (YHA vs. OLA) and the dependent measures were Conservation Accuracy, Response Time and Length Bias.

Procedure

Number: Children were shown a brief physical demonstration in which the experimenter placed two rows of 5 counters each on a table in front of the child. Children were told that the ‘real’ task, a computer-based task, would follow the same theme. Children were asked to indicate whether one of the rows had a greater number of counters, or whether they both had the same number. Once the child made their response, they were told that the experimenter would then do something to one of the rows. The experimenter illustrated one possible action by compressing one row of counters. The child was asked again to indicate whether one row had a greater number or whether they were equal in number.

Each child then watched an instructional video played on the monitor. In this video, the child’s attention was directed to the important areas of the screen: the area where the coins appeared and three buttons at the bottom of the screen that were to be used to indicate choices. These buttons were, “left (more)”, “same” and “right (more)”.

Children were shown that they would first be presented with two rows of coins and that they would have to decide which row had more coins, or whether each row had the same number. They were asked to respond as quickly as possible, making as few errors as possible, by touching the button that matched their choice. Once a button was pressed, they would then see a hand appear and change one of the sides (just as in the physical demonstration). The video depicted an animation of a hand moving across the screen and subtracting one coin from the row on the left. The video then explained that the child would need to press a button to indicate again which side had more, or whether the two sides had the same number. Full verbatim instructions are given in Appendix G.

A total of 4 computer-based practice trials, each using six coins, followed the instruction video (these problems were only encountered in the practice trials). Each problem randomly presented one of four possible transformations to one of the rows (i.e., adding, subtracting, elongating, or compressing). Once the practice trials were completed, children were then given a total of 16 experimental trials. Table 7-1 depicts the properties of the stimuli before (start) and after (end) transformation. The table shows 4 problems of four types where stimuli either started perceptually and conceptually equal (SS), or perceptually and conceptually different (DD) and ended in different perceptual and conceptual combinations (i.e., SD, DD, SD, SS).

For the purpose of this study, it is trial type (i) in Table 7-1 that is of key interest, as these problems provided the measurements of conservation accuracy, RT and Length Bias. Conservation was assessed as the proportion of correct responses on these trials where the transformation had altered the perceptual but not the conceptual properties of the stimuli. RT was measured as the amount of time taken to respond on correct trials. Length Bias was assessed using error data on these trials. For example, if a transformation elongated one of two rows containing equal numbers of coins and the child indicated the longer (transformed) row had a greater number, this would indicate a bias of length. Similarly, if the transformation compressed one of two rows containing equal numbers of coins and the child selected the longer (untransformed) row this would also indicate a bias for length. Each of the other trial types presented in Table 7-1 were distractors. The numbers of coins used in this task ranged from 8-12.

Table 7-1 Problem types for conservation of number task.

Type	*Start**	*Transformation*	*End**	*N trials*
i)	SS	elongate/compress	DS	4
ii)	SS	add/subtract	DD	4
iii)	DD	elongate/compress	SD	4
iv)	DD	add/subtract	SS	4

*SS=Perceptually-Same AND Conceptually-Same,
DD=Perceptually-Different AND Conceptually-Different,
SD=Perceptually-Same AND Conceptually-Different,
DS=Perceptually-Different AND Conceptually-Same

Liquid: This task followed the same format as the number task. Children were first shown a short physical demonstration in which the experimenter introduced two identical 500ml measuring-containers filled with water, on a table in front of the child. Children were asked to indicate whether one of the containers held more water, or whether they both held the same amount[1]. Once the child made their initial response the experimenter then poured the liquid from one of the original two containers into a third container. The child was asked again to indicate which container held more water or whether they were equal. Again, children were told that the computer-based task would follow the same format.

An instructional video was then played via the laptop on to the monitor. The video illustrated a trial in which two identical containers were filled with equal levels of "liquid". They listened to the video prompt asking whether the containers were equal or not and watched as three buttons were highlighted on the bottom of the screen. Again these buttons corresponded to "left (more)", "same" and "right (more)". Children were shown that once a button was pressed, one of the containers would be poured into a third container. This section of the video showed an animation of the pouring from one container to a new container. A sound file (recording of "pouring liquid") was synchronised with the pouring action. The prompt was played again asking for a decision on which container held more. Verbatim instructions are provided in Appendix G.

A total of 4 practice trials followed the instruction video. These consisted of 2 examples of problems starting with identical quantities and appearances and ending in identical quantities and appearance and 2 examples beginning and ending with different quantities and appearances. Children then completed the experimental trials consisting of a total of 36 problems. Type (i) of Table 7-2 provided the trials used to assess conservation accuracy, RT and Height Bias (the equivalent of the Length Bias for liquid). As in the Number task, conservation was assessed as the proportion of correct responses on these trials where the transformation had altered the perceptual but not the conceptual properties of the stimuli. Height Bias was assessed in the following way. If following a conserving transformation (either pouring to a

---

[1] Even though almost exactly even quantities of water were given in the initial presentation, children often claimed one container held more. Thus, a dropper was used to "even up" the level of one container until the child was satisfied the two containers were equal.

narrower and taller, or wider and shorter container) the child indicated that the container showing the higher level of liquid contained more liquid, then this was taken to indicate a bias of height. Each of the other trial types presented in Table 7-2 were distractors.

Table 7-2 Problem types for conservation of liquid task

Type	*Start**	*Transformation*	*End**	*N trials*
i)	SS	narrow+tall/wide+short	DS	6
ii)	SS	poured into identical	SS	6
iii)	DD	narrow+tall/wide+short	SD	6
iv)	DD	poured into identical	DD	6

*SS=Perceptually-Same AND Conceptually-Same,
DD=Perceptually-Different AND Conceptually-Different,
SD=Perceptually-Same AND Conceptually-Different,
DS=Perceptually-Different AND Conceptually-Same

Figure 7.2 and Figure 7.3 present screenshots of the key stages within the animated number and liquid tasks, respectively. Tile 1 of each figure illustrates the starting screen where children were asked to decide which side had greater number of coins / greater amount of liquid, or whether both sides were equal. Children responded by pressing one of three on-screen buttons (‘left more’, ‘same’, ‘right more’).

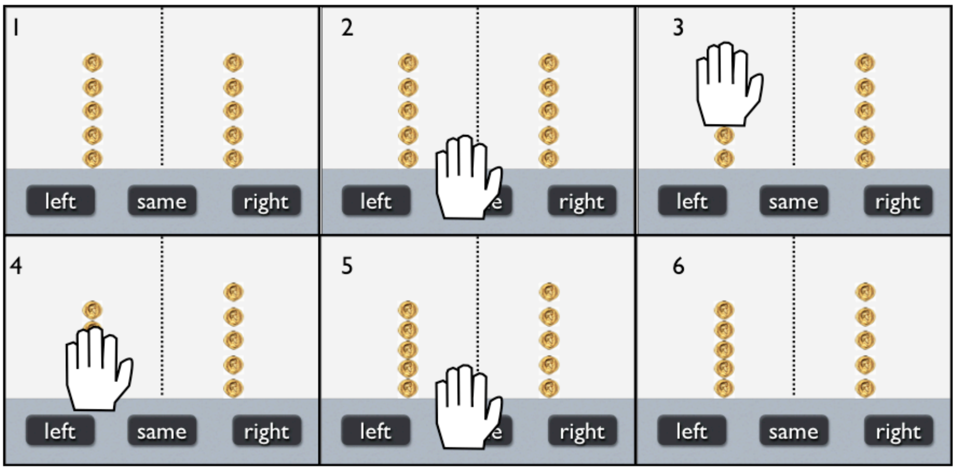


Figure 7.2 Screenshots of the computer-based version of the conservation of number task. In this example, a transformation compresses the rows of coins on the left-hand side, but leaves their number unchanged.

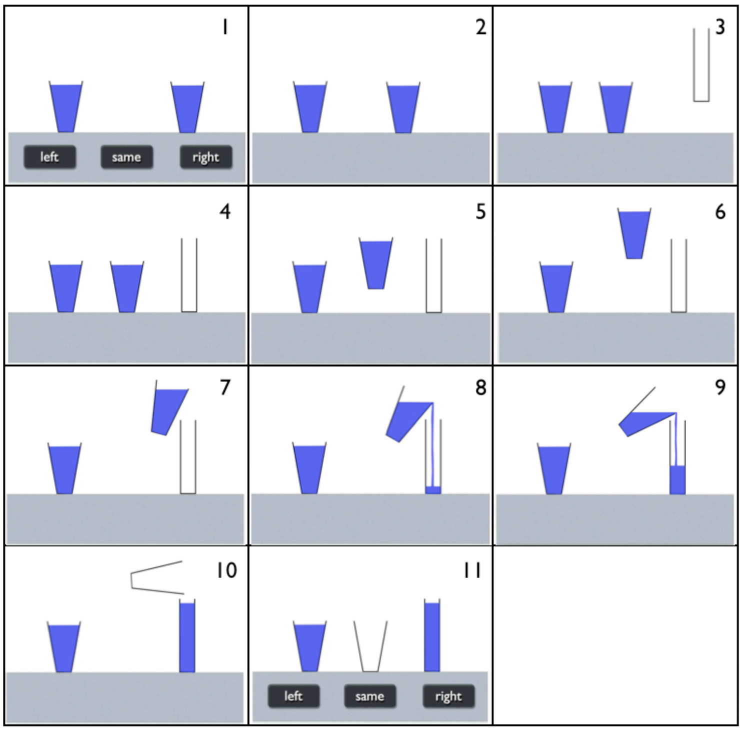


Figure 7.3. Screenshots of the computer-based conservation of liquid task. In this example, a transformation occurs where the liquid from one of two identical containers is poured into a longer, narrower container. The volume has not changed.

Materials

The physical demonstrations required the use of 10 large plastic counters and three measuring containers (two initially filled with water, the third empty). These were not used for any other aspect of the task. The experimental tasks were coded in MATLAB^{TM} by Frank Baughman and required the use of a touchscreen display and external speakers attached to an Apple Mac G4 iBook laptop computer (1.33 GHz, 1GB RAM).

Treatment of missing and incomplete data

Number: All participants completed this test. However, a total of 10 children in the younger Primary School sample failed all 4 of the conservation of number problems. Therefore, their accuracy was 0% and their RT for correctly solved conservation problems was undefined. As data existed for the errors made by these children (and were therefore relevant to assessing Length Bias), the decision was taken to replace their RT with the group average RT (accuracy remained 0%). This ensured these cases were not excluded during the analysis. The same procedures were applied for 2 children in the Primary School older children group. No data were replaced in the Secondary School sample.

Liquid: Data were missing on all measures for 6 younger children in the Primary School sample. Because there were no error data for these individuals, these data were not replaced and were excluded from the analysis. The final samples compared within the Categorical analysis are as follows: YHA=11, OLA=14; and in the Continuous analysis: YHA=14, OLA=20. No data were missing or replaced in the Secondary School sample. In the case of the categorical comparisons, removing these children did not affect the MA-matching. YHA and OLA groups remained matched on an MA of 8.2 years.

As we shall see, the effect of these missing data, in combination with the higher proportion of scores at floor, leads to the slightly odd outcome where the YHA appear to be *more* accurate on Liquid task than the Number task. This is odd because children typically demonstrate superior ability on conservation of number tasks, before they reach equivalent competence on conservation of liquid tasks (see e.g., Goodnow, 1966). We return to the issue of task validity in the discussion

Results

The data reported here are on conserving problems only. That is, the measures of RT, accuracy and bias effects relate to those problems where trials started with equal stimuli and their number/volume did not change following transformations. The other conditions provided the context so that it was unpredictable whether the target property would be conserved and a variety of transformations were experienced. Results are first presented for the Primary School MA-matched sample (YHA n=11; OLA n=14), then the Secondary School MA-matched sample (YHA=16; OLA=19) and then the full Primary and Secondary School samples combined. Before presenting the results, I briefly highlight two issues. Firstly, the ages of children in the Primary OLA group and both Secondary YHA and OLA groups are *above* the age typically associated with conserving[2]. We may therefore expect accuracy to be at ceiling on these tasks for older children. In this case, we may expect response times to offer sensitivity in discriminating group performance. Secondly, in this chapter, the results of the analysis using MA-CA disparity as a covariate are omitted in the full samples of Primary School and Secondary School children. For two reasons. First, the small number of conserving problems in each task (Number=4, Liquid=6) provided too limited a range for the prediction of accuracy using MA-CA disparity (i.e., performance scores were limited to a restricted range in the Number task of 0%, 25%, 50%, 75% or 100%, and in the Liquid task to 0%, 17%, 33%, 50%, 67%, 83% or 100% and in each task there were a number of participants at ceiling and floor which artificially affected regression lines). Second, though measurements of RT in principle has sensitivity and is suitable for use with MA-CA disparity as a covariate, the results of the analysis showed no MA-CA disparity x RT relationship.

---

[2] Piaget described these abilities generally emerging around the age of 7, during the concrete operation stage (1954).

Primary School MA-matched group comparisons

The performances of YHA and OLA groups on each of the dependent measures (overall accuracy, response times and length and height bias effects) are shown separately for the Number and Liquid tasks in Table 7-3. Here we can see that on the Number task, the YHA were less accurate compared to the OLA (YHA mean accuracy=23%, se=10%; OLA mean accuracy=63%, se=9%). This was also true for the Liquid task (YHA mean accuracy=42%, se=10%; OLA mean accuracy=85%, se=9%). The table further shows that compared to the OLA, the YHA showed a greater length bias (*Number*: YHA mean bias=77%, se=10%; OLA mean bias=36%, se=9%) and height bias (*Liquid*: YHA mean bias=49%, se=10%; OLA mean bias=12%, se=9%). On RT, Table 7-3 shows the YHA were faster than the OLA group on the Number task (YHA mean RT=1.90s, se=0.46; OLA mean RT=2.20s, se=0.40) and the Liquid task (YHA mean RT=2.73s, se=0.54; OLA mean RT=2.99s, se=0.48). While this combination of lower accuracy and faster response times often indicates a speed-accuracy trade-off, a t-test on each group's inverse efficiency scores (i.e., RT/Accuracy), showed no reliable speed-accuracy differences. Using a rule assessment methodology approach (e.g., Siegler, 1998), analysis of children's preference for height/length bias did not change, or add to the results presented here.

Table 7-3. Summary table of performance on Number and Liquid tasks for Primary MA-matched groups

Measure	*Task*	*Group*	*Mean*	*Std. error*	*diff* [a]	*Std. error*	
Accuracy	Number	YHA	23%	(10%)	40%	(10%)	*
		OLA	63%	(9%)			
	Liquid	YHA	42%	(10%)	42%	(10%)	
		OLA	85%	(9%)			
RT (secs)	Number	YHA	1.90	(0.46)	0.30	0.43	
		OLA	2.20	(0.40)			
	Liquid	YHA	2.73	(0.54)	0.26	0.51	
		OLA	2.99	(0.48)			
Bias	Number	YHA	77%	(10%)	-42%	(10%)	*
		OLA	36%	(9%)			
	Liquid	YHA	49%	(10%)	-37%	(10%)	
		OLA	12%	(9%)			

[a] The difference between OLA and YHA

* indicates reliable group differences

This table presents the results of the analyses for Number and Liquid tasks combined. Three 2 (Group: YHA vs. OLA) x 2 (Task: Number vs. Liquid) repeated measures ANOVAs were performed to test Accuracy, RT and Bias effects. These analyses showed reliable main effects of Group on: Accuracy (F(1,23)=14.92, p=.001, η^2=.393), which stemmed from higher overall accuracy in the OLA group (mean accuracy=74%, se=7%) compared to the YHA (mean accuracy=33%, se=8%); and Bias (F(1,23)=14.92, p=.001, η^2=.393). This was due to a greater Bias effect in the YHA (mean Bias=63%, se=8%) compared to the OLA (mean Bias=24%, se=7%). No group differences on overall response times were observed (mean RT=2.3s, se=.32; OLA mean RT=2.6s, se=.28).

Reliable main effects of Task were found on Accuracy (F(1,23)=5.66, p=.026, η^2=.197) and Bias (F(1,23)=9.70, p=.005, η^2=.296). These effects stemmed from higher overall accuracy on the Liquid task (64%, se=7%) compared to the Number task (43%, se=7%) and a greater bias effect on the Number task (mean Bias=57%, se=7%) compared to the Liquid task (mean bias=30%, se=7%). No reliable effect of Task on Response Times was observed (*Number*: mean RT=2.1s, se=.31; *Liquid*: mean RT=2.9s, se=.36) and no Group x Task interactions were observed for any of the dependent measures. This indicates that the relative differences between groups were similar across the two tasks.

Secondary School MA-matched group comparisons

Table 7-4 presents a summary of the performances of the YHA and OLA groups separately for the Number and Liquid tasks. Whereas in the Primary School data, the OLA showed higher accuracy than the YHA, in the Secondary School data Table 7-4 shows higher accuracy in the YHA group on both Number (YHA mean accuracy=97%, se=3%; OLA mean accuracy=95%, se=3%) and Liquid (YHA mean accuracy=90%, se=8%; OLA mean accuracy=75%, se=7%) tasks. Of the number of errors made in the Number task, Table 7-4 shows a similar proportion were attributable to length bias in the YHA and OLA groups (*Number*: YHA mean bias=3%, se=3%; OLA mean bias=4%, se=3%). However, in the Liquid task the proportion of errors due to height bias was greater in the OLA compared to the YHA (*Liquid*: YHA mean bias=2%, se=6%; OLA mean bias=17%, se=6%). Table 7-4 also shows that compared to the YHA, the OLA were faster to respond on the Number task (YHA mean RT=1.96s, se=0.21; OLA mean RT=1.30s, se=0.20) and the Liquid task (YHA mean RT=2.39s, se=0.35; OLA mean RT=1.78s, se=0.32). Once more, the data indicate that the group with the more accurate performance responded more slowly.

Table 7-4. Summary table of performance on Number and Liquid tasks for Secondary MA-matched groups.

Measure	*Task*	*Group*	*Mean*	*Std. error*	*diff* [a]	*Std. error*	
Accuracy	Number	YHA	97%	(3%)	-2%	(3%)	
		OLA	95%	(3%)			
	Liquid	YHA	90%	(8%)	-14%	(7%)	
		OLA	75%	(7%)			
RT (secs)	Number	YHA	1.96	(0.21)	-0.67	0.21	*
		OLA	1.30	(0.20)			
	Liquid	YHA	2.39	(0.35)	-0.60	0.34	
		OLA	1.78	(0.32)			
Bias	Number	YHA	3%	(3%)	1%	(3%)	
		OLA	4%	(3%)			
	Liquid	YHA	2%	(6%)	15%	(6%)	
		OLA	17%	(6%)			

[a] The difference between OLA and YHA

* indicates reliable group differences

The results of three 2 x 2 repeated measures ANOVA on Accuracy, RT and Bias effect, revealed a reliable Group difference on RT (F(1,33)=6.50, p=.016, η^2=.164). This difference stemmed from faster overall responses in the OLA (mean RT=1.54s, se=0.17) compared to the YHA (mean RT=2.17s, se=0.18). This was the only reliable group difference that was found. However, a reliable main effect of Task on Accuracy (F(1,33)=5.59, p=.024, η^2=.145) showed higher overall scores were obtained on the Number task (*Number*: mean accuracy=96%, se=2%; *Liquid*: mean accuracy=83%, se=5%).

Combined Primary and Secondary School Results

Three 4 x 2 repeated-measures ANOVAs were performed on combined Primary and Secondary School data, using Group (Primary-Younger, Primary-Older, Secondary-Younger, Secondary-Older) as the between-participants factor and Task (Number, Liquid) as the within-participant factor. These analyses found reliable main effects of Group on Accuracy ($F(3,56)=20.99$, $p<.001$, $\eta^2=.529$), RT ($F(3,56)=4.32$, $p=.008$, $\eta^2=.188$) and Bias effect ($F(3,56)=23.21$, $p<.001$, $\eta^2=.554$). They also showed a reliable main effect of Task on Bias effect ($F(1,56)=5.13$, $p=.027$, $\eta^2=.084$). Figure 7.4 depicts the source of these effects on Accuracy (Tile A) and Bias (Tile B). Figure 7.4B shows, for example, that when taken together the entire sample showed a higher bias effect on the Number task (mean length bias=30%, se=3.10), compared to the Liquid task (mean height bias=20%, se=3.80).

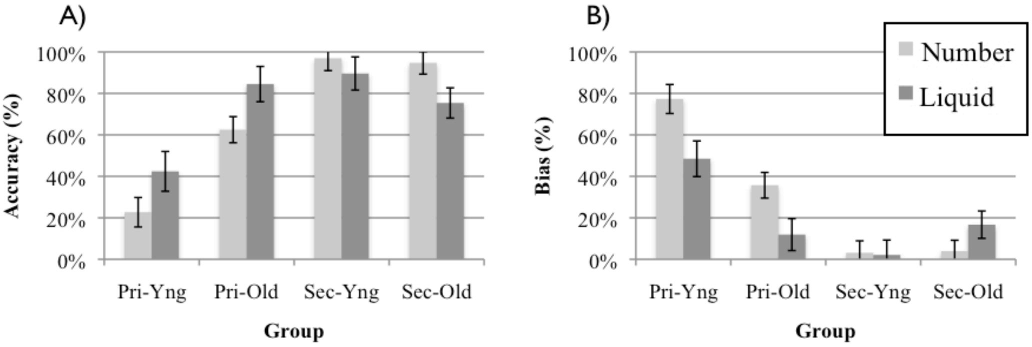


Figure 7.4 Mean accuracy (Tile A) and Bias (Tile B) on Number and Liquid tasks for younger and older groups within Primary and Secondary School combined data.

In addition, the analysis revealed a reliable main effect of Task on RT (F(1,56)=4.93, p=.030, η^2=.081). Figure 7.5 illustrates that taken together the entire sample was on average faster to respond to Number (mean RT=1.84s, se=.15) than Liquid (mean RT 2.47s, se=.21).

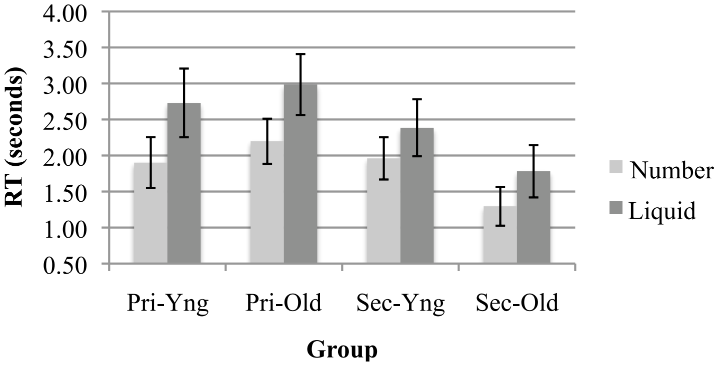


Figure 7.5 Mean response times on Number and Liquid tasks for younger and older groups within Primary and Secondary School combined data.

Discussion

Because the ability to conserve has been argued to emerge at around 7 years of age (Piaget, 1954), ceiling scores were expected in the samples of Secondary School children. At this school level, the computer-based versions of the two conservation tasks offered no challenge and the results showed no sensitivity to measures of accuracy or bias effect. Measures of RT, however, did discriminate the groups at this level, indicating greater sensitivity. The same outcome might also have been expected of the older group of Primary School children, where ages ranged between 10-11years old. However, this was not the case. The analyses of the Primary MA-matched groups showed these groups were not at ceiling and they differed reliably on both accuracy and bias effects, thus suggesting differences between groups in their ability to represent quantities (at least as represented in the computerised format). The Primary School groups also showed no difference in the overall time taken to respond correctly. These two findings stand in contrast to results in previous chapters where RT was consistently found to differ between groups at this school level, but measures of information processing were not found to differ.

These results suggest that there are indeed differences in the underlying processing between groups of Primary School MA-matched younger and older children. More specifically, in Piagetian terms, these data indicate that groups who are of the same overall ability but of different ages are at different developmental levels. This outcome is more in line with the pattern we might have predicted given Spitz's (1982) review, in which different strengths and weaknesses were demonstrated in the cognitive profiles of similar MA-matched younger and older children. Thus the results from the Primary School analysis suggest, for the first time so far, that *differences in intelligence and differences in cognitive development are not the same*.

Turning to the data from the Secondary School level, the analyses showed an inverse pattern of results compared to the Primary School sample. Here, the results showed MA-matched groups were not different in measures of accuracy or bias effect. At these ages, the tasks appeared to offer no sensitivity to the measures, suggesting groups were equal in their understanding of quantities. Instead, it was response times that separated the performance of the younger and older groups. This too contrasts with the pattern of results described in previous chapters. What might explain the lack of differences in RT at Primary School level and the reliable differences at Secondary School level favouring OLA? Considering the influence of the task design, it is possible that one source of differences stems from children's sensitivity towards the timecourse of events within the tasks. In both conservation tasks, children were presented with a picture of two stimuli and were required to make a decision regarding their equivalence by touching one of three on-screen buttons. Children's responses were followed by an animated component (i.e., in the number task a hand moved across the screen transforming one of the rows of coins; in the liquid task the contents of one cup were poured into another). Children were then asked again to indicate the equivalence of the two sets of stimuli. The response buttons appeared only once the animated component had finished, effectively imposing a deadline when responses could be made. It is possible that the presence of RT differences in the Secondary School were due to a greater sensitivity in the OLA group to the schedule of those events. That is, they may have been more aware of the length of each trial and thus aware of when to expect buttons to re-appear. Combined with the fact that the position of the buttons did not change, the older children may have been more ready to respond (e.g., finger nearer to the screen where buttons were expected to appear) when trials ended. By contrast, the absence of RT differences in the Primary school groups may be due to lower awareness in both groups of the timecourse of events. It is possible that children in these groups initiated their responses after the animated component had fully finished. This possibility could be tested by re-designing the task so that the buttons remained on-screen during the animated component.

Overall, it appears that the results from the combined Primary and Secondary School data favour an experience-based view of Piagetian stage-like progression. For example, the data suggest that at Primary School level, the greater experience of the older group offers advantages to these children that greater intelligence and less experience does not. That is, older children who have had more time and experience dealing with, interacting with and manipulating real objects, have developed a better understanding of the properties of those objects and the effects of certain types of transformations. At these ages the results suggest that individual differences and cognitive development are different forms of cognitive variability. Yet, by the time children reach adolescence the advantage of just a few more years' experience is negligible. The Secondary School data suggests that the amount of experience that both younger higher ability and older lower ability children have obtained is sufficient to perform at similar levels on the tasks. At these ages, the results suggest no differences between individual differences and cognitive development. While this interpretation appears to account for the pattern of results, it is made tentatively. This is because the small number of conservation problems did not allow for a full analysis of accuracy using each child's MA-CA disparity. Therefore, we cannot exclude the possibility that one's intelligence also has a modulatory effect on conservation performance. If a greater number of these problems were administered and MA-CA was still found not to modulate performance reliably, then this would add strength to the above interpretation. Alternatively, if MA-CA disparities were found to modulate accuracy, then the experience-based view of Piagetian stage-like progression would be challenged.

Using an existing developmental computational model of the conservation of number task (Shultz, 1998), Richardson, Forrester, Baughman and Thomas (2006) investigated the influence of neurocomputational mechanisms on variability in conservation performance. They targeted a range of model parameters including ones relating to computational complexity (hidden layers), capacity (numbers of units in a hidden layer), plasticity (learning rate) and the ability of the model to make categorical, or rule-like distinctions (sigmoid function). Within the literature, the influences of these mechanisms have been explored with relation to variability in typical and atypical development (see e.g., Thomas & Karmiloff-Smith, 2003). From Richardson et al. (2006), the results of the manipulations to the sigmoid function are particularly relevant. Briefly, within learning systems, steeper sigmoid functions generally allow for sharper category boundaries to be developed. These may be especially useful on tasks where rule-like decisions are required to learn relations. Shallower sigmoid functions by contrast offer the potential of learning a greater number of fine-grain distinctions. Richardson et al. (2006) showed that during the early stages of a model's learning, the effects of a shallow sigmoid had a significant impact on the ability of the model to acquire conservation compared to normal. This was evidenced by larger than normal bias effect and more training events needed for the model to reach normal levels of accuracy. Importantly though, at later stages, this model showed a convergence in performance compared to the normal model. That is, the model reached ceiling levels of accuracy and was then indistinguishable compared to the normal model. This resembles the pattern of conservation performance found between the Primary School younger and older and the Secondary School younger and older groups in this chapter. Thus, it is possible that differences in the ability to learn category, rule-like boundaries offer a potential source of differences in younger and older children who are matched on MA.

In summary, previous chapters uncovered little evidence of the kinds of differences between younger and older MA-matched groups that Spitz described. The pattern of results from Chapters 4, 5 and 6 was consistent with the idea that intelligence and cognitive development are similar forms of cognitive variability. However, the findings presented in this chapter suggest this is too simplistic an account. The results of the conservation tasks suggest that important differences do underlie the reasoning processes of younger and older MA-matched groups. In the following chapter, I present the results of the balance scale task (Inhelder & Piaget, 1958) in order to pursue further the question of group differences. The balance scale task forms a logical progression of our investigation because, like the conservation tasks, it requires an integration of knowledge from more than one dimension in order to perform successfully. In the case of the balance scale task, the dimensions that children must integrate relate to the quantities of *weights* and their respective *distances* from a fulcrum. The task offers two key improvements over the conservation tasks. Firstly, the balance scale task is able to discriminate accuracy over a greater range of ages (including into adulthood). Secondly, the presence of a large number of problems that are organised into sets consisting of unique combinations of weight and distance dimensions allows children's understandings of these different problem types to be assessed independently. Indeed, these different problem types have been used extensively to evidence the emergence of more complex forms of rule-like behaviour.

Given these improvements, the balance scale task should offer greater sensitivity to accuracy and RT in the groups of MA-matched Primary and Secondary School children and once again allow us to test the contribution of MA-CA disparities on performance. The central focus of the following chapter, then, is to examine further the possibility that groups of younger and older MA-matched children may differ in their ability to integrate and thus process information. If, as the results of the conservation tasks appear to suggest, the YHA and OLA do differ in their ability to make rule-like distinctions, then we might expect to find reliable group differences as problems demand greater ability to integrate knowledge of the two dimensions. The balance scale task may also give us greater insight into the RT differences we observed for the conservation tasks.

Chapter 8 The Balance Scale Task

Introduction

The original version of the balance scale task was devised by Piaget's colleague Inhelder (1958) and it involved presenting children with a scale from which cradles were hung. Weights were added to the cradles and children moved the cradles to different positions on the scale in order to achieve a balance. Piaget believed that in order to get the scale to balance properly, children must understand the relationship between two key variables (or dimensions) constraining the task – the number of *weights* on each arm of the balance, and their respective *distances* from the fulcrum. Administering the task to children of different ages, he noted that younger children typically reasoned about only one of the dimensions and that they appeared to apply this reasoning to a range of problems, often resulting in incorrect solutions. Later, around the ages of 11-12 years old, Piaget argued that children begin to acquire the concept of *proportionality* – the understanding that change on one variable is directly proportional to change on another (see e.g., Normandeau, et al., 1989) – a vital ingredient to successful task performance. However, children of this age also frequently showed a basic lack of understanding on particular problems where the information from both dimensions had to be integrated. From this Piaget asserted that a kind of 'naïve physics' biased the reasoning of children in this stage – a formal understanding of the problem constraints was not yet achieved.

Siegler (1976) later modified the basic design of the balance scale task, such that the cradles were substituted for pegs on which weights could be placed. Supports were also added to prevent the scale from tipping. Figure 8.1 illustrates Siegler's revised design.

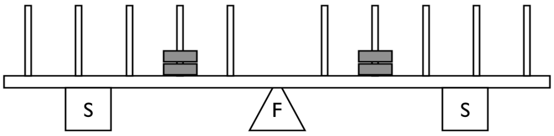


Figure 8.1. Depicting Siegler's (1976) revised balance scale apparatus. A number of weights (in grey) are placed on pegs on either side of a fulcrum (F). The child is asked to predict the outcome of removing the supporting blocks (S).

Under Siegler's revised version of the balance scale task, the aim was to predict the outcome of removing the supporting blocks (i.e., to state whether the balance scale would tip left, balance, or tip right), as means of demonstrating an understanding of proportions was the same. Siegler (1976) delineated six unique types of problem that were derived from different combinations of the weight and distance dimensions. These problem types, illustrated in Figure 8.2, are: (A) *Balance*: both the number of weights and their distance from the fulcrum is identical and the scale balances; (B) *Distance*: the number of weights on each side is identical, but their respective distances from the fulcrum differ and the side with the greatest distance wins; (C) *Weight*: the numbers of weights on each side differ, their respective distances from the fulcrum are equal and the side with the greatest number of weights wins; (D) *Conflict Balance*: in which the numbers of weights and the distances from the fulcrum on each side are both dissimilar, but the scale balances; (E) *Conflict Distance*: the scale contains a larger number of weights closer to the fulcrum on one side and a smaller number of weights farther from the fulcrum on the other and the side with the greatest distance wins; and (F) *Conflict Weight*: one side has a greater number of weights closer to the fulcrum, the other side has fewer weights farther from the fulcrum and the side with the greatest number of weights wins.

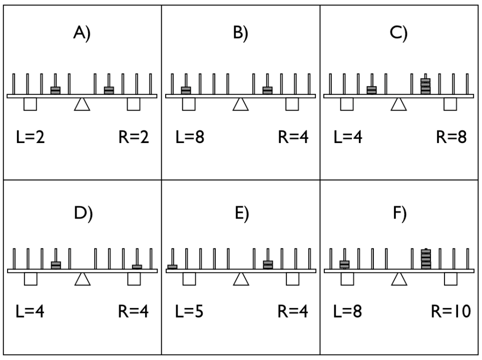


Figure 8.2. Illustrating six problem types on the balance scale. Each cell shows the amount of torque on each side of the fulcrum. (See text for details).

From his observations of children's performance on these problem types, Siegler proposed that four core rules typified the reasoning of children at different stages of development (1976). According to the rule assessment methodology that Siegler developed (1978), during earlier years (around the ages of 4-5 years-old) children are able to attend only to the numbers of weights (*Rule I*), then they begin to incorporate distance information but only when the number of weights is equal (*Rule II*), then the child understands weight and distance are important in all cases but 'muddles through' without a proper understanding of the way to combine this information (*Rule III*), before the child finally integrates information relating to weight and distance, applying the rule of torque (*Rule IV*) – see Equation 2 below. Though the descriptions and the number of rules that Sielger proposed have been challenged (see e.g., Jansen & van der Maas, 2002; McFadden, Dufresne, & Kobasigawa, 1987; Shultz & Takane, 2007), the view is held by a number of researchers that rules do underlie complex reasoning on tasks such as the balance scale task (e.g., Laird, 2008). Such theoretical frameworks can be found embodied within various computational approaches in which high-level cognitive processes are assumed to rely on a growing set of more complex rules (for a recent review of these see e.g., Langley, Laird, & Rogers, 2008).

$$\tau = r \text{ x } F$$

Equation 2. The explicit rule for calculating torque (τ) = distance (r) x force (F)

At a behavioural level, Siegler's analysis of errors offers a baseline from which age-related normative levels of performance may be established. In Figure 8.3 below, I plot the trends reported by Siegler (1976) to illustrate the typical profiles for the age groups he tested. Along the X-axis, problem types have been organised according to their observed level of difficulty with each group's mean level of accuracy (proportion correct) plotted on the Y-axis. This figure shows relatively high levels of accuracy on problems involving only one dimension (i.e., manipulations to either numbers of weights, *or* distance from fulcrum) and relatively lower levels of accuracy on problems where both dimensions are combined (i.e., manipulations to both numbers of weights and distances from the fulcrum).

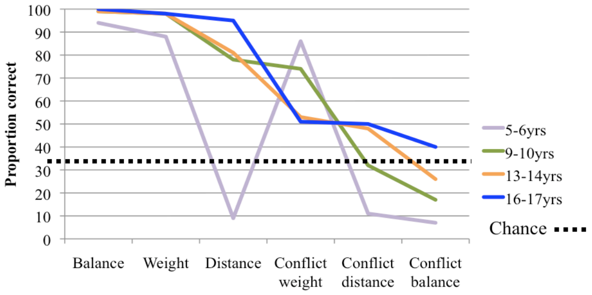


Figure 8.3. Siegler's (1976) developmental trends on the balance scale task: Plotting the percentage of each problem type predicted correctly. Dashed line indicates 33% level of chance.

More recently, van der Maas and Jansen (2003) included an analysis of response times on balance scale performance of children of different age groups. Below, in Figure 8.4, I re-plot their data showing both accuracy (Tile A) and RT (Tile B).

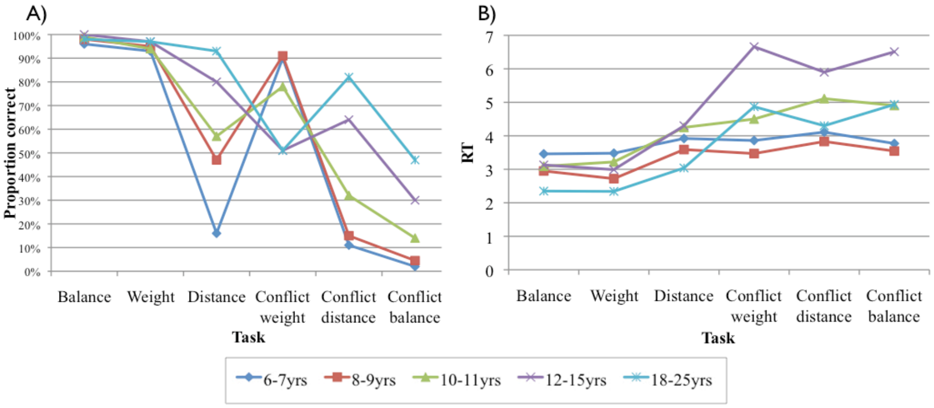


Figure 8.4. Plotting van der Maas and Jansen's (2003) data showing Tile A: Accuracy, Tile B: Response Times in seconds, for different age groups.

While the age groups used by Siegler (1976) and van der Maas et al. (2003) were not identical, the respective profiles of these samples can be seen to share a great deal of similarity across the six problem types (see Figure 8.3 and Tile A of Figure 8.4, respectively)[1]. Additionally, van der Maas and Jansen's analysis of RT, shown in Tile B of Figure 8.4, depicts the finding that RT in some conditions actually *increases* with age. For example, the response times of the 6-7 year-olds and the 8-9 year-olds show little change across problem type. However, the RT for the 12-15 year-olds shows a clear increase. Using RT, van der Maas et al. (2003) demonstrated

---

[1] Note: For consistency with the analysis I collapse the two sub-types of balance conflict items that van der Maas et al. (2003) used to examine the compensation rule.

that RT can reliably predict which rules children appear to be using on the different problem types[2]. Although calculating rule usage is not covered in this chapter, the plots of data from Siegler and van der Maas et al. will become relevant later, when comparing the performance of YHA and OLA MA-matched groups.

While on one hand the developmental trends that Siegler observed might be taken to offer support for Piaget's stage-like theory of cognitive development, on the other hand, there have been a number of important criticisms that are relevant to note here. Several of these criticisms have been argued to lend weight to *nativist* accounts – the position that the cognitive abilities that we come to possess as adults are relatively independent modules, present from birth and a product of evolution (see e.g., Fodor, 1983). These include for example, the claim that children appear to be competent in a given domain *earlier* than Piaget assumed (e.g., Spelke, et al., 1995), that children often do not show uniform abilities across different domains and that not all children (or even adults) appear to reach the final stage of formal operations (see e.g., Flavell, 1982). Other criticisms have focused on logical issues. For example, Fodor (1983) argued it is not possible for a system to be capable of determining new, more complex information requires a new schema to be created, if the structure for processing the more complex information is not already in place. (For a fuller review of several key criticisms of Piaget's theory see e.g., Brainerd, 1974; Flavell, 1982). Later, in Part 6 of the thesis we will consider some of the ways modularity accounts have influenced theories of intelligence[3]. However, the key point that relates to these criticisms concerns the feasibility of learning mechanisms in bringing about the marked differences in cognitive ability seen between children of different ages. Thomas et al. (2002) point out one of the problems with such debates is that evaluating competing verbal accounts of development can be difficult when there is no consensus on exactly how to characterise a given 'learning mechanism'. They argue that computational methods offer a clear advantage over verbal theories in the sense that they require a formal specification of theory. Their

---

[2] RT is predicted as follows: Rules underlie responses; problems that are more complex comprise rules with more complex set of steps to execute. The time taken to produce a response is the sum of the steps needed to produce it.

[3] In Chapter 11, I explore alternative views concerning how cognition may be organised (i.e., its functional architecture) by developing a series of dynamical systems models. This approach is useful in that it offers a simplified framework within which questions that relate to development can be tested.

argument is supported by a review of a series of connectionist and generative connectionist models of various cognitive tasks. They choose, as examples, models of the balance scale task (McClelland & Morris, 1989), seriation (Mareschal & Shultz, 1999) and conservation tasks (Shultz, 1998). In each example, they explain how the model in question captures key empirical phenomena. Importantly, in the case of the balance scale task the simulation of the stage-like changes and abrupt shifts in behaviour seen in children is explained as the accumulation of small, graded changes between the connection weights over time (see also Raijmakers, van Koten, & Molenaar, 1996). In this way, Thomas et al. (2002) argue that computational modelling may also offer a way to reconcile the seemingly opposite views taken by Piagetian and Fodorian theorists.

Richardson, Baughman, Forrester and Thomas (2006) extended the original model of McClelland et al. (1989) to examine the consequences of changes to a number of the model's parameters. They focused on changes to the computational properties of the network (e.g., the number of hidden layers, the number of hidden units and the learning rate), the problem encoding, and the effects of more or less experience of the problem via the environment. Richardson et al. (2006) argued for the importance of studying variability for the following three reasons: (1) intra-individual variability in performance appears to accompany changes between ability on more complex problem types (i.e., there is a period of instability around transitions); (2) that the variability of people within the same age provides information on differences in intelligence; and (3) that the variability represented by a divergence from the normal developmental profile may shed light on qualitative differences in reasoning within disorder groups.

Within the literature there appears to be a paucity of work focusing on variability on the balance scale task (see e.g., Richardson, Baughman, et al., 2006). Two notable exceptions include work by Siegler and work by Jansen and van der Maas. Briefly, in both cases, the focus of these researchers is on the intra-individual variability that occurs within short timeframes (e.g., on a single trial of the balance scale task). For example, according to Sielger's (2006) *overlapping waves theory*, intra-individual variability can arise from individuals knowing and choosing among a variety of possible strategies, *within a given task*. Siegler (2007) illustrates this form of variability citing evidence of the use of multiple strategies in infants (Adolph, 1997), toddlers (Chen & Siegler, 2000), preschoolers (Tunteler & Resing, 2002) and adults (Alibali, 1999). Over the longer course of development, Siegler proposes that children adapt their use of strategies based on the previous effectiveness of those strategies. Indeed, support for the use of more advanced strategies over age is prevalent within the literature (see e.g., Shrager & Siegler, 1998). However, one question concerns how exactly these new strategies emerge. That is, what are the mechanisms that drive change? In Siegler's overlapping waves theory these mechanisms are not formally specified. Thus, by and large, it appears primarily a verbal theory.

Formal specifications of intra-individual variability in behaviour have been offered in work by Jansen and van der Maas (2001). They applied catastrophe theory within computational models to illustrate patterns of discontinuous (i.e., catastrophic) change on problems, such as the balance scale task. They explained the models behaviour as the result of a dynamic interaction between the system's current state and parameters they call 'control variables'. In comparing the ability of McClelland's connectionist model of the balance scale using a set *catastrophe flags* (i.e., predictions that are made on the properties of the model. See van der Maas & Raijmakers, 2009), van der Maas et al. (2009) demonstrate that the connectionist model fails and argue connectionist models in general do not offer a plausible approach to studying higher level cognitive processes. However, Thomas, McClelland, Richardson, Schapiro and Baughman (2009) have shown that the failure of McClelland's original model was due to the specific objectives and attendant simplifications that motivated that particular model, rather than an inability of connectionist approaches per se. Additionally, Schapiro and McClelland (2009) have demonstrated that a modified version of McClelland's (1989) model can exhibit the catastrophe flags highlighted by van der Maas et al.

The focus within this chapter and of the thesis as a whole is on variability. However, it is more specifically focused on contrasting variability across age with variability within age. Thus, it is concerned with exploring the second type of variability highlighted by Richardson et al. (2006). MA-matching looks at intelligence and cognitive development at the same time. By virtue of the rich empirical and theoretical literature on the balance scale task, applying this task to groups of younger and older groups of MA-matched children should elucidate further the relationship between intelligence and cognitive development.

Current aims

The aims of this chapter are to assess the degree to which MA-matched younger and older children are similar in their reasoning ability on the balance scale task. In doing so, I aim to gain a clearer picture of the relationship between intelligence and cognitive development. That is, if these groups show no underlying differences in their performance on the task, this would be evidence that, with respect to proportional reasoning, intelligence and cognitive development are variations that lie along the same dimension. On the other hand, reliable differences between MA-matched younger and older groups would suggest differences in intelligence and differences in cognitive development are not the same thing. Importantly, in this study, the balance scale procedure was adapted so that reaction time data could be collected, allowing a greater degree of sensitivity in comparing the performance of the MA-matched groups (see also van der Maas & Jansen, 2003, for use of reaction times on the balance scale task).

Method

Participants

See ‘Participants’ section, Chapter 3 General Methodology. All participants completed this task.

Design

The between-participants variable was Group (YHA vs. OLA) and the within-participants variable was Problem Type (Balance, Weight, Distance, Conflict Weight, Conflict Distance and Conflict Balance). The dependent variables were accuracy and response time (measured in seconds) for correct responses.

Procedure

Participants were shown an apparatus of a real balance scale and two examples where the experimenter made the scale balance and two where the scale tipped to either side. They were then introduced to a 2-D representation of a simple balance problem (one weight, nearest peg right vs. one weight, nearest peg left) on a monitor and asked if they recognised the similarity between the 2-D version and the real version. All children stated that they saw the similarity and all children correctly predicted the simple problem shown in the 2-D version would balance.

Participants sat directly in front of a touchscreen display, at a comfortable arm’s length (approximately 30-50cm) and then watched an instruction video outlining the task requirements. In this video, children were shown that a sequence of 2-D representations of the balance scale would appear on the screen and they would need to decide which side the scale would tip, or whether it would balance when “supports” were removed, by touching a “button” on the screen (see Figure 8.5). In the video, an animated hand moved to one of the buttons (illustrating a possible response). This was followed by an animated component where the supporting blocks moved away from the fulcrum with the appropriate outcome of either the scale tipping to one side, or balancing (verbatim instructions given in the video can be found in Appendix B).

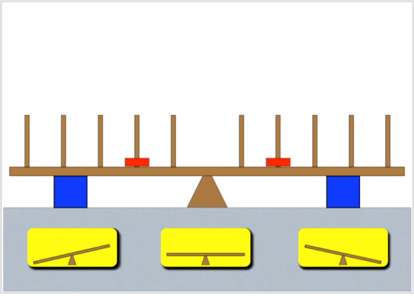


Figure 8.5. The computer-based presentation of the balance scale with three on-screen response buttons for indicating the consequence of removing the blue supporting blocks.

Children then completed a short test trial to ensure they understood the task requirements and how to interact with the touchscreen. In this trial version, children were given 6 simple problems showing each of the possible outcomes: *tip left* (2 weights nearest left vs. 1 weight nearest right, 2 weights second-nearest left vs. 1 weight second-nearest right), *balance* (1 weight nearest left and nearest right, 2 weights nearest left and nearest right) and *tip right* (the reverse of ‘tip left’ examples). As instructed, children made their selection by pressing the button on the touchscreen corresponding to their choice. Animated and audio feedback accompanied their response. Once the button was touched it would become highlighted, the blue supporting blocks would move to the sides and the balance scale would either tip to one side, or remain in balance. Depending on the participant’s choice, an audio wav file was synchronised to play either “tada” for correct responses, or “wrong buzzer” for incorrect ones. Feedback was present for both the practice and the experimental trials. A total of 96 unique problems were presented in randomised order (see Appendix A). These consisted of 16 of each of the problem types: Balance, Distance, Weight, Conflict Balance, Conflict Distance and Conflict Weight. In each problem type, problems were counter-balanced for outcome (tip-left, balance and tip-right) and for torque. The total set of problems was administered in 8 blocks of 12, allowing for short breaks when needed.

Materials

Apparatus included a real balance scale with two weights (for instructional use only). The experimental tasks were coded in MATLAB^{TM} by Frank Baughman and required the use of a touchscreen display and external speakers attached to an Apple Mac G4 iBook laptop computer (1.33 GHz, 1GB RAM).

Primary School results

Results are first presented for the MA-matched YHA (n=14) and OLA (n=14) groups, then the full sample of younger (n=20) and older (n=20) children. Log transformations of RT data were examined. However, the results of this analysis were identical in outcome to untransformed data. Here, data are presented for the original, untransformed RT data.

Primary School MA-matched group comparisons

Accuracy*:* A 2 x 6 repeated-measures ANOVA with Group (YHA vs. OLA) as the between-participants factor and Problem Type (Balance, Weight, Distance, Conflict Weight, Conflict Distance, Conflict Balance) as the within-participants factor, revealed no reliable overall group differences on accuracy (YHA mean=57%, se=3 vs. OLA mean=64%, se=3). A reliable main effect of Problem Type was found ($F(3.29,85.52)=90.68$, $p<.001$, $\eta^2=.777$)[4]. Figure 8.6 shows the source of this effect was higher overall accuracy on the problems involving only one-dimension (i.e., Balance, Distance and Weight) and lower overall accuracy on problems where the two dimensions were combined (i.e., Conflict Balance, Conflict Distance and Conflict Weight).

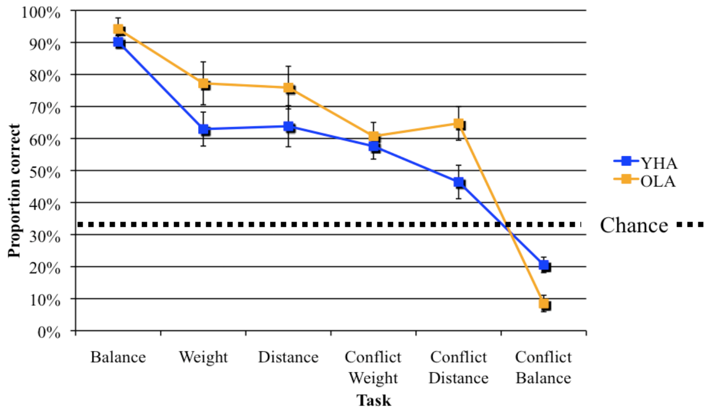


Figure 8.6. Mean accuracy across problem types for MA-matched YHA and OLA Primary School groups. Error bars show standard error of the mean. Dashed line indicates 33% chance level.

---

[4] Mauchly's test of sphericity showed significant differences in the variances on Problem Type (chi-square=32.52, p=.004), thus statistics relating to Greenhouse-Geisser are reported.

Of particular interest was the significant Group x Condition interaction (F(3.29,85.52)=4.04, p=.008, η^2=.134) which revealed reliable differences between the YHA and the OLA in their respective accuracies across the six problem types. Figure 8.6 indicates that this effect was due to a mixture of superior YHA and OLA performances. Post-hoc ANOVAs examining these differences revealed only one to be reliable following bonferonni corrections. This was between the YHA and OLA on Conflict Balance problems (F(1,27)=11.66, p=.002)[5]. Examining this result further, the accuracies of the YHA and OLA were tested against the likelihood that their performance was due to random choice of the three possible outcomes. Post-hoc t-tests showed that each group's scores were reliably different to 33% predicted by chance on both Conflict Balance (YHA: t(13)=-4.84, p<.001, 2-tailed; OLA: t(13)=-10.15, p<.001, 2-tailed) and Conflict Distance (YHA: t(13)=2.55, p=.024, 2-tailed; OLA: t(13)=6.07, p<.001, 2-tailed).

---

[5] For example, although Figure 8.6 shows a relatively large separation between error bars for the YHA and OLA on Conflict Distance, correcting for multiple post hoc comparisons showed the differences were not reliable (uncorrected: F(1,27)=6.08, p=.021).

Response times: A 2 (Group) x 6 (Problem Type) repeated-measures ANOVA performed on the RT data showed a reliable main effect of Group (F(1,26)=9.15, p=.006, η^2=.260)[6]. This was due to the fact that the overall performance of the OLA group was faster than the YHA group (YHA=3.69s, se=.17; OLA=2.96, se=.17). A significant main effect of Problem Type was also found (F(3.15,81.95)=8.66, p<.001, η^2=.250). Figure 8.7 shows that as difficulty of Problem Type increased, so too did the overall time taken to respond correctly. No Group by Problem Type interaction was found indicating that the OLA were consistently faster in all conditions.

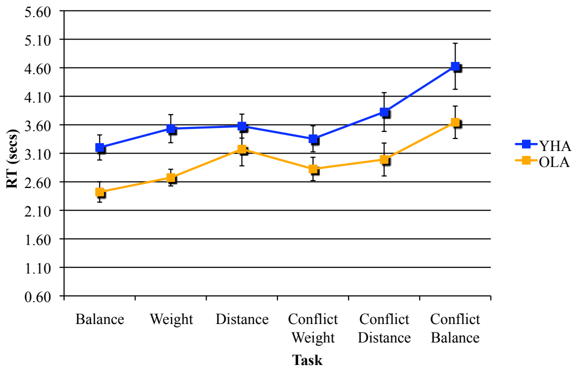


Figure 8.7. Mean RT across problem types on balance scale for MA-matched YHA and OLA Primary School groups. Error bars show standard error of the mean.

Primary School full sample analysis

Accuracy: Including MA-CA disparity as the covariate within a 2 (Group) x 6 (Problem Type) ANCOVA showed MA-CA disparity did reliably modulate overall accuracy (F(1,36)=5.74, p=.022, η^2=.138). Furthermore, now using the full sample of Primary School children, the analysis showed a main effect of Group (F(1,36)=17.87, p<.001, η^2=.332), which stemmed from greater levels of accuracy in the older group compared to the younger group (younger=50%, se 4; older=72%, se=3). Group and MA-CA disparity were not found to interact. In only one individual problem type did the main effect of disparity remain reliable following a bonferroni correction for multiple comparisons. This was on Conflict Balance (F(1,38)=8.07, p=.007, η^2=.175). Figure 8.8 plots MA-CA disparity data against accuracy for this condition only. It depicts separately the trendlines for the YHA (in

---

[6] Mauchly's test of sphericity showed significant differences in variance in task condition (chi-square=38.13, p<.001), thus Greenhouse-Geisser are reported.

blue) and OLA (in red) groups and a combined trendline (in green) and illustrates the mixture of separate age and advantage effects.

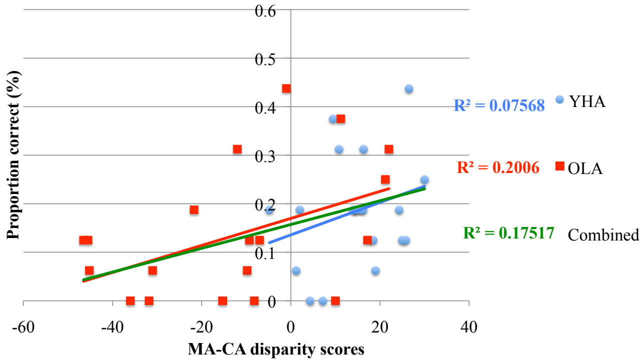


Figure 8.8. Accuracy by MA-CA disparities on Conflict Balance for full Primary School sample

Response times: A second 2 (Group) x 6 (Problem Type) ANCOVA showed neither Group nor MA-CA disparity scores were reliable in predicting overall RT, or separately on any of the individual problem types. Again, no Group by MA-CA interaction was found.

Summary of balance scale: Primary School data

Within the MA-matched group comparisons, overall accuracy was not reliably different between the younger and older children. Looking at the six problem types individually, the YHA and OLA groups differed reliably only on one problem type: Conflict Balance problems. Here, the YHA group showed they were significantly more accurate compared to the OLA group. This pattern was not found to be reflected in the RT data, where the OLA group showed they were uniformly faster across each of the six conditions. When MA-CA disparity scores were used as a covariate, no reliable overall effects were found for MA-CA on overall RT or overall accuracy. However, individual post-hoc ANCOVAs on the six problem types did reveal reliable effects of MA-CA scores on Conflict Balance accuracy.

Secondary School results

Results are first presented for the MA-matched YHA (n=16) and OLA (n=19) groups, then the full sample of younger (n=16) and older (n=19) children. Again, log transformations of response time data were examined. However, here too the results of analysis were identical in outcome to untransformed data. The original, untransformed data are presented.

Secondary School MA-matched group comparisons

Accuracy: As was the case in the Primary School data, a 2 (Group) x 6 (Problem Type) repeated-measures ANOVA revealed no reliable overall differences in accuracy between groups. The analysis of the Secondary School groups did show a reliable main effect of Problem Type (F(2.85,94.05)=111.78, p<.001, η^2=.772)[7]. This conforms to the expected pattern of highest overall performance on Balance problems and lowest overall performance on Conflict Balance problems (Siegler, 1976). No Group by Problem Type interaction was observed, indicating that the YHA and OLA Secondary groups performed similarly across problem types. Figure 8.9 shows the mean proportion correct for YHA and OLA groups across each of the six problems types. Note, that the performances of the Secondary YHA and OLA are at chance levels for the Conflict Balance problems. I return to consider the significance of this within the discussion.

---

[7] Mauchly's test of sphericity showed significant differences in the variances on Problem Type (chi-square=61.13, p<.001), thus statistics relating to Greenhouse-Geisser are reported.

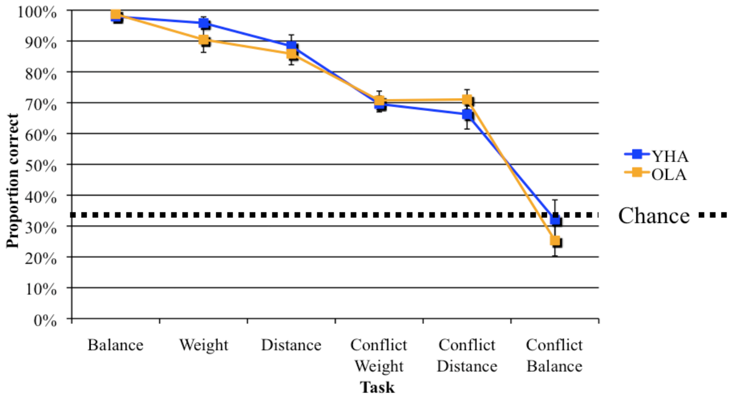


Figure 8.9. Mean Accuracy on problem types for YHA and OLA Secondary School groups. Error bars show standard error of the mean. Dashed line indicates 33% chance level.

<u>Response times</u>: Unlike the findings from the Primary School data, a 2 (Group) x 6 (Problem Type) repeated measures ANOVA performed on the Secondary RT data showed no reliable differences in RT between Secondary School YHA and OLA groups. The analysis did reveal a main effect of Problem Type (F(2.84,93.55)=46.72, $p<.001$, $\eta^2=.586$)[8]. Figure 8.10 plots the RT data for YHA and OLA groups and shows that the reliable main effect of problem type reflected an increase in response times with problem difficulty. However, no reliable Group by Problem Type interaction was observed. The effect of problem type on RT was similar in both groups. Lastly, while Figure 8.10 shows some divergence between Groups in RT for Weight, Distance and Conflict Balance problems, post-hoc tests found none were reliable following bonferonni corrections.

---

[8] Mauchly's test of sphericity showed significant differences in variance in task condition (chi-square=61.16, p<.001), therefore Greenhouse-Geisser are reported.

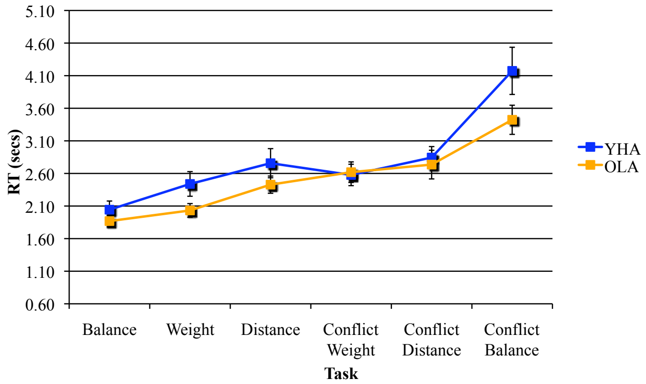


Figure 8.10. Mean RT on problem types for YHA and OLA Secondary School groups. Error bars show standard error of the mean.

Secondary School full sample comparisons

Accuracy: Using MA-CA disparity scores as the covariate in a 2 (Group) x 6 (Problem Type) ANCOVA, the full Secondary School sample was analysed. MA-CA disparity was reliable in modulating overall accuracy (F(1,33)=14.45, p=.001, η^2=.305) and MA-CA disparity was found to interact reliably with Problem Type (F(2.89,95.49)=4.571, p=.005, η^2=.122).

Figure 8.11A-C shows scatterplots for three conditions in which effects remained reliable following bonferroni corrections. These were: (A) Distance (F(1,33)=9.35, p=.004, η^2=.221); (B) Weight (F(1,33)=8.86, p=.005, η^2=.212); and (C) Conflict Balance (F(1,33)=11.24, p=.002, η^2=.254). In each scatterplot, trendlines depict the combined (in green) and separate performances of YHA (in red) and OLA (in blue) relative to MA-CA disparity scores.

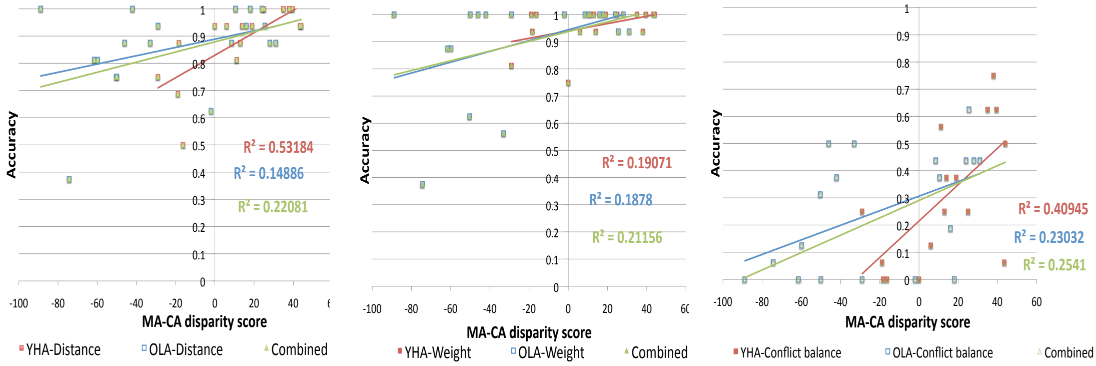


Figure 8.11A-C Full Secondary School dataset: Accuracy by MA-CA disparities on (A) Distance problems; (B) Weight problems; and (C) Conflict Balance problems

While a number of children are at ceiling in their accuracy on Distance and Weight problems, the scatterplots in Figure 8.11 suggest that children with more negative MA-CA disparity scores are more likely to score lower than children with more positive MA-CA disparities. Performance on Distance and Conflict Balance problems appears to be modulated by age and ability effects, whereas performance on Weight problems appears modulated primarily by ability.

Response Times: MA-CA disparity scores did not predict overall RT and no interaction was present on Problem Type. Furthermore, RT was not reliably predicted by MA-CA disparity on any of the individual problem types.

Summary of balance scale: Secondary School data

For the MA-matched groups, the Secondary School sample showed no differences in their overall accuracy or RT. This contrasted with the results in the Primary School data, where younger and older groups were different in their response times. When MA-CA disparity scores were used as a covariate for the Secondary School data, MA-CA scores were found to reliably modulate accuracy scores on three problem types: Distance, Weight and Conflict Balance problems. By contrast, MA-CA disparity was not found to modulate RT on any of the problem types.

Combined Primary and Secondary School Results

Combining both the Primary School and Secondary School datasets, two 4 x 6 repeated measures ANCOVAs were performed on accuracy scores and then solution time. Within these ANCOVAs, Group (Primary Younger, Primary Older, Secondary Younger, Secondary Older) was the between-participants factor and Problem Type (Balance, Distance, Weight, Conflict Balance, Conflict Distance, Conflict Weight) was the within-participants factor. The results of the analysis on solution time showed neither Group nor MA-CA disparity modulated performance. Thus, the details of this analysis are not presented.

Accuracy: Consistent with the previous ANCOVAs performed separately on Primary School and Secondary School data, the ANCOVA on combined data revealed a reliable main effect of MA-CA disparity on accuracy (F(1,67)=14.49, p<.001, η^2=.178). Figure 8.12 plots MA-CA disparity by accuracy on just the Conflict Balance problem type for the entire sample of Primary and Secondary children. This figure shows that higher levels of accuracy were associated with more positive MA-CA disparity scores. The results of the ANCOVA further revealed a reliable Group x MA-CA x Problem Type interaction (F(11.51, 257.04)=2.06, η^2=.084) which suggested that MA-CA modulated accuracy differentially in younger versus older children across the six problem types.

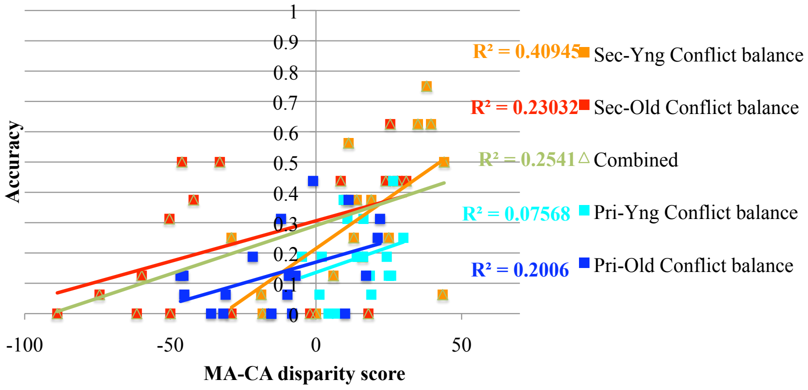


Figure 8.12. Primary School and Secondary School younger and older groups MA-CA disparity scores by accuracy on Conflict-Balance problems.

The analysis also revealed a main effect of Group (F(3,67)=12.53, p<.001, η^2=.359). Figure 8.13 illustrates that this effect was due largely to the poorer performance of the youngest group (Pri-Yng). On the individual problem types, both Group and MA-CA disparity were found to interact with Problem Type in modulating performance (i.e., MA-CA x Problem Type: F(3.84, 257.74)= 6.51, p<.001, η^2=.089; Group x Problem Type: F(11.51, 257.74)= 3.80, p<.001, η^2=.146). Lastly, an overall main effect of Problem Type (F(3.84, 257.74)= 141.12, p<.001, η^2=.678) was found that stemmed from highest overall levels of accuracy on Balance problems (mean accuracy=94%, se=1) and lowest overall accuracy on Conflict Balance (mean accuracy=20%, se=3).

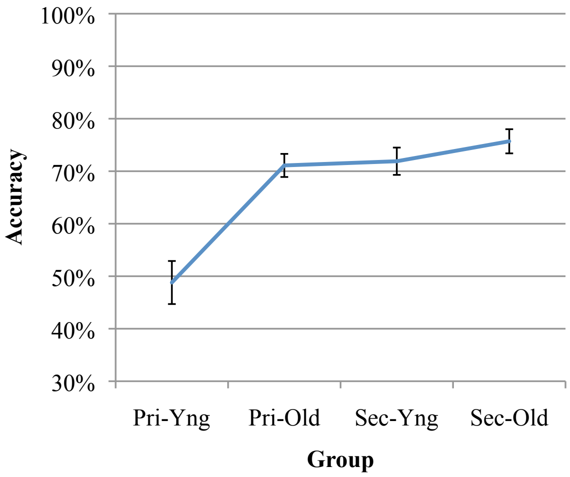


Figure 8.13. Mean overall accuracy for Primary School younger (Pri-Yng), Primary School older (Pri-Old), Secondary School younger (Sec-Yng) and Secondary School older (Sec-Old) children.

We now turn to a discussion of these findings and what they may suggest about the relationship between intelligence and cognitive development.

Discussion

The balance scale task was chosen because of its potential to demonstrate basic differences in children's ability to integrate information. While in the previous chapter on conservation, issues were raised relating to the sensitivity of the accuracy and response time measures (due in part to the low numbers of trials used), this did not seem to be an issue in the balance scale task. Here, the large number of problems showed performance was differentiated among the groups of children tested.

In comparing the profiles observed in the Primary School YHA and OLA groups (Figure 8.6) with those of Siegler (1976; Figure 8.3), the following points are highlighted. Firstly, neither the YHA (5-6 year olds) nor the OLA (10-11 year olds) groups appeared to follow the trends for their given ages. Rather, their profiles fell more evenly between the profiles Siegler found for 5-6 and 9-10 year-olds. This corresponds well with the fact that the overall mental ages of YHA and OLA groups were 8.2 years. For example, the performance of the YHA appeared similar to the performance of Siegler's 9-10 year-olds for Conflict Balance and Conflict Distance. On Conflict Weight, the performance of the YHA was similar to both the performances of Siegler's 13-15 and 16-17 year-olds, where both groups showed the same level of accuracy. In contrast to the profiles Siegler found, the YHA did not show the same poor performance for Weight problems typical of their age, but rather showed a level somewhere in between that typical for their age and that typical for 9-10 year olds. Looking at the performance profiles of the OLA group, we can see that their overall performance on Conflict Balance problems was closest to the level Siegler found for 5-6 year-olds. However, on Conflict Distance their performance was above average for their age group and on the remaining tasks the OLA showed performance levels that were approximately in line with levels typical for their age. On RT, the Primary YHA and OLA groups (Figure 8.7) showed profiles that appeared closest to the performance levels van der Maas et al. (2003) found for children aged 8-9 and 10-11 year-olds, respectively (see Figure 8.4B). Thus, while the OLA responded at levels typical for their age, the YHA appeared to respond at levels nearer that for 8-9 year olds.

Turning to the results of the analysis of the Primary School data, a pattern was uncovered similar to what we might have predicted given Spitz's original study

(1982). That is, the same overall level of accuracy in younger and older groups was comprised of different strengths and weaknesses on the underlying component problem types. On the individual problem types, the YHA group showed the same levels of accuracy as the OLA group on Balance and Conflict Weight. However, they showed superior performance compared to the OLA on Conflict Balance problems – problems considered more complex because of the need to integrate information relating to weight *and* distance (Siegler & Chen, 2002).

Although the mean accuracy of each group was lowest on Conflict Balance and Conflict Distance problems, post-hoc tests confirmed that these levels were reliably different to chance. Both group's poorer performance on the Conflict Balance problems suggest that stable, Rule IV-like behaviour had not been attained by either group. By contrast, the OLA group showed better performance than the YHA on Weight, Distance and Conflict Distance problems. Although none of these differences were reliable, it suggests the OLA group performed better on problems with fewer dimensions, or where distance was the important dimension. In line with Spitz (1982), these data suggest that qualitatively different forms of reasoning underlie the thinking of younger and older MA-matched groups. That is, greater age may provide an advantage in processing problems where there are limited and non-conflicting dimensions, whereas greater intelligence may provide an advantage in integrating information from more than one dimension.

The pattern of differences in accuracy of the Primary groups, however, was not reflected in the response time data where the YHA were uniformly slower than the OLA. A number of explanations may be considered. For example, one might suppose that the superior yet slower performance of the YHA group compared to the OLA indicates differences in speed-accuracy trade-off strategies (e.g., Luce, 1986). That is, the YHA group may have compromised their speed of response in order to maximise their accuracy, while the OLA compromised accuracy to maximise their speed. This possibility was tested comparing the RT of YHA and OLA on Conflict Balance (where the YHA were more accurate) and Conflict Distance (where the OLA were more accurate). The results showed no reliable interaction between RT in the two groups ($F(1,26)=.07$, $p=.801$, $\eta^2=.002$). Thus, compared to the OLA, the YHA were not taking any longer on problems where they were more accurate, compared to problems where they were less accurate. A second possibility is that the consistently slower responses of the YHA compared to the OLA were due to differences in basic motor control. That is, older children may have advantages on tasks involving motor responses (e.g., pressing a button) because they have had greater experience controlling motor movements. The results from the full Primary School analysis offer further insights into these possibilities.

Noteworthy of the results relating to the full Primary School data, were two findings in which the direction of significance changed. Firstly, whereas in the analysis of the Primary School MA-matched groups, the YHA and OLA showed no reliable differences in accuracy, in the full Primary School samples, the groups were found to differ reliably in their overall accuracy and this favoured the older children. Thus, the older more able children and younger less able children who had been removed to obtain the MA-matched groups, served to increase the disparity between the groups in their overall accuracy. Secondly, while the MA-matched YHA and OLA had shown differences in their overall RT, in the full sample the younger and older groups were no different in their overall RT. That is, adding in the previously excluded children led to a decrease in the disparity between the group's RT. However, if age offered an advantage in RT by way of better motor control, then it seems reasonable to expect those differences to be revealed in the full sample also. Instead, these findings suggest that irrespective of age, more intelligent children show greater understanding of the problems involving integrating of more than one dimension and yet they take longer to process information relating to the two dimensions of weight and distance. Thus, from the perspective of tbe balance scale task, these data indicate cognitive development and intelligence are not the same thing. This interpretation is further supported by the results from the full Primary School sample analysis that showed MA-CA disparity score was a reliable predictor of accuracy. For example, on the Conflict Balance problems the results showed that more positive MA-CA disparity scores were associated with higher accuracy scores. Taken together the Primary School findings suggest that ability, more than age, plays a key role in balance scale performance. We return to consider these results in the final discussion (Chapter 11).

In contrast to the Primary School profiles, the profiles of Secondary School groups were indistinguishable on accuracy (Figure 8.9) and RT (Figure 8.10). Comparing the accuracy profiles of these groups to the trends Siegler (1976) reported, we can see the greatest similarity is with Siegler's 13-14 year old group (see earlier Figure 8.3). Again, this is consistent with the fact that the groups shared an overall mental age of 13.8 years. It was further noted that the YHA and OLA Secondary School children were both around the 33% level for accuracy for Conflict Balance problems. While this might indicate children were guessing on these types of problems, the comparatively lower accuracy in the Primary School groups suggests that performance is on a trajectory of improvement.

On RT, the Secondary School groups appeared to resemble the performance profile that van der Maas et al. (2003) found for 10-11 year olds (see Figure 8.4B). The results of the MA-matched analysis of Secondary School YHA and OLA groups confirmed the groups were not different in their accuracy or RT on any of the balance scale problems. Although there was some indication that the YHA were slower than the OLA on Conflict Balance problems, these differences turned out not to be statistically reliable. One finding that did mirror the Primary School results was that MA-CA disparity reliably modulated accuracy scores in the full Secondary sample and this held when the Primary and Secondary School samples were combined. In sum, Secondary School data demonstrated that Problem Type did modulate performance, but in just the same way for MA-matched groups. Here, intelligence and cognitive development appears to be the same thing.

Why the difference between Primary and Secondary Schools results

In attempting to account for the differences in results between Primary and Secondary School levels, a number of possibilities may be considered. Firstly, bearing in mind that the Primary School and Secondary School samples differed with respect to the range of ages that comprised younger and older groups, the first issue to consider is sampling. That is, the groups that were contrasted at Primary School level were comprised of 14 younger children with an age range of 6.0-6.8 years and 14 older children, aged between 10.0-10.8 years. There was not the same CA gap in the Secondary School samples where the groups where the groups comprised 16 younger children with an age range of 11.8-13.9 years and 19 older children, aged between 14.0-15.9. Thus, it is possible that the absence of differences found at Secondary School level was influenced by the fact that these children were closer in chronological age. Yet, if sampling errors were not to account for the change in patterns observed between Primary and Secondary levels, this raises an additional intriguing possibility that the contributions of cognitive development and intelligence may change across age. One possibility is that rather than differences in children's capacities to perform (or learn), it may be differences in *rates of change of capacity* that account for the differences we observe between groups at younger ages but not at older ages. This explanation is consistent with brain imaging studies that have reported differences in the rates of change in cortical thickness in more intelligent versus less intelligent individuals (see Shaw, et al., 2006). Based on the pattern of findings revealed in this chapter, Figure 8.14 aims to illustrate different rates of change of capacity, using two hypothetical trajectories. The grey highlighted areas indicate the ages where data were collected. For example, the area of grey on the left-hand side of the figure represents the data collected at the Primary Schools. Here, older children aged around 10-11 (depicted by the blue marker-point) and younger children aged around 5-6 years old (red marker-point) were matched on a mental age of approximately 8 years. The disparity between the red and blue markers within this area reflects the reliable group differences on the problem types presented earlier. The right-hand area of highlighted grey represents the Secondary School samples. Again, these groups were matched on a MA of around 14 years. However, their performance on the balance scale task was the same, thus no disparity is shown between red and blue markers. These groups appear to be approximately equal in their level of reasoning. The blue and red dashed-lines outside of the grey areas represent hypothetical trajectories for each group.

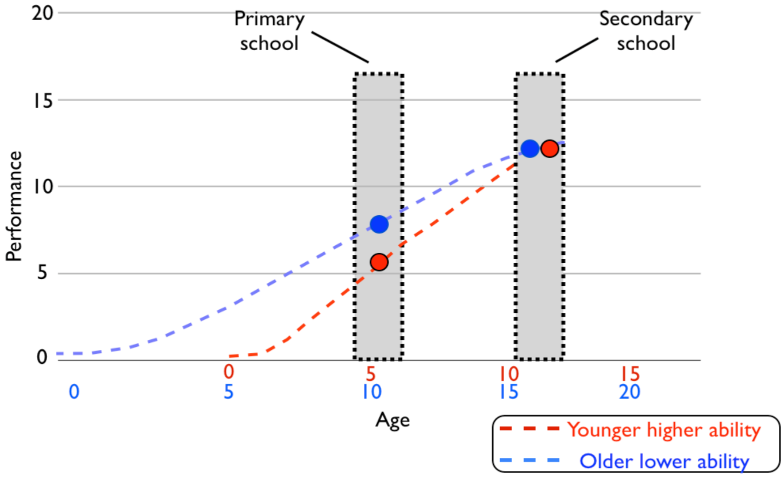


Figure 8.14. MA-matched older lower ability (blue dashed-line) and younger higher ability children (red dashed-line) are tested at different school levels. Their performances on cognitive tests are different at Primary School level, but the groups are virtually indistinguishable at Secondary School level.

Figure 8.14 suggests that the contributions of intelligence and cognitive development change over time such that at earlier ages they appear as different forms of variability (and hence cannot be described as variability on a single dimension) whereas at older ages they appear similar (and hence may be described as variability on a single dimension). Thus, it is possible that the question of dimensionality may itself depend on the phase of development we are considering. Furthermore, we do not know what the ongoing influences of intelligence and cognitive development might be. It is possible that beyond the ages of the children tested (i.e., 16 years old+), MA-matched individuals of different ages would once again show a divergence in their performances. For example, does the performance of the YHA continue to increase past the ages tested? Or does their rate of change plateau to show similar levels of performance to OLA in later years? To shed light on these questions, future work might aim to establish the true trajectories for these groups and for age groups beyond those tested here.

Computational perspectives

Given that there have been numerous computational studies investigating learning on the balance scale task (e.g., Dawson & Zimmerman, 2003; McClelland & Morris, 1989; Shultz, Mareschal, & Schmidt, 1994; van Rijn, van Someren, & van der Maas, 2003), these may provide useful platforms from which notions such as rate of change of capacity may be investigated. To take one example, Thomas, Richardson, Forrester and Baughman (submitted) focused on charting the effects of altering several model parameters within a version of McClelland's (1989) balance scale model. Their investigations included assessing the effect of the numbers of hidden layers in their model versus the number of hidden units within each layer. Increasing the number of hidden units within a single layer is broadly equivalent to increasing the *capacity* of a model to learn representations of a given complexity, whereas increasing the number of hidden layers generally increases the *complexity* of problems a model can represent. Of particular relevance here were Thomas et al.'s (submitted) findings that increasing the number of hidden units resulted in models reaching key developmental transitions associated with learning *faster* than models with fewer hidden units. By contrast, increases to the number of hidden layers showed patterns that did not follow the same transitions typical in learning on the balance scale problem. The nature of the development appeared altered. One other manipulation Thomas et al. (submitted) did not investigate yet which might yield important insights is the effect of changing the model's sigmoid functions. Recall that this parameter was discussed in Chapter 7 with reference to differences between younger and older children on the conservation tasks (see Richardson, Forrester, et al., 2006). Thus, using the Thomas et al. model, the above parameters might be tested with the aim of determining whether the patterns of accuracy for the YHA and OLA groups on the different problem types could be simulated.

Conclusion

In comparing the performances of younger and older MA-matched groups on the balance scale task, results were found to resemble the pattern of differences described by Spitz (1982) at Primary School but not Secondary School level. At Primary School level similar overall levels of ability masked different underlying strengths and weaknesses in younger more able and older less able groups of children. This pattern of differences between groups was not previously captured in the previous chapters on the BAS-II (Chapter 4), the Stroop task (Chapter 5), or the semantic priming task (Chapter 6). However, the findings did reinforce the earlier interpretation on conservation (Chapter 7) that, at least at some ages, cognitive development and intelligence do not appear to be the same thing. In the following chapter, I present the results of comparisons between younger and older MA-matched groups on the Tower of London task. This problem-solving task was chosen for its potential to give deeper insights into the reasoning processes of younger and older MA-matched children. The next chapter focuses in particular on understanding whether the groups differed in their ability to inhibit and in their use of sub-goals. Furthermore, an analysis of some of the errors that were made on the Tower of London task provides a potentially richer understanding of children's different abilities. This allows us to pursue further the possible influence of the dimensionality of the problem. Perhaps in a more challenging domain we will see some differences emerging between groups at the Secondary School level.

Chapter 9 The Tower of London Task

Introduction

The Tower of London task is a problem-solving task designed by Shallice (1982) for the purpose of investigating the processes underlying goal-directed behaviour. These processes, more formally referred to as *executive functions*, are generally assumed to be responsible for modulating higher-level cognitive behaviour (Miyake, et al., 2000). However, there is disagreement over the exact role these processes take. For example, the literature shows a significant split between theorists who take the view that processes are regulated by a unitary executive component versus those who argue that a range of diverse processes interact in modulating behaviour (see e.g., Zelazo, et al., 1997). Nevertheless, several candidate executive functions have frequently been put forward. These include: (1) *inhibition of prepotent responses* ('inhibition'); (2) *shifting of mental sets* ('shifting'); and (3) *updating of working memory* ('updating'). Much of the evidence relating these processes to goal-directed behaviour has come from neuropsychological studies of patients with brain damage. In these studies, deficits in executive functions (also referred to as executive 'dysfunctions') have been repeatedly linked to areas in the frontal lobes and in particular to the left frontal anterior region (e.g., Bocková, et al., 2007; Braver, et al., 2006; Duncan, et al., 1996; Haier, 2003; Morris, et al., 1993; Newman, et al., 2003; Roth, et al., 2006; Shallice, 1982). Additionally, recent brain imaging studies have found patterns of increased prefrontal cortical activation associated with superior performance on the Tower of London task (Cazalis, et al., 2003). Further research has showed that this region is also one that undergoes protracted development during later childhood and adolescence and its development coincides with the emergence of more complex cognitive abilities, such as reasoning and planning (see e.g., Diamond, 2006; Pennington & Ozonoff, 1996; Roth, et al., 2006). The focus within this chapter is to evaluate whether differences in problem solving on the Tower of London task exist between groups of younger and older children who are matched on mental ability. In this introduction, I describe the origins and background of this task and review how it has been used to investigate differences in executive functions. I then detail the measures that are obtained from the Tower of London task and describe how these will be used to cast light on the central question of this thesis: are intelligence and cognitive development the same thing?

The origins of the Tower of London task

The Tower of London task was derived from the Tower of Hanoi task in order to assess problem solving abilities in younger children and individuals with neurological impairments (Shallice, 1982). In both tasks the basic aim is to move objects from a *current state*, to a defined *goal state.* That is, participants must transform the initial state to match a goal state. In the case of the Tower of Hanoi this involves moving a stack of different sized disks from one outer peg to another outer peg in a series of individual moves (see Figure 9.1, left panel). An important restriction that applies to this task is that for each move that is made, a disk may only be placed on top of another disk if the disk underneath is *larger*. The greater the number of disks that are used, the harder the problem becomes[1] and the more moves needed to solve the problem[2]. However, Shallice (1982) noted that the Tower of Hanoi task produced frequent failures in clinical groups and concluded that maintaining "in mind" the rule of 'smaller disk on top of larger' was an additional task demand that likely put extra load on working memory. The Tower of London task was designed to reduce these demands, thus offering the potential to assess executive functions more directly.

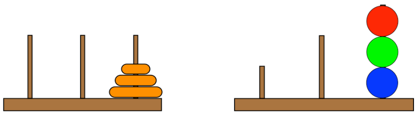


Figure 9.1. Illustrating the Tower of Hanoi (left) and Tower of London (right) tasks

---

[1] There is a simple (yet not well-known) recursive strategy that guarantees success on problems with any number of disks. For problems with an odd number of disks: on every second turn, move the smallest available disk to the nearest free place to the left. Then move the next available larger disk to the only free position. A similar pattern is performed for problems with an even number of disks, except the direction the smaller disk should be moved is to the right. Using this recursive strategy, the disks are eventually all moved to the target peg. Few participants know or discover this strategy during normal testing.

[2] The minimum number of moves is found to be a function of $2^n - 1$ (where n is the number of disks).

In the Tower of London task, disks are replaced with different coloured balls (see Figure 9.1, right panel). Participants are typically shown a picture representing a desired goal state and are then asked to move the coloured balls around on the pegs to match the goal. The physical setup of the task imposes a key constraint that is intended to reduce the load on working memory, thereby making it more suitable to special populations (Shallice, 1982). Specifically, as illustrated in Figure 9.1 (right), the heights of the individual pegs limit the number of balls that may be placed on them to 1, 2, and 3 balls, respectively. However, in contrast to the Tower of Hanoi task, the Tower of London task contains no restriction on the order of the placement of balls. That is, as long as there is space on a peg for a ball, it may be placed there regardless of the colours of the balls underneath. These features of the task distinguish it from the Tower of Hanoi task. Yet in keeping with the Tower of Hanoi task, two rules remain: (1) balls may only be moved individually and (2) they must be placed on a peg (i.e., they may not be kept in hand, or left on the table).

Types of problems on the Tower of London task

The literature on the Tower of London task shows a variety of measures have been used to assess executive functions underlying performance. These measures include but are not limited to: the time taken to plan and execute individual moves, the total time to solve problems, the number of correct solutions, whether correct solutions are completed in the minimum number of moves (i.e., perfect solutions), the number of additional moves made, the number of rule breaks, the number of attempts and the number of immediate hits. Additionally, several variations exist in the types of problems administered on the Tower of London task. Two broad divisions are made here between problems that manipulate the *configuration* of the goal state and those that manipulate the *complexity* of problems. For example, goal states can either be of a *flat-ending* or a *tower-ending* configuration (see Figure 9.2).

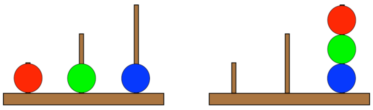


Figure 9.2. Illustrating flat-ending (left) vs. tower-ending (right) configurations on the Tower of London task.

Klahr and Robinson (1981) reported variability in performance on different configurations, with performance often appearing adversely affected in flat-ending problems. One explanation offered for these findings relates to the degree to which problems can be broken into unambiguous sub-goals, referred to as the *goal hierarchy* (Kaller, et al., 2004). For example, on tower-ending tasks the sub-goal ordering results in an unambiguous goal hierarchy, as one ball must always be placed first in the bottom position on the target peg. However, on flat-ending tasks there is not necessarily a particular order for placement of balls – they all occupy a bottom position. For this reason, the goal hierarchy of flat-ending tasks has been referred to as *ambiguous* (see e.g., Kaller, et al., 2004; Klahr & Robinson, 1981; Zelazo, et al., 1997).

Problem complexity generally refers to how straightforward the solution path is from the initial state to the goal state. The simplest types of problems are those with the shortest solution paths (e.g., a problem requiring only one move). Generally speaking the more complex the problem, the longer the solution path (Shallice, 1982). Thus, task complexity is formally manipulated via the *minimum number of moves* needed to solve a problem (see, Newman & Pittman, 2007). By comparing an individual's *actual* number of moves made to the minimum number of moves, it is possible to infer something about the efficiency of his or her problem solving. For example, Figure 9.3A depicts a 3-move problem in which a minimum of 3 moves is required to move the balls from the initial state (shown on the left-hand side) to the goal state (shown on the right-hand side). Figure 9.3B shows a more complex problem requiring a minimum of 6 moves. In practice, an individual performing on either of these examples may take a great many more moves than the minimum number required. However, an indicator of more efficient problem solving, say between two individuals, would be a smaller disparity between the actual number of moves made and the minimum number of moves needed. By the same measure, a greater disparity between the actual number of moves made and the minimum number needed would indicate less efficient problem solving.

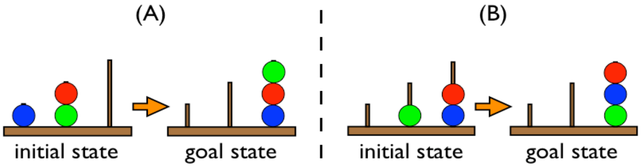


Figure 9.3. Illustrating two tower-ending problems where a minimum of 3 moves (A) and 6 moves (B), respectively are needed to move the balls from their initial states to their goal states.

Goal hierarchies: Getting from the initial state to the goal state

Using the typical setup of three different coloured balls and three different sized pegs, a total of 36 unique configurations are possible on the Tower of London task. Figure 9.4 illustrates each possible configuration and shows the solution paths available between any two given states. For example, within Figure 9.4 the most efficient (i.e., shortest) solution path between the configuration marked ‘start’ (bottom middle-left) and the configuration marked ‘goal’ (top-right) comprises a total of 5 moves (the sequence of moves are numbered in the figure from 1-5). However the figure also shows there are a variety of other less efficient paths that still enable the solution to be achieved.

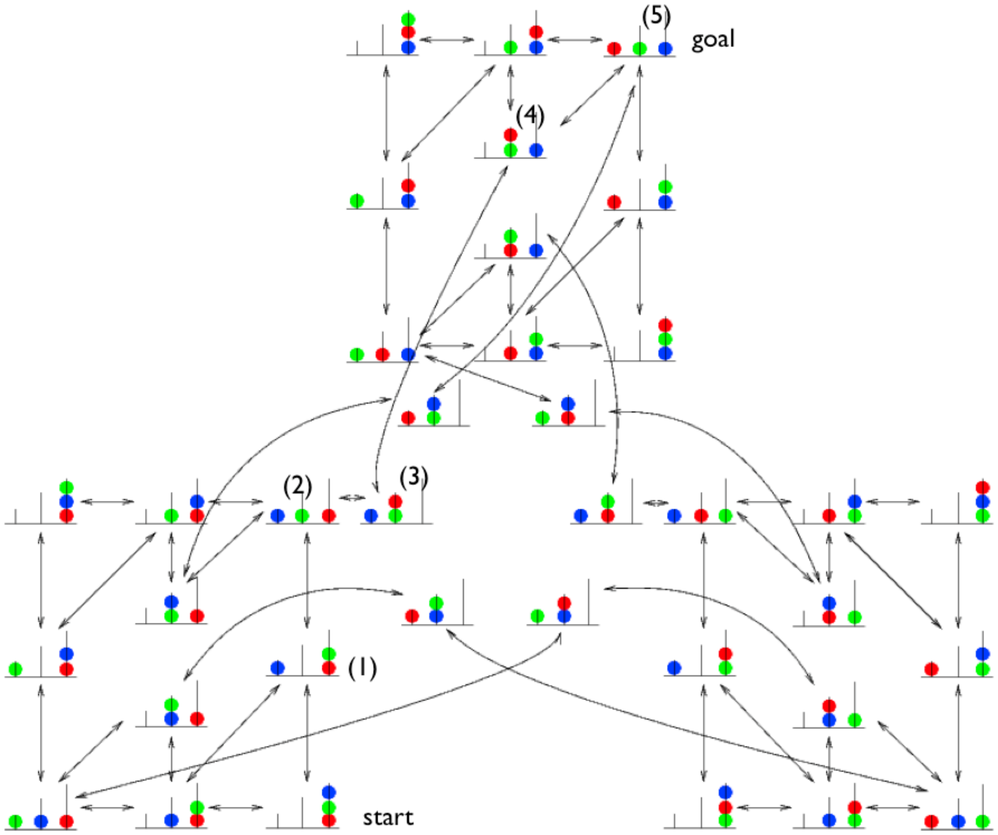


Figure 9.4. Mapping the full problem space of 36 unique configurations on the Tower of London task. The sequence of moves 1-5, depict the shortest solution path between the configuration marked 'start' and the configuration marked 'goal'.

A considerable amount of research has focused on exposing the way in which the solution paths, or goal hierarchies are constructed. For example, Newell and Simon influenced many early information-processing accounts with their seminal work on the General Problem Solver (GPS; Newell & Simon, 1972). Within their model, means-ends analysis plays a key role in reducing the disparity between the current state and goal state of a problem. It does this through the use of *problem spaces* (states in which differences are examined) and *operators* that allow states to be modified by sets of *If-Then* rules – called *production rules* (Newell & Simon, 1972). Given the right set of rules, the GPS demonstrates complex problem solving through searching the paths, or branches of each possible permutation of each possible outcome within an expanding problem space. Figure 9.5 illustrates a goal hierarchy in which three outcomes are possible for each action that may be taken. The GPS constructs the full problem space and then through a series of elimination, it identifies the correct solution path in the hierarchy (here highlighted by red dotted lines).

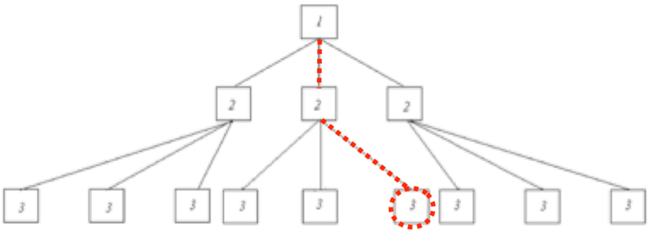


Figure 9.5. Illustrating a problem space with a growing number of possible outcomes. A depth-first solution path is found, depicted by the red dotted lines.

Although means-ends-analysis provides an effective way of solving problems, its plausibility as a human cognitive process is undermined by the amount of cognitive resources it assumes in its exhaustive search of the problem space. These limitations have encouraged researchers to consider the alternative of heuristics – strategies that aid problem solving and yet impose fewer demands on cognitive resources.

Some of the heuristics examined include, for example, simple perceptual strategies (see e.g., Altmann & Trafton, 1999; Bull, Espy, & Senn, 2004; Simon, 1975) and look-ahead strategies (Klahr, 1985). Direct perceptual biases can offer fast goal completion on some problems when both the target position for a ball is free and the target ball is free to move. In this instance, no forward planning is needed to remove obstacles and the ball may be placed in position immediately, also referred to as an 'immediate hit' (see Baughman & Cooper, 2007). Klahr reported that younger children often appear to be overly-influenced by this type of perceptual feature and this can lead them to erroneous solutions, especially on problems with ambiguous goal hierarchies (Klahr & Robinson, 1981). Recently, Baughman and Cooper (2007) examined the effects of inhibiting an immediate-hits strategy within a computational model of the Tower of London task. They developed two 'child' models within which two key strategies were implemented in parallel. These were a direct perceptual bias and a one-move look-ahead strategy. The perceptual bias worked by continually analysing the problem space and influencing[3] actions (e.g., pick up Ball X, or move Ball X to Peg A) based on the availability of balls and the availability of

---

[3] An interactive-activation network was used to allow influence to build up over cycles such that when a threshold level was reached the action would be performed.

positions. If a ball was free to move and the position for that ball was free, then the influence to move the ball to its position would increase[4]. The one-move look-ahead worked by evaluating the outcome of moving each ball that could be moved, up to one move and then evaluating that state's perceptual similarity to the goal state. Baughman and Cooper (2007) showed that by inhibiting the perceptual bias in one of those models, they could simulate several key characteristics of the performances of 3-4 and 4-5-year-old children. Thus, their work suggests that at some stages of development, increases in task performance may not just be a matter of acquiring more advanced strategies, but rather inhibiting less efficient ones. Carder, Handley and Perfect (Carder, Handley, & Perfect, 2004) similarly proposed that "*problem difficulty is a function, not of planning efficiency, but of the ability to successfully inhibit inappropriate move selections at specific points within a solution path*" (p.1460). Figure 9.6 illustrates a scenario in which the use of an immediate hit strategy has resulted in a 'dead end' state – a state in which the problem cannot be fully solved unless balls that currently occupy their correct positions are moved. The right-hand side of Figure 9.6 shows the goal state, the left-hand side shows the current state. Within the current state, the green ball has been moved to its correct position on the left peg. However, without moving the green out of position and backtracking, the problem cannot be solved correctly – there is no way (short of breaking task rules) to achieve correct placements for the red and blue balls. In these instances, younger children may opt to start over, or move on to the next problem, rather than undo a move that has resulted in one correct placement.

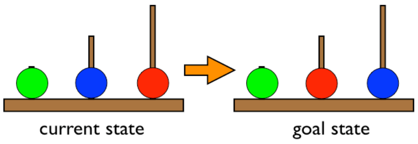


Figure 9.6. Illustrating a deadend state on a flat-ending problem on the Tower of London task. In the current state (left) the Green is depicted in position. The only way of achieving the goal state (right) is to temporarily move the Green out of its position.

---

[4] The modelling framework used was COGENT (Cooper, 2002). COGENT operates in a cyclic manner, analogous to timesteps.

The *look-ahead* strategy offers a more directed, top-down approach to solving problems. This strategy involves identifying the various actions that are possible given a problem and then mentally transforming a representation of the current state in order to evaluate the outcomes of those candidate actions. Figure 9.7 shows a current state on a flat-ending problem. The Red ball already occupies its goal position, however, both it and the Green ball may be moved to one of two positions. Thus, four resulting states are possible given the current state (illustrated by options 1-4 on the right-hand side). Using a one-move look-ahead strategy, an individual faced with this problem might determine that options 1 & 2 do not help in reducing the disparity between the current and the goal states. Option 3 reduces the perceptual differences (that is, it transforms the current state to the correct flat-ending configuration) but there are no balls that may then be placed directly in position (i.e., no immediate hits follow this one-move look-ahead). The final option where the Green is moved to the centre peg, allows a subsequent immediate hit for the Blue ball. Thus, the fourth option is the optimal move.

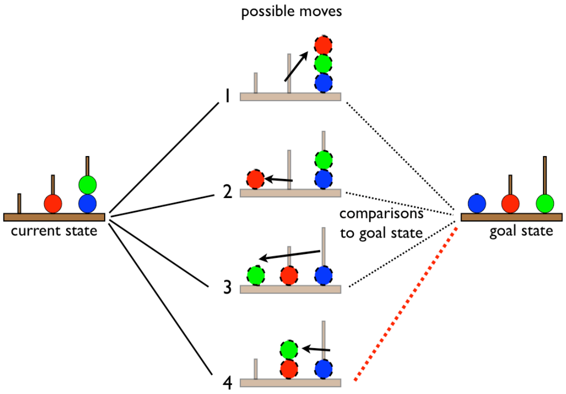


Figure 9.7. Illustrating four possible outcomes of a one-move look-ahead strategy in a 3-move flat-ending problem. See text for details.

Limited look-ahead strategies, in which only the first one or two possible moves of a problem are considered, have received support in the problem solving literature (see e.g., Bull, et al., 2004; Goel, Pullara, & Grafman, 2001; Newman & Pittman, 2007). For example, young children demonstrate the ability to move one object to a different location in order to reach a second (e.g., Gratch, 1975; Klahr & Robinson, 1981) and adults appear capable of a larger number of look-ahead moves. However, in both instances, there is evidence that the search of candidate options is not exhaustive, with searches terminating when one possible action yields a somewhat beneficial outcome (Gilhooly, 1982). This implies possible limits of working memory in maintaining sequences of hypothetical moves. Due to the strong planning component involved in the Tower of Hanoi and Tower of London tasks and the possible role of inhibition in resisting moves biased by perceptual similarity, both tower tasks have become popular tests of executive functions (Newman & Pittman, 2007).

Variability in performance on the Tower of London task

Studies of variability in performance on the Tower of London task have tended to focus on differences between atypically developing and typically developing groups, or between patients with brain damage or mental illness, and controls; rather than on individual differences in task performance. For instance, the Tower of London task has been used to examine executive functions in patients with mental health disorders such as schizophrenia (e.g., Penadés, et al., 2000) and addictions (e.g., Goudriaan, et al., 2006), individuals with developmental disorders such as Asperger's and autism (e.g., Ozonoff, 1998), Williams syndrome (e.g., Vicari, Bellucci, & Carlesimo, 2001), ADHD (e.g., Geurts, et al., 2005) and Downs syndrome (e.g., Vicari, Bellucci, & Carlesimo, 2000) and children with observed learning difficulties, such as those with poor arithmetic skills (e.g., Sikora, et al., 2002). One reason for the paucity of literature focusing on individual differences on the Tower of London task is due to the fact that early studies suggested there was little relationship between executive functions and intelligence. For example, Shallice (1982) reported that brain-damaged patients with deficits in performance on the Tower of London task showed no impairments on tests of intelligence (see also Damasio & Anderson, 1985). Similarly, on the Tower of Hanoi task, Welsh, Pennington, Groisser and Green (1991) found only age and not measures of intelligence correlated with task performance. This view has been challenged by evidence from a number of later studies. For example, Luciano, Wright, Smith, Geffen, Geffen and Martin (2001) found reliable proportions of covariance between the genetic contributions underlying measures of intelligence and working memory. Additionally, Salthouse, Atkinson and Berish (2003) found measures of inhibition and intelligence were related in a large sample of adults aged 18-84 years old. They also reported weak evidence for distinct executive functions. Indeed, some theoretical accounts further argue that inhibition plays a vital role in the development of intelligence (e.g., Anderson, 2001). Currently, evidence concerning the relationship between executive functions and intelligence remains inconsistent (see e.g., Friedman, et al., 2006; Friedman, et al., 2008).

Given that the Tower of London task has been used predominantly within developmental contexts, it seems surprising that the literature shows that only a few studies have attempted to establish normative, age-related profiles of performance on it. Two studies that have attempted this come from Anderson, Anderson and Lajoie (1996) and Huizanga, Dolan and van der Molen (2006). Anderson et al. (1996) administered the Tower of London task to 376 children, aged 7 to 13 years old. A total of 12 items were given to each child. These comprised: two 2-move, two 3-move, four 4-move and four 5-move problems. They reported a general increase in the ability to solve problems in the minimum number of moves and a general decrease in time taken over age (although the results of analyses on solution times were not reported). Anderson et al. also found no relation of age to the number of failed attempts. Figure 9.8 shows the accuracy data Anderson et al. (1996) observed in children's performance across each of the age groups[5]. This figure illustrates the largest incremental increases in performance occurred between children aged 7-8 and 8-9 years old. Based on the relatively small changes in performance after age 12, Anderson et al. (1996) proposed that the functions underlying the Tower of London task were near adult levels around this age.

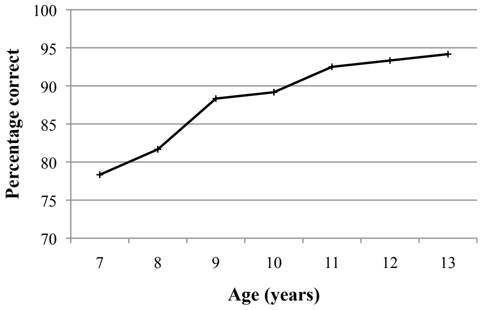


Figure 9.8. Developmental trends in accuracy on the Tower of London task (Anderson et al., 1996).

Huizinga et al. (2006) later reported on a similarly large-scale study in which a computer-based version of the Tower of London task was used to test 386

---

[5] In this figure, data from Anderson et al. (1996) are converted from mean numbers out of 12, to percentages.

individuals across 4 age groups (7, 11, 15 and 21 years old), on a range of 4, 5 and 6-move problems. Huizinga et al. examined each group's performance on: (1) the proportion of *perfect solutions* (i.e., the percentage of problems solved in the minimum number of moves); (2) the number of *additional moves* (the number of moves made over-and-above the minimum possible for an item); and (3) *planning time* (the time interval between an item appearing on-screen and the first mouse click to move a ball). Figure 9.9 depicts two key results from Huizinga et al. (2006). Tile A shows the proportion of perfect solutions and Tile B shows the mean number of additional moves made. (Note: standard errors of the mean are shown, though in Tile A these are not clearly visible). Huizinga et al. reported that the increments in numbers of perfect solutions depicted in Tile A were reliable between each of the age groups they tested. However, differences in the numbers of additional moves (and similarly on their measure of planning times – not depicted here) were reliable only between 7-11 and 11-15, but not between 15-21-years-old.

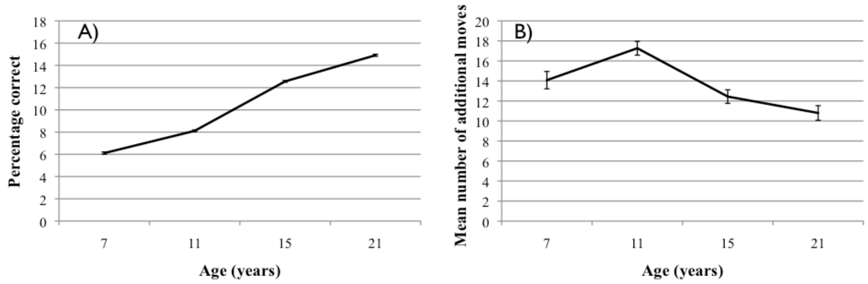


Figure 9.9. Plotting the developmental trends on the tower of London from Huizinga et al. (2006). Tile A depicts the mean number of perfect solutions obtained. Tile B shows the mean number of additional moves made, by each group.

Huizinga et al. (2006) argued that their findings corresponded well with those of a number of other studies proposing that adult levels of executive functions are obtained by around the age of 15 years. For example, Welsh (2006) reported evidence suggesting shifting and inhibition appeared to be at adult levels by around 15 years[6]. Based on their results relating to numbers of additional moves, this claim appears justified. However, Huizinga et al. do not clarify how this view might be

---

[6] Note, that in other studies, it has been claimed that adult levels of shifting and updating may be obtained as young as 12-years-old (see e.g., Cepeda, Kramer, & De Sather, 2001; Klenberg, Korkman, & Lahti-Nuuttila, 2001).

reconciled with their finding that the proportion of perfect solutions continues to increase into adulthood.

While Anderson et al. (1996) and Huizinga et al. (2006) offer platforms from which performance on the Tower of London task may be standardised, neither approach is ideally suited given the aims of this thesis. For instance, although Anderson et al. recorded solution times for each trial, they did not specify how time taken to solve problems changed over development. Instead, solution times were converted to item scores and the number of failed attempts was subtracted from these scores to derive summary scores of overall ability[7]. However, it is not clear that combining accuracy and speed in this way is optimal for assessing ability. For instance, using this method one might determine the following two children were equal in their overall abilities: Child A solved all 3 and 4-move problems quickly and on their first attempt, but failed all 5 and 6-move problems. Child B solved all 3, 4, 5 and 6-move problems, but made two attempts on each problem. Huizinga et al, on the other hand, did examine differences in planning time across development. However, this included only the interval between a problem first appearing on-screen and the first click of the mouse button to move an object.

In contrast, central to the approach taken within this thesis, is the view that solution time (or response time) is a key dependent measure for examining potential differences in underlying processing. Therefore, converting solution times to summary scores, or measuring only the first part of performance on a task may miss subtle variations that are suggestive of differences in ability. Secondly, although complexity was manipulated in the problems administered by Anderson et al. (i.e., 2, 3, 4 and 5-moves) and Huizinga et al. (4, 5 and 6-moves), neither study demonstrated how performance on problems of different complexity might change at each of the ages they tested. Although we might expect, for instance, that younger children obtained a higher proportion of 2-move problems compared to 5-move problems, we cannot substantiate this given their data. Lastly, it is not clear whether the end configuration of items (i.e., tower-ending or flat-ending) were manipulated in either

---

[7] Anderson et al. (1996) scored solution times according to the following boundaries: <5secs=9 points, 6-10secs=8 points, 11-20secs=7 points, 21-40secs=6 points, 41-60secs=5 points, >60secs=0 points. If a child completed a problem within 5secs, they were awarded a score of 9, minus the number of failed attempts (e.g., if this child also had 2 failed attempts before succeeding, then their final score would be 9-2=7).

the Anderson et al. (1996) or Huizinga et al. (2006) studies. These problem types have been argued to place different demands on the cognitive system in achieving solution paths for more or less ambiguous problems (e.g., Klahr & Robinson, 1981). Therefore, based on the data from Anderson et al. and Huizinga et al. we also cannot determine how age-related changes in the ability to sub-goal may proceed over development. In what follows I describe the measures that are adopted for use within this study to examine the differences between MA-matched groups of younger and older children.

Current aims

The aim of this chapter is to examine whether MA-matched younger and older children differ in their ability to reason on problems that vary in complexity (i.e., 3, 4, 5 and 6-move problems) and which are either unambiguous or ambiguous in their goal hierarchies (i.e., tower-ending vs. flat-ending). Specifically, I focus on potential differences between MA-matched groups in their ability to inhibit and sub-goal. Towards this end, the following dependent measures are used: (1) *accuracy* (the proportion of correct solutions); (2) *solution time* (the time in seconds to correctly solve a trial); (3) *number of moves* (the total number of moves made on correct solutions); (4) *immediate hits* (the number of additional times a ball is placed in its goal position on correctly solved trials[8]); and (5) *number of attempts* (the number of times a problem is attempted, independent of accuracy). Inhibition is conceptualised as the ability to suppress the response of making an immediate hit (i.e., to avoid a move that places one ball in its correct position, but which prevents successful goal completion for other balls). Sub-goaling ability is assessed using the remaining measures jointly. That is, more efficient sub-goaling ability may be discerned by greater accuracy, shorter solution times, fewer moves and fewer attempts. Conversely, less efficient sub-goaling may be discerned by lower accuracy, longer solution times, greater number of moves and greater number of attempts.

---

[8] On each problem correctly solved, there is a minimum of three immediate hits (i.e., one for each ball that is placed in goal position). The number of times a ball is placed in its goal position then moved (e.g., to avoid a dead-end state) and then replaced in its goal position, thus represents the immediate hit measure.

Method

Participants

See 'Participants' section, Chapter 3 General Methodology. All participants completed this task.

Design

The between-participants factor was Group (YHA vs. OLA) and the within-participants factors were Configuration (tower-ending vs. flat-ending) and Complexity (3, 4, 5 and 6-move). The dependent variables were accuracy, solution time, number of moves, immediate hits, and number of attempts.

Procedure

Children were shown a physical demonstration of the Tower of London task in which the experimenter moved balls one at a time to various pegs. It was shown that each peg was of a different height, thus allowing a different number of balls to be placed on each. Peg A allowed space for only 1 ball, Peg B for 2 balls and Peg C for 3 balls. The experimenter explained that a ball could only be picked up if it was the top ball on a peg and it could only be placed on a peg where there was room for it (placing a ball on a peg without room would result in the ball falling off). Participants then sat in front of a touchscreen display, at a comfortable arm's length (approximately 30-50cm) and watched an instruction video outlining the task requirements. In this video, children were shown a 2-D representation of a simple problem (requiring only one ball to be moved to complete the problem). The video directed children's attention to two key areas of the screen: a large 'work-space' area (this was the area within which children would move balls around on screen) and a 'goal' area (a small area in the uppermost left-hand corner of the screen that depicted the goal state). A sample display is shown in Figure 9.10. Children were told that the aim of the activity was to match the picture shown in the goal area by moving the larger balls around on screen. This was done by touching the picture of the ball and moving their finger (i.e., 'dragging' the ball) across the screen. When they thought they had finished a problem they were told to press the "done" button. They were also told that if they got stuck and wanted to try a problem again, they could press the "re-start" button. Figure 9.10 provides a screen shot taken from the instruction video.

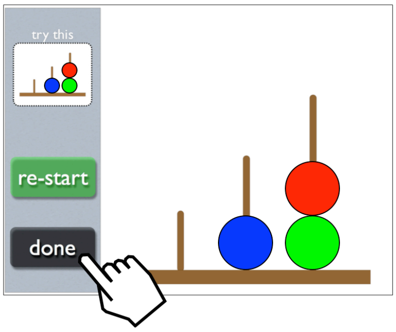


Figure 9.10. Illustrating a practice trial on the computer-based version of the Tower of London task.

In the video, an animated hand illustrated the entire process for a sample problem by moving a ball into its correct position and then pressing the done button (verbatim instructions given in the video can be found in Appendix C). Children were shown that the task echoed physical constraints of the problem in that (a) a ball had to come off the top of a peg, (b) this could only happen if there was no ball(s) above it and (c) balls could only go on a peg where there was space for the ball. All participants completed a short practice set of 6 problems not administered in the experimental trial. These comprised 3 tower-ending and 3 flat-ending problems (each consisting of one 1-move and two 2-move problems). In the experimental trial, a total of 16 problems were presented in random order. These consisted of 4 problems of 4 levels of complexity (3, 4, 5 and 6 moves minimum). Within each level of complexity, the four problems were counter-balanced with half ending in a tower configuration and half ending in a flat configuration. Overall, therefore, there were 8 tower-ending problems and 8 flat-ending problems. The full set of 16 problems is given in Appendix E.

Materials

Apparatus included a physical version of the Tower of London task (for instructional use only). The experimental tasks were coded in MATLAB^{TM} by Frank Baughman and required the use of a touchscreen display and external speakers attached to an Apple Mac G4 iBook laptop computer (1.33 GHz, 1GB RAM).

Primary School results

In this section, results are presented for the categorical analysis (ANOVA) of Primary School MA-matched YHA (n=14) and OLA (n=14) groups. No results are presented for the continuous analysis (ANCOVA) using the full sample because MA-CA disparity did not predict any of the individual dependent variables.

Primary School MA-matched group comparisons

For each of the dependent measures, I report the results of a 2 (Group: YHA vs. OLA) x 2 (Configuration: tower-ending vs. flat-ending) x 4 (Complexity: 3, 4, 5, and 6-move) repeated measures ANOVA. I report first the test statistics of key interest relating to Group differences and interactions involving Group, followed by main effects for each of the factors.

Number of attempts: An ANOVA performed on the number of attempts showed no reliable main effect of Group. However, a reliable Group x Complexity interaction was found (F(2.0,51.5)=3.56, p=.036, η^2=.120)[9]. This stemmed from a greater number of attempts being made in the YHA compared to the OLA groups as problems became more difficult. The effect of Complexity was also observed when data were collapsed across Group and Configuration (F(2,51.5)=5.93, p<.001, η^2=.469). Figure 9.11 illustrates these effects and shows the largest separation between YHA and OLA was on the 6-move problems (note: the dotted horizontal line represents the minimum of one attempt of 1 per problem). While Figure 9.11 also suggests that Configuration had some effect on the YHA and OLA for 6-move problems (i.e., relatively large differences appear for YHA vs. OLA on 6-move flat-ending problems), the results of a post-hoc t-test showed these differences were not reliable. Additional post-hoc t-tests also showed that within each group the number of attempts on tower-ending and flat-ending problems were not reliably different.

---

[9] Mauchly's test of sphericity showed significant differences in the variances on Problem Complexity (chi-square=26.26, p<0.001), therefore statistics relating to Greenhouse-Geisser are reported.

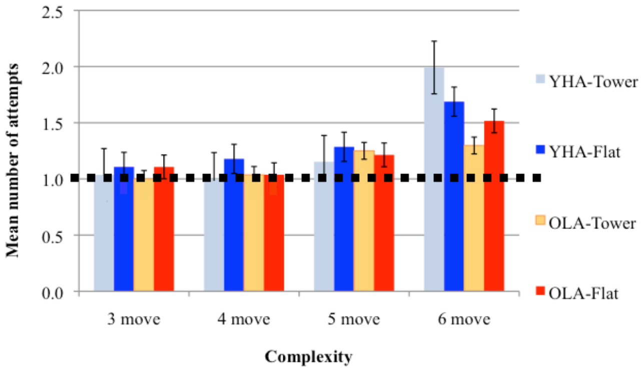


Figure 9.11. Mean number of attempts on the Tower of London task by the Primary School YHA and OLA groups on flat-ending and tower-ending problems at each level of complexity. The dotted line represents the minimum number on each problem. Error bars show standard error of the mean.

Accuracy: The ANOVA on accuracy data showed that while the YHA obtained a marginally higher proportion of correct solutions compared to the OLA (YHA=90%, se=3.3; OLA=84%, se=3.3), these differences were not reliable. The analysis also showed no reliable interactions for Group x Complexity x Configuration, or Group x Complexity, suggesting that the effects of these conditions were similar within each group. The ANOVA did reveal a marginally reliable Group x Configuration interaction (F(1,25)=4.10, p=.053, η^2=.136), indicating that the YHA and OLA differed in their ability to correctly solve tower-ending versus flat-ending problems. Figure 9.12 depicts accuracy for the YHA and OLA. It shows the performance of both groups is at ceiling on the 3 and 4-move problems and that accuracy becomes differentiated on the more difficult 5-move and 6-move problems. The figure further shows that the reliable Group x Configuration interaction stems from the relatively large differences between the YHA and OLA on the 6-move flat-ending problems. A t-test comparing groups on just those problems revealed the differences were marginally reliable (t(26)=2.02, p=.054, 2-tailed). However, examination of the raw data showed Group differences on the 6-move flat-ending problems were due to restricted range of performance. Specifically, differences were due to a combination of: (A) the errors of just one child in the YHA who obtained 0% correct (i.e., 0 out of a possible 2 and all other YHA at ceiling); and (B) the errors of four children in the OLA who obtained 0% correct and two children who obtained 50% correct (with all other OLA at ceiling). Finally, the analysis also showed a reliable Configuration x Complexity interaction (F(2.3,59.3)=4.24, p=.008, η^2=.140), demonstrating that accuracy on tower-ending and flat-ending configurations was moderated by task difficulty (i.e., 3, 4, 5, or 6-moves)[10]. While overall differences in Configuration were not reliable, a reliable main effect of Complexity (F(1.7,43.0)=16.46, p<.001, η^2=.388) was found. This is depicted in Figure 9.12 which shows a general decline in accuracy as problems become more difficult.

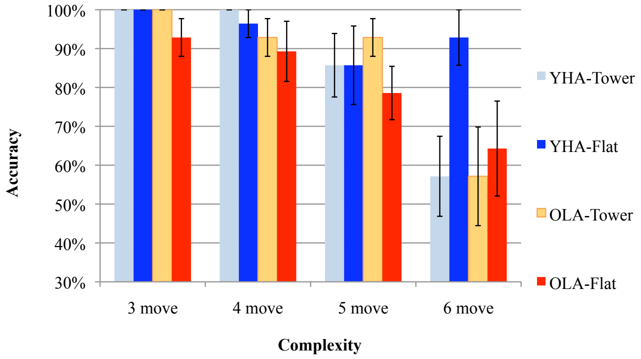


Figure 9.12. Mean Accuracy for Primary School MA-matched YHA and OLA groups on tower-ending and flat-ending problems across each level of complexity. Error bars show standard error of the means.

Solution times: An ANOVA performed on log transformed Solution Times (ST) showed a reliable overall main effect of Group (F(1,26)=35.17, p<.001, η^2=.575) that stemmed from faster solution times by the OLA (mean ST=23.6s, se=2.0) compared to the YHA (mean ST=40.0s, se=2.0). None of the interactions involving Group were found to be reliable. However, in the analysis on untransformed ST data, it was found that the effect of Complexity interacted with group performance (Group x Complexity: F(3,78)=14.94, p<.001, η^2=.365). This was due to the YHA becoming progressively slower as problems became more difficult. Indeed, on the 6-move problems (collapsing across both types of configuration) the YHA took approximately twice as long to complete problems compared to the OLA group (YHA=79.1s, se=4.5; OLA=39.3, se=4.5). Figure 9.13 shows the mean solution times taken by the YHA and OLA groups on each of the 4 levels of problem complexity for tower-ending and flat-ending problems.

---

[10] Mauchly's test of sphericity showed significant differences in the variances on Problem Complexity (chi-square=30.3, p<0.001), therefore statistics relating to Greenhouse-Geisser are reported.

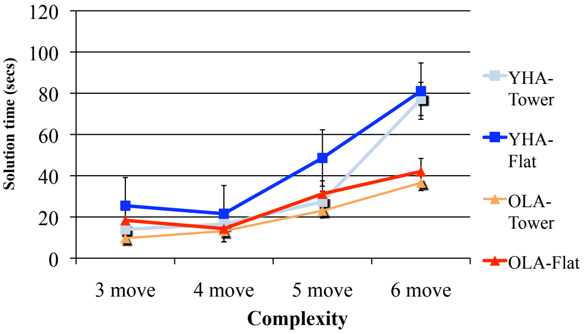


Figure 9.13. Mean solution times for Primary School MA-matched YHA and OLA groups on tower-ending and flat-ending problems across each level of complexity. Error bars show standard error of the mean.

Number of moves: The results of an ANOVA on the number of moves found no reliable differences between groups. Furthermore, no 3-way interaction was found for Group x Configuration x Complexity, nor between any of the 2-way analyses involving Group. Reliable main effects were found for Configuration (F(1,26)=14.73, p=.001, η^2=.362) and Complexity (F(2.17,30.37)=114.99, p<.001, η^2=.816)[11] indicating that as the complexity of problems increased, so too did the number of moves taken to correctly solve them. However, these were not found to interact reliably. The mean numbers of moves made by the YHA and OLA groups on both problem types for each level of complexity are shown in Figure 9.14.

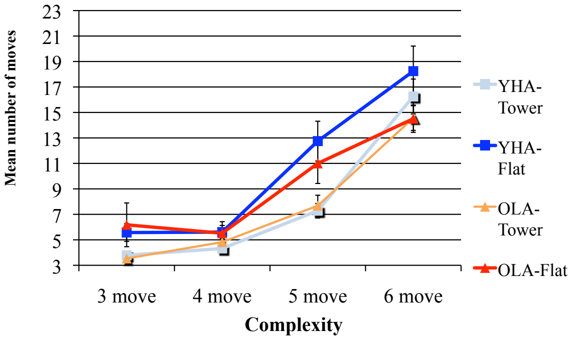


Figure 9.14. Mean number of moves taken by Primary School MA-matched YHA and OLA groups on tower-ending and flat-ending problems across each level of complexity. Error bars show standard error of the mean.

---

[11] Mauchly's test of sphericity showed significant differences in the variances on Problem Complexity (chi-square=13.94, p<0.05), therefore statistics relating to Greenhouse-Geisser are reported

Focusing again on the performances of the two groups on the 6-move problems Figure 9.14 shows the mean numbers of moves for YHA and OLA were very similar. While the YHA were more accurate and took longer on 6-move problems, their mean number of moves was not reliably different to the OLA.

Immediate hits: An ANOVA on the number of additional immediate hits revealed no reliable differences between Group and no interactions involving Group by Configuration or Group by Complexity. However, reliable main effects of Configuration (F(1,26)=25.55, $p<.001$, $\eta^2=.496$) and Complexity (F(2.3,59.8)=7.36, $p<.001$, $\eta^2=.221$) were found and these were also found to interact (Configuration x Complexity: F(2.2,56.0)=6.00, $p<.005$, $\eta^2=.188$)[12]. Figure 9.15 plots the immediate hits made by YHA and OLA at each level of Complexity for both tower-ending and flat-ending problem types. Once again, relatively large group differences can be seen on the 6-move problems where the YHA made more immediate hits overall compared to the OLA. The numbers of immediate hits made by both groups were largest on flat-ending problems. However, two t-tests focusing on YHA and OLA performance on tower-ending and flat-ending 6-move problems showed these differences were not reliable (tower-ending: t(26)=1.01, p=.322; flat-ending: t(26)=1.67, p=.107).

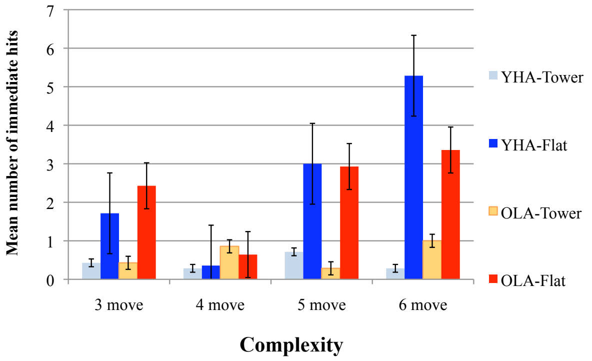


Figure 9.15. Mean number of additional immediate hits made by Primary School MA-matched YHA and OLA groups on tower-ending and flat-ending problems across each level of complexity. Error bars show standard error of the mean.

---

[12] Mauchly's test of sphericity showed significant differences in the variances on Problem Complexity (chi-square=15.87, $p<0.05$), therefore statistics relating to Greenhouse-Geisser are reported.

Primary School full sample comparisons

Five ANCOVAs were performed on accuracy, solution time, number of moves, immediate hits and number of attempts. In each case, the between-participants factor was Group (younger vs. older) and the within-participant factors were Configuration (tower-ending vs. flat-ending) and Complexity (3, 4, 5 and 6-moves). These analyses showed no reliable main effect of MA-CA disparity and no MA-CA disparity x Group interactions. Furthermore, MA-CA disparity was not found to modulate performance on the different configurations, or the different levels of complexity. The same outcome was found when the data for just the 5 and 6-move problems were analysed (i.e., the problems where the greatest amount of variability was observed). Therefore, the results of these tests are not detailed.

Summary of Primary School data

The experimental manipulations, Complexity and Configuration, both had strong effects and broadly the groups showed a similar response to these manipulations. However, there were some more subtle differences. The results of the categorical analysis of MA-matched groups showed that the majority of YHA and OLA completed the 3, 4 and 5-move problems on their first attempt. Groups were found to differ on the number of attempts made on the 6-move problems. On these most difficult problems, the YHA made a greater number of attempts than the OLA. This was also found to co-occur with longer solution times and higher levels of accuracy on the flat-ending problems. The results showed that the groups were not different in the number of immediate hits. However, there was a hint that on the 6-move problems the YHA made more immediate hits compared to the OLA. For example, the data showed the numbers of immediate hits made by the YHA increased steadily as problems became more challenging, whereas in the OLA only small changes in the number of immediate hits were observed. When data from the full Primary School sample were analysed, MA-CA disparity was not found to modulate performance on any of the dependent measures.

Secondary School results

Results are first presented for the categorical analysis of MA-matched YHA (n=16) and OLA (n=19) groups. This is followed by the results of the continuous analysis where MA-CA disparity was used as the covariate in the full sample of younger (n=16) and older (n=19) children.

Secondary School MA-matched group comparisons

This section follows the format used in the Primary School section. Results are reported for a series of 2 (Group: YHA vs. OLA) x 2 (Configuration: tower-ending vs. flat-ending) x 4 (Complexity: 3, 4, 5 and 6-moves) repeated measures ANOVAs. Statistics relating to Group differences and interactions involving Group are given first, followed by a summary of subsequent reliable main effects relating to the other factors.

Number of attempts: The ANOVA on number of attempts showed no reliable differences between YHA and OLA groups. Furthermore, Group was not found to interact with either Configuration or Complexity. Main effects were observed for Configuration (F(1,33)=8.67, p=.006, η^2=.208) and Complexity (F(1,33)=17.83, p<.001, η^2=.351) however these did not interact. Figure 9.16 depicts the mean number of attempts for the YHA and OLA Secondary School groups. The figure shows that the overall number of attempts increased as problems became more difficult and the number of attempts was typically higher for flat-ending than tower-ending problems.

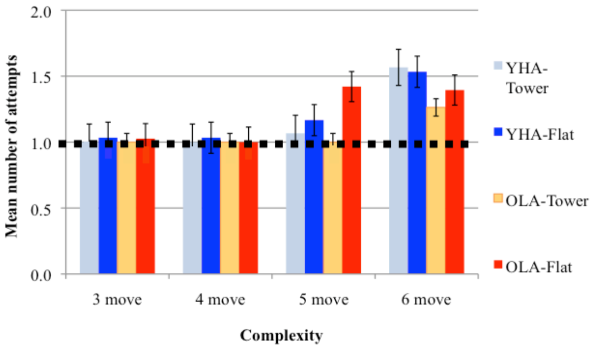


Figure 9.16. Mean number of attempts on the Tower of London task by the Secondary School YHA and OLA on flat-ending and tower-ending problems at each level of complexity. The dotted line represents the minimum number on each problem. Error bars show standard error of the mean.

Accuracy: The YHA and OLA were both at ceiling in their accuracy for each level of Complexity and for both types of Configuration. Subsequently, the ANOVA performed on accuracy found no reliable main effects or interactions. Figure 9.17 illustrates each group's accuracy for each condition.

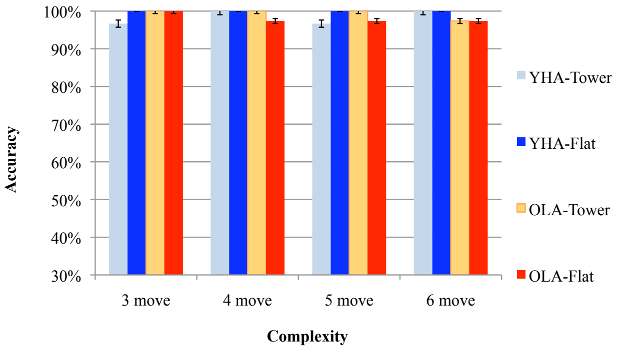


Figure 9.17. Mean accuracy for Secondary School MA-matched YHA and OLA groups on Tower and Flat-ending problems across each level of complexity. Error bars show standard error of the mean.

Solution times: An ANOVA performed on log transformed solution times found no overall effect of Group. Furthermore, no 3-way Group x Configuration x Complexity interaction was found, indicating that these conditions had similar effects on the time taken to solve problems in both YHA and OLA groups. A significant Group x

Configuration interaction was found (F(1,33)=12.70, p=.001, η^2=.278)[13]. This stemmed from a smaller disparity between solution times for tower-ending vs. flat-ending configurations in the YHA (mean ST tower-ending=19.6secs, se=1.3; mean ST flat-ending=21.1secs, se=1.7) compared to the OLA (mean ST tower-ending=15.3secs, se=1.2; mean ST flat-ending=24.7secs, se=1.5). A further significant Configuration x Complexity interaction (F(2.4,80.6)=6.83, p<.001, η^2=.172) was also found, indicating that as problem difficulty increased so too did the amount of time taken to correctly solve problems. Lastly, reliable main effects were also present for Configuration (F(1,33)=44.99, p<.001, η^2=.577) and Complexity (F(1.7,55.8)=202.86, p<.001, η^2=.860). Figure 9.18 illustrates these effects.

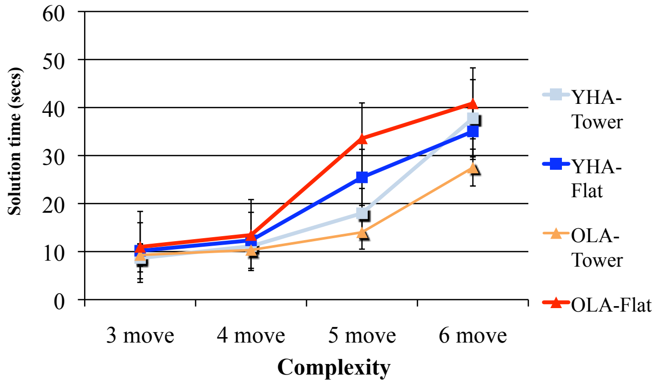


Figure 9.18. Mean solution times taken by Secondary School MA-matched YHA and OLA groups on tower-ending and flat-ending problems across each level of complexity. Error bars show standard error of the mean.

<u>Number of moves</u>: An ANOVA performed on number of moves found no main effect of Group and no interactions involving Group. A significant interaction was found between Configuration and Complexity (F(1.91,63.09)=7.52, p<.001, η^2=.186). Figure 9.19 shows that this interaction stems from complexity having greater modulatory effects for flat-ending versus tower-ending problems. Main effects were also found to be reliable for these separately (Configuration: F(1,33)=17.17, p<.001, η^2=.342; Complexity: F(2.17,71.53)=164.55, p<.001, η^2=.833).

---

[13] Mauchly's test of sphericity showed significant differences in the variances on Complexity (chi-square=45.69, p<0.05), therefore statistics relating to Greenhouse-Geisser are reported.

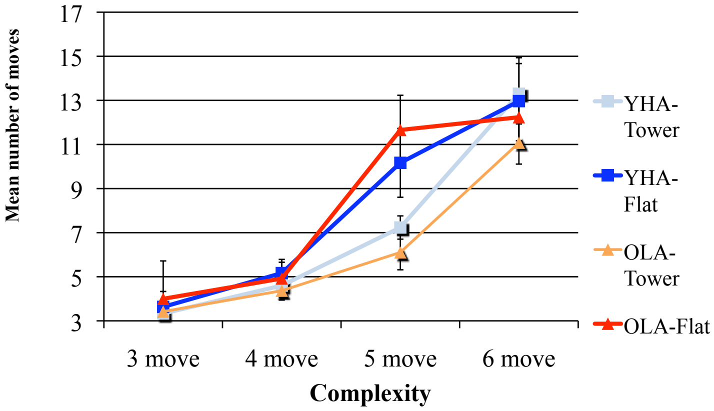


Figure 9.19. Mean number of moves made by Secondary School MA-matched YHA and OLA groups on tower-ending and flat-ending problems across each level of complexity. Error bars show standard error of the mean.

Immediate hits: The ANOVA on the number of immediate hits also revealed no main effect of Group and none of the interactions involving Group were reliable. A significant Configuration x Complexity interaction was found (F(2.07,68.36)=18.23, $p<0.001$, $\eta^2=.356$) which stemmed from a greater increase in number of attempts as the level of difficulty increased on flat-ending compared to tower-ending problems. Reliable main effects were also found separately for Configuration (F(1,33)=45.31, $p<0.001$, $\eta^2=.579$) and Complexity (F(2.43,80.06)=17.75, $p<0.001$, $\eta^2=.350$). See Figure 9.20.

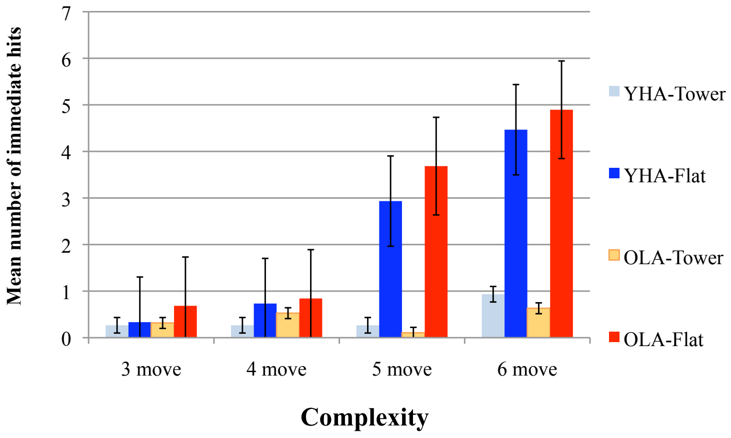


Figure 9.20. Mean number of immediate hits made by Secondary School MA-matched YHA and OLA groups on tower-ending and flat-ending problems across each level of complexity. Error bars show standard error of the mean.

Secondary School full sample comparisons

As was the case in the Primary School data, the Secondary School data showed little variation between the younger and older groups on the 3 and 4-move problems. Therefore, in this section, the data for the simpler problems are omitted in order to highlight any potential interactions that MA-CA disparity may have with Group for tower-ending versus flat-ending problems. ANCOVAs were performed on solution time, number of moves, immediate hits and number of attempts. For accuracy, the data were at ceiling and thus did not offer the possibility of revealing modulatory effects of MA-CA disparity. For number of attempts, MA-CA disparity failed to modulate performance on any of the dependent measures. Thus, only details for the analyses on solution time, number of moves and immediate hits are presented here.

Solution time: The ANCOVA on log-log transformed solution times showed reliable interactions for MA-CA x Group x Configuration x Complexity (F(1,31)=5.40, p=.027, η^2=.148) and Group x Configuration x Complexity (F(1,31)=5.17, p=.030, η^2=.143). Figure 9.21 plots the untransformed solution time data and shows that in the younger age group, more positive MA-CA disparity scores appear associated with faster solution times for 6-move tower-ending problems (dark blue trendline; R^2=.38). However, MA-CA disparity does not appear to modulate solution times for the same problems in the older group (red trendline, R^2=.02), nor for the easier 5-move tower-ending problems in either younger (light blue trendline, R^2=0) or older (orange trendline, R^2=.00) groups.

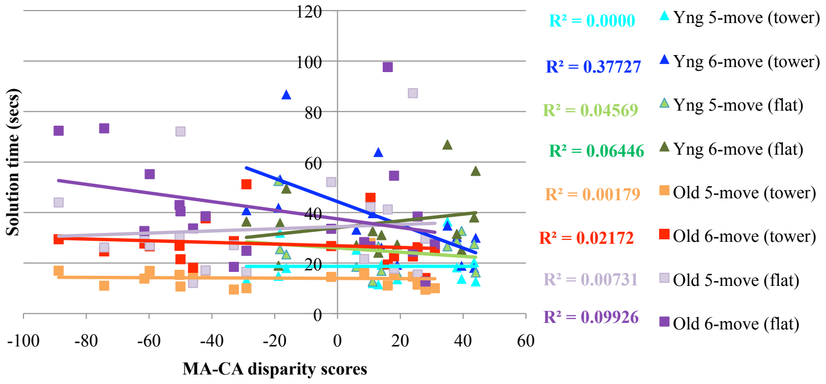


Figure 9.21. Secondary School children's solution times by MA-CA disparity on 5 and 6-move tower-ending and flat-ending problems.

Number of moves: The ANCOVA on number of moves also showed a reliable 4-way MA-CA x Group x Configuration x Complexity interaction (F(1,31)=5.83, p=.022, η^2=.158) and reliable 3-way interactions for MA-CA x Configuration x Complex (F(1,31)=5.04, p=.032, η^2=.140) and MA-CA x Group x Configuration (F(1,31)=5.91, p=.021, η^2=.160). Figure 9.22 plots MA-CA disparity by numbers of moves for tower-ending and flat-ending 5 and 6-move problems in the younger and older groups. The figure shows the majority of trendlines on these conditions are near zero, and thus suggest no modulatory effect of MA-CA. However, the trendlines relating to the 6-move problems in the younger group again show gradients different to zero. For example, on tower-ending problems, as the number of moves made decreases, the more positive the MA-CA disparity (dark blue trendline, R^2=.66). However, on the flat-ending problems the reverse is observed. There, greater numbers of moves appear to be associated with more positive MA-CA disparity scores (dark green trendline, R^2=.30).

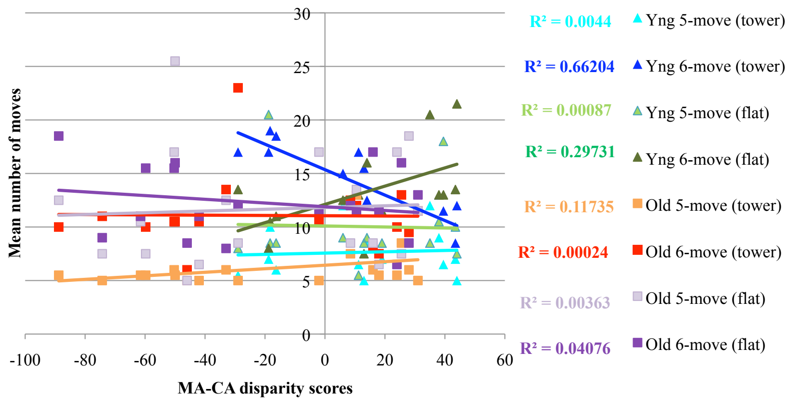


Figure 9.22. Secondary School children's mean number of moves on 5 and 6-move tower-ending and flat-ending problems.

Immediate hits: The ANCOVA on the immediate hits data found reliable interactions for MA-CA x Group x Configuration x Complexity (F(1,31)=5.11, p=.031, η^2=.142), MA-CA x Group x Complexity (F(1,31)=4.41, p=.044, η^2=.125) and MA-CA x Complexity (F(1,31)=5.10, p=.031, η^2=.141). Figure 9.23 shows that the source of these interactions involving MA-CA disparity comes primarily from its effects in the younger group. For example, the figure shows greater modulatory effects of MA-CA disparity on 6-move flat-ending problems, in the younger group (dark green trendline, R^2=.33). In the older group, the trendlines depicting MA-CA disparity by immediate hits show gradients closer to zero for both the 5 and 6-move problems on both the flat-ending problems (light purple and dark purple trendlines) and tower-ending problems (orange and red trendlines).

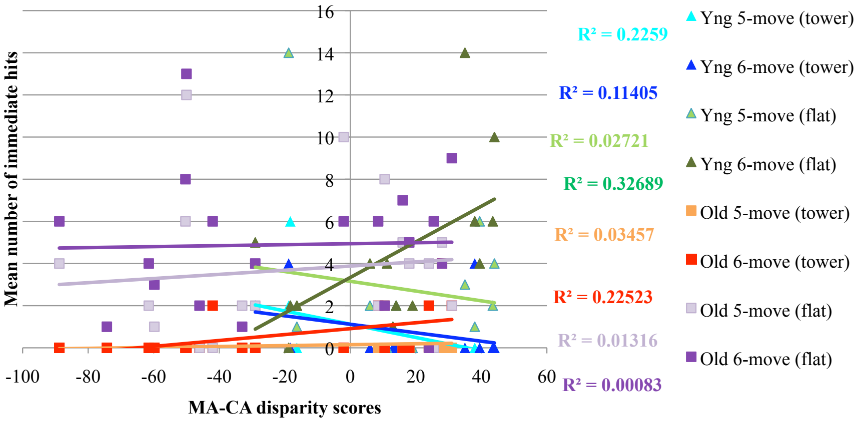


Figure 9.23. Secondary School children's MA-CA disparity by number of immediate hits on 5 and 6-move tower-ending and flat-ending problems

Summary of Secondary School data

The results of the categorical analysis involving the MA-matched younger and older children showed very little difference between groups. Groups were equally matched across the range of measures including, number of attempts, accuracy (where both groups were at ceiling on each condition), number of moves and immediate hits. The one measure that groups were found to differ on was solution time. While the solution times within both groups were found to increase as problems became more difficult, the greatest increases were found in the OLA on flat-ending problems. When the full sample was used with MA-CA disparity entered as the covariate, the ANCOVAs on solution time, number of moves and immediate hits showed that MA-CA disparity had the greatest modulatory effects in the younger group of children. Furthermore, the effects of MA-CA appeared associated most closely to the 6-move problems. However, at this level of complexity MA-CA disparity showed differential effects. For example, on the flat-ending problems, more positive MA-CA disparity scores were associated with greater number of moves and greater number of immediate hits. However, on the 6-move tower-ending problems, more positive MA-CA disparity scores were found associated with lower number of moves, lower number of immediate hits as well as quicker solution times.

Combined Primary and Secondary School results

Using the full samples of Primary School and Secondary School children, a series of 4x2x2 ANCOVAs were performed using MA-CA disparity as the covariate. These analyses focused on the 5 and 6-move problems, where the greatest amount of variability was observed in children's data. Within each ANCOVA, the between-participants factor was Group (Yng-Pri, Old-Pri, Yng-Sec, Old-Sec) and the within-participants factors were Configuration (tower-ending vs. flat-ending) and Complexity (5 and 6-moves). The results of these tests showed MA-CA disparity interacted to modulate performance on only one measure: number of moves. Here a reliable MA-CA x Group x Configuration interaction was found (F(3,67)=3.27, p=.027, η^2=.127). Figure 9.24 depicts this interaction, collapsing across both 5 and 6-move problems. As Figure 9.24 shows, in the Older Primary group MA-CA disparity appears to affect the numbers of moves for flat-ending (dark green trendline) differently than for tower-ending (dark blue trendline) problems. On the tower-ending problems, fewer numbers of moves were associated with more positive MA-CA disparity scores (dark blue trendline). However, on flat-ending problems, the more one's MA exceeded one's CA, the greater the number of moves were made (dark green trendline). This differed from the Older Secondary group for whom fewer numbers of moves on flat-ending problems was associated with a more positive MA-CA disparity score (dark purple trendline).

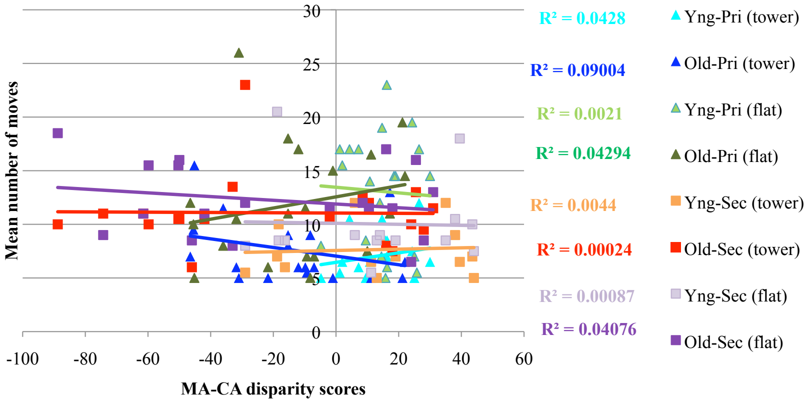


Figure 9.24. Mean number of moves by MA-CA disparity for Primary and Secondary School groups on the different configurations for 5 and 6-move problems.

Discussion

This chapter set out to examine whether groups of younger and older children who were matched on overall mental age differed in their ability to sub-goal and inhibit on the Tower of London task. The results showed that the range of measures used to assess performance did not lead to cleanly dissociable group profiles. For example, at both Primary and Secondary School levels neither the younger nor older groups showed consistent profiles of: (a) greater accuracy, shorter solution times, fewer moves and fewer attempts; or, (b) lower accuracy, longer solution times, greater number of moves and greater number of attempts. Nevertheless, at both school levels the analyses did reveal differences that were reliable and suggestive of differential effects of ability and age. For example, taking the results first from the Primary School and focusing on the findings pertaining to measures of sub-goaling ability, the analysis on number of moves showed that both YHA and OLA were more efficient at identifying the shortest solution paths on easier, but not more difficult problems. That is, the results indicated that the look-ahead ability of both groups at Primary School level was rather limited, and both groups made many more moves than the minimum necessary on the most difficult 6-move problems. Yet the results further showed that while the MA-matched groups were at similar levels of accuracy on the 3, 4 and 5-move problems, the YHA obtained a greater proportion of correct solutions on the most difficult 6-move flat-ending problems – problems previously argued to be more challenging due to the greater number of moves needed to solve them and the ambiguous nature of their sub-goal hierarchy (e.g., Kaller, et al., 2004). While the superior performance of the YHA might have stemmed from the use of a greater number of moves in testing different solutions (or parts of solutions), the results showed that the numbers of moves made by the YHA and the OLA were not reliably different. In addition, the results showed that the YHA took considerably longer than the OLA to solve 6-move problems. Jointly, these findings appear to indicate that the YHA were capable of a greater number of look-ahead moves than the OLA. Thus, by considering candidate solution paths in greater depth, the YHA were able to achieve a greater proportion of correct solutions. One qualification to this interpretation relates to the fact that the YHA also made a greater number of attempts than the OLA on the 6-move problems. Consequently, it is possible that in repeating problems the YHA gained familiarity with those problems, which increased their chances of success. One way of avoiding this possible confound in future studies, is to remove the option to re-start problems, thereby giving participants one attempt per problem. This issue aside, the results from the Primary School children showed that when faced with complex problems older less able children are more likely to give up than younger more able children.

Another finding that stood out from the Primary School analysis relates to the differences in the YHA performance on the 6-move problems of tower-ending and flat-ending configurations. Given the YHA superior performance on the more challenging flat-ending problems, it seems surprising that on the less ambiguous, tower-ending problems they were no different than the OLA in their accuracy. Looking at Figure 9.4 and the number of solution paths that are available suggests that one source of the differences in accuracy may be due to differences associated with backtracking on the two problem types. Specifically, tower-ending and flat-ending problems differ with respect to the chances that backtracking from a deadend state involves moving a ball that is already in position.

The following examples illustrate this hypothesis. Given a tower-ending problem where: (1) all three balls occupy a position on the tallest peg and (2) only one ball is in its correct position and (3) the correctly placed ball is either in the centre or top position, then backtracking necessarily involves removing the correctly placed ball. That is, there is a 2-in-3 chance that the correctly placed ball would need to be moved. By contrast, given a flat-ending problem where currently: (1) each ball occupies a bottom position on the three pegs and (2) only one ball is in its correct position and (3) the correctly placed ball is either on the middle-longest peg or the longest peg, then backtracking does not require the correctly placed ball to be moved. That is, in a flat-ending configuration where only one ball is in place, there is a 1-in-3 chance that the correctly placed ball needs to be moved. Thus, for younger children, for whom backtracking poses greater challenges (see e.g., Klahr, 1985), the greater odds associated with moving correctly placed balls when backtracking from deadend states may explain the difference in accuracy for tower-ending versus flat-ending problems.

Turning to the measure of inhibition, the results of the analysis on the Primary School children's immediate hits hinted that as problems became more challenging, the YHA had greater difficulty inhibiting moves that increased only the perceptual similarity of the problem. This type of pattern is consistent with evidence that inhibitory control is a skill that remains under development until later childhood (e.g., Asato, Sweeney, & Luna, 2006; Mitchell & Poston, 2001; Novikova & Stroganova, 2006). Thus, small influences of lower inhibitory control may have greater influence on the performance of the younger children than the older children. Finally, when the full sample of Primary School children was included and MA-CA disparity scores used as a covariate in the analysis, the results showed that one's advantage (i.e., one's increase in MA over CA) did not reliably modulate performance on any of the dependent measures.

Similar to the results of the Primary School, the results of the Secondary School analysis showed that on number of moves, both YHA and OLA were more efficient at identifying the shortest solution paths on easier problems. These results suggest that, like the Primary School children, the Secondary School children were still limited in the number of look-ahead moves they were capable of. However, the overall number of moves made by the Secondary School children was fewer than the Primary School children, thus indicating an increase in sub-goaling efficiency over age.

Overall, there was little to distinguish the performances of the MA-matched Secondary School groups. For example, both the YHA and OLA were at ceiling in terms of their accuracy at all levels of complexity for the tower-ending and flat-ending problems. They were also no different in the number of moves they made or their number of immediate hits. However, the groups were found to differ in their solution times on the different configurations. In contrast to the results in the Primary School data, in the Secondary School sample it was the OLA who took longest on the flat-ending problems. If longer solution times are taken to indicate a deeper processing of candidate moves, then this would suggest the OLA are capable of a greater number of look-ahead moves. However, given that both groups were at ceiling in their accuracy, the current set of problems may not have allowed the greater sub-goaling efficiency of the OLA to be revealed. To test this, a range of more complex problems is needed for which the solution requires a greater number of minimum moves. Also in contrast to the Primary School data, the covariate analysis on the Secondary School data showed MA-CA was a reliable influence on several of the dependent measures. Specifically, these were solution time, number of moves and immediate hits. However, most of these reliable MA-CA interactions were found to stem from the YHA group only, where more positive MA-CA disparities were found to modulate faster solution times and fewer moves, but also greater number of immediate hits on the 6-move flat-ending problems. While it seems sensible that one's greater advantage might modulate faster solution times and fewer numbers of moves, why should this be accompanied by increased numbers of errors in the form of immediate hits? One possibility, is that children engaged in a speed-accuracy trade-off and, by going faster to solve problems, they were temporarily led off course by the perceptual features of some configurations. Finally, in the combined analysis the results showed interactions between age, ability and number of moves. However, the results did not reveal clearly in which groups the effects of MA-CA disparity originated.

Although the data from the Primary School and Secondary School samples offer somewhat mixed results, one issue that appears relatively clear is the lack of influence that inhibition has on overall performance. That is, in as far as immediate hits offer a reliable measure of inhibition, the data show no role of inhibition in explaining differences between younger more able and older less able children. Thus, while at younger ages (e.g., 3-4 and 5-6 year olds) previous studies have found evidence suggesting that differences in inhibitory control may account for differences in performance on the Tower of London task (see e.g., Baughman & Cooper, 2007), at older ages inhibition appears less influential. Given that the results showed that performance improved with age (i.e., accuracy improved, solution time decreased and numbers of moves were fewer), differences appeared to be related primarily to differences in the ability to sub-goal. One method of further exploring this notion is to use the models developed by Baughman et al. (2007) and increase the model's look-ahead capacity. In this way, the models might be compared to the experimental data to see if patterns of solution time, accuracy, number of moves and immediate hits appear similar to those reported here.

Several additional factors require that caution be applied to the interpretations offered above. One such factor relates to the potential of the experimental design to reveal variability between groups. For example, the results of the accuracy data for the Secondary School children showed no variability on any of the levels of complexity, for either tower-ending or flat-ending problems. Similarly, at the Primary School level the 3 and 4-move problems offered little in the way of challenge to even the youngest children. It was only the 5 and 6-move problems that appeared successful at differentiating abilities in the groups tested. However, given that there were only two tower-ending and two flat-ending problems administered to children at each level of complexity, the sensitivity of the measures based on these data is rather limited. A more effective design would include a greater number of problems ranging from a minimum of 5-moves upwards.

Another issue highlighted by the current study concerns the variability in number of solution paths available on easier versus more difficult problems. For example, tracing the solutions in Figure 9.4 on the simpler 3-move problems showed there was typically only one solution path that allowed goal completion in the minimum number of moves. However, on the more complex 5 and 6-move problems often two or three possible solution paths were possible, each leading to the goal state within the minimum number of moves. While each such solution path is equally valid, there were also differences in the proportion of instances where immediate hits were possible. In other words, some solution paths led to situations where there were greater opportunities to be led astray by perceptual influences. One question this raises, is what factors influence the early choice moves for the different solution paths? It is possible that more might be learned about these factors through a refined analysis of problems in which, for example, the solution paths are pre-specified for a limited number of problems and the path chosen is identified on the basis of the precise moves made. Analysing the series of moves children make may also further clarify whether, upon evaluating the different possible solution paths, children remain committed to one solution path or, move between paths, in their attempts to solve problems.

With respect to the question of the relationship between intelligence and cognitive development, the results presented here suggest that at younger ages one’s chronological age offers a stronger influences on one’s ability to inhibit perceptual biases, and that one’s mental ability (or advantage) offers a stronger influence on one’s ability to sub-goal. Age and ability appear to interact over time, so that when older children attempt to process information quickly they become vulnerable to influences of perceptual biases.

This pattern is illustrative of a differential involvement of the mechanisms underlying planning and problem solving and thus indicates that intelligence and cognitive development are not exactly the same thing. Finally, the pattern of findings presented in this chapter begs a broader consideration of how the Tower of London task is used to understand executive functions. For example, one benefit that the measures used in this chapter appear to highlight is their potential to reveal differences in cognitive processing that standard measures do not capture. For example, if accuracy or number of moves had been measured according to standard binary outcomes (e.g., ‘Were problems completed?’: Yes/No. Or, ‘Were problems solved in the minimum number of moves?’: Yes/No), then the subtle pattern of differences found between MA-matched groups reported here would not have been revealed.

***

With the last of the chapters on the experimental studies now complete, we next turn to study of uneven cognitive profile using a series of computational models. In this modelling work, we will consider some of the conditions and outcomes associated with presenting small disparities early on in the development of a cognitive system.

Part 6

Preface to computational work

The focus of previous chapters was on providing an analysis of the performance profiles of groups of younger and older children of the same mental age. In those chapters I highlighted differences in each group's behaviour on the BAS II and on a range of computer-based cognitive tasks. I then linked those findings to various mechanistic accounts (offered by a number of computational models each relating to a specific cognitive domain) and suggested how subsequent computational work could help to narrow the role of proposed mechanisms underlying variability. Yet, a focused account of the mechanisms of variability within specific cognitive domains should be complemented by a consideration at a broader level of how cognitive domains relate to each other within development itself (e.g., Karmiloff-Smith, 1998). In the following chapter, I explore this argument using a series of dynamical systems models that aim to investigate the following question: to what extent does the overarching cognitive architecture influence development within multi-component systems to produce uneven cognitive profiles? This modelling work forms an important part of the thesis because it allows the causal influences of small amounts of change to the starting state of a system to be charted within the full architecture comprising multiple component processes (e.g., Mareschal, et al., 2007).

Chapter 10 A computational framework for exploring uneven cognitive profiles

Introduction

As children develop, their abilities within various cognitive domains can be observed to fluctuate. In typically developing children, for instance, the relative differences between abilities may be subtle. However, in the case of some developmental disorders, the unevenness of cognitive profiles may be rather more pronounced. Questions concerning the origin of these uneven profiles and the extent to which deficits can be truly restricted to single domains lie at the centre of ongoing debate (for a recent discussion, see Bullinaria, 2007). At one extreme of the debate are arguments that the existence of uneven cognitive profiles supports a view that the functional organisation of cognition is largely *modular*. Cognitive neuropsychology, for example, frequently appeals to modular views on the basis of evidence from studies of adult focal brain damage or disease in which behaviours are found to disassociate. The observed pattern of cognitive breakdown in patients, referred to as a ‘fractionation’ of abilities, have thus provided the basis for models of normal functional cognition in adults (see e.g., Shallice, 1988). However, a point of contention for many theorists lies in the application of such models to *developmental disorders* (e.g., Temple, 1997). This is because such an extension would seem to require at least two assumptions: (1) that the child’s cognitive system is also modular; and (2) that domain-specific deficits can persist without compensation from or spread to other causally linked cognitive abilities that are undergoing processes of development. As we shall see shortly, both these assumptions have been questioned. The issue of modularity in cognition is a large and complex theoretical debate, informed by evidence at cognitive, brain and genetic levels and there is insufficient space to do it justice here. For present purposes, I briefly describe some of the key points relating to modular accounts.

Modularity in cognition

According to Fodor (1983), modules are essential properties underlying development within cognitive systems. He suggested that modules might consist of the following attributes: modules are specific to particular cognitive domains; they are fast and automatic in their operation; they rely on dedicated knowledge bases; they are instantiated within dedicated neural structures; and they are innately prespecified. While Fodor (1983) initially believed these modules operated within low-level cognitive processes, other researchers have extended the proposal to include higher-level cognition – a position referred to as the 'massive modularity' hypothesis (for discussion see e.g., Sperber & Dupoux, 2001). Other related accounts involving modular perspectives may be highlighted. For example, in Chapter 2, I described Anderson's (1992) theory of minimal cognitive architecture, in which he proposed that a mixture of modules characterised development. He argued that some of these were the product of experience, while other lower-level modules were innately prespecified. Overall, one of the basic premises that modular accounts share is that uneven cognitive profiles may be explained by the failure, or the delay of specialised functional modules. This failure is often characterised as occurring in an otherwise normally developing cognitive system – a position Thomas and Karmiloff-Smith (2002) referred to as 'Residual Normality'.

Alternative views of cognitive architecture

Within the area of developmental disorders, however, assumptions concerning modularity in cognition have been challenged. For example, using a computational approach, Thomas and Karmiloff-Smith (2002) have argued that the assumption of Residual Normality is lacking. Specifically, they focus on showing how initially unaffected components within a model, where a single component has experienced a deficit, can compensate for the damaged process and be affected by spread of damage.

Nevertheless, within the literature there is some agreement with the view that some form of modularity characterises normal cognition in adults (see e.g., Barrett & Kurzban, 2006; Bullinaria, 2007; Goswami, 1998; Karmiloff-Smith, 1995). However, there is considerable disagreement about the factors that produce it, and at what point in development modules emerge (e.g., Karmiloff-Smith, 1998). For example, Karmiloff-Smith (1995) has proposed that 'progressive modularisation' occurs as a process of competition amongst *domain-relevant* biases over development. Support for this claim is offered from a variety of sources including neuroscience, where studies have reported only slight differences across the cortex in brain chemistry, neuronal density and type and orientation of neurons (see

Karmiloff-Smith, 1995). Karmiloff-Smith argues that these subtle differences in the starting state of a system provide initial, information-processing biases that allow neural areas to become specialised for particular functions. According to Karmiloff-Smith (1995), domain-specific modules emerge over the course of development and under constraints imposed by the environment.

In contrast to modular theories (and at the other extreme of the debate on cognitive unevenness) are views that postulate the cognitive system as highly distributed (see e.g., Stiles, et al., 2005). For example, distributed theories (of the sort inspired by McClelland & Rumelhart, 1988) lead to doubts that any deficit, however domain specific to begin with, could remain so across development. In such theories, cognitive abilities are graded and interactive, relying on the contribution of many different processes (and brain regions). Evidence from the neurosciences supports the view that early in development the brain is highly interactive and capable of compensation following some forms of damage (Thomas, et al., 2002). Additionally, between the extremes of fully modular and fully distributed theories lie various other accounts that propose more limited degrees of cognitive differentiation. For example, one might view the emergence of laterality effects in language after unilateral brain damage (Bates & Roe, 2001) as suggestive of hemispherical architecture. Other accounts focus on the importance of a central executive connected to many specialised subsystems (see e.g., Baddeley, 1996), while still other accounts describe uni-directional vs. multi-directional processes (see e.g., Cohen, 1985). Finally, one may also find accounts that emphasise a hierarchical organisation in cognition (e.g., Anderson & Lebiere, 1998). Figure 10.1 provides simple depictions of each of these views.

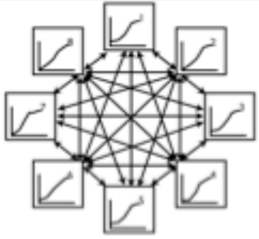
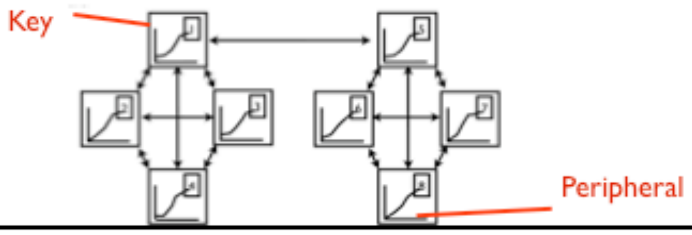
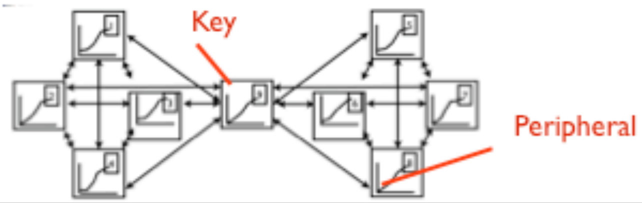
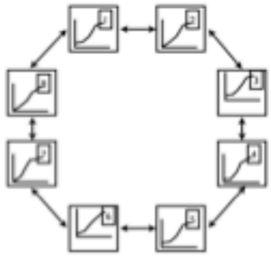
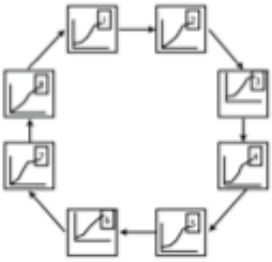
Architecture	Representation
A) Fully Connected	
B) Hemispheric	
C) Central Processor	
D) Bi-directional Loop	
E) Unidirectional Loop	
F) Hierarchical	
G) Fully Modular	

Figure 10.1 Depictions of alternative cognitive architectures. Nodes show growth curves representing developmental state of processing component over time. Arrows between nodes represent causal connections. See text for details.

A computational approach to the study of uneven cognitive profiles

Given the range of views concerning the possible underlying architecture of cognition, the following question may be asked: *Is each architecture equally capable of supporting uneven profiles over development, similar to those observed in atypical, as well as typical development?* The answer to this question is difficult to predict as soon as one moves away from a modular account within which each ability develops in isolation (and so, presumably, can be affected in isolation). One response to this difficulty is to appeal to formal computational modelling of developmental systems. In this way, the downstream effects of early anomalies can be quantitatively charted. However, little work has explored the development of whole cognitive systems at this broad scale. One reason is that computational models of development have, in the main, focused on the acquisition of specific domains rather than the development of large scale systems with multiple and heterogeneous interacting components. In this chapter, I address this omission by constructing a series of multi-component systems, each representing a different view of the architecture of the cognitive system. I implement this series of models within a dynamical systems framework that evaluates causal influences of small changes to the initial state within a global architecture. This set of models allows the effects of change to be traced over development within different global architectures comprising a number of other component processes. Specifically, I explore two types of changes to the initial states of the cognitive architectures. I examine the effects of: (1) *pure focal deficit* and (2) *combined focal advantage and general deficits*. The inspiration for these approaches stems broadly from reported cases of uneven cognitive profiles within developmental disorders relating to 'specific language impairment'[1] and 'savant syndrome'[2]. A schematic developmental profile that a pure focal deficit might resemble is offered in Figure 10.2. The blue line in this figure depicts the case of typical development,

---

[1] The term 'specific language impairment' (SLI) is used to refer to an impairment in some aspect of language (e.g., phonology, vocabulary, or grammar), that is not the result of any known sensory, neural or intellectual deficit (see e.g., Shafer, et al., 2007). The disorder is "specific" because the language deficit is proposed to exist in an otherwise normally developing cognitive system. (Note, that in the field of language development, the true degree of specificity of the disorder remains controversial and the disorder is widely viewed as heterogeneous.)

[2] Savant syndrome is broadly taken to refer to a profile of generally low cognitive functioning, but where one ability stands out as remarkably better than the rest (see e.g., Treffert, 1989).

while the red dashed lines depicts the effects of a focal deficit where there is no spread or compensation offered by the remainder of the cognitive system. The upper-most red dashed line portrays the trajectories of unaffected processes and the lower-most red dashed line depicts the trajectory of affected process.

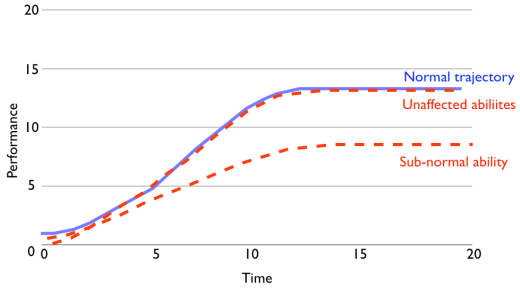


Figure 10.2. Depicting a developmental profile of pure focal deficits (values on x and y axes are arbitrary). See text for details.

For the combined focal advantage and general deficits models, I take the example of savant profile as motivation for the models. Figure 10.3 offers two schematic developmental profiles that are inspired by Treffert’s (1989) distinction between: (1) “talented savants” – individuals who exhibit an ability that is remarkable, given their other generally low cognitive functioning, but which is within the normal range for chronological age; and (2) “prodigious savants” – individuals who show an ability that is exceptional by normal standards (i.e., above-average ability compared to normal populations) but whose functioning in other areas is lower than normal.[3] Illustrative profiles for talented savant and prodigious savant are depicted in Figure 10.3 left and Figure 10.3 right, respectively. In both cases, the blue line depicts the case of typical development, the upper-most red dashed line depicts the case of focal advantage and the lower-most red dashed line depicts general impairments. Again, the effects of focal advantage and general deficit are portrayed as unfolding in a system where development is not characterised by a pattern of spread or compensation.

---

[3] The talented savant might show, for example, normal or lower than normal verbal memory in addition to extremely low levels of ability in other domains, whereas the prodigious savant might exhibit special skills (e.g., memorisation, learning of multiple languages, mastery of music or art, mathematical or calendrical computations), but be otherwise well below average in other domains (see e.g., Treffert, 1989).

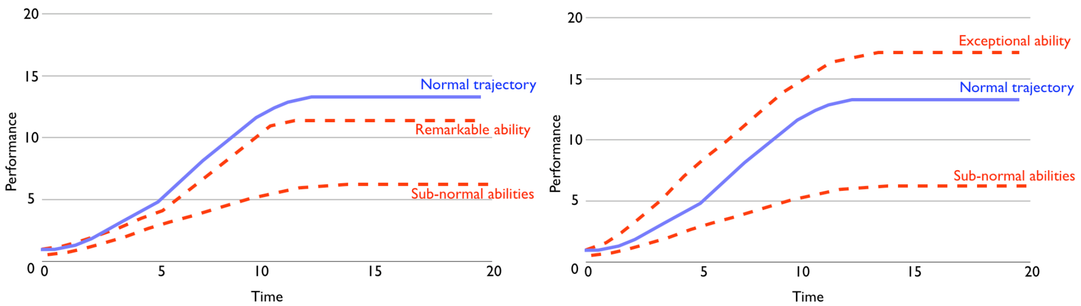


Figure 10.3 Depicting 'talented' savant (left) and 'prodigious' savant (right) developmental profiles (values on x and y axes are arbitrary). See text for details.

It is clear that the profiles depicted in Figure 10.2 and Figure 10.3 above represent very different cases of uneven cognitive profiles in development. The important point about these profiles is that they are intended to be *illustrative* of uneven cognitive profiles found in cases such as specific language impairment and savant syndrome – disorders that have been used to argue the view that the functional organisation of cognition is modular (see e.g., Smith & Tsimpli, 1995). Next, I set out the rationale for adopting a dynamical systems theory framework to explore conditions of uneven profiles in development.

Dynamical systems theory

The primary reason for adopting a dynamical systems framework in this thesis to explore questions of uneven cognitive profiles originates in recent work on the development of intelligence by van der Maas, Dolan, Grasman, Wicherts, Huizenga and Raijmakers (2006). Before we take a closer look at that work, let us briefly review what dynamical systems theory is and how it relates to the study of uneven cognitive profiles in this thesis.

Dynamical systems: An example from population dynamics

Dynamical systems research includes a range of related approaches that use mathematics to study changes within dynamic, non-linear systems over time (see e.g., Spencer & Schoner, 2003). For example, within the field of mathematical biology, dynamical systems theory is used to study changes in populations of species that hold complex relationships with each other and with their environments. For example, the Lotka-Volterra model (see e.g., Murray, 2003; Wangersky, 1978) has been used extensively to model changes in populations of predators and prey. To illustrate its use, let us briefly consider the scenario in which there is a population of foxes (predators) and a population of rabbits (prey). In nature, where these populations exist, one could expect that if resources are plentiful (e.g. lots of food) and when there are few threats (e.g., few foxes), then the population of rabbits will grow. However, as the number of rabbits grows, their greater number also provides greater opportunity for the foxes that prey on them. Thus, the fox population also grows. The growing number of foxes eventually results in a reduction to the rabbit population, which in turn leads to a downturn in the number of foxes. This is because the declining rabbit population cannot continue to sustain a large fox population, and competition amongst the fox population for fewer resources leads to natural decline in their numbers. Figure 10.4 depicts the patterns of change in these populations over time, via two oscillating 'growth curves' representing foxes (red line) and rabbits (green line). Growth curves represent the developmental state of a population over time.

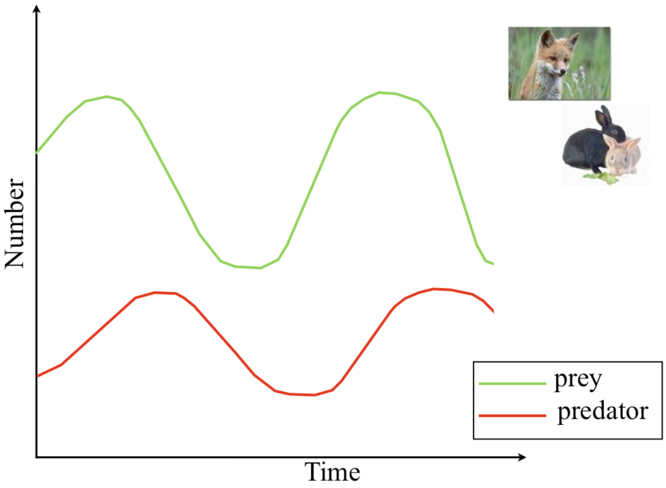


Figure 10.4. Oscillating growth curves depicting the dynamic interactions between populations of prey and predator in the Lotka-Volterra model.

The Lotka-Volterra model captures this pattern of non-linear dynamic change within populations using coupled differential equations. The model specifies precisely how change in one population will effect change in the other. Equation 3 below shows one form of the Lotka-Volterra model. In this equation the size of two populations are given: x (the prey; in this case rabbits) and y (the predators; in this case foxes). Additionally, the population of the prey grows at an exponential rate a, and is constrained by resources in the environment K (referred to as the carrying capacity) and the frequency with which it comes into contact with and is killed by its predator *Myx*. The population of foxes, *y*, grows at rate *b* and is constrained by the resources available (i.e., rabbits) and the effectiveness of its predation on rabbits *Mbx* (i.e., the proportion of times in which an encounter with a rabbit results in dinner for the fox).

Prey: $$\frac{dx}{dt} = ax(1 - x/K) - Myx$$

Predator: $$\frac{dy}{dt} = (Mbx - c)y$$

Equation 3. Lotka-Volterra predator-prey model. See text for details.

Isolating the factors underlying growth curves

In the case of predator-prey models such as the Lotka-Volterra model, populations are intricately linked in a cycle of competition that results in the oscillatory pattern of growth curves depicted earlier in Figure 10.4. However, by isolating just one population, we can examine precisely how factors such as rate of growth, capacity of the environment and the population's initial starting size interact in influencing change. We can thus trace the causal influence of changing the initial state of one or more of these properties within a global architecture and trace its effects over time. For example, Equation 4 focuses on just one part of the Lotka-Volterra model and describes how change (*d*) in one population is dependent on its initial value (*x*), how much time has passed (*t*), its growth rate (*a*) and its capacity (or asymptote) level (*K*).

$$\frac{dx}{dt} = ax(1 - x/K)$$

Equation 4. Equation specifying a non-linear growth curve in a single population.

Taking Equation 4 and providing different values to each of the parameters (*a*, *K* and *x*), Figure 10.5 depicts four sample growth curves capturing a range of shapes within the parameter space. Thus, in a cognitive context, growth curves can be regarded as summarising the outcome of some underlying experience-dependent, developmental process.

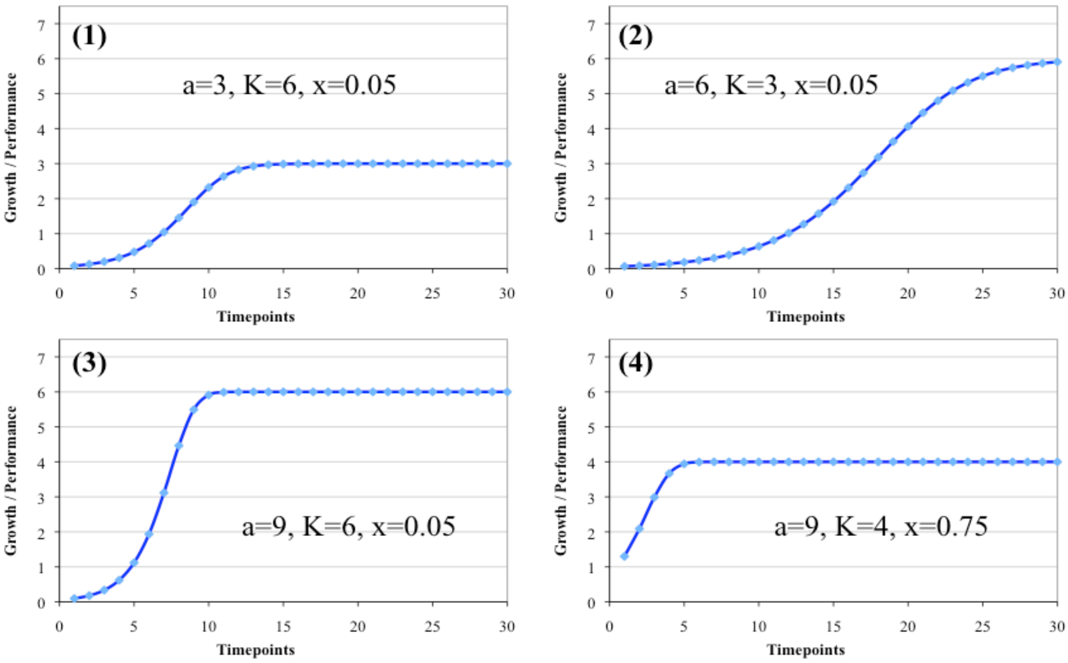


Figure 10.5. Four sample growth curves with different settings for growth *a*, starting state *x* and asymptote level, *K*. Parameter values are included within each tile.

A dynamical model of intelligence (van de Maas et al., 2006)

Using a dynamical systems model, van de Maas et al. (2006) set out to account for two key findings from the literature on intelligence. These were as follows: (1) cognitive performance in different domains is not well correlated in early childhood but becomes correlated over time (referred to as the 'positive manifold'); and (2) factor analysis typically reveals a single higher-order factor from tests of intelligence (labelled the *g factor*). Both of these findings have led to the hypothesis that a real substantive property exists that influences individual variability in cognitive development (see e.g., Jensen, 1998), so that the statistical construct is explained by a biological factor. However, van der Maas et al. demonstrated these empirical findings could be explained in terms of the developmental interactions between initially uncorrelated processes, instead of invoking a single underlying property governing development.

The model that van der Maas et al. (2006) put forward was based on the Lotka-Volterra model. However, it differed to the Lotka-Volterra model in a number of important ways. Firstly, the van der Maas et al. (2006) model was comprised of a large number of component processes that interacted in a fully distributed fashion. Figure 10.6 depicts the design of a model showing only 8 fully connected processes – in their model, van der Maas et al. (2006) used 16 processes. The figure below shows each node is connected to every other node and each component process can be seen to exhibit an individual growth curve. Thus, the behaviour of each process influences and is influenced by all other processes.

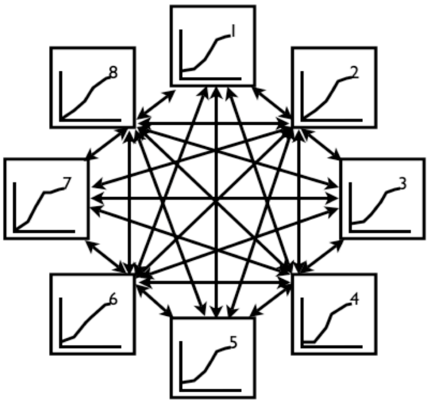


Figure 10.6. Depicting the mutualism model by van der Maas et al. (2006). Here a fully connected dynamical systems model showing growth curves of each process. Note: the number of actual processes used was 16. These have been reduced here for clarity of illustration.

While unique parameters drive the development of an individual process, so too do the development of all the other processes in the system to which it is connected. Specifically, a second fundamental difference between the van der Maas model and the Lotka-Volterra model is that in the van der Maas model the interactions between processes result in mutually beneficial and positive influences over development. Hence, the model is referred to as the ‘mutualism’ model. van der Maas et al. (2006) explain part of the justification derives from evidence that shows that children appear able to make use of a number of unrelated processes to facilitate their performance in some domain. For example, a child might use verbal processes to talk through and facilitate their reasoning in an abstract reasoning task (see van der Maas, et al., 2006). In van der Maas et al.’s model, the following equation specifies this interaction:

$$a_i \sum_{\substack{j=1 \\ j \neq i}}^{W} M_{ij} x_j x_i / K_i \quad (\text{for i,j} = 1:\text{W})$$

Equation 5. Summing the interactions of all other processes within the mutualism model (van der Maas et al., 2006).

In Equation 5, *M* represents a matrix specifying the connectivity between processes. For example, *M*=[0 1 1;1 0 1;1 1 0], specifies a three-component system in which each component is connected to each other component, but a component is not connected to itself. *W* is the number of components and *i* is the component currently summing interactions from component *j* (for all j from 1 to W). Putting the previous equations together, Equation 6 specifies the entire dynamics of the mutualism model. It states that at each point in time (*t*) the change in the performance level *x* of a given process *i* (dx_i) is a product of the sum of the strength of connection (or interaction weight) of each process *j* with which it is functionally connected ($M_{ij}x_jx_i$), multiplied by the rate of growth of process *i* (a_i) times the current level of performance of process x_i, divided by the asymptote level for that process (K_i). Changes in x_i at each time step are constrained by the performance (and thus the individual properties) of all other processes to which it is connected.

$$\frac{dx_i}{dt} = a_i x_i (1 - x_i / K_i) + a_i \sum_{\substack{j=1 \\ j \neq i}}^{W} M_{ij} x_j x_i / K_i \quad (\text{for i,j} = 1:\text{W})$$

Equation 6. The mutualism equation

The success of van der Maas et al.’s model in accounting for a range of empirical phenomenon relating to intelligence therefore offers a useful foundation for exploring a number of related questions concerning intelligence and cognitive development. Specifically, it is possible to make the initial parameters of the different components more or less uneven, and then explore how these parameters – in conjunction with the causal connectivity between components – can influence cognitive processes across development. Because these parameters are relatively few (i.e., *a*, *K* and *x*) and because the functional architecture can be explicitly specified via a matrix of functional connectivity (*M*), it offers a valuable framework for investigating issues surrounding uneven cognitive profiles within various architectures. By stipulating each of the different global architectures described earlier (i.e., fully distributed, hemispheric, central processor, bi-directional, uni-directional, hierarchical, and modular), it is then possible to examine the potential causal origin of uneven cognitive profiles within an explanatory developmental setting.

Extending the mutualism (van der Maas, 2006) model

In this chapter I make a number of assumptions similar to those made by van der Maas et al. (2006). I assume the development of cognitive processes can be simulated using growth curves of the sort found in non-linear learning systems of a given domain. In doing this, I can therefore focus on how processes interact across development. Modelling necessarily involves simplification in order to target the phenomenon under study. I further assume that the process of some experience-dependent developmental process may be represented by a growth curve. In simplifying the process of development within cognitive components, I can thus focus on the dynamics of interactions *between* components across development. By sampling the values of a, K and x to vary around means with pre-specified standard deviations, a range of performance can then be created for a particular individual component in the population. Combining this variability with Equation 4 then allows us to determine the upper and lower boundaries around normal performance, for each process. These trajectories are illustrated in Figure 10.7 below.

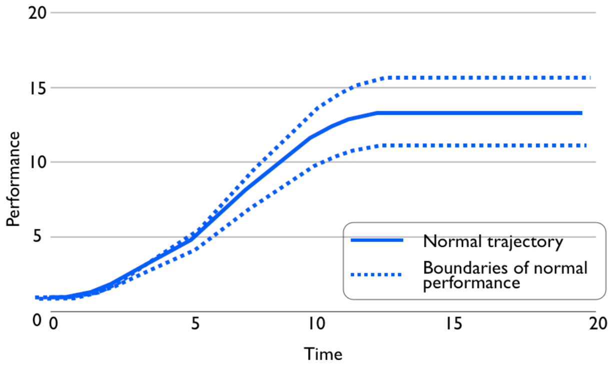


Figure 10.7. A non-linear growth curve showing upper and lower boundaries of normal performance

Similarly, the different components within a given individual may have a range of initial parameter values (which may, or may not be correlated). In the following simulation work, we first consider the uneven profiles caused by initial localised damage to a single component in each of the cognitive architectures that were depicted in Figure 10.1. We then consider uneven profiles where initial impairments exist within multiple components but where performance of one component is markedly superior to the others. Respectively, these two approaches will be referred to as '*Focal Deficit*' and '*Combined Advantage and Deficit*' models.

Method

Normal models

Processes

Each architecture contained 16 components (17 in the case of the central processor). Pilot simulations indicated that the results were not especially sensitive to the number of components included, with one exception (see later). The development of each cognitive component was defined by a growth curve (summarising the outcome of development for a given cognitive process) with the following 3 parameters, *onset*, *rate*, and *asymptote*. In the mutualism model, these parameters were assumed to vary both within and between individuals. This variability was implemented by sampling the parameters from the following normal distributions: onset: mean=0.05, standard deviation (SD)=.01; rate: mean=6, SD=.5; and asymptote: mean=3, SD=.5. These values were based on the settings used by van der Maas et al. (2006) and chosen because the small initial values offered small and uncorrelated performances on each of the processes at early stages.

Global architectures

A series of M-matrices defined each of the architectures shown in Figure 10.1 (the full set of matrices specifying these architectures are provided in Appendix J). The links between the component processes were defined by a connectivity matrix ($M_{ij}x_jx_i$) that determined the amount of influence each causally connected component exerted on the other. Within a given architecture, matrix values were fixed and were invariant across the population of runs. Differences were applied to the values that were set for this matrix within the hierarchical system in order to attenuate the differences between processes over development. This made trajectories easier to discriminate, yet left the basic relationship between processes unaltered. For the hierarchical system, matrix values were set to 0.25, on all others the value were set to 0.05. Investigation of the strength of interaction weights that were possible in each model revealed different boundary ranges. Table 10-1 shows the boundary ranges for each of the architectures. The values in the middle column show the values that were used in the model and the values in the right-hand column show the range of interaction weights that were possible given that architecture.

Table 10-1. The interaction weights M_{ij} used in the models and the boundary ranges that applied within each architecture.

Architecture	M_{ij} value in models	M_{ij} range
Fully connected	0.05	0.01-0.07
Hemispheric	0.05	0.01-0.14
Central Processor	0.05	0.01-0.10
Bi-directional Loop	0.05	0.01-0.50
Uni-directional Loop	0.05	0.01-0.90
Hierarchical	0.25	No limit
Fully Modular	N/A	N/A

N/A= Not applicable

Ranges above the boundary values resulted in a catastrophic failure in the model. Figure 10.8 illustrates the consequence of exceeding the boundary range in the fully distributed model. Here, an interaction weight value of 0.10 was used with the resulting effect depicted. This highlights the fact that each model operated within a narrow range of values.

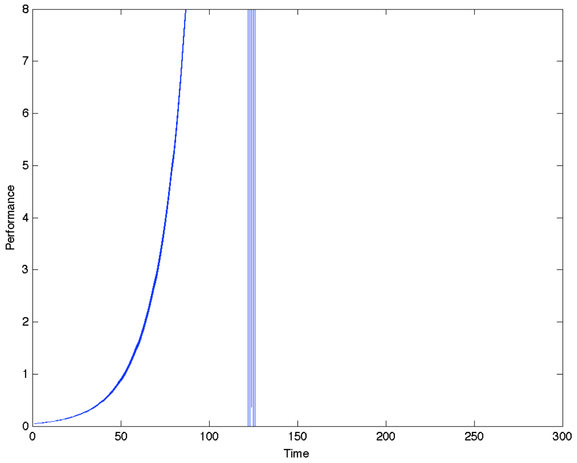


Figure 10.8. Illustrating the typical outcome associated with exceeding the boundary range of the M interactions. The above is an example of the range being exceeded in the fully distributed model.

In the following simulations, I assume that a given M-matrix defines the population, and that other parameter variations correspond to intra-individual heterogeneity in cognitive mechanisms (and their relative strengths), as well as inter-individual variations in ability. I did not consider variations in the M-matrix as a means of simulating disorders, although clearly the model provides the opportunity to explore the possibility of deficits arising through disconnection between processes.

Populations

Populations of 200 individuals were generated for each condition and mean performance was calculated. The development of each individual system was simulated for 300 time steps. For impaired models, systems could be run for more time steps until a stable state was reached[4].

Impaired models

Focal Deficit models: In this selection of models, I applied an initial focal deficit to a single component, either to its onset (*x*), growth rate (*a*), final asymptote (*K*), or combinations of these three (i.e., *a*, *K*, *x*, *aK*, *ax*, *Kx*, or *aKx*).

Combined Advantage and Deficit models: The approach taken for this set of models involved carrying out manipulations on all processes, simultaneously. This comprised a basic method of increasing the startstate of one process by a specific proportion and decreasing the startstates of the other remaining processes by the same proportional amount. Thus, the aim was to attempt to choose a set of initial manipulations that offered the best chances of producing the target profile. For example, if the startstate for onset (*x*) was increased by 75% in one process, it was decreased by 75% in all other processes. These manipulations were performed on each of the parametric combinations (i.e., *a*, *K*, *x*, *aK*, *ax*, *Kx*, and *aKx*).

For each of the architectures shown in Figure 10.1, and for both atypical profiles, three levels of change were assessed (25%, 50% and 75% of normal values). Several levels were used in order to probe for possible non-linearities or threshold effects in the subsequent trajectories of processes. However, within both set-ups, linear changes in the amount of parameter change demonstrated approximately linear effects in the consequent trajectories. Therefore, for clarity, I present the results and analyses only for the highest level of manipulation (75%) across the different architectures. For each scenario, I then trace the effects of initial focal deficits or advantages over the full architecture as development proceeded.

---

[4] Stable states were defined as states in which the rate of change over the entire model was less than an average of 0.001 for 10 timesteps.

For fully distributed, bi-directional, uni-directional and modular architectures, only a single condition was run, since all components are functionally equivalent. For hemispheric and central processor models, I distinguished between *key* processes and *peripheral* processes. For the hemispheric model, the key processes were those that communicated between hemispheres. For central processor, the key process was the central processor itself. In both cases, peripheral processes constituted the remainder. For the hierarchical system, I investigated the consequences of damaging the hierarchy at the lowest, intermediate and highest levels. These distinctions were marked earlier, in Figure 10.1.

Results of Focal Deficit simulations

To assess the effects of early damage, it is necessary to quantify the difference between growth curves in normal and affected systems. For the Focal Deficit studies, I present two metrics for this purpose. The first focuses on the endstate performance level reached by each process. Where this is lower after damage, the system has experienced a *deficit*. The second metric looks at the area under the curve of each process, thereby assessing the trajectory towards the endstate. Where the area is reduced after damage, the system has experienced a *delay*. Both deficit and delay are possible within the same process. Delay is possible without final deficit, but a final deficit is not possible without delay. The focus of these measures is to shed light on the extent to which the overall process of development that operates within each architecture alters the pattern of impairment. Development may ameliorate the deficit in the damaged component via *compensation* from other initially unimpaired components. Development may exaggerate the impact on the wider system via *spread* of deficit to other processes. To assess compensation and spread, I begin by measuring the normal level of performance in each architecture, both in terms of the mean area under the curve for the growth trajectories of its component processes and the endstate levels of the component processes. These values are shown in Table 10-2. The scale of these numbers is to some extent arbitrary. The values merely reflect the amounts of activation cycling around each type of system, and the values will be naturally higher in systems with more interactivity. However, the values serve as a baseline for analysing each kind of architecture and in the Focal Deficit studies proportional changes in the values allow for comparisons between architectures.

Table 10-2. Normal performance for each architecture in terms of the area under the growth curves (representing how long development takes) and endstate levels (indicating final performance)

	Normal Area	Normal Level
Fully distributed	10883	12.0
Hemispheric (peripheral)	4277	4.6
Hemispheric (key)	4488	4.9
Central processor (peripheral)	4756	5.2
Central processor (key)	6552	7.1
Bi-directional loop	3096	3.3
Unidirectional loop	2936	3.2
Hierarchical (beginning)	2792	3.0
Hierarchical (middle)	3703	4.0
Hierarchical (end)	3703	4.0
Fully modular	2792	3.1

To derive a measure of compensation for a damaged component, we need to know what level of performance might be expected from it if the component was damaged and no compensation from other processes was possible. The modular architecture permits consideration of this situation and so generates the predicted pure impairment for a damaged process. The normal system provides information about the performance expected for the process when there is no damage. These two values (predicted normal performance and predicted performance after damage with no compensation) give us the upper and lower bounds against which to gauge actual compensation. Formally, I measure the range of predicted damage (the normal performance N minus the predicted pure impairment P) and evaluate what proportion of that range has been closed by the observed performance A (derived by subtracting the predicted damaged performance P from the actual performance A). This value is expressed as a percentage.

$$\frac{A-P}{N-P} \times 100$$

Equation 7. Normal (N), Predicted pure impairment (P), Actual (A) performance

If the observed performance is fully compensated and therefore appears normal, A=N and Equation 7 yields 100%. If there is no compensation, A=P and Equation 7 yields 0%. Figure 10.9 shows the mean amount of *compensation* that each architecture offered for 75% damage and collapsed across all 7 combinations of parameter change (onset, rate, asymptote and combination of these) for area (delay) and final level (deficit) metrics. First, delay and deficit are tightly coupled. Secondly, the architecture within and the point at which initial damage occurred both have an effect. The fully distributed model offered the greatest degree of compensation to the damaged process on both metrics, with performance around 70% above the level predicted by the damage. Within more differentiated architectures, points of higher connectivity experienced greater compensation than those of lesser, shown in the hemispheric and central processor architecture. Only downstream processes experienced compensation in the hierarchical system, but middle or last position made no difference. By definition, the modular system could experience no compensation.

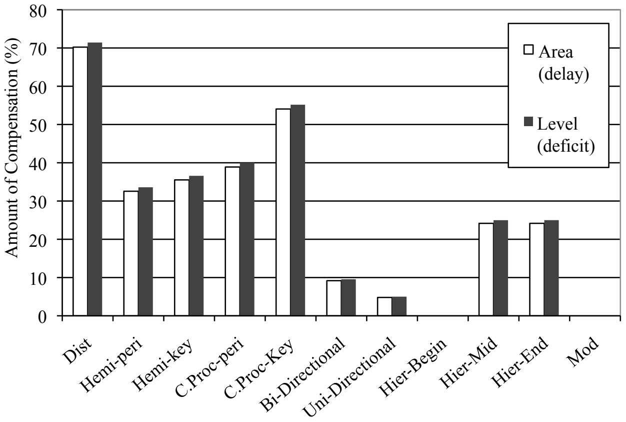


Figure 10.9 Compensation after early process-specific damage for each architecture. *Area* assesses rate of development and *level* measures endstate performance. (Y-axis labels denotes the following: Dist=distributed; Hemi-peri=Hemispheric-peripheral; Hemi-key=Hemispheric-key; C.Proc-peri=Central processor-peripheral; C.Proc-Key=Central processor-key; Bi-Directional=Bi-directional loop; Uni-Directional= Uni-directional loop; Hier-Begin=Hierarchical-beginning; Hier-Mid=Hierarchical-middle; Hier-End=Hierarchical-end; and Mod=Modular.)

The *spread* of deficit is derived by measuring how much performance was delayed for the initially undamaged processes, or reduced at the end of development, as a proportion of normal performance. These values are shown in Figure 10.10. The pattern across architectures more or less mirrors that seen in Figure 10.9. For example, the distributed system that exhibited the most compensation also showed the most spread of deficit, its processes dropping in their final performance by 2.0% and the area reducing by 2.6%, corresponding to slower development. For all cases of spread, delay was more salient than deficit. One point is particularly notable. The degree of spread was much lower than that of compensation. In the 16-process models, the fully distributed system experienced on average 70% compensation for the damaged process but only 2% spread of deficit to initially intact processes.

This differential turned out to be the one result that was sensitive to the number of processes in the model. While the amount of deficit spread stayed roughly constant with changes in process number (at the level observed above), compensation varied from 70% with 16 processes to 51% with 12 processes, to 31% with 8 processes, and 14% with 4 processes. While the impact of a damaged process on the rest of the system depends only on the connectivity, the potential for compensation also depends on the number of contributory processes.

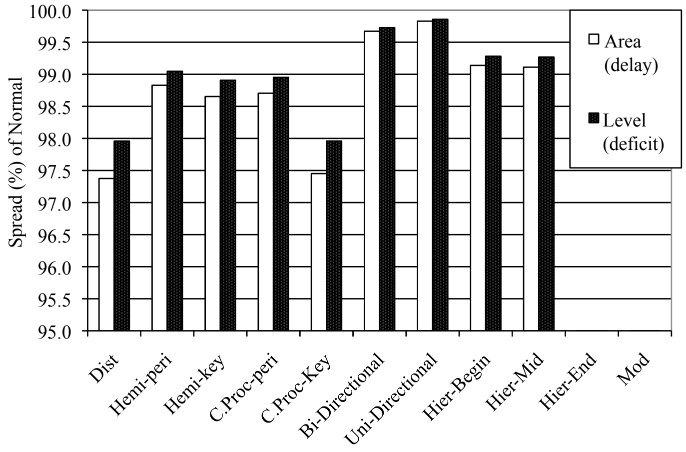


Figure 10.10. Spread of deficit after early process-specific damage, assessed as the proportional decline in performance of initially undamaged processes. Area measures rate of development and level measures endstate performance.

Results of Focal Advantage and Deficit simulations

For these models, the central interest is in determining whether a given architecture supports the savant-like profiles that were depicted earlier in Figure 10.3. In this results section, I present the data for each of the seven architectures introduced previously and again chart the effects of manipulation to each model's parameters at the highest level applied (75%). As was the case in the Focal Deficit models, I discriminate between *peripheral* and *key* processes (where these are relevant; see Figure 10.1) and investigate each separately to examine the effects of targeting each as the *exceptional* process. That is, the single process receiving a surplus in its initial values is the exceptional process, while all other processes that receive a deficit to their initial values are the impaired processes. I assess each model for its fit with the profiles depicted earlier in Figure 10.3. These loosely follow Treffert's (1989) distinction of talented savants (*T-savant*) and prodigious savants (*P-savant*). The following basic conditions are assumed necessary to simulate these profiles. To simulate the T-savant profile: models must exhibit a range of processes that are below average in endstate levels of performance (*other* < *normal*) and exhibit one process with an endstate that is reliably better compared to all the others (*exceptional* > *other*), but which is not reliably higher than normal levels of performance (*exceptional* ≤ *normal*). To simulate the P-savant profile: models must exhibit below average endstate levels of performance in the majority of processes (*other* < *normal*) and exhibit above average endstate performance for one process (*exceptional* > *normal*).

Table 10-3 summarises the *z* score results for each architecture and each parameter manipulation (*a*, *k*, *x*, *ak*, *ax*, *kx*, *akx*). To assess each model's fit for the T-savant profile we must use all three rows: "except. vs. norm", "other vs. norm." and "except. vs. other". Where "except. vs. norm." contains a *z* score that is not greater than 1.60, and where "other vs. norm." shows a *z* score that is negative and greater than -1.60 and where "except. vs. other" is greater than 1.60, then this model can be described as simulating the T-savant profile. To assess each models fit for P-savant profile we must look to the rows labelled: "except. vs. norm" and "other vs. norm". Where "except. vs. norm." contains a *z* score that is positive and greater than 1.60 ($p < 0.05$) and where "other vs. norm." shows a *z* score that is negative and greater than

-1.60 for “other vs. norm.”, then this model can be described as simulating the P-savant profile.

On the basis of these conditions, Table 10-3 shows: (1) T-savant profiles emerged in Central Processor-Peripheral and Central Processor-Key models; and (2) P-savant profiles in Hemispheric-Peripheral, Hemispheric-Key, Bi-directional, Uni-directional, Hierarchical-Beginning / Middle / End and Modular models. The table also shows these profiles are consistently in conditions relating to manipulation of *K*, the asymptote levels.

Table 10-3. Z-score results comparing: (1) the performance of the exceptional process in the combined Advantage and Deficit model to its counter-part process in the normal model; (2) the performance of the other impaired processes in the savant model to their counter-part processes in the normal model; (3) the performance of the exceptional process compared to the performance of the impaired processes in the savant model.

Architecture	**Except Process**		**A**	**K**	**X**	**AK**	**AX**	**KX**	**AKX**
		except. vs. norm	-0.60	-6.29	-0.04	-6.33	-0.74	-6.49	-6.41
Fully Dist	Peripheral	*other vs. norm*	-0.90	-12.13	-0.02	-12.33	-0.85	-12.16	-12.38
		except. vs other	1.32	25.63	-0.12	26.24	0.64	25.01	26.15
			--	--	--	--	--	--	--
		except. vs. norm	0.08	2.26	-0.07	2.57	-0.13	2.40	2.48
Hemispheric	Peripheral	*other vs. norm*	-0.48	-6.42	-0.01	-6.59	-0.53	-6.43	-6.60
		except. vs other	1.66	33.85	-0.82	35.75	1.05	34.41	35.42
			--	P	--	P	--	P	P
		except. vs. norm	-0.21	1.96	-0.11	1.83	-0.19	1.88	1.90
Hemispheric	Key	*other vs. norm*	-0.52	-6.33	0.00	-6.52	-0.52	-6.35	-6.51
		except. vs other	3.13	34.80	1.49	35.05	3.21	34.57	35.31
			--	P	--	P	--	P	P
		except. vs. norm	-0.39	1.32	-0.06	1.13	-0.22	1.45	1.23
Central Proc	Peripheral	*other vs. norm*	-0.52	-7.16	0.07	-7.36	-0.47	-7.16	-7.35
		except. vs other	0.43	35.04	-0.59	35.04	0.95	35.57	35.42
			--	T	--	T	--	T	T
		except. vs. norm	-0.31	-0.95	0.09	-1.02	-0.17	-0.77	-0.75
Central Proc	Key	*other vs. norm*	-0.48	-6.72	0.00	-6.88	-0.53	-6.71	-6.85
		except. vs other	15.79	38.55	15.65	38.91	16.65	39.35	40.03
			--	T	--	T	--	T	T
		except. vs. norm	0.02	3.90	-0.07	4.27	0.00	4.10	4.65
Bi-Directional	Peripheral	*other vs. norm*	-0.40	-4.87	0.00	-5.03	-0.38	-4.89	-5.02
		except. vs other	1.97	34.32	0.05	36.29	1.79	35.12	37.69
			--	P	--	P	--	P	P
		except. vs. norm	0.07	3.78	0.13	4.18	0.06	4.16	4.37
Uni-Directional	Peripheral	*other vs. norm*	-0.32	-4.64	-0.01	-4.79	-0.35	-4.65	-4.79
		except. vs other	1.35	33.88	0.33	36.10	1.41	35.45	36.88
			--	P	--	P	--	P	P
		except. vs. norm	-0.04	3.84	0.00	4.23	0.04	3.98	4.28
Hierarchical	Begin	*other vs. norm*	-0.44	-5.74	-0.04	-5.90	-0.47	-5.74	-5.91
		except. vs other	-5.67	32.04	-7.15	34.36	-5.26	32.65	34.62
			--	P	--	P	--	P	P
		except. vs. norm	0.07	3.20	0.10	3.37	0.06	3.13	3.30
Hierarchical	Middle	*other vs. norm*	-0.46	-5.60	-0.05	-5.76	-0.51	-5.61	-5.77
		except. vs other	2.43	34.82	0.91	36.10	2.55	34.54	35.85
			--	P	--	P	--	P	P
		except. vs. norm	-0.14	2.65	-0.20	2.85	-0.15	2.58	2.81
Hierarchical	End	*other vs. norm*	-0.46	-5.81	-0.04	-5.98	-0.46	-5.83	-5.98
		except. vs other	2.21	35.12	0.32	36.58	2.21	34.87	36.44
			--	P	--	P	--	P	P
		except. vs. norm	0.03	4.36	0.07	4.76	0.06	4.23	4.66
Modular	Peripheral	*other vs. norm*	-0.34	-4.69	0.00	-4.86	-0.38	-4.70	-4.85
		except. vs other	1.41	36.85	0.22	39.15	1.68	36.36	38.71
			--	P	--	P	--	P	P

T=T-savant profile
P=P-savant profile
-- Denotes no match to either T-svant or P-savant profles

Discussion

Cases of uneven cognitive profiles in development have presented something of a challenge for developmental theorists (see e.g., Bullinaria, 2007). Modular accounts have argued the view that uneven cognitive profiles provide evidence of the developmental independence and isolation of processes within the cognitive system. However, many researchers have questioned the degree of supposed independence between cognitive domains, especially in early development (Thomas, et al., 2002). A view persists that cognitive processes and the brain are highly interactive during early years. The lack of integration between these approaches has led to doubts over how cases of uneven profiles might emerge under different views of the cognitive system, and just how *specific* deficits can be.

The main focus within this chapter was to investigate these issues by studying the causal influences of initial disparities within the starting state of a system and then charting the effects of that change over development in a number of different cognitive architectures. Using a dynamical systems framework to model multi-component systems, this chapter offers a first step towards reconciling seemingly opposite views that exist in the literature on uneven cognitive profiles in development. Importantly, as the models demonstrate, it is not necessary to assume a modular architecture in order to account for patterns of uneven cognitive profiles. Rather, a diverse range of underlying architectures is capable of exhibiting such unevenness.

For example, the results of the Focal Deficit simulations showed that a range of architectures was able to simulate uneven cognitive profiles. This was often possible following only a single manipulation to the asymptote level, with more serious consequences following damage to combinations of the parameters including asymptote. The results of the Focal Advantage and Deficit simulations likewise showed that a range of architectures could produce uneven cognitive profiles. These shared some similarity with the distinctions Treffert made between the two types of savant profile (i.e., Treffert, 1989). Again, manipulations to the asymptote levels proved to be a key factor in producing these patterns.

At a more detailed analysis, the simulation work revealed at least three findings. These were: (a) the density of connectivity at the point of damage, as well as positioning in hierarchical systems, is influential in determining spread and compensation; (b) the number of processes interacting to generate a behaviour affects compensation but not spread; and (c) damage to growth curve asymptotes (the developmental equivalent of the capacity of a process) is more serious than damage to its rate (equivalent to plasticity). However, combining the results from the two simulation approaches raises a number of questions. For example, within both the Focal Deficit and Focal Advantage and Deficit simulations, why were manipulations to asymptote so influential while initial changes to growth rate and to onset had so little end impact? In the Focal Advantage and Deficit simulations, why did some of the models show only talented savant profiles while others showed only prodigious savant profiles? Why did we not find architectures capable of showing both talented *and* prodigious profiles (say for example, in manipulations to peripheral vs. key processes)?

One answer to the question of why manipulations to asymptote had such a great effect and why manipulations to growth and onset had less effect is simply that models were not constrained in the amount of time that was available to them. That is to say, slowing the growth rate of a process simply lengthened the time that was needed for that process to get to a normal level. Using an analogy to illustrate this, two computers with different clock speeds (but which are otherwise identical) are each capable of running similar applications – one will just take slightly longer. Thus, in the absence of any further constraints, no deficits in endstates are possible given changes to growth alone. Between growth and onset, growth showed the most serious consequences following change. This was because once the change was made to growth rate of a process, that value was fixed and remained at this value over the course of development. Changing the onset (or starting level of a process), however, had severe effects initially, but its effects were limited to early stages as the level of that process changed over development.

Considering two earlier questions: (1) why in the Focal Advantage and Deficit studies did some architectures display one profile and others not; (2) why different profiles were not found in the same model (e.g., as a result of manipulations to peripheral vs. key processes)? I propose that a comparison between the hemispheric and central processor models may hold part of the answer. Looking at the organisation of these two architectures it can be seen that they are largely similar in their design (see earlier illustrations of hemispheric and central processor models in

Figure 10.1C and Figure 10.1D, respectively). In the simulations, both hemispheric and central processor models comprise two clusters of 8 processes that are fully connected within each cluster. In the case of the hemispheric model, one process in each cluster is connected to the other and thus can spread its influence (and the combined influence of all processes within that cluster) throughout the opposite cluster. Additionally, in the case of the central processor model, all processes within each cluster are connected to a 17th 'central' process. Comparing the outcomes of changes in these two architectures shows that the hemispheric model achieves the more extreme P-savant profile, but the central processor model produces only the T-savant profile. This suggests that connectivity is a critical factor in producing the different profiles. In this case, the point of sensitivity between T-savant vs. P-savant profiles was around 8 processes. Thus, in architectures where the degree of connectivity is lower than 8, P-savant profiles should be found. For the other architectures, the results showed this also to be the case. The P-savant profile was an outcome in the bi-directional (where connectivity around an exceptional process was limited to 2), uni-directional (connected to 1 other process), hierarchical beginning (connected to no others), hierarchical middle (directly affected by previous, with previous process experiencing graded effects of 6 others), hierarchical end (directly affected by previous process, but with previous process experiencing graded effects of 14 others), and modular (no connectivity) models.

Although modular accounts do not offer the only framework for explaining cases of uneven profiles in development, it remains a widely accepted idea that modularity plays an important part in later development. For example, Elman (1996) has focused on demonstrating how emergent forms of behaviour may be constrained by genes, although not in a domain-specific way to produce specialised learning systems. In a similar vein, while Karmiloff-Smith and Thomas (Karmiloff-Smith & Thomas, 2003) have suggested that cases of adult brain damage may offer reasonable evidence of modularity in adults, they have argued against assuming modularity as a starting point. Karmiloff-Smith (see e.g., Karmiloff-Smith, 1995) has additionally argued that modularity emerges during development, as a process of *progressive modularisation*, comprising interactions at many different levels (e.g., genetic, environmental, neurophysiological).

The simulation work contained here does not include a role for emergent modularity and specialisation. Next, I briefly outline a research strategy for extending this work in order to explore how modularity might emerge within different systems comprising multiple components.

Future work

Keeping within the current dynamical systems framework, the most obvious starting point for exploring how any kind of emergent modularity could be produced would be to investigate changes to the matrix specifying each architecture's connectivity (recall, this was handled by the M matrix, in Equation 6). In the models described here, M was fixed and so the architecture remained fixed over development. One extension of the current work, then, would be to investigate how changing M (e.g., from fully distributed to a more modular design) would change the pattern of profiles described so far. One way of achieving this would be to introduce a simple decay algorithm into the M matrix. In this way, over time the influence of each process on all other processes would become smaller thus resulting in an encapsulated and modular organisation. Another approach, more closely linked to evidence of actual neurophysiological processes, would be to make reductions in the connections between processes dependent on interactions between *real learning systems*. In its most complete form, this could involve merging connectionist approaches within the current dynamical systems framework. Each process might then be implemented as a neural network with a variety of problems given to each system to learn so that processes that fire together, wire together (Hebb, 1949). In this way, problems of similar classes may then group to form clusters of related processes (see Thomas & Richardson, 2006, for work on the computational principles guiding the emergence of modules in learning systems).

Of relevance to this suggested work, can be found within research into ecosystems (Kondoh, 2003) and more recently (Uchida & Drossel, 2007). Here, the influence of adaptation between populations is under active exploration using the Lotka-Volterra model (the model from which the mutualism model is derived). Thus, there may already exist pointers as to how an adaptive version of M may be implemented in psychological theory.

In sum, the present chapter demonstrates the utility of computational approaches in the study of development. The same ability to trace the consequences of specific and small-scale changes within large multi-component interactive cognitive system is not amenable using other approaches that focus on a particular domain of learning using individual models (for instance using models of the Stroop task, or the balance scale task). Although the work in this chapter is only a small step towards identifying the functional architecture of the cognitive system, it is a step in the direction of a potentially important research avenue for examining the causes of uneven cognitive profiles in development. Future work might aim to clarify the precise architectures is representative of the cognitive system.

***

In the next (and last) chapter, I present a final discussion to summarise the findings of this research project. I evaluate the success of the research aims in investigating the relationship between intelligence and cognitive development and point to some future avenues of work that are raised.

Part 7

Chapter 11 Final Discussion

Restating the aims of the thesis

This thesis set out to explore the dimensionality of the relationship underlying differences in intelligence and differences in cognitive development. At the beginning of this thesis the following question was posed: *Is intelligence like having 'a little more' cognitive development?*

The results presented within this thesis suggest that the answer to this question is, broadly, '*Yes*'. The findings from the range of tests administered to the Primary and Secondary groups suggest that intelligence and cognitive development are largely overlapping forms of cognitive variability. That is, when matched on mental age neither the Primary YHA and OLA groups, nor the Secondary YHA and OLA groups showed the kinds of differences in cognitive profiles that one might have predicted given Spitz (1982). The degree of similarity between groups of MA-matched children at both school levels suggest that intelligence *is* like having a little more cognitive development. Overall, experimental manipulations caused various modulations in behaviour, but these modulations were largely the same in the older and younger MA-matched groups. Thus, the findings are largely consistent with unidimensional views. At a more detailed level, however, there were a number of subtle differences that stood out within the results and that were suggestive of variability between groups in their information-processing. Next, I summarise the key findings of this thesis.

Summary of key findings

The results of the BAS II (Chapter 4) showed that Primary YHA and OLA were no different on the Core scales, but were different on only one of the Diagnostics: Speed of Processing. The continuous analysis showed that the speed of processing differences was predicted by one's age and not one's advantage (MA-CA disparity). Interestingly, Digits Backward was predicted by advantage. Contrasting these two findings simplistically, suggests that speed of processing changes over age and working memory is modulated by intelligence. At Secondary School level, the YHA and OLA were indistinguishable. This is the first indication that the answer to our question may depend on *age.*

On the Stroop task (Chapter 5), the task was successful in manipulating interference on incongruent colour-naming trials and the Primary OLA was reliably faster than the YHA at responding. By contrast there was no speed difference at the Secondary School level. However, groups showed the same patterns of interference at both Primary and Secondary levels.

On the primed lexical decision task (Chapter 6), the Primary OLA was faster than YHA but the groups were no different in the size of priming effect. However, at Secondary level, the YHA and OLA groups did differ in the size of priming, with the YHA showing the greater effect.

On the conservation of number and liquid tasks (Chapter 7), a restricted range of problems led to poor sensitivity of the experimental measures. Secondary groups were at ceiling. However, the Primary OLA was more accurate at conserving than the YHA, indicating a possible advantage of experience over intelligence.

On the Balance Scale task (Chapter 8) problem difficulty had largely uniform effect. However, the Primary YHA showed superior accuracy on the Conflict Balance problems. The OLA were, by contrast, uniformly faster than the YHA. At Secondary level, once more there were no differences between the YHA and OLA groups.

On the Tower of London task (Chapter 9), once more problem complexity had large effects that were uniform across groups. However, with respect to the hardest 6-move flat-ending problems, the Primary YHA were more accurate than the OLA, yet took longer to solve these problems. This was not apparently to be due to a speed-accuracy trade-off because the YHA were equally slow on problems where they were also less accurate, compared to the OLA. In the Secondary level analysis, the YHA and OLA as before did not differ.

In the chapter using computational models (Chapter 10), the emphasis was on examining the conditions under which uneven cognitive profiles may emerge in development. Several different cognitive architectures were distinguished using formal mathematical methods. This chapter highlighted the utility of computational approaches in determining the causal influence of small changes to the initial state of a system. These models allowed different types of deficit to be distinguished and the consequences of these deficits to be quantitatively charted over time. This chapter showed that a range of cognitive architectures was able to produce uneven profiles and that in some instances these could be differentiated by the pattern of spread and compensation of early disparities within the cognitive system.

Why the disparity with Spitz (1982)?

Spitz's argument that intelligence and cognitive development are *not* equivalent was based on Merrill's (1924) findings in which clear differences were shown in the cognitive profiles of younger and older children who were matched on mental age. For example, Spitz showed that the younger children demonstrated superior performance to the older children on tasks tapping verbal abilities and abstract reasoning, and conversely that the older children were superior to the younger children on tasks tapping maturation and experience. This pattern of relative strengths and weaknesses between younger and older MA-matched groups was not replicated in this thesis. One possible explanation for the disparity between Spitz's results and those reported here might be due to differences in the samples used. For example, the range Spitz reported between the younger and older children's mental ages given their chronological ages was greater than the range obtained in this study. By comparison, Spitz's groups exhibited more extreme forms of variability in intelligence. The express aim of this research project, however, was to sample children from the normal range. Thus, children who had been classified with special educational needs were purposefully not included. It is possible that in Merrill's more extreme groups, there were children who would be classified as having learning difficulties. In IQ terms, Merrill's older lower ability groups scored 67 (see Chapter 4 'Discussion', for more detail). These differences may explain the disparity between Spitz's findings and the findings presented in this thesis. Additionally, while effect sizes for each of the computerised tasks showed the opportunity was present for groups to differ, the overall power of tests was low given the small sample size (i.e., 14 children in both the Primary School YHA and OLA groups, once matched for MA). To increase the power of this test to 0.80 or better (and for the same effect size) one would need to test a total of 50 children (i.e., 25 per group).

Another possible explanation for the disparity between Spitz's findings and those presented here concerns the difference in standardised tests that were used.

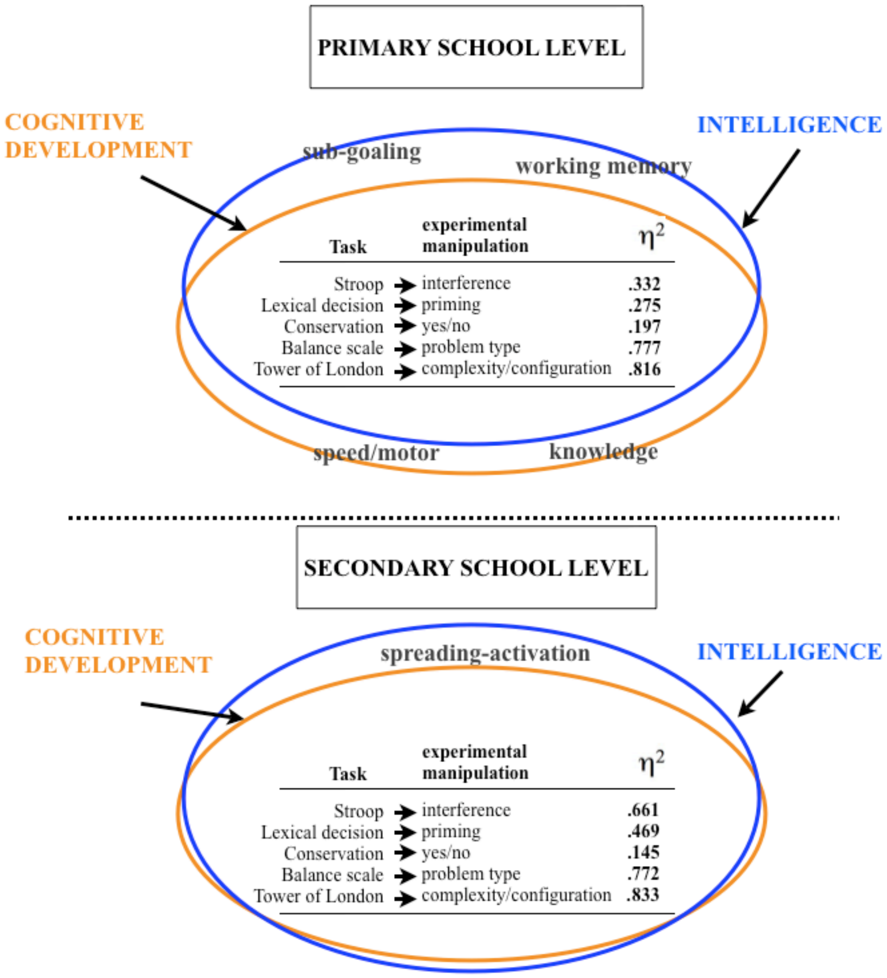
Recall that Merrill used the earliest version of the Stanford-Binet (1916) and that this battery has been claimed to suffer limitations due to a high verbal loading (see e.g., Becker, 2003). Furthermore, whereas in the Stanford-Binet there were tasks designed to tap maturational advantage (e.g., ‘date’ and ‘tying one’s shoes’), there do not seem to be the equivalent in the BAS II (Elliot, et al., 1997). Thus, it is possible that the tasks included in the BAS II might not have been ideal for providing the opportunity of distinguishing the advantages of age versus ability. With this same point in mind, we may ask: ‘Did the computer-based cognitive tasks allow the opportunity for groups to differ?’ Next, I address this question using two Venn diagrams that describe the extent to which experimental manipulations provided the opportunity for groups to differ.

The overlap between intelligence and cognitive development

The overlap between groups of younger and older children who were matched on mental age at both Primary and Secondary School levels was striking. While there was a broad degree of overlap at Primary School level, there was an even greater overlap at Secondary School level. These findings are characterised as two overlapping areas within Figure 11.1. In this figure, I illustrate the patterns of overlap within the Primary (Figure 11.1 Top) and Secondary (Figure 11.1 Bottom) School levels. In both Venn diagrams, I use orange circles to depict cognitive development and blue circles to depict intelligence. Within the shared areas in both diagrams, a table provides the effect sizes of the main experimental manipulations within each computer-based task. As these tables show, at both Primary School and Secondary School levels (with the exception of the conservation tasks), the battery of computer-based tasks provided ample opportunity for the younger and older groups to diverge in their performance. That is, the experimental manipulations within the Stroop task (interference), the lexical decision task (priming effect), the balance scale task (problem type) and the Tower of London task (complexity), were successful in influencing overall performance. Thus, given those tasks where the younger and older groups did not show differences in response to these manipulations, this suggests that the influences of intelligence and cognitive development are largely the same.

In the Venn diagram corresponding to the Primary School level (Figure 11.1 Top), in the non-overlapping areas, ‘working memory’, ‘sub-goaling’, ‘knowledge’ and ‘speed/motor’ have been positioned in the non-overlapping areas. This reflects the following results: (1) ‘working memory’ indexes the fact that the results of the Digits Backward subtest in the BAS II showed MA-CA disparity was a reliable predictor of performance; (2) ‘sub-goaling’ indexes the fact that on the Tower of London task the YHA were more accurate than the OLA on the hardest problem type; (3) ‘Knowledge’ indexes the empirical effect that the OLA were better on conservation tasks; and (4) ‘speed/motor’ indexes the fact the OLA generated faster response in all speeded tasks and on the Speed of Processing (BAS II).

In the diagram corresponding to the Secondary School level (Figure 11.1 Bottom), the overlap between intelligence and cognitive development is now greater. None of the results suggested unique influences of cognitive development. However, in the area unique to intelligence, Figure 11.1 Bottom shows ‘spreading-activation’. This stems from the larger priming effect that was found in the YHA group on the lexical decision task. The absence of such an effect at Primary level and the presence of this effect at Secondary level suggests, that at older ages, the spread of activation through cognitive systems is greater depending on one’s intelligence.



PRIMARY SCHOOL LEVEL

Task	experimental manipulation	η^2
Stroop	interference	.332
Lexical decision	priming	.275
Conservation	yes/no	.197
Balance scale	problem type	.777
Tower of London	complexity/configuration	.816

SECONDARY SCHOOL LEVEL

Task	experimental manipulation	η^2
Stroop	interference	.661
Lexical decision	priming	.469
Conservation	yes/no	.145
Balance scale	problem type	.772
Tower of London	complexity/configuration	.833

Figure 11.1. Venn diagrams depicting the possible overlap between intelligence (blue circles) and cognitive development (orange circles), at Primary School and Secondary School levels. Mina effects of experimental manipulations are shown in the centre of each Venn diagram. See text for details.

Figure 11.1 illustrates that at younger ages intelligence and cognitive development are broadly similar and that over time the amount of overlap increases. Two other explanations for the increase in overlap between Primary and Secondary levels may be considered. Firstly, might the increase be due to differences in the samples? For example, does the increase in overlap arise due to the smaller CA disparity between the two groups at Secondary versus Primary level? One way to answer this question would be to look at the predictive power of MA-CA. That is, if MA-CA is found to be predictive of performance on a cognitive task, then it is picking up variance in ability at those ages and would therefore demonstrate the *opportunity* for children at that age to differ. This was indeed found to be the case. At the Secondary School level, MA-CA disparity was found to be predictive of performance on the lexical decision task, the balance scale task and the Tower of London task. Thus, it does not seem that the closeness of the group's CAs accounts for the increase in overlap between Primary and Secondary Scool levels.

Secondly, might the source of the increased overlap at Secondary School level be due to a lack of sensitivity of the tests? For example, does the increase in overlap arise due to the inability of the tests to distinguish groups at this school level? Here, again, this does not seem to be the case. Inspection of the effect sizes for each of the main experimental manipulations showed that the tasks provided ample opportunity for children to differ.

I offer the view that the profiles of abilities that were obtained for younger more able and older less able children were accurate reflections of their current intellectual *and* developmental levels. However, the smaller degree of overlap in the Primary School data also fits with the sense that at those ages there was more room for the effects of greater intelligence or greater development to be revealed. That is, it is possible that whatever intelligence lends to development and whatever development lends to intelligence is most apparent during earlier years. But, that by the time children get to adolescence they have obtained most of skills or abilities they need (one way or another) to perform similarly on cognitive tests. One way of exploring this argument further would be to replicate the design with Secondary School children where there is a greater CA disparity (e.g., a 4-year age disparity).

Theoretical relevance of this research

The interpretation that having more intelligence is broadly equivalent to having more cognitive development, and vice versa, has direct relevance to a number of research areas. For instance, notions of unidimensionality extend beyond the confines of accounting for variability within the normal range. It is also found within numerous accounts of developmental disorders where performance of disorder groups are portrayed as extreme scores on the normal continuum (see e.g., Kovas, et al., 2007). While a unidimensional account may offer the advantage of parsimony in explaining the variability in intelligence between people of the same age, people of different ages and between control and disordered groups, empirical research to test the question of whether these differences represent variation on the same dimension, has largely gone unaddressed (see e.g., Thomas & Karmiloff-Smith, 2003).

At a broader level, we might ask how these findings relate to wider fields of research. For example, should intelligence researchers be interested in the findings from developmental research and should developmental researchers be interested in the findings from intelligence research? Additionally, within some fields of research, individual difference methods are used to address developmental questions concerning the origins of variability (behavioural genetics being one example). In both these cases, it seems the answer is crucial as to whether the relationship between intelligence and cognitive development is unidimensional.

Practical relevance of this research

Although the primary focus within this thesis has not been to address practical issues of intelligence and education, nevertheless clearly its findings are relevant for instance to educational techniques. With respect to the Secondary School in Portsmouth that pioneered the ‘Ability not Age’ approach, this research suggests that at the Secondary School level, teaching classes of children based on their overall ability is a viable option, at least with respect to cognitive ability. The ‘Ability not Age’ approach, which is aimed to give schools more freedom to accelerate the learning of more able students while providing greater assistance to lower ability children, may show in time to be an effective method of improving the quality of children’s learning.

At the Primary School level the source of the differences on the Speed of Processing task remains unclear. It is possible that these differences reflect either a difference in information-processing and/or a difference in motor control. This is a pressing question that we turn to shortly. However, even if the Speed of Processing differences stem from nothing more than differences in basic motor control, this could still cause practical problems in accommodating children with different speeds on the same tasks. For example, this might be that tasks in class may be finished sooner by older children, leaving younger children to catch up. Consequently, it is of some concern that in the UK educational advisors have recently advocated the ‘Ability not Age’ approach should be adopted within the next 5-years across Primary and Secondary level. However, the source of the evidence underlying the viability of this proposal is not clear. For example, as one newspaper journalist pointed out, such a system “could result in primary school children regularly starting secondary classes at the age of nine – rather than 11 – and pupils taking their GCSEs at 14" (Paton, 2008).

Of course, this says nothing on the question of whether those children sitting their GCSEs at age 14 are also emotionally, or physically mature enough to enter A-levels, University and careers at an earlier stage also.

Future avenues of research

One avenue of research to explore following this research would be to use targeted measures to focus on the areas where the Venn diagram did not show an overlap at the Primary and Secondary School levels. An example is the speed of response and Speed of Processing (BAS II) differences that were highlighted. One straightforward way to test this is to repeat the design comparing inspection time with response times.

A second strand of research might aim to extend the design to include the study of older individuals in their late teens and adulthood. One question this work might address is where in development ability levels begin to asymptote. If cognitive development increases during childhood and adolescence, but asymptotes in early adulthood, a unidimensional account would argue intelligence would also asymptote then. This would predict that intelligence has no predictive power in adulthood. This proposition seems counter-intuitive. Thus, future work looking at the relationship between intelligence and cognitive development in later years might shed light on this issue. However, one obstacle in replicating the design described here, is that standardised tests of ability show less differentiation on tests in adulthood. Other tests would have to be found.

A third area that future work might address, is the sources of the subtle disparities as well as the larger commonalities found in this thesis, using formal computational models. Recall, that each cognitive task was selected because a normative model already exists. Therefore these models provide convenient frameworks for exploring the neurocomputational basis of variability.

Conclusion

There is, to some extent, a sense that educational research and research relating to development and intelligence exist in their own spheres and without enough integration between these fields of the key issues, or driving interests. A clear case for this is the recent proposals to adopt the ‘Ability not Age’ approach in Primary and Secondary School across the UK. This is surprising given there does not appear to be a body of empirical evidence to support the idea that this will work. In conclusion, this research provides further impetus for a greater integration of research between educational, developmental and individual differences approaches and for a re-evaluation of assumptions concerning the relative independence of individual differences and cognitive development approaches. These fields of study may share more than initially has been supposed.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working Memory and Intelligence: The Same or Different Constructs? *Psychological Bulletin, 131*(1), 30-60.

Adolph, K. (1997). Learning in the Development of Infant Locomotion. *Monographs of the Society for Research in Child Development, 62*(3).

Alibali, M. (1999). How Children Change Their Minds: Strategy Change Can Be Gradual or Abrupt. *Developmental Psychology, 35*, 127-145.

Altmann, E., & Davidson, D. (2001). *An integrative approach to Stroop: Combining a language model and a unified cognitive theory*.

Altmann, E., & Trafton, J. (1999). *Memory for goals: An architectural perspective.* Paper presented at the Cognitive Science, Vancouver.

Anastasi, A. (1992). What counselors should know about the use and interpretation of psychological tests. *Journal of Counseling & Development, 70*(5), 610-615.

Anderson, J. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 261-295.

Anderson, M. (1988). Inspection time, information processing and the development of intelligence. *British Journal of Developmental Psychology, 6*(1), 43-57.

Anderson, M. (1992). *Intelligence and development: A cognitive theory*: Blackwell Publishing.

Anderson, M. (1998). Mental retardation general intelligence and modularity. *Learning and Individual Differences, 10*(3), 159-178.

Anderson, M. (2001). Conceptions of intelligence. *Journal of Child Psychology and Psychiatry, 42*(3), 287-298.

Anderson, M., & Miller, K. L. (1998). Modularity, mental retardation and speed of processing. *Developmental Science, 1*(2), 239-245.

Anderson, M., Reid, C., & Nelson, J. (2001). Developmental changes in inspection time: what a difference a year makes. *Intelligence, 29*(6), 475-486.

Anderson, P., Anderson, V., & Lajoie, G. (1996). The Tower of London Test: Validation and standardization for pediatric populations. *Clinical Neuropsychologist, 10*(1), 54-65.

Andreasen, N. C., Flaum, M., Swayze, V. d., O'Leary, D. S., Alliger, R., Cohen, G., Ehrhardt, J., & Yuh, W. T. (1993). Intelligence and brain structure in normal individuals. *American Journal of Psychiatry, 150*(1), 130-134.

Asato, M. R., Sweeney, J. A., & Luna, B. (2006). Cognitive processes in the development of TOL performance. *Neuropsychologia, 44*(12), 2259-2269.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychological Review, 113*(3), 628-647.

Baughman, F. D., Anderson, M., Reid, C., & Thomas, M. S. C. (in prep). Speed of processing differences between MA-matched groups of younger and older children.

Baughman, F. D., & Cooper, R. P. (2007). Inhibition and young children’s performance on the Tower of London task. *Cognitive Systems Research, 8*(3), 216-226.

Becker, K. (2003). History of the Stanford-Binet Intelligence Scales: Content and Psychometrics. *Nelson, SB5 Assessment Service Bulletin*(2).

Benoit, K. E., McNally, R. J., Rapee, R. M., Gamble, A. L., & Wiseman, A. L. (2007). Processing of emotional faces in children and adolescents with anxiety disorders. *Behaviour Change, 24*(4), 183-194.

Binet, A., & Simon, T. (1905). The development of intelligence in children. *L'Année Psychologique, 12*, 191-244.

Bland, J., & Altman, D. (1994). Regression towards the mean. *BMJ: British Medical Journal, 308*(6942), 1499.

Blewitt, P., & Toppino, T. (1991). The Development of Taxonomic Structure in Lexical Memory. *Journal of Experimental Child Psychology, 51*(2), 296-319.

Bocková, M., Chládek, J., Jurák, P., Halámek, J., & Rektor, I. (2007). Executive functions processed in the frontal and lateral temporal cortices: Intracerebral study. *Clinical Neurophysiology, 118*(12), 2625-2636.

Boring, E. (1923). Intelligence as the test tests it. *New Republic, 6*, 35-37.

Brainerd, C. J. (1974). Postmortem on judgments, explanations, and Piagetian cognitive structures. *Psychological Bulletin, 81*(1), 70-71.

Braver, T. S., Ruge, H., Cabeza, R., & Kingstone, A. (2006). Functional neuroimaging of executive functions *Handbook of functional neuroimaging of cognition (2nd ed.).* (pp. 307-348): MIT Press.

Brekke, B. (1976). Conservation of Weight With the Gifted. *Journal of Genetic Psychology, 129*(2), 179-184.

Bub, D. N., Masson, M. E. J., & Lalonde, C. E. (2006). Cognitive control in children: stroop interference and suppression of word reading. *Psychological Science, 17*(4), 351-357.

Bull, R., Espy, K. A., & Senn, T. E. (2004). A comparison of performance on the Towers of London and Hanoi in young children. *Journal of Child Psychological Psychiatry, 45*(4), 743-754.

Bullinaria, J. A. (2007). Understanding the emergence of modularity in neural systems. *Cognitive Science: A Multidisciplinary Journal, 31*(4), 673-695.

Burns, N. R., Nettelbeck, T., & Cooper, C. J. (1999). Inspection time correlates with general speed of processing but not with fluid ability. *Intelligence, 27*(1), 37-44.

Butterworth, B. (2005). Developmental dyscalculia. *Handbook of Mathematical Cognition*, 455–467.

Carder, H., Handley, S., & Perfect, T. (2004). Deconstructing the Tower of London: Alternative moves and conflict resolution as predictors of task performance. *The Quarterly Journal of Experimental Psychology Section A, 57*(8), 1459-1483.

Carroll, J., Davies, P., & Richman, B. (1971). *The American Heritage Word Frequency Book*: Houghton Mifflin.

Carson, S., Peterson, J., & Higgins, D. (2003). Decreased latent inhibition is associated with increased creative achievement in high-functioning individuals. *Journal of Personality and Social Psychology, 85*(3), 499-506.

Case, R. (1992). Neo-Piagetian theories of child development. *Intellectual development*, 161-196.

Case, R. (1998). The development of conceptual structures. *Handbook of child psychology, 2*, 745–800.

Cepeda, N., Kramer, A., & De Sather, J. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology, 37*(5), 715-729.

Chen, E. Y. H., Wong, A. W. S., Chen, R. Y. L., & Au, J. W. Y. (2001). Stroop interference and facilitation effects in first-episode schizophrenic patients. *Schizophrenia Research, 48*(1), 29-44.

Chen, Z., & Siegler, R. (2000). *Across the Great Divide: Bridging the Gap Between Understanding of Toddlers' and Older Children's Thinking*: Blackwell Publishers.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review, 97*(3), 332-361.

Cohen, R. (1985). *The Development of Spatial Cognition*: Lawrence Erlbaum Associates.

Cooper, R. (2002). *Modelling High-Level Cognitive Processes*: Lawrence Erlbaum Associates.

Cox, W. M., Fadardi, J. S., & Pothos, E. M. (2006). The Addiction-Stroop Test: Theoretical Considerations and Procedural Recommendations. *Psychological Bulletin, 132*(3), 443-476.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: simulating semantic priming. *Cognitive Science: A Multidisciplinary Journal, 23*(3), 371 - 414.

Damasio, A., & Anderson, S. (1985). The frontal lobes. *Clinical neuropsychology*, 339-375.

Davis, H., & Anderson, M. (1999). Individual differences and development--One dimension or two? *The development of intelligence.* (pp. 161-191): Psychology Press/Taylor & Francis (UK).

Davis, H., & Anderson, M. (2001). Developmental and individual differences in fluid intelligence: Evidence against the unidimensional hypothesis. *British Journal of Developmental Psychology, 19*(2), 181-206.

Dawson, M. R. W., & Zimmerman, C. (2003). Interpreting the Internal Structure of a Connectionist Model of the Balance Scale Task. *Brain and Mind, 4*(2), 129-149.

Deary, I. J. (2000). *Looking down on human intelligence*: Oxford University Press New York.

Deary, I. J., & Caryl, P. G. (1997). Neuroscience and human intelligence differences. *Trends in Neurosciences, 20*(8), 365-371.

Dempster, F. N. (1991). Inhibitory Processes: A Neglected Dimension of Intelligence. *Intelligence, 15*(2), 157-173.

Detterman, D., & Daniel, M. (1989). Correlations of Mental Tests with Each Other and with Cognitive Variables Are Highest for Low IQ Groups. *Intelligence, 13*(4), 349-359.

DeVries, R. (1971). Evaluation of Cognitive Development with Piaget-Type Tests: Study of Young Bright, Average, and Retarded Children. Final Report.

Diamond, A. (2006). The early development of executive functions. *Lifespan cognition: Mechanisms of change*, 70ñ95.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the Frontal Lobe: The Organization of Goal-Directed Behavior. *Cognitive Psychology, 30*(3), 257-303.

Edwards, J., & Lahey, M. (1996). Auditory lexical decisions of children with specific language impairment. *Journal of Speech and Hearing Research, 39*(6), 1263.

Elliot, C., Smith, P., & McCulloch, K. (1997). *The British Ability Scales II.* London: nferNelson.

Elman, J. L., Bates, E. A., Johnson, M. H., & Karmiloff-Smith, A. (1996). *Rethinking innateness: A connectionist perspective on development*: The MIT Press.

Feldman, D. H., & Fowler, R. C. (1997). The nature (s) of developmental change: Piaget, Vygotsky, and the transition process. *New Ideas in Psychology, 15*(3), 195-210.

Fischer, K., & Silvern, L. (1985). Stages and individual differences in cognitive development. *Annual Review of Psychology, 36*(1), 613-648.

Flavell, J. H. (1982). On Cognitive Development. *Child Development, 53*(1), 1-10.

Fodor, J. A. (1983). The Modularity of Mind: Bradford Book.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*: The MIT Press.

Friedman, N., Miyake, A., Corley, R., Young, S., DeFries, J., & Hewitt, J. (2006). Not All Executive Functions Are Related to Intelligence. *Psychological Science, 17*(2), 172-179.

Friedman, N., Miyake, A., Young, S., DeFries, J., Corley, R., & Hewitt, J. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology General, 137*(2), 201.

Galton, F. (1892). *Hereditary Genius: An Inquiry into Its Laws and Consequences. 1869*. London: Watts and Co.

Gardner, H. (1983). *Frames of mind*: Basic Books New York.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*: Basic Books.

Geake, J. G., & Hansen, P. C. (2005). Neural correlates of intelligence as revealed by fMRI of fluid analogies. *Neuroimage, 26*(2), 555-564.

Geurts, H., Verte, S., Oosterlaan, J., Roeyers, H., & Sergeant, J. (2005). ADHD subtypes: do they differ in their executive functioning profile? *Archives of Clinical Neuropsychology, 20*(4), 457-477.

Gignac, G., Vernon, P. A., Wickett, J. C., & Helmuth, N. (2003). Factors Influencing the Relationship Between Brain Size and Intelligence *The Scientific Study of General Intelligence (First Edition)* (pp. 93-106). Oxford: Pergamon.

Gilhooly, K. (1982). *Thinking: Directed, undirected, and creative*: Academic Press Inc.

Goel, V., Pullara, D., & Grafman, J. (2001). A computational model of frontal lobe dysfunction: working memory and the Tower of Hanoi task. *Cognitive Science, 25*(2), 287-313.

Goodnow, J., & Bethon, G. (1966). Piaget's tasks: the effects of schooling and intelligence. *Child Dev, 37*(3), 573-582.

Goswami, U. (1998). *Cognition in children*: Psychology Press/Erlbaum (UK) Taylor & Francis.

Gottfredson, L. S. (2005). Suppressing intelligence research: Hurting those we intend to help. *Destructive trends in mental health: The well-intentioned path to harm*, 155–186.

Goudriaan, A., Oosterlaan, J., de Beurs, E., & van den Brink, W. (2006). Neurocognitive functions in pathological gambling: a comparison with alcohol dependence, Tourette syndrome and normal controls. *Addiction, 101*(4), 534.

Gratch, G. (1975). Recent studies based on Piaget’s view of object concept development. *Infant perception: From sensation to cognition, 2*, 51–99.

Gray, J. R., & Thompson, P. M. (2004). Neurobiology of intelligence: science and ethics. *Nature Review of Neuroscience, 5*(6), 471-482.

Haier, R. J. (2003). *Brain Imaging Studies of Intelligence: Do We Finally Know Where Intelligence Is in the Brain?, Haier Symposium, Proc. International Society for Intelligence Research, Newport Beach, CA.*

Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2004). Structural brain variation and general intelligence. *Neuroimage, 23*(1), 425-433.

Haier, R. J., White, N. S., & Alkire, M. T. (2003). Individual differences in general intelligence correlate with brain function during nonreasoning tasks. *Intelligence, 31*(5), 429-441.

Hale, S. (1990). A global developmental trend in cognitive processing speed. *Child Development, 61*(3), 653–663.

Halford, G. (1999). The development of intelligence includes the capacity to processing relations of greater complexity. In M. Anderson (Ed.), *The development of intelligence.* (pp. 193-213). Hove: Psychology Press.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*: Wiley.

Herba, C. M., Tranah, T., Rubia, K., & Yule, W. (2006). Conduct Problems in Adolescence: Three Domains of Inhibition and Effect of Gender. *Developmental Neuropsychology, 30*(2), 659-695.

Hood, H. (1962). An experimental study of Piaget's theory of the development of number in children. *British Journal of Psychology, 53*, 273-286.

Houdé, O. (2000). Inhibition and cognitive development: object, number, categorization, and reasoning. *Cognitive Development, 15*(1), 63-73.

Huizinga, M., Dolan, C., & van der Molen, M. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*, 2017-2036.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. *Trans. A. Parsons & S. Milgram., Basic Books, New York.*

Jansen, B. R. J., & Van der Maas, H. L. J. (2001). Evidence for the Phase transition from rule I to rule II on the balance scale task. *Developmental Review, 21*(4), 450-494.

Jansen, B. R. J., & van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology, 81*(4), 383-416.

Jensen, A. (2002). Psychometric g: Definition and substantiation. *The general factor of intelligence: How general is it*, 39–53.

Jensen, A. R. (1985). *Methodological and statistical techniques for the chronometric study of mental abilities*. New York:: Plenum.

Jensen, A. R. (1993). Why is reaction time correlated with psychometric g? *Current Directions in Psychological Science, 2*(2), 53-56.

Jensen, A. R. (1998). *The g factor: The science of mental ability*: Praeger Publishers/Greenwood Publishing Group.

Johnson, M. H., Munakata, Y., & Gilmore, R. O. (2002). *Brain Development and Cognition: A Reader*: Blackwell Publishing.

Kail, R. V. (1996). Source of Age Differences in Speed of Processing. *Child Development, 57*(4).

Kaller, C. P., Unterrainer, J. M., Rahm, B., & Halsband, U. (2004). The impact of problem structure on planning: insights from the Tower of London task. *Cognitive Brain Research, 20*(3), 462-472.

Kane, M., & Engle, R. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology General, 132*(1), 47-70.

Kang, H., & Simpson, G. (1996). Development of semantic and phonological priming in a shallow orthography. *Developmental Psychology, 32*(5), 860-866.

Karmiloff-Smith, A. (1995). *Beyond Modularity: A Developmental Perspective on Cognitive Science*: MIT Press.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences, 2*(10), 389-398.

Karmiloff-Smith, A., & Thomas, M. (2003). What can developmental disorders tell us about the neurocomputational constraints that shape development? The case of Williams syndrome. *Development and Psychopathology, 15*(4), 969-990.

Karmiloff-Smith, A., Tyler, L. K., Voice, K., Sims, K., Udwin, O., Howlin, P., & Davies, M. (1998). Linguistic dissociations in Williams syndrome: evaluating receptive syntax in on-line and off-line tasks. *Neuropsychologia, 36*(4), 343-351.

Kaufman, A., & Kaufman, N. (2004). KABC-IImanual. *Circle Pines, MN: AGS Publishing*.

Klahr, D. (1985). Solving Problems with Ambiguous Subgoal Ordering: Preschoolers' Performance. *Child Development, 56*(4), 940-952.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology, 13*(1), 113-148.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception: ECVP Abstract Supplement, 36*.

Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3-to 12-year-old Finnish children. *Developmental Neuropsychology, 20*(1), 407-428.

Kondoh, M. (2003). Foraging Adaptation and the Relationship Between Food-Web Complexity and Stability (Vol. 299, pp. 1388-1391).

Kovas, Y., Haworth, C., Dale, P., & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development, 72*(3).

Laird, J. (2008). *Extending the Soar cognitive architecture.* Paper presented at the Arificial General Intelligence.

Langley, P., Laird, J., & Rogers, S. (2008). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research, 10*(2), 141-160.

Laver, G., & Burke, D. (1993). Why do semantic priming effects increase in old age? A meta-analysis. *Psychology and Aging, 8*(1), 34.

Leech, R. (2006). A variant on the lexical decision task used by Nation and Snowling (1999). Birkbeck, University of London.

Ligon, E. M. A. (1932). Genetic study of color naming and word reading. *American Journal of Psychology, 44*, 103-121.

Loewenthal, K. M. (2001). *An Introduction to Psychological Tests and Scales*: Psychology Press.

Luce, R. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*: Oxford University Press, USA.

Luciano, M., Wright, M., Smith, G., Geffen, G., Geffen, L., & Martin, N. (2001). Genetic covariance among measures of information processing speed, working memory, and IQ. *Behavior Genetics, 31*(6), 581-592.

Mabbott, D. J., Noseworthy, M., Bouffet, E., Laughlin, S., & Rockel, C. (2006). White matter growth as a mechanism of cognitive development in children. *Neuroimage, 33*(3), 936-946.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychology Bullitin, 109*(2), 163-203.

MacLeod, C. M. (1992). The Stroop task: The 'gold standard' of attentional measures. *Journal of Experimental Psychology: General, 121*(1), 12-14.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 126–135.

MacLeod, D., & Prior, M. (1996). Attention deficits in adolescents with ADHD and other clinical groups. *Child Neuropsychology, 2*(1), 1-10.

Mainela-Arnold, E., Evans, J., & Alibali, M. (2006). Understanding Conservation Delays in Children With Specific Language Impairment: Task Representations Revealed in Speech and Gesture. *Journal of Speech, Language, and Hearing Research, 49*(6), 1267.

Mandler, J. (1994). Precursors of Linguistic Knowledge. *Philosophical Transactions of the Royal Society B: Biological Sciences, 346*(1315), 63-69.

Mareschal, D., Johnson, M., Sirios, S., Spratling, M., Thomas, M., & Westermann, G. (2007). *Neuroconstructivism: How the brain constructs cognition*: Oxford: Oxford University Press.

Mareschal, D., & Shultz, T. R. (1999). Development of children's seriation: A connectionist approach. *Connection Science, 11*(2), 149-186.

Martel, M., Nikolas, M., & Nigg, J. T. (2007). Executive function in adolescents with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(11), 1437-1444.

McClelland, J. L. (1989). Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data. In J. A. Reggia, E. Ruppin & D. Glanzman (Eds.), *Disorders of brain, behavior, and cognition: The neurocomputational perspective* (pp. 75-80). Oxford: Elsevier.

McClelland, J. L., & Morris, R. G. M. (1989). Parallel distributed processing: Implications for cognition and development *Parallel distributed processing: Implications for psychology and neurobiology.* (pp. 8-45): Clarendon Press/Oxford University Press.

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*: The MIT Press.

McFadden, G. T., Dufresne, A., & Kobasigawa, A. (1987). Young children's knowledge of balance scale problems. *Journal of Genetic Psychology, 148*(1), 79-94.

McNamara, T. (2005). *Semantic Priming: Perspectives From Memory and Word Recognition*: Psychology Press (UK).

McNamara, T. P. (1992). Theories of priming. I: associative distance and lag. *Journal of experimental psychology. Learning, memory, and cognition, 18*(6), 1173-1190.

Merrill, M. A. (1924). On the relation of intelligence to achievement in the case of mentally retarded children. *Comparative Psychology Monographs*(2), 1-100.

Meyer, D., & Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227-234.

Minzenberg, M., Ober, B., & Vinogradov, S. (2002). Semantic priming in schizophrenia: a review and synthesis. *Journal of the International Neuropsychological Society, 8*(05), 699-720.

Mitchell, C. L., & Poston, C. S. L. (2001). Effects of inhibiting of response on Tower of London performance. *Current Psychology: Developmental, Learning, Personality, Social, 20*(2), 164-168.

Miyake, A., Friedman, N. P., Emerson, M. J., 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.

Morris, R. G., Ahmed, S., Syed, G. M. S., & Toone, B. K. (1993). Neural correlates of planning ability: Frontal lobe activation during the Tower of London test. *Neuropsychologia, 31*(12), 1367-1378.

Moss, H., Hare, M., Day, P., & Tyler, L. (1994). A distributed memory model of the associative boost in semantic priming. *Connection science(Print), 6*(4), 413-427.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 863-883.

Murray, J. (2003). *Mathematical biology*: Springer Verlag.

Myerson, J., Ferraro, F., Hale, S., & Lima, S. (1992). General slowing in semantic priming and word recognition. *Psychology and Aging, 7*(2), 257–270.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: evidence from semantic priming. *Cognition, 70*(1), B1-13.

Neely, J. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. *Basic processes in reading: Visual word recognition*, 264-336.

Neisser, U., Boodoo, G., Bouchard, T., Boykin, A., Brody, N., Ceci, S., Halpern, D., Loehlin, J., Perloff, R., & Sternberg, R. (1996). Intelligence: Knowns and Unknowns. *American Psychologist, 51*(2), 77-101.

Nelson, D., McEvoy, C., & Schreiber, T. (1994). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, 36*(3), 402.

Nettelbeck, T. (1987). Inspection time and intelligence. *Speed of information processing and intelligence*, 295-346.

Nettelbeck, T. (2001). Correlation between inspection time and psychometric abilities: A personal interpretation. *Intelligence, 29*(6), 459-474.

Neubauer, A. C., Grabner, R. H., Freudenthaler, H. H., Beckmann, J. F., & Guthke, J. r. (2004). Intelligence and individual differences in becoming neurally efficient. *Acta Psychologica, 116*(1), 55-74.

Newell, A., & Simon, H. A. (1972). *Human problem solving*: Englewood Cliffs, NJ: Prentice-Hall.

Newman, R., & German, D. (2002). Effects of lexical factors on lexical access among typical language-learning children and children with word-finding difficulties. *Language and Speech, 45*(3), 285.

Newman, S. D., Carpenter, P. A., Varma, S., & Just, M. A. (2003). Frontal and parietal participation in problem solving in the Tower of London: fMRI and computational modeling of planning and high-level perception. *Neuropsychologia, 41*(12), 1668-1682.

Newman, S. D., & Pittman, G. (2007). The tower of London: A study of the effect of problem structure on planning. *Journal of Clinical and Experimental Neuropsychology, 29*(3), 332-342.

Nichelli, F., Scala, G., Vago, C., Riva, D., & Bulgheroni, S. (2005). Age-related trends in Stroop and conflicting motor response task findings. *Child Neuropsychology, 11*(5), 431-443.

Normandeau, S., Larivée, S., Roulin, J.-L., & Longeot, F. (1989). The balance-scale dilemma: Either the subject or the experimenter muddles through. *Journal of Genetic Psychology, 150*(3), 237-250.

Novikova, S. I., & Stroganova, T. A. (2006). Action planning by 5-6-year-olds: Age-specific and individual differences in "The Tower of London" task. *Voprosy Psychologii, 4*, 36-46.

Ozonoff, S. (1998). Assessment and remediation of executive dysfunction in autism and Asperger syndrome. *Asperger syndrome or high-functioning autism*, 263-289.

Pasnak, R., Willson-Quayle, A., & Whitten, J. (1998). Mild Retardation, Academic Achievement, and Piagetian or Psychometric Tests of Reasoning. *Journal of Developmental and Physical Disabilities, 10*(1), 23-33.

Paton, G. (2008, 5th May). Classes 'based on ability not age in five years'. *The Telegraph*.

Penadés, R., Boget, T., Lomeña, F., Bernardo, M., Mateos, J. J., Laterza, C., Pavìa, J., & Salamero, M. (2000). Brain perfusion and neuropsychological changes in schizophrenic patients after cognitive rehabilitation. *Psychiatry Research: Neuroimaging, 98*(2), 127-132.

Pennington, B., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37*(1), 51-87.

Phaf, R. H., Christoffels, I. K., Waldorp, L. J., & den Dulk, P. (1998). Connectionist investigations of individual differences in Stroop performance. *Perceptual and Motor Skills, 87*(3), 899-914.

Piaget, J. (1950). *The Psychology of Intelligence*: Routledge.

Piaget, J. (1954). The construction of reality in the child (M. Cook, Trans.). *New York: Ballantine.(Original work published 1937).*

Piaget, J. (1972). Intellectual development from adolescence to adulthood. *Human Development, 15*, 1-12.

Piaget, J., Inhelder, B., & Chilton, P. (1997). *Mental imagery in the child*: Routledge New York.

Plaut, D. (1995). *Semantic and associative priming in a distributed attractor network*.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107*(4), 786-823.

Plomin, R., DeFries, J., & McClearn, G. (2008). McGuffin, p. Behavioral genetics: Worth Publishing New York, NY:.

Plomin, R., Fulker, D. W., Corley, R., & DeFries, J. C. (1997). Nature, nurture, and cognitive development from 1 to 16 years: A parent-offspring adoption study. *Psychological Science, 8*(6), 442-447.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. *Information processing and cognition: The Loyola symposium*, 55-85.

Posthuma, D., & de Geus, E. (2006). Progress in the molecular-genetic study of intelligence. *Current Directions in Psychological Science, 15*(4), 151-155.

Posthuma, D., De Geus, E. J. C., Baare, W. F. C., Pol, H. E. H., Kahn, R. S., & Boomsma, D. I. (2002). The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience, 5*(2), 83-84.

Raduege, T., & Schwantes, F. (1987). Effects of Rapid Word Recognition Training on Sentence Context Effects in Children. *Journal of Reading Behavior, 19*(4), 395-414.

Raijmakers, M. E. J., van Koten, S., & Molenaar, P. C. M. (1996). On the validity of simulating stagewise development by means of PDP networks: Application of catastrophe analysis and an experimental test of rule-like network performance. *Cognitive Science, 20*(1), 101-136.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510-510.

Richardson, F. M., Baughman, F. D., Forrester, N. A., & Thomas, M. S. C. (2006). *Computational Modeling of Variability in the Balance Scale Task.* Paper presented at the Proceedings of the 7th International Conference of Cognitive Modeling.

Richardson, F. M., Forrester, N., Baughman, F. D., & Thomas, M. S. C. (2006). Computational Modeling of Variability in the Conservation Task. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 26–29.

Roth, R. M., Randolph, J. J., Koven, N. S., Isquith, P. K., & Dupri, J. R. (2006). Neural Substrates of Executive Functions: Insights from Functional Neuroimaging *Focus on neuropsychology research.* (pp. 1-36): Nova Science Publishers.

Rowe, J., Lavender, A., & Turk, V. (2006). Cognitive executive function in Down's syndrome. *British Journal of Clinical Psychology, 45*(1), 5.

Rubia, K., Smith, A., & Taylor, E. (2007). Performance of children with attention deficit hyperactivity disorder (ADHD) on a test battery of impulsiveness. *Child Neuropsychology, 13*(3), 276-304.

Salthouse, T., Atkinson, T., & Berish, D. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of experimental psychology. General, 132*(4), 566.

Schapiro, A. C., & McClelland, J. L. (2009). Continuous or Discontinuous Change? A connectionist model of developmental transition in the balance scale task. In J. Spencer, M. S. C. Thomas & J. L. McClelland (Eds.), *Toward a new unified theory of development: Connectionism and dynamical systems theory re-considered*. Oxford: Oxford University Press.

Shafer, V. L., Garrido-Nag, K., Hoff, E., & Shatz, M. (2007). The neurodevelopmental bases of language *Blackwell handbook of language development.* (pp. 21-45). Malden, MA US: Blackwell Publishing.

Shallice, T. (1982). Specific Impairments of Planning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 298*(1089), 199-209.

Shallice, T. (1988). *From neuropsychology to mental structure*: Cambridge University Press.

Shapiro, L., Swinney, D., & Borsky, S. (1998). Online Examination of Language Performance in Normal and Neurologically Impaired Adults. *American Journal of Speech-Language Pathology, 7*(1), 49.

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., & Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature, 440*(7084), 676-679.

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*(5), 405-410.

Shultz, T. R. (1998). A computational analysis of conservation. *Developmental Science, 1*(1), 103-126.

Shultz, T. R., Mareschal, D., & Schmidt, W. C. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning, 16*(1), 57-86.

Shultz, T. R., & Takane, Y. (2007). Rule following and rule use in the balance-scale task. *Cognition, 103*(3), 460-472.

Siegler, R. (2006). Microgenetic analyses of learning. In W. Damon, R. Lerner, D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology* (Vol. 2, pp. 464-510).

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*(4), 481-520.

Siegler, R. S. (1978). The origins of scientific reasoning *Children's thinking: What develops?* (pp. 109-149): Lawrence Erlbaum Associates, Inc.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46*(2), 84-84.

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology, 28*(3), 225-273.

Siegler, R. S. (2007). Cognitive variability. *Developmental Science, 10*(1), 104-109.

Siegler, R. S., & Chen, Z. (2002). Development of rules and strategies: Balancing the old and the new. *Journal of Experimental Child Psychology, 81*(4), 446-457.

Siegler, R. S., Granott, N., & Parziale, J. (2002). Microgenetic studies of self-explanation *Microdevelopment: Transition processes in development and learning.* (pp. 31-58): Cambridge University Press.

Sikora, D., Haley, P., Edwards, J., & Butler, R. (2002). Tower of London test performance in children with poor arithmetic skills. *Developmental Neuropsychology, 21*(3), 243-254.

Simon, H. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7*(2), 268-288.

Smith, N., & Tsimpli, I. (1995). *The Mind of a Savant: Language Learning and Modularity*: Blackwell Publishers.

Smith, P., & Traynelis, J. F. (unpublished). NFER Spatial Memory Test. Slough, England: National Foundation for Educational Research.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15*(2), 201-293.

Spelke, E., Mehler, J., & Franck, S. (1995). Initial knowledge: Six suggestions *Cognition on cognition.* (pp. 433-447): The MIT Press.

Spencer, J. P., & Schoner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science, 6*(4), 392-412.

Sperber, D., & Dupoux, E. (2001). In defense of massive modularity *Language, brain, and cognitive development: Essays in honor of Jacques Mehler.* (pp. 47-57): The MIT Press.

Spitz, H. H. (1982). Intellectual extremes, mental age, and the nature of human intelligence. *Merrill-Palmer Quarterly, 28*(2), 167-192.

Stern, W. (1912). *Die psychologische Methoden der Intelligenzprufung.* Liepzig: Barth.

Sternberg, R. (1987). Intelligence. In R. L. Gregory (Ed.), *The Oxford companion to the mind* (pp. 375-379). Oxford: Oxford University Press.

Sternberg, R. (2005). The Theory of Successful Intelligence. *Interamerican Journal Of Psychology, 39*(2), 189.

Sternberg, R., Grigorenko, E., & Kidd, K. (2005). Intelligence, race, and genetics. *American Psychologist, 60*(1), 46-59.

Sternberg, R. J. (2000). *Handbook of Intelligence*: Cambridge University Press.

Stiles, J., Reilly, J., Paul, B., & Moses, P. (2005). Cognitive development following early brain injury: Evidence for neural adaptation. *Trends in Cognitive Sciences, 9*(3), 136-143.

Stroop, J. R. (1935). Studies of interferences in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Szameitat, A. J., Schubert, T., Muller, K., & von Cramon, D. Y. (2002). Localization of Executive Functions in Dual-Task Performance with fMRI. *Journal of Cognitive Neuroscience, 14*(8), 1184-1199.

Taylor, S. F., Kornblum, S., & Tandon, R. (1996). Facilitation and interference of selective attention in schizophrenia. *Journal of Psychiatric Research, 30*(4), 251-259.

Telford, C. W. (1930). Differences in responses to colors and their names. *Journal of Genetic Psychology 37*, 151-159.

Temple, C. M. (1997). Cognitive neuropsychology and its application to children. *J Child Psychol Psychiatry, 38*(1), 27-52.

Thomas, M., & Karmiloff-Smith, A. (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behaviour Brain Science, 25*(6), 727-750; discussion 750-787.

Thomas, M. S. C. (1997). Connectionist networks and knowledge representation: The case of bilingual lexical processing. Unpublished D.Phil. thesis. Oxford University.

Thomas, M. S. C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research.*

Thomas, M. S. C., Dockrell, J. E., Messer, D., Parmigiani, C., Ansari, D., & Karmiloff-Smith, A. (2008). Speeded naming, frequency and the development of the lexicon in Williams syndrome. *Language and Cognitive Processes, 21*(6), 721 - 759-721 - 759.

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Connectionist models of development, developmental disorders and individual differences. In R. J. Sternberg, J. Lautrey & T. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 133–150): American Psychological Association.

Thomas, M. S. C., Karmiloff-Smith, A., & Goswami, U. (2002). Modeling typical and atypical cognitive development: Computational constraints on mechanisms of change *Blackwell handbook of childhood cognitive development.* (pp. 575-599): Blackwell Publishing.

Thomas, M. S. C., McClelland, J. L., Richardson, F. M., Schapiro, A. C., & Baughman, F. D. (2009). Dynamical and Connectionist Approaches to Development: Toward a Future of Mutually Beneficial Co-evolution. In J. Spencer, M. S. C. Thomas & J. L. McClelland (Eds.), *Toward a new unified theory of development: Connectionism and dynamical systems theory re-considered*. Oxford: Oxford University Press.

Thomas, M. S. C., & Richardson, F. (2006). Atypical representational change: Conditions for the emergence of atypical modularity. *Processes of change in brain and cognitive development: Attention and performance XXI*, 315-347.

Thomas, M. S. C., Richardson, F. M., Forrester, N. A., & Baughman, F. D. (submitted). Modelling individual variability in cognitive development. *Connection Science*.

Thorndike, R. L., Hagen, E. P., & France, N. (1986). *Cognitive abilities test administration manual*. Windsor, UK: NFER-Nelson.

Treffert, D. (1989). *Extraordinary People: Understanding 'Idiot savants'*. New York: Bantam.

Tunteler, E., & Resing, W. C. M. (2002). Spontaneous analogical transfer in 4-year-olds: A microgenetic study. *Journal of Experimental Child Psychology, 83*(3), 149-166.

Tyler, L., Karmiloff-Smith, A., Voice, J., Stevens, T., Grant, J., Udwin, O., Davies, M., & Howlin, P. (1997). Do individuals with Williams syndrome have bizarre semantics? Evidence for lexical organization using an on-line task. *Cortex, 33*(3), 515-527.

Uchida, S., & Drossel, B. (2007). Relation between complexity and stability in food webs with adaptive behavior. *Journal of Theoretical Biology, 247*(4), 713-722.

Ungoed-Thomas, J. (2005, 4th September). School pioneers the all-age class. *The Sunday Times*.

van der Maas, H. (1995). Beyond the metaphor? *Cognitive Development, 10*, 621-642.

van der Maas, H. L., & Raijmakers, M. E. J. (2009). Transitions in cognitive development: prospects and limitations of a neural dynamic approach. In J. Spencer, M. S. C. Thomas & J. L. McClelland (Eds.), *Toward a new grand theory of development: Connectionism and dynamical systems theory reconsidered*. Oxford: Oxford University Press.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review, 113*(4), 842-861.

van der Maas, H. L. J., & Jansen, B. R. J. (2003). What response times tell of children’s behavior on the balance scale task. *Journal of Experimental Child Psychology, 85*(2), 141-177.

van Mourik, R., Oosterlaan, J., & Sergeant, J. (2005). The Stroop revisited: a meta-analysis of interference control in AD/HD. *Journal of Child Psychology and Psychiatry, 46*(2), 150-165.

van Rijn, H., van Someren, M., & van der Maas, H. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science, 27*(2), 227-257.

Vicari, S., Bellucci, S., & Carlesimo, G. (2000). Implicit and explicit memory: a functional dissociation in persons with Down syndrome. *Neuropsychologia, 38*(3), 240-251.

Vicari, S., Bellucci, S., & Carlesimo, G. (2001). Procedural learning deficit in children with Williams syndrome. *Neuropsychologia, 39*(7), 665-677.

Wangersky, P. J. (1978). Lotka-Volterra Population Models. *Annual Review of Ecology and Systematics, 9*(1), 189-218.

Wechsler, D. (1997). WAIS-III administration and scoring manual. *San Antonio, TX: The Psychological Corporation*.

Wechsler, D. (2004). *WISC-IV: Wechsler Intelligence Scale for Children: Technical and Interpretive Manual*: Psychological Corporation.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive Development: Foundational Theories of Core Domains. *Annual Review of Psychology, 43*(1), 337-375.

Welsh, M., Friedman, S., & Spieker, S. (2006). Executive Functions in Developing Children: Current Conceptualizations and Questionsfor the Future. *Blackwell handbook of early childhood development*, 167.

Welsh, M., Pennington, B., Groisser, D., & Green, L. (1991). A Normative-Developmental Study of Executive Function: a Window on Prefrontal Function in Children. *Developmental Neuropsychology, 7*(2), 131-149.

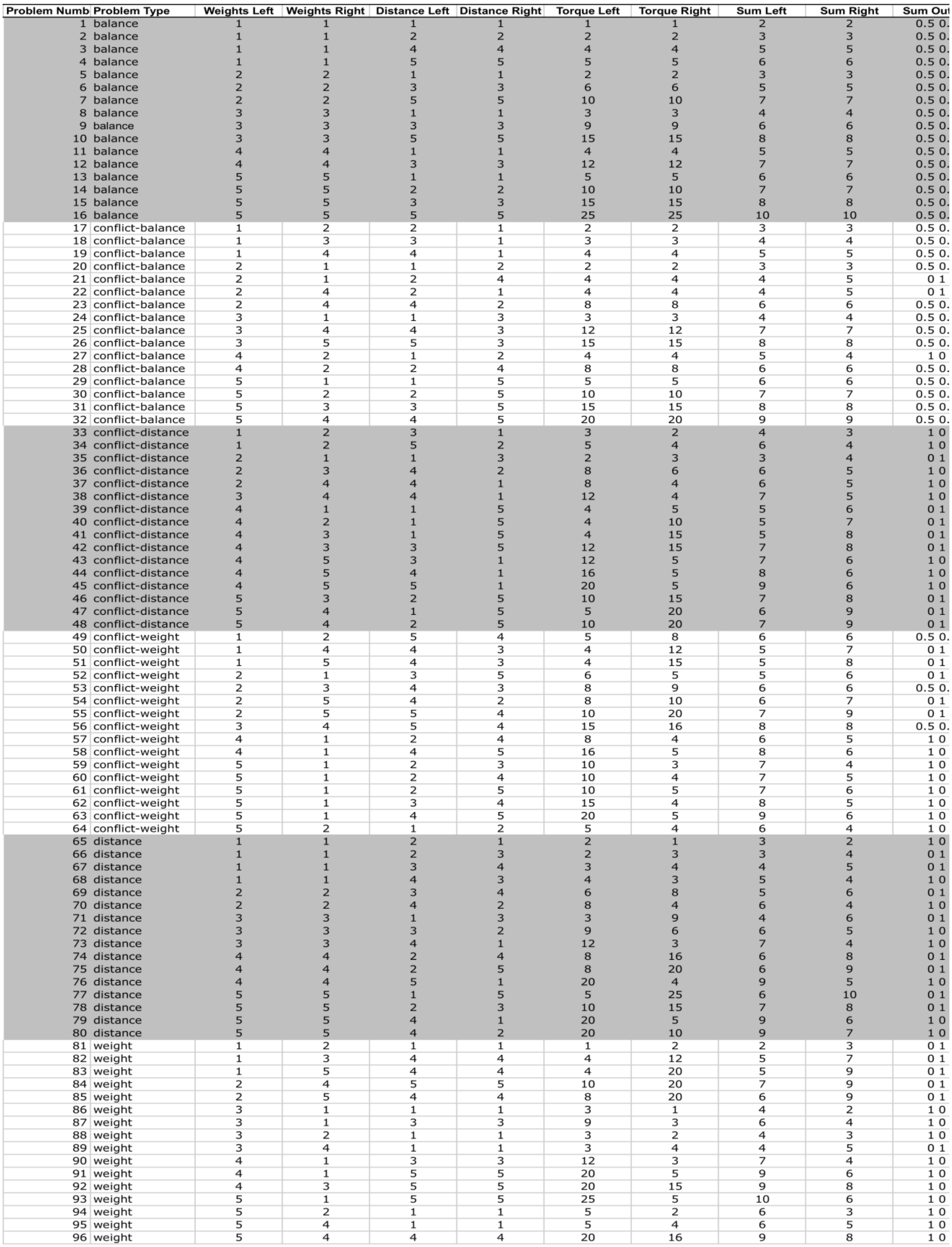
Williams, R. H., Zimmerman, D. W., Zumbo, B. D., & Ross, D. (2003). Charles Spearman: British Behavioral Scientist. *Human Nature Review, 3*, 114-118.

Wright, M., De Geus, E., Ando, J., Luciano, M., Posthuma, D., Ono, Y., Hansell, N., Van Baal, C., Hiraishi, K., Hasegawa, T., Smith, G., Geffen, G., Geffen, L., Kanba, S., Miyake, A., Martin, N., & Boomsma, D. (2001). Genetics of cognition: outline of a collaborative twin study. *Twin Research, 4*(1), 48-56.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1*(2), 198-226.

Appendices

Appendix A Problem Types used on the balance scale task (Chapter 8)



Problem Numb	Problem Type	Weights Left	Weights Right	Distance Left	Distance Right	Torque Left	Torque Right	Sum Left	Sum Right	Sum Out
1	balance	1	1	1	1	1	1	2	2	0.5 0.
2	balance	1	1	2	2	2	2	3	3	0.5 0.
3	balance	1	1	4	4	4	4	5	5	0.5 0.
4	balance	1	1	5	5	5	5	6	6	0.5 0.
5	balance	2	2	1	1	2	2	3	3	0.5 0.
6	balance	2	2	3	3	6	6	5	5	0.5 0.
7	balance	2	2	5	5	10	10	7	7	0.5 0.
8	balance	3	3	1	1	3	3	4	4	0.5 0.
9	balance	3	3	3	3	9	9	6	6	0.5 0.
10	balance	3	3	5	5	15	15	8	8	0.5 0.
11	balance	4	4	1	1	4	4	5	5	0.5 0.
12	balance	4	4	3	3	12	12	7	7	0.5 0.
13	balance	5	5	1	1	5	5	6	6	0.5 0.
14	balance	5	5	2	2	10	10	7	7	0.5 0.
15	balance	5	5	3	3	15	15	8	8	0.5 0.
16	balance	5	5	5	5	25	25	10	10	0.5 0.
17	conflict-balance	1	2	2	1	2	2	3	3	0.5 0.
18	conflict-balance	1	3	3	1	3	3	4	4	0.5 0.
19	conflict-balance	1	4	4	1	4	4	5	5	0.5 0.
20	conflict-balance	2	1	1	2	2	2	3	3	0.5 0.
21	conflict-balance	2	1	2	4	4	4	4	5	0 1
22	conflict-balance	2	4	2	1	4	4	4	5	0 1
23	conflict-balance	2	4	4	2	8	8	6	6	0.5 0.
24	conflict-balance	3	1	1	3	3	3	4	4	0.5 0.
25	conflict-balance	3	4	4	3	12	12	7	7	0.5 0.
26	conflict-balance	3	5	5	3	15	15	8	8	0.5 0.
27	conflict-balance	4	2	1	2	4	4	5	4	1 0
28	conflict-balance	4	2	2	4	8	8	6	6	0.5 0.
29	conflict-balance	5	1	1	5	5	5	6	6	0.5 0.
30	conflict-balance	5	2	2	5	10	10	7	7	0.5 0.
31	conflict-balance	5	3	3	5	15	15	8	8	0.5 0.
32	conflict-balance	5	4	4	5	20	20	9	9	0.5 0.
33	conflict-distance	1	2	3	1	3	2	4	3	1 0
34	conflict-distance	1	2	5	2	5	4	6	4	1 0
35	conflict-distance	2	1	1	3	2	3	3	4	0 1
36	conflict-distance	2	3	4	2	8	6	6	5	1 0
37	conflict-distance	2	4	4	1	8	4	6	5	1 0
38	conflict-distance	3	4	4	1	12	4	7	5	1 0
39	conflict-distance	4	1	1	5	4	5	5	6	0 1
40	conflict-distance	4	2	1	5	4	10	5	7	0 1
41	conflict-distance	4	3	1	5	4	15	5	8	0 1
42	conflict-distance	4	3	3	5	12	15	7	8	0 1
43	conflict-distance	4	5	3	1	12	5	7	6	1 0
44	conflict-distance	4	5	4	1	16	5	8	6	1 0
45	conflict-distance	4	5	5	1	20	5	9	6	1 0
46	conflict-distance	5	3	2	5	10	15	7	8	0 1
47	conflict-distance	5	4	1	5	5	20	6	9	0 1
48	conflict-distance	5	4	2	5	10	20	7	9	0 1
49	conflict-weight	1	2	5	4	5	8	6	6	0.5 0.
50	conflict-weight	1	4	4	3	4	12	5	7	0 1
51	conflict-weight	1	5	4	3	4	15	5	8	0 1
52	conflict-weight	2	1	3	5	6	5	5	6	0 1
53	conflict-weight	2	3	4	3	8	9	6	6	0.5 0.
54	conflict-weight	2	5	4	2	8	10	6	7	0 1
55	conflict-weight	2	5	5	4	10	20	7	9	0 1
56	conflict-weight	3	4	5	4	15	16	8	8	0.5 0.
57	conflict-weight	4	1	2	4	8	4	6	5	1 0
58	conflict-weight	4	1	4	5	16	5	8	6	1 0
59	conflict-weight	5	1	2	3	10	3	7	4	1 0
60	conflict-weight	5	1	2	4	10	4	7	5	1 0
61	conflict-weight	5	1	2	5	10	5	7	6	1 0
62	conflict-weight	5	1	3	4	15	4	8	5	1 0
63	conflict-weight	5	1	4	5	20	5	9	6	1 0
64	conflict-weight	5	2	1	2	5	4	6	4	1 0
65	distance	1	1	2	1	2	1	3	2	1 0
66	distance	1	1	2	3	2	3	3	4	0 1
67	distance	1	1	3	4	3	4	4	5	0 1
68	distance	1	1	4	3	4	3	5	4	1 0
69	distance	2	2	3	4	6	8	5	6	0 1
70	distance	2	2	4	2	8	4	6	4	1 0
71	distance	3	3	1	3	3	9	4	6	0 1
72	distance	3	3	3	2	9	6	6	5	1 0
73	distance	3	3	4	1	12	3	7	4	1 0
74	distance	4	4	2	4	8	16	6	8	0 1
75	distance	4	4	2	5	8	20	6	9	0 1
76	distance	4	4	5	1	20	4	9	5	1 0
77	distance	5	5	1	5	5	25	6	10	0 1
78	distance	5	5	2	3	10	15	7	8	0 1
79	distance	5	5	4	1	20	5	9	6	1 0
80	distance	5	5	4	2	20	10	9	7	1 0
81	weight	1	2	1	1	1	2	2	3	0 1
82	weight	1	3	4	4	4	12	5	7	0 1
83	weight	1	5	4	4	4	20	5	9	0 1
84	weight	2	4	5	5	10	20	7	9	0 1
85	weight	2	5	4	4	8	20	6	9	0 1
86	weight	3	1	1	1	3	1	4	2	1 0
87	weight	3	1	3	3	9	3	6	4	1 0
88	weight	3	2	1	1	3	2	4	3	1 0
89	weight	3	4	1	1	3	4	4	5	0 1
90	weight	4	1	3	3	12	3	7	4	1 0
91	weight	4	1	5	5	20	5	9	6	1 0
92	weight	4	3	5	5	20	15	9	8	1 0
93	weight	5	1	5	5	25	5	10	6	1 0
94	weight	5	2	1	1	5	2	6	3	1 0
95	weight	5	4	1	1	5	4	6	5	1 0
96	weight	5	4	4	4	20	16	9	8	1 0

Appendix B Verbatim instructions in instruction video for balance scale problems (Chapter 8)

“In this activity, you will see a beam. Underneath the beam are two blocks. These blocks stop the beam from moving when weights are put on each side.

If these blocks are taken away, the beam can move. If one side of the beam is heavier, it will tip to that side. The beam will not always tip to one side. Sometimes it will balance.

The challenge is to decide whether the beam will balance or whether it will tip to one side. Try and decide this as quickly as possible, making as few mistakes as possible.

Touch the button that matches your choice. For example, if you think that this beam would tip to the left, you should touch the button on the left. If you think that this beam would tip to the right, you should touch the button on the right. If you think it would balance, you should touch the button in the middle.

Let's have a practice. Remember you just need to decide whether the beam will balance, or whether it will tip to one side.”

[TEST ITEMS....]

“Remember, decide whether the beam will balance or whether it will tip to one side as quickly as possible, making as few mistakes as possible. Touch the button that matches your choice.

Are you ready?

Then let's begin...”

Appendix C Verbatim instructions in instruction video for Tower of London task (Chapter 9)

“In this activity, you will see in the top left hand corner of the screen a small picture that has 3 balls on a board with 3 pegs. You should try and match this picture by moving these larger balls around on screen.
You can move a ball from one peg to another by touching the picture of the ball and moving your finger across the screen.
The challenge is to match the smaller picture as quickly as possible, making as few mistakes as possible.
For example, if you think that the RED ball should be moved on top of the GREEN ball to match the smaller picture, you should touch the RED ball and move it to the peg where the GREEN is.
When you think you have finished, touch this button that says DONE.
If you get stuck at any time and want to try again, touch this button that says RE-START.
Let's have a practice. Remember you need to match the picture in the top left hand corner by moving the larger balls around on-screen”

[TEST ITEMS....]

“That was good.
Remember you need to match the picture in the top left hand corner by moving the larger balls around on-screen. Try and do this as quickly as possible, making as few mistakes as possible.
Are you ready? Then let's begin...”

Appendix D Tower of London task: Full problem set (Chapter 9)

				START				*GOAL*	
Trial no.	Configuration	Min moves	RED	BLUE	GREEN		RED	BLUE	GREEN
1	Tower	3	**R1**	**C1**	**C2**		**R1**	**R2**	**R3**
2	Flat	3	**R1**	**C1**	**R2**		**L1**	**C1**	**R1**
3	Tower	3	**C1**	**L1**	**C2**		**R3**	**R2**	**R1**
4	Flat	3	**C1**	**R2**	**R1**		**C1**	**R1**	**L1**
5	Tower	4	**C2**	**C1**	**L1**		**R3**	**R2**	**R1**
6	Flat	4	**L1**	**C1**	**C2**		**R1**	**L1**	**C1**
7	Tower	4	**R1**	**C2**	**C1**		**R3**	**R1**	**R2**
8	Flat	4	**C2**	**L1**	**C1**		**C1**	**R1**	**L1**
9	Tower	5	**C1**	**R1**	**R2**		**R2**	**R3**	**R1**
10	Flat	5	**R1**	**R2**	**C1**		**L1**	**C1**	**R1**
11	Tower	5	**R1**	**C1**	**R2**		**R3**	**R2**	**R1**
12	Flat	5	**R1**	**R3**	**R2**		**L1**	**R1**	**C1**
13	Tower	6	**R2**	**C1**	**R1**		**R3**	**R1**	**R2**
14	Flat	6	**R1**	**C1**	**C2**		**L1**	**R1**	**C1**
15	Tower	6	**R1**	**R2**	**L1**		**R2**	**R1**	**R3**
16	Flat	6	**C1**	**R2**	**R1**		**L1**	**R1**	**C1**

Note: A letter representing the peg and a number representing the position on that peg is used to represent the positions of balls. For example, L1 represents the first (bottom) position on the Left peg. C2 represents the second (top) position on the Centre peg. R3 represents the third (top) position on the Right peg.

Appendix E Verbatim instructions (audio and visual presentation) used on the Lexical decision task (Chapter 6)

“You will hear a woman say a word.
Then you will hear a man say another word.
The word will be a real word like dog or else a made-up silly word like glorp.

Here’s what to do:

If the man says a real word, press the green button.
If the man says a made-up word, press the red button.
Sometimes what the woman says and what the man says will seem strange. Don’t worry about this, your job is just to decide if the word the man says is a real word or a silly word.

Always use the pointing finger of the hand you write with.
Try to press the buttons as fast as you can without making a mistake. You will have a chance to practice the task before the real experiment begins.

The experiment will take about 5 minutes. Have fun!”

Appendix F Primes and targets for semantically related word pairs used in Lexical Decision task[a] (Chapter 6)

	Block A			Block B	
Condition	Item	Item	Condition	Item	Item
PCA	bat	ball	UCA	belt	ball
PCA	salt	pepper	UCA	christmas	pepper
PCA	moon	stars	UCA	kitchen	stars
PCA	king	queen	UCA	farm	queen
PCA	dog	cat	UCA	butcher	cat
PCA	brother	sister	UCA	beach	sister
UCA	grill	ship	PCA	boat	ship
UCA	umbrella	pen	PCA	pencil	pen
UCA	shampoo	chair	PCA	table	chair
UCA	bow	saucer	PCA	cup	saucer
UCA	hammer	brush	PCA	comb	brush
UCA	kettle	hat	PCA	coat	hat
PCNA	cow	goat	UCNA	hospital	goat
PCNA	aeroplane	train	UCNA	zoo	train
PCNA	pig	horse	UCNA	war	horse
PCNA	nose	head	UCNA	circus	head
PCNA	lake	mountain	UCNA	lounge	mountain
PCNA	green	pink	UCNA	market	pink
UCNA	knife	skirt	PCNA	jumper	skirt
UCNA	oven	magazine	PCNA	book	magazine
UCNA	party	desk	PCNA	bed	desk
UCNA	broom	balloon	PCNA	kite	balloon
UCNA	string	wall	PCNA	roof	wall
UCNA	fridge	guitar	PCNA	violin	guitar
PFA	butcher	meat	UFA	green	meat
PFA	beach	sand	UFA	cow	sand
PFA	belt	trousers	UFA	aeroplane	trousers
PFA	christmas	tree	UFA	pig	tree
PFA	kitchen	sink	UFA	nose	sink
PFA	farm	animal	UFA	lake	animal
UFA	cup	hair	PFA	shampoo	hair
UFA	pencil	tea	PFA	kettle	tea
UFA	comb	nail	PFA	hammer	nail
UFA	coat	toast	PFA	grill	toast
UFA	boat	arrow	PFA	bow	arrow

UFA	table	rain
PFNA	war	army
PFNA	circus	lion
PFNA	lounge	sofa
PFNA	market	vegetables
PFNA	hospital	doctor
PFNA	zoo	penguin
UFNA	roof	cheese
UFNA	violin	parcel
UFNA	bed	potato
UFNA	book	floor
UFNA	kite	music
UFNA	jumper	bread
NW1	lane	balras
NW1	mud	kep
NW1	rug	ralt
NW1	chain	wuth
NW1	steam	trantor
NW1	owl	mimber
NW1	bay	sladding
NW1	frost	twesk
NW1	frog	pell
NW1	barrel	werp
NW1	ankle	puct
NW1	bubble	sammer
NW1	plug	tafflest
NW1	eye	glistow
NW1	fisherman	drist
NW1	page	weast
NW1	road	sprool
NW1	stove	bannifer
NW1	stew	toag
NW1	jar	pleck
NW1	bridge	teep
NW1	peach	lats
NW1	hill	stopograttic
NW1	elbow	tegwop
NW1	valley	klat
NW1	daylight	flipple

PFA	umbrella	rain
UFNA	salt	army
UFNA	moon	lion
UFNA	king	sofa
UFNA	dog	vegetables
UFNA	brother	doctor
UFNA	bat	penguin
PFNA	fridge	cheese
PFNA	string	parcel
PFNA	oven	potato
PFNA	broom	floor
PFNA	party	music
PFNA	knife	bread
NW2	valley	balras
NW2	daylight	kep
NW2	smoke	ralt
NW2	fat	wuth
NW2	ditch	trantor
NW2	pet	mimber
NW2	blanket	sladding
NW2	shower	twesk
NW2	gun	pell
NW2	ticket	werp
NW2	mirror	puct
NW2	lady	sammer
NW2	sock	tafflest
NW2	sting	glistow
NW2	cake	drist
NW2	island	weast
NW2	letter	sprool
NW2	drink	bannifer
NW2	uncle	toag
NW2	stable	pleck
NW2	square	teep
NW2	coal	lats
NW2	branch	stopograttic
NW2	pocket	tegwop
NW2	lane	klat
NW2	mud	flipple

NW1	fat	speeb
NW1	smoke	vater
NW1	ditch	yote
NW1	pet	dippler
NW1	shower	glort
NW1	blanket	sply
NW1	gun	newper
NW1	ticket	jendol
NW1	mirror	degs
NW1	lady	sarl
NW1	sock	flun
NW1	sting	hend
NW1	cake	rubid
NW1	island	felly
NW1	letter	mosp
NW1	drink	waip
NW1	uncle	hinshink
NW1	branch	lerman
NW1	square	poil
NW1	coal	niz
NW1	stable	fenner
NW1	pocket	merly

NW2	rug	speeb
NW2	chain	vater
NW2	steam	yote
NW2	owl	dippler
NW2	bay	glort
NW2	frost	sply
NW2	frog	newper
NW2	barrel	jendol
NW2	ankle	degs
NW2	bubble	sarl
NW2	eye	flun
NW2	plug	hend
NW2	fisherman	rubid
NW2	page	felly
NW2	road	mosp
NW2	stove	waip
NW2	stew	hinshink
NW2	jar	lerman
NW2	bridge	poil
NW2	peach	niz
NW2	hill	fenner
NW2	elbow	merly

[a] Items taken from Nation and Snowling (1999).

Codes correspond to

PCA category related - associated
UCA unrelated (control for category related - associated)
PCNA category related - non-associated
UCNA unrelated (control for category related - non-associated)
PFA functionally related - associated
UFA unrelated (control for functionally related - associated)
PFNA functionally related - non-associated
UFNA unrelated (control for functionally related - non-associated)
NW1 distractor trial block 1
NW2 distractor trial block 2

Appendix G. Instructions for Conservation of Number Task (Chapter 7Chapter 1)

“In this activity you will see some gold coins on each side of a dotted line.

You will first need to decide which side has more coins, or whether they both have the same number. Touch the button that matches your choice. For example, if you think the side on the left has more gold coins, you should touch the button on the left.

If you think the side on the right has more, you should touch the button on the right. If you think both sides have the same number of coins, you should touch the button in the middle.

Once you have made your choice, a hand will appear and change one of the sides. You will then need to decide again which side has more, or whether they both have the same number of coins.

If you think the side on the right has more, you should touch the button on the right. If you think both sides have the same number of coins, you should touch the button in the middle.

The challenge is to decide whether they have the same number of coins, or whether one side has more, as quickly as possible making as few mistakes as possible.

Let's have a practice”

Appendix H. Instructions for Conservation of Liquid Task (Chapter 7)

“In this activity you will see some pictures of cups with water.

You will first need to decide which cup has more water in it, or whether they both have the same amount. Touch the button that matches your choice. For example, if you think the side on the left has more water, you should touch the button on the left.

If you think the side on the right has more, you should touch the button on the right. If you think both sides have the same amount of water, you should touch the button in the middle.

Once you have made your choice, one of the cups will move and the water will be poured out into a new cup. You will then need to decide again which cup has more water, or whether they both have the same amount.

If you think the side on the right has more, you should touch the button on the right. If you think both sides have the same amount of water, you should touch the button in the middle.

The challenge is to decide whether they have the same amount, or whether one side has more, as quickly as possible making as few mistakes as possible.

Let's have a practice?”

Appendix I. Specifying the architectures with M matrices in the dynamical systems models.

Fully connected

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	
1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	;
2	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	;
3	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	;
4	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	;
5	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	;
6	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	;
7	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	;
8	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	;
9	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	;
10	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	;
11	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	;
12	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	;
13	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	;
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	;
15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	;
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	]

Hemispheric

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	
1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	;
2	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	;
3	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	;
4	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	;
5	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	;
6	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	;
7	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	;
8	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	;
9	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	;
10	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	;
11	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	;
12	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	;
13	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	;
14	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	;
15	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	;
16	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	]

Central
processor

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1
2	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1
3	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	1
4	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	1
5	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	1
6	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	1
7	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1
8	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1
9	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
10	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1
11	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1
12	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	1
13	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1
14	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1
15	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1
16	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	1
17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0

Bi-directional
loop

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	;
2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	;
3	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	;
4	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	;
5	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	;
6	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	;
7	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	;
8	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	;
9	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	;
10	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	;
11	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	;
12	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	;
13	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	;
14	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	;
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	;
16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	]

Uni-directional loop

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	;
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	;
5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	;
6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	;
7	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	;
8	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	;
9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	;
10	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	;
11	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	;
12	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	;
13	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	;
14	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	;
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	;
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	]

Hierarchical

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	;
5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	;
6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	;
7	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	;
8	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	;
9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	;
10	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	;
11	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	;
12	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	;
13	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	;
14	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	;
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	;
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	]

Modular

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	;
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	]