VISUAL PROBLEM SOLVING IN AUTISM, PSYCHOMETRICS, AND AI: THE CASE OF THE RAVEN'S PROGRESSIVE MATRICES INTELLIGENCE TEST

A Dissertation
Presented to
The Academic Faculty

By

Maithilee Kunda

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Computer Science

Georgia Institute of Technology

May, 2013

VISUAL PROBLEM SOLVING IN AUTISM, PSYCHOMETRICS, AND AI: THE CASE OF THE RAVEN'S PROGRESSIVE MATRICES INTELLIGENCE TEST

Approved by:

Dr. Ashok K. Goel, Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. Gregory Abowd
School of Interactive Computing
Georgia Institute of Technology

Dr. Nancy Nersessian
School of Interactive Computing
Georgia Institute of Technology

Dr. Brian Scassellati
Department of Computer Science
Yale University

Dr. Eric Schumacher
School of Psychology
Georgia Institute of Technology

Dr. Isabelle Soulières
Département de Psychologie
Université du Québec à Montréal

Date Approved: March 27, 2013

ACKNOWLEDGMENTS

This research was supported by an NDSEG fellowship from the Office of Naval Research, an NSF GRFP fellowship, NSF IIS grant (#0534622) titled, “Multimodal Case-Based Reasoning in Modeling and Design,” NSF RI grant (#1116541) titled, “Addressing Visual Analogy Problems on the Raven's Intelligence Test,” by Google through an Anita Borg scholarship, and by the Georgia Tech GVU Center through a Foley scholarship.

This dissertation would not have been possible without the support and inspiration of many individuals on many fronts. First and foremost, I would like to thank my advisor, Dr. Ashok Goel, for his mentorship over the past several years. From the very first class project that I undertook, he has always been open and encouraging towards my ideas, challenging me to meet high intellectual expectations in my research, and guiding me forwards with unfailing patience and enthusiasm.

Each of the members of my committee contributed significantly to this dissertation and to my growth as a researcher. Dr. Gregory Abowd introduced me to the growing field of work at the intersection of technology and autism, and I am continually inspired by his compassionate and influential approach to doing this kind of research. Dr. Nancy Nersessian helped to free my cognitive representations from being stuck inside my head, and she contributed many significant scholarly insights to guide this work. Dr. Eric Schumacher has been an invaluable liaison to the world of psychology and cognitive neuroscience, and his input has helped this dissertation be far more relevant with respect to these disciplines. Dr. Brian Scassellati was extremely helpful in advising on situating the AI aspects of this work, and his cheerful confidence in this line of research has been greatly appreciated. Last but not least, I thank Dr. Isabelle Soulières, who has been unfailingly supportive, all the way from the earliest stages of this research through to its conclusions, and also incredibly generous in sharing her time, expertise, and data.

I owe much gratitude also to Rosa Arriaga and Agata Rozga at Georgia Tech, who have taught me a considerable amount about autism and developmental psychology, to Fred Shic and Eli Kim at Yale, fellow computer scientists and lunchmates at autism conferences, and to Laurent Mottron and Michelle Dawson at the Université de Montréal, whose ground-breaking autism research has been both intellectually and personally inspirational.

I would also like to thank past and current members of the Design & Intelligence Lab who have provided a fun, stimulating, and wonderfully friendly atmosphere in which to conduct this work. Thanks especially to Josh Jones and Swaroop Vattam, our venerable lab elders; to Keith McGreggor, co-conspirator for all this crazy "visual thinking" stuff; to Michael Helms, qualifier study buddy and fellow weaver of fundamental theories of cognition; and to Bryan Wiltgen, whose contagious laughter has brightened many a day and who selflessly gave up many hours to be the "independent coder" for the work in Chapter 5 of this dissertation.

Finally, I would like to thank my family and friends for all of their love and support: my husband and best friend Jim Neidhoefer; my parents, who have always supported and encouraged me and taught me the value of getting an education; Prakash Kunda, the most excellent of big brothers; and Vivek Rao, Adam Roddy, Jacque Tio, and Karen Whitworth, longtime friends with whom to talk, laugh and/or commiserate, as needed, about life, the universe, and everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial intelligence
ASTI	Affine and Set Transformation Induction model
APM	Advanced Progressive Matrices
ASD	Autism Spectrum Disorder
AU/AUT	Autism
CPM	Colored Progressive Matrices
D	Difference error
ED	Executive Dysfunction
EPF	Enhanced Perceptual Functioning
FSIQ	Full-scale IQ
IC	Incomplete correlate error
NSGD	No significant group differences
NV	Nonverbal
PDD-NOS	Pervasive Developmental Disorder—Not Otherwise Specified
PIQ	Performance IQ
R	Repetition error
RPM	Raven's Progressive Matrices
RT	Reaction time
SPM	Standard Progressive Matrices
TD	Typically developing
TiP	Thinking in Pictures
V	Verbal

VMA	Verbal mental age
VIQ	Verbal IQ
WCC	Weak Central Coherence
WP	Wrong principle error

SUMMARY

Much of cognitive science research and almost all of AI research into problem solving has focused on the use of verbal or propositional representations. However, there is significant evidence that humans solve problems using different representational modalities, including visual or iconic ones. In this dissertation, I investigate visual problem solving from the perspectives of autism, psychometrics, and AI.

Studies of individuals on the autism spectrum show that they often use atypical patterns of cognition, and anecdotal reports have frequently mentioned a tendency to "think visually." I examined one precise characterization of visual thinking in terms of iconic representations. I then conducted a comprehensive review of data on several cognitive tasks from the autism literature and found numerous instances indicating that some individuals with autism may have a disposition towards visual thinking.

One task, the Raven's Progressive Matrices test, is of particular interest to the field of psychometrics, as it represents one of the single best measures of general intelligence that has yet been developed. Typically developing individuals are thought to solve the Raven's test using largely verbal strategies, especially on the more difficult subsets of test problems. In line with this view, computational models of information processing on the Raven's test have focused exclusively on propositional representations. However, behavioral and fMRI studies of individuals with autism suggest that these individuals may use instead a predominantly visual strategy across most or all test problems.

To examine visual problem solving on the Raven's test, I first constructed a computational model, called the Affine and Set Transformation Induction (ASTI) model, which uses a combination of affine transformations and set operations to solve Raven's

problems using purely pixel-based representations of problem inputs, without any propositional encoding. I then performed four analyses using this model.

First, I tested the model against three versions of the Raven's test, to determine the sufficiency of visual representations for solving this type of problem. The ASTI model successfully solves 50 of the 60 problems on the Standard Progressive Matrices (SPM) test, comparable in performance to the best computational models that use propositional representations. Second, I evaluated model robustness in the face of changes to the representation of pixels and visual similarity. I found that varying these low-level representational commitments causes only small changes in overall performance. Third, I performed successive ablations of the model to create a new classification of problem types, based on which transformations are necessary and sufficient for finding the correct answer. Fourth, I examined if patterns of errors made on the SPM can provide a window into whether a visual or verbal strategy is being used. While many of the observed error patterns were predicted by considering aspects of the model and of human behavior, I found that overall error patterns do not seem to provide a clear indicator of strategy type.

The main contributions of this dissertation include: (1) a rigorous definition and examination of a disposition towards visual thinking in autism; (2) a sufficiency proof, through the construction of a novel computational model, that visual representations can successfully solve many Raven's problems; (3) a new, data-based classification of problem types on the SPM; (4) a new classification of conceptual error types on the SPM; and (5) a methodology for analyzing, and an analysis of, error patterns made by humans and computational models on the SPM. More broadly, this dissertation contributes significantly to our understanding of visual problem solving.

1 INTRODUCTION

Stated most broadly, this dissertation is about problem solving, and it lies at the intersection of three areas of scientific research: autism, psychometrics, and artificial intelligence (AI). Within each area, I argue that there is a gap in the literature having to do with theories of problem solving, and in particular with theories of visual problem solving. This dissertation aims to fill in some of these gaps.

Problem solving has long been identified as a key component of intelligence, for humans, machines, and animals alike (Newell & Simon, 1972). It is intimately tied together with notions of goals and agency, and can be loosely defined as the process of acting from some set of starting conditions to achieve a specified end goal (Barbey & Barsalou, 1999). From an information processing standpoint, problem solving can be defined more specifically in terms of representations and reasoning. An agent begins with some starting representation(s) of a problem (and whatever other potentially relevant representations the agent can access, e.g. from the environment or from its internal memory), and then uses available reasoning mechanisms to operate over these representations to eventually reach the end goal, which is also represented in some fashion.

This description is left intentionally vague to remain inclusive towards the incredible diversity of problem solving that abounds in intelligent agents of all kinds, even from the perspective of a strict information processing view. Representations can be internal to an agent or external, and they can exist in numerous modalities, such as propositional, visual, spatial, auditory, haptic, proprioceptive, interoceptive, etc. (Markman, 1999). Reasoning, too, can take many forms, from classical notions of deductive and inductive reasoning to less formal variants such as analogical reasoning, reasoning using simulation or models, etc. (Leighton & Sternberg, 2004).

A considerable amount of research on problem solving, especially as it relates to human or human-level intelligence, has focused on problem solving using propositional representations (e.g. Newell & Simon, 1972). There are many reasons for this focus, three of which I mention here. The first reason is that propositional representations are extremely powerful and versatile. Propositional accounts of problem solving have been successful across many different domains, from logic and puzzles to communication and planning. Propositions also support composition and can represent distal information, for instance as pointers into separate data structures (Newell, 1994), and propositional systems like the theoretical Post production system have been proven to be Turing-universal (Minsky, 1967).

The second reason is that human language is fundamentally a propositional form of representation, and for eons, language has extensively shaped human thought and communication.[1] As a result, when we verbalize (whether internally or externally) our introspective perceptions of how we have solved a problem, we first necessarily convert whatever other forms of representations we might have actually used into propositional form. (Consider how our communications about problem solving might be different if human beings came equipped with small projectors on their bellies that could convey visual representations of our introspection!)

[1] The extent to which language influences thought has long been a controversial topic in the cognitive sciences. Opinions vary from one extreme position—that language is merely a tool for communication and is a transduced form of the stuff of which "thought" is made—to the other extreme position—that thought is fundamentally rooted in and composed of linguistic representations. Regardless of where on this spectrum science eventually leads, there is no question that language does play a critical role in our cognitive and communicative abilities.

The third reason is that conventional digital computation also uses fundamentally propositional forms of representation, and for the better part of a century, this particular computational paradigm has exerted enormous influence on our beliefs about the nature of intelligence. From a theoretical standpoint, the analogy of "mind-as-computer" helped conceptualize information processing as a framework within which to explain and understand intelligence.[2] However, it is a fallacy to equate "information processing" with "propositional information processing." Information of any kind may be computed upon, but the dominance of our digital, propositional computers has biased the kinds of information processing frameworks that we consider towards strictly propositional ones.

As a result, then, of the sheer efficacy of propositional representations, our propositional language, and our propositional computers, the vast majority of accounts of problem solving, especially the sorts of high-level problem solving that we associate with human-level intelligence, are propositional in nature. However, these accounts do not tell the whole story. One missing piece of the puzzle of intelligence can be found in the notion of visual problem solving. Accounts of visual problem solving have been in existence for a long time and surface in a variety of contexts, ranging from high-level domains like creativity and scientific discovery to simple tasks such as short term memory recall in children who have not yet developed verbal fluency (e.g. Clement, 2008; Hitch et al., 1989b; Nersessian, 2008).

However, despite the prevalence of descriptive accounts of visual problem solving, these accounts have historically not been as rigorously defined or as deeply investigated as comparable propositional accounts (Glasgow and Papadias, 1992; Larkin and Simon,

[2] The "mind-as-computer" analogy suggests, in its weakest form, that intelligence can be usefully modeled to some extent as a set of computational processes, and in its strongest form, that intelligence itself consists of computation.

1987; Naryanan, Glasgow, & Chandrasekaran, 1995). Again, there are many reasons for this lack of focus, of which I provide three here. First, the knowledge contained in a representation of any modality can be re-represented, to an arbitrarily good degree of precision, in propositional form. As a result, visual representations have often been dismissed as peripheral at best and purely phenomenological at worst, in either case being clearly subservient to propositional representations. For instance, while most information processing accounts would agree that inputs to an intelligent agent are often visual or perceptual in nature, and outputs are likewise perceptual (e.g. motor commands), the dominant view of the processing that goes on between these two endpoints has assumed that perceptions are converted into or out from propositional representations in order for high-level information processing to take place. This versatility of propositional representations also drove much of the debate on whether mental imagery really exists or is simply an illusory phenomenon that overlies propositional computation, because any processing characteristics of an account using visual representations could be simulated using propositional representations, and so positing a visual account was considered to be un-parsimonious, and indeed unnecessary.

The second reason is that, throughout the history of research on problem solving, it has not always been clear what is meant by visual representations. Different research threads adopt different definitions, ranging anywhere from external, high-level, diagrams that contain textual information to internal, low-level, neuron-like representations that contain only patterns of light-based activation. While there is similar diversity within the family of propositional representations, propositions enjoy the comfort of being directly specifiable at their core using formal mathematical definitions, whereas no similar fundaments of general visual representations have come into widespread agreement and use. As a result, different definitions of and assumptions about visual representations are often conflated in the literature, which has led to an overall lack of cohesion and integration among theories of visual representation and problem solving.

The third reason is that, from a practical standpoint, visual processing algorithms are relatively inefficient when implemented on propositional computers. Biologically, visual processing is massively parallel, yet the majority of digital computers in use today are serial or have at most a handful of parallel processing streams. Only recently have serial processors become fast enough to handle visual computations of any significant size, and the development of computing hardware that even comes close to matching the levels of parallelism found in biological visual systems is far from fruition.

As a result, then, of the tendency to subsume visual representations into more general propositional accounts, the lack of formal specification of visual representations, and our serial-processing computers, visual problem solving has not received the same kind of focused and sustained research attention that propositional problem solving has received. These twin biases in research—towards propositional accounts of problem solving and away from visual accounts of problem solving—would perhaps not be such a problem if it were acknowledged that visual problem solving remains an unexplored factor in many problem-solving domains. However, when combined with another unfortunate tendency in problem-solving research, these biases create a very significant problem indeed.

Consider the fact that there can be multiple ways to successfully solve the same problem. To explore this idea, I first define the notion of a **problem-solving strategy**:

Definition: A **problem-solving strategy** consists of some combination of particular forms of representations and reasoning that together attempt to solve a certain type of problem.

A problem can be considered to have an associated set of problem-solving strategies that can solve it to varying degrees of success. Within this set of all possible strategies, certain highly successful strategies may be preferred for various reasons. For instance, in the context of human cognition, a preferred strategy might be one used by most people, for evolutionary, neurobiological, or cultural reasons. In the context of intelligent machines, a preferred strategy might be one that is readily implemented in algorithmic form.

The majority of research into problem solving, across many different problem domains and within contexts of both human cognition and AI, has determinedly pursued the identification and study of preferred problem-solving strategies. This is entirely reasonable, given our goal of understanding general processes of problem solving in human cognition and in artificial intelligent agents. However, a major conceptual pitfall in the study of problem solving has been the assumption that the **preferred** strategy for solving a given problem is the **only** strategy for solving that problem, without taking into account other potentially relevant and successful strategies that may exist.

Within the context of the imbalance in research between visual and propositional accounts of problem solving, this general conceptual problem has specifically manifested itself in the following way:

Dissertation Problem Statement:

> **Many information processing accounts of problem solving, having first identified a preferred propositional strategy for solving a certain problem, fail to consider the possibility of visual strategies that can also solve the same problem.**

For the remainder of this chapter, I first identify more precisely the kinds of propositional and visual representations that this dissertation examines, particularly within each individual context of artificial computational systems (which I henceforth refer to as AI systems) and human cognition. Then, for each of the three research areas that I mentioned earlier—autism, psychometrics, and AI—I discuss how the general dissertation problem statement is motivated by research in this area, and I present the research questions, hypotheses, and research designs that guide my work.

1.1 Visual Versus Propositional Representations

This dissertation contrasts information processing accounts of problem solving that use either visual representations or propositional representations. I define these two types of representation according to the approach suggested by Nersessian (2008), in which these two representational types can be characterized along two different dimensions:

1) **Iconic vs. propositional:** The term *iconic* refers to representations that are analogical, in the sense that they carry some structural correspondence to what they represent (Chandrasekaran, 2011). *Propositional* representations, on the other hand, carry no such correspondence between format and content.
2) **Modal vs. amodal:** Symbols used in an iconic representation can be either modal or amodal, depending on whether they are rooted in perceptual states. I focus on modal representations in the visual modality.

There are four possible types of representations that emerge from this categorization:

1) **Amodal propositional:** A linguistic description of the shapes and relations in a problem would constitute an amodal propositional representation. For example: **is-left-of(triangle, circle)**.
2) **Amodal iconic:** A diagram indicating the spatial layout of shapes with each shape described linguistically would constitute an amodal iconic representation, in that the representation does show some structural correspondence with the problem, but the linguistic symbols are not themselves directly related to perceptual inputs. For example: **triangle ─ circle**.
3) **Modal propositional:** A linguistic description of the relationships between visual shapes, using the visual shapes themselves as the representation's

constituent symbols, could be considered to be a modal propositional representation. For example: **is-left-of(▲, ●)**. However, this would not really be a propositional representation, because such representations by definition cannot contain any non-propositional information, which would include modal symbols.

4) **Modal iconic:** An image showing the spatial layout of shapes as well as their visual appearance would constitute a modal iconic representation, in that the representation shows structural correspondence with the problem as well as with the visual perceptual state generated by looking at it. For example: ▲ — ●.

Of these four types, I focus on contrasting amodal propositional representations with modal iconic representations.

1.1.1 Representation in AI systems and in human cognition

In the context of AI systems, it is a straightforward matter to examine a system's internal representations in order to evaluate into which category they fall, because any representation will be implemented as a data structure that can be directly inspected. An argument could be made to trivialize representational distinctions by noting that since digital computers are fundamentally propositional devices, any representation used by such a system is also propositional. However, in classifying the representations used by an AI system, we must examine the representation at the level of abstraction at which its semantics are task-relevant, in other words looking at Marr's "algorithmic and representational" level instead of the "implementation" level (Marr, 1982).

Pixel-based images are one example of a modal iconic representation. Such images are clearly iconic, in that the patterns of color in a 2D pixel array have a structural correspondence with the visual appearance of whatever is shown in the image. Images are also modal, if we consider their input into the computer to be a form of perception; whether scanned, imported from a camera, or drawn on a digital canvas, the resulting image file remains in the same format in which it was initially "perceived" by the system.

Most of the familiar data structures in AI systems embody amodal propositional representations. Examples include logical predicates, semantic networks, frames, scripts, productions, etc. The key aspect of these representations is that they are manipulated according to syntactic rules (Newell & Simon, 1972). Because these representations have an arbitrary relationship to what they represent, the actual encoding can bear no influence on how they are used; any content-specific directives must be represented in additional propositional expressions.

In the context of human cognition, it is more difficult to specify and classify the internal representations that are in use, because they are largely unobservable and certainly less precisely defined than their computational counterparts. However, considerable progress has been made along this front with the advent of advanced neuroscience and particularly neuroimaging techniques.

Unlike early theories of cognition supposing that, like AI systems under the physical symbol system hypothesis, some universal amodal store maintained abstract "deep" forms of conceptual knowledge, knowledge representation in the brain is now thought to be largely (if not entirely) modal (e.g. Barsalou, 1999), in the sense that a particular concept is distributed across a network of activation that includes modality-specific, perceptual regions related to the representation of that concept. (While this view seems to have growing consensus among the research communities of cognitive psychology and neuroscience, the philosophical stance on the nature of mental representation, and especially of “pictorial” or other modal forms of mental representations, remains an area of some controversy; see Thomas, 2010, for a summary.) Knowledge is thought to be represented in an attribute- or feature-specific manner and may potentially have category-specific representational seats as well (Thompson-Schill, 2003). Again, we can distinguish between the representational/algorithmic level, at which representations are semantically relevant to the cognitive task at hand, and the implementation level, at which neural representations are instantiated within a myriad of neuronal and network properties. At some level of abstraction, knowledge may seem to require amodal representation, such as high-level conceptual knowledge, but this form of knowledge has been hypothesized to result from the convergence of many modalities of underlying knowledge, instead of the absence of modality; what we think of as amodal representations may be derived from this convergence of modalities in an online, as-needed fashion (Binder & Desai, 2011).

Most of the task domains that I examine, including the Raven’s Progressive Matrices test, involve mental representations in short term memory. I follow the theoretical framework laid out in Baddeley’s two-stage model of working memory, in which the phonological loop represents a short-term buffer for storing and rehearsing verbal phonological information, and the visuospatial sketchpad represents a corresponding short-term buffer for visuospatial information (Baddeley, 2003). Following the work of Kosslyn, among others (e.g. Kosslyn, Thompson, & Ganis, 2006), I conceptualize visual representations in the visuospatial sketchpad as being a form of mental imagery, with these representations exhibiting structural isomorphism with the elements that they represent. Section 3.1.3 discusses mental imagery in more detail, Section 2.3.5 addresses visual and verbal mental representations in long-term semantic memory, and Section 2.3.6 explores the potential implications of overtly linguistic versus non-linguistic propositional representations used for representing foundational concepts.

1.1.2 Terminology

Thus, when discussing AI systems, I distinguish between visual and propositional representations, and when discussing human cognition, I distinguish between visual and verbal representations. These terms and their associated properties, as I use them throughout this dissertation, are summarized in Table 1.

Table 1. Specification of representational terms used throughout this dissertation.

Context	Representation	Type	Primary example
AI system	Propositional	Amodal propositional	Propositions
	Visual	Modal iconic	Pixel-based images
Human cognition	Verbal	Amodal propositional	Verbal/linguistic representations
	Visual	Modal iconic	Mental imagery

In the three following sections, I discuss how the visual/verbal or visual/propositional divide has affected theories of cognition and problem solving in the fields of autism, psychometrics, and artificial intelligence, respectively.

1.2 Autism

Research into autism spectrum disorders (ASDs) has yielded many persistent and striking observations of atypical cognitive functioning across a variety of domains, but we still do not have a clear understanding of the depth or organization of these cognitive differences, their etiological significance, or their effects on the day-to-day experiences of affected individuals. As the prevalence of ASDs continues to rise, more precise characterizations of the cognition of these individuals will be crucial to continued efforts into basic research as well as the development of new methods for intervention, communication, and education.

A considerable amount of accumulated evidence suggests that certain individuals with autism exhibit a bias towards "thinking visually." This evidence comes from introspective accounts (e.g. Grandin, 2006; Hurlburt, Happé, & Frith, 1994), anecdotal observations by family members, therapists, and teachers (e.g. Quill, 1997), and empirical research into behavior (e.g. Heaton et al. 2008; Joseph et al. 2005; Whitehouse et al. 2006) and neurobiological functioning (e.g. Mottron et al., 2006). I call this the "Thinking in Pictures" (TiP) hypothesis about cognition in autism (Grandin, 2006). This hypothesis has two main parts:

Assumption: Typically developing (TD) individuals use both visual and verbal mental representations.

Hypothesis: A subset of individuals on the autism spectrum exhibits a disposition towards using visual mental representations and a corresponding bias against using verbal mental representations.

In cognitive research in autism, this hypothesis seems to have received limited focused and sustained consideration, despite the frequency with which accounts of visual thinking have emerged within the autism community. In particular, different types of problems are frequently investigated in autism research to identify patterns of cognitive strengths and weaknesses, but for a given problem, overall performance is assumed to rest on the same cognitive processes in individuals with autism as in TD individuals:

Problem Statement for Autism:

> **It is often assumed that individuals with autism solve a given problem in the same way as typically developing (TD) individuals, which precludes any study of the possibility that individuals with autism may have a specialized bias towards using visual strategies, as is suggested by numerous anecdotal accounts.**

To address this problem, I present two research questions:

1) To what extent is published research on cognitive tasks consistent with the TiP account, i.e. individuals with autism exhibit a bias towards using visual mental representations and away from using verbal ones?
2) To what extent does the TiP account provide a better explanation of published research on cognitive tasks than do other current theories of cognition in autism?

I studied both of these questions using the research design of narrative literature review. Chapter 2 describes details of the specific predictions that I tested, the methods used to conduct the literature reviews, and my results.

In particular, for question 1, I reviewed empirical studies that looked at different cognitive tasks in autism: the *n*-back task, serial recall, dual task studies, Raven's Progressive Matrices, semantic processing, false belief tasks, visual search, spatial recall, and visual recall. Results of this study are mixed. Certain task domains offer evidence that is highly consistent with and well explained by the TiP hypothesis, including: (1) the *n*-back task, (2) serial recall, (3) dual tasking, (4) Raven's Progressive Matrices, (5) semantic processing, and (6) false belief tasks. Other task domains, while not inconsistent with the TiP hypothesis, are not directly explained by it either, namely: (7) visual search. Finally, there are task domains whose data seem to contradict the TiP hypothesis, which are: (8) spatial recall, and (9) visual recall. The main finding of this study is that, at least across several task domains, there is a significant amount of evidence that is highly consistent with the TiP hypothesis, which empirically substantiates the anecdotal evidence for visual thinking that has long been common in the autism community.

For question 2, I reviewed the literature on four different existing theories about cognition in autism—Mindblindness, Executive Dysfunction, Weak Central Coherence, and Enhanced Perceptual Functioning. For each of these theories, I find that no theory explicitly posits the differences in representational strategy use in autism that I observed during the first study, and TiP does explain many of the findings presented in support of each theory.

The main contributions of this work are listed below:

1) I have compiled comprehensive literature reviews of several cognitive tasks that have been studied in autism research.

2) I have provided a concrete definition of the Thinking in Pictures (TiP) hypothesis and generated specific empirical predictions arising from it.
3) I have performed a systematic and empirical evaluation of this hypothesis using existing published data.
4) I have identified key areas in which the predictions of TiP differ from the predictions of other theories, which therefore warrant further study to better distinguish among these theories.
5) I have clearly identified certain experimental design and data interpretation pitfalls, in particular showing many instances in which published explanations are not consistent with data from additional studies.

There are many important areas for future work, including identifying the individuals or subsets of individuals with autism or other ASDs to whom the TiP account does or does not apply, investigating more deeply the similarities and differences in predictions made by TiP and other cognitive theories of autism, and understanding the role that TiP might play in neurobiological and developmental accounts of autism.

1.3 Psychometrics

Psychometrics is one of the three research traditions in the study of mental phenomena that had its roots in the late 1800s, together with experimental psychology and psychophysics. These traditions aimed to study mental phenomena in a systematic, empirical way, without relying on introspection as a means of inquiry. Psychophysics focused on relationships between physical properties of stimuli and the perceptions they would engender in human subjects. Experimental psychology was mainly concerned with the specific behaviors exhibited by subjects in controlled experimental conditions. Psychometrics, as the name suggests, aimed to develop scales of measurement to quantify individual differences in mental capabilities, particularly intelligence.

However, in pursuing a mapping from individual differences in intelligence onto numerical scales, psychometrics has flattened the notion of individual differences to imply only differences of degree rather than differences of degree and of kind. Thus, the numerical measurements obtained from conventional psychometric instruments are not often used to represent qualitative differences in strategy and especially differences in representation, despite evidence that humans can use different strategies to accomplish the same task.

Interestingly, the importance of strategy variations in intelligent behavior was recognized early on by Alfred Binet, who created the first intelligence test in its modern form. For instance, based on extensive observations and experimentation with his two young daughters, Binet explicitly pondered the possibility for and role of individual differences in strategy when solving the same problems. He made similar observations in his studies of savant-type individuals, finding, for example, that of two calculating prodigies, one seemed to perform his calculations using auditory mental representations, while the other seemed to use visual mental imagery (Fancher, 1985). However, as intelligence testing became more widely adopted, Binet's nuanced position was lost in favor a view of intelligence as a unitary, unidimensional construct, especially during the promotion of intelligence testing as part of the eugenics movement (e.g. Goddard, 1922).

This conceptualization of intelligence and psychometric testing implicitly contains one of two assumptions about their relationship. The weaker assumption is that it does not matter *how* an individual solves particular test items; it just matters *whether* they can solve them. The stronger assumption is that test items themselves elicit very particular strategies, and these strategies taps into some facet of "general intelligence." In other words, all individuals solve psychometric tests in the same way, and the resulting scores measure quantitative variations in a unidimensional cognitive ability.

With respect to the stronger assumption, psychology is slowly discovering that it is not the case that all individuals exhibit the same qualitative forms of cognition, as exemplified by the earlier discussion of atypical cognition in autism. With respect to the weaker assumption, one way to conceptualize the existing divide is in terms of "correlational and experimental approaches to human cognitive activity" (Keating, 1984, p. 17), where the former refers to traditional psychometric approaches and the latter refers to information-processing accounts of cognition. Understanding the strategies that underlie individual performance on psychometric tests will not only give better insight into what cognitive abilities particular tests are actually measuring, but will also shed light on the nature of individual differences in cognition, beyond just numerical variations on a unidimensional scale.

Some strides have been made towards trying to incorporate strategy differences into psychometric theory. Keating & Bobbitt admit that many sources of variance can contribute to measures of mental ability, including: 1) quantitative differences in cognitive processing efficiency, 2) differences in strategy, or 3) differences in metacognitive abilities of strategy selection, though they only attempt to experimentally address the first of these (Keating & Bobbitt, 1978). In the context of sentence-picture verification tasks, Hunt addresses in some detail the impacts of strategy differences on resulting behavioral measures, though he does observe that the question of how the existence of strategy differences can be reconciled with the fairly robust statistical evidence of a *g* factor for intelligence is an important open question (Hunt, 1980). Finally, Mislevy and Verhelst have attempted to explicitly incorporate strategy differences into item response theory, though this theoretical approach has not been widely applied (Mislevy & Verhelst, 1990).

Problem Statement for Psychometrics:

> **Psychometrics generally assumes that quantitative variations in test scores reflect quantitative variations in cognitive ability among individuals, often neglecting the possibility that qualitative strategy variations, such as between visual and verbal strategies, can also play a role in the score achieved by an individual.**

To investigate this problem, I focus on one particular psychometric test: Raven's Progressive Matrices (RPM). The RPM is a collection of widely-used standardized intelligence tests consisting of analogy problems in which a matrix of geometric figures is presented with one entry missing, and the correct missing entry must be selected from a set of answer choices. Correlation studies have found the RPM to be the single best measure of intelligence among any psychometric instrument, barring standard multi-domain measures that contain items in many different formats (Snow, Kyllonen, & Marshalek, 1982). Due to its ease of administration and scoring, as well as the fact that it requires little verbal instruction or explicit verbal comprehension, the RPM is widely used as a test of intelligence in clinical, educational, occupational, and scientific settings.

There is considerable evidence from neuroimaging and behavioral studies that humans use both visual and verbal strategies for solving RPM problems. However, the most widely cited computational model of information processing on the RPM posits a purely propositional strategy, with individual differences modeled as variations within this strategy (Carpenter, Just, & Shell, 1990), and most other accounts of information processing on the RPM are likewise propositional.

To investigate the nature and existence of visual problem solving strategies on the RPM, I present three research questions:

1) To what extent can a purely visual strategy, implemented as a computational model, be successful on the RPM tests?
2) How can this model be used to classify problems on the RPM according to their information processing demands?
3) To what extent can errors made on the RPM serve as behavioral markers to indicate the use of a visual versus verbal strategy?

For question 1, I constructed a computational model, called the Affine and Set

Tranformation Induction (ASTI) model, which uses affine transformations and set operations on purely visual, pixel-based representations of RPM problems to generate solutions. I tested this model against the three main tests in the RPM family: the Standard Progressive Matrices (SPM), the Colored Progressive Matrices (CPM), and the Advanced Progressive Matrices.

For question 2, I conducted computational experiments to test how systematic ablations of the ASTI model, removing various functionalities in different combinations, affect performance on individual RPM problems. Using this approach, I was able to define minimally sufficient sets of visual reasoning mechanisms necessary to solve particular problems, which provides a problem classification in terms of the information processing demands of each problem.

For question 3, I performed an observational study of the error patterns made by typically developing individuals, individuals with autism, and the ASTI model. I first developed a classification of errors on the Standard Progressive Matrices (SPM), one version of the RPM tests, using qualitative coding by two independent raters. Then, I analyzed the types of errors made on the SPM among the three groups.

The main contributions of this work are listed below:

1) I have provided a proof by construction that a visual strategy is sufficient for solving a large proportion of problems on the RPM tests.
2) I have tested a single computational model against all three major RPM tests, and I have provided a new classification of problem types across this family of tests, based on whether and which of a particular set of visual reasoning mechanisms is sufficient for solving a particular problem.

3) I have developed a new classification of error types found on the SPM.
4) I have performed the first comparison of errors made on the SPM among TD individuals and individuals with autism, and also the first comparison of errors made among human groups and a computational model.

There are many important areas for future work, including identifying what other types of visual reasoning might be successful on RPM problems and other psychometric instruments, the role that strategy differences play in the predictive functionality of psychometric measures, and how such tests might be augmented to include more explicit measures of qualitative variations in the representations used among individuals taking the test.

1.4 Artificial Intelligence

The field of artificial intelligence has long focused on building systems that use propositional representations to behave intelligently (for some definition of intelligence). This arises partially from the pragmatic reasons outlined earlier, having to do with the ease with which propositional strategies can be implemented on conventional digital computers, but also comes from the philosophical stance on intelligence best exemplified by the physical symbol system hypothesis. This hypothesis states that certain properties in a system, namely being physically instantiated and having the ability to process symbols in a syntactic fashion, are necessary and sufficient for implementing intelligent behavior (Newell & Simon, 1976). Partly stemming from this philosophical stance and partly propagating it has been the tendency in AI to focus on problem domains for which propositional representations can be quite successful (e.g. chess, logic) and neglect those for which propositional representations do not provide a ready answer (e.g. navigation, social reasoning).

However, this approach neglects much of what we know about visual representations and problem solving. From the perspective of human cognition, while there is considerable evidence for the use of mental-imagery-based problem solving strategies, few AI systems have attempted to reason using visual representations (see Glasgow & Papadias, 1992 for one example). Many AI systems that work in visual domains first convert problem inputs into propositional form and then solve the problem (e.g. Chandrasekaran et al., 2011; Davies, Goel, & Nersessian, 2009; Larkin & Simon, 1987). From the perspective of building intelligent machines, the AI focus on propositional representations ignores evolutionary arguments for the primacy and importance of perceptual representations for very basic forms of intelligent behavior (Brooks, 1991). In fact, the difficulty AI has faced in problem domains like navigation and visual recognition suggests that perhaps propositional representations may *not* be sufficient for general intelligence. One important exception can be found in how, in recent decades, the field of computer vision has emerged from its early roots in AI to focus on the processing and understanding of visual information, though computer vision does not address how visual representations might be used in high-level problem solving.

Problem Statement for Artificial Intelligence:

> **AI problem-solving systems built to date have focused almost exclusively on using propositional representations, even for the representation of visual information, but these accounts do not model the uses of mental imagery observed in human problem-solving.**

To investigate this problem, I consider the same problem domain described in the previous section: the Raven's Progressive Matrices (RPM) test. As I mentioned, there is evidence from behavioral and neuroimaging studies of humans for the use of both visual and verbal strategies on the RPM, and yet the vast majority of computational models of the RPM have focused purely on propositional representations of the visual knowledge found on the test.

This work builds upon a long line of related research on analogical reasoning. These studies have shown how functional and causal knowledge of physical systems enables analogical reminding and transfer in both within-domain analogies (Goel, Bhatta, &

Stroulia, 1997; Goel & Chandrasekaran, 1988) and cross-domain analogies (Goel & Bhatta, 2004; Griffith, Nersessian, & Goel, 2000). In this work, functional and causal knowledge was represented propositionally.

Later work showed that visual knowledge and reasoning alone can address some classes of analogy problems that had been assumed to require causal knowledge and reasoning (Davies & Goel, 2001; Davies, Goel, & Yaner, 2008). This research also showed how visual analogies can account for several aspects of creative problem solving in scientific discovery (Davies, Nersessian, & Goel, 2005) and engineering design (Davies, Goel, & Nersessian, 2009). All of these studies, however, used propositional representations; while the content of knowledge was visuospatial, the form of representation was still propositional.

In the growing cognitive science literature on analogy, several other lines of research have explored visual analogies (e.g. Casakin & Goldschmidt, 1999; Clement, 2008; Croft & Thagard, 2002; Davies & Goel, 2008; Evans, 1968; Hofstadter, 1995; Leyton, 2001; Nersessian, 2008; Ojha & Indurkhya, 2009; Stafford, 2001; Yaner & Goel, 2006). Many factors explain this emphasis on visual analogy: for example, the requirements of task and domain, explanations of behavioral data, and consistency with theories of mental imagery. Another important reason is that visual analogies support the construction of representations as well as re-representations, as advocated, for instance, by Indurkhya (1998) and Kokinov (1998). Indeed, Chalmers, French, and Hofstadter (1992) view much of analogy as high-level perception in which representations are constructed rather than assumed as given. These various theories of visual analogy, however, differ along dimensions of modal/amodal and iconic/propositional representations.

To investigate problem solving on the RPM using visual representations from a computational perspective, I present two research questions:

1) To what extent can a purely visual strategy, implemented as a computational model, be successful on the RPM tests?
2) How do changes in the underlying representational commitments of the model affect its behavior?

For question 1, as mentioned in the previous section, I constructed a computational model called the "ASTI model" to solve RPM problems using purely visual representations. For question 2, I conducted computational experiments to test how changing the representations used by the ASTI model, in particular through varying the model's perceptual interpretation of figure-ground separation and through altering its similarity function, affect its performance on the SPM.

The main contributions of this work are listed below:

1) I have developed and implemented a model for visual reasoning using affine and set transformations for high-level problem solving.
2) I have tested this model against unedited and uninterpreted versions of the RPM tests, which is the first instance of a computational model that can solve RPM problems using images scanned directly from the paper test booklets.

There are many important areas for future work, including identifying what other classes of problems the ASTI model might be able to address, exploring the extent to which the ASTI model provides a model of the types of visual reasoning humans perform on the RPM, and augmenting the ASTI model with additional visual reasoning capabilities to further evaluate the problem-solving power provided by a visual strategy.

1.5 Dissertation Overview

Table 2 summarizes the problem statements, research questions, and studies presented in this introductory chapter. The remainder of this dissertation proceeds as follows:

- Chapter 2 presents the results of the two studies that I performed to evaluate the Thinking in Pictures (TiP) hypothesis about cognition in autism.
- Chapter 3 first summarizes data on strategy use on the Raven's Progressive Matrices (RPM) tests from both computational and behavioral studies. I then describe the representations, reasoning, and problem solving architecture of the ASTI model, and presents results generated from testing the ASTI model against the RPM tests and from experiments that vary the representational commitments made by the model.
- Chapter 4 summarizes existing approaches to problem type categorization on the RPM tests and then presents results of ablation experiments performed using the ASTI model to generate a new classification of RPM problems based on their information processing demands.
- Chapter 5 summarizes existing approaches for analyzing the errors made by individuals on the RPM, describes a new classification of error types on the Standard Progressive Matrices (SPM) test, and then presents results of an observational study comparing error patterns on the SPM among TD individuals, individuals with autism, and the ASTI model.
- Chapter 6 summarizes the conclusions and synthesis of this work.

Table 2. Summary of problems, research questions, and studies in this dissertation.

Problem	Research questions	Studies
It is often assumed that individuals with autism solve a given problem in the same way as typically developing (TD) individuals, which precludes any study of the possibility that individuals with autism may have a specialized bias towards using visual strategies, as is suggested by numerous anecdotal accounts.	To what extent is published research on cognitive tasks consistent with the Thinking in Pictures (TiP) account, i.e. individuals with autism exhibit a bias towards using visual mental representations and away from using verbal ones?	Chapter 2: Narrative literature review of nine cognitive tasks
	To what extent does the TiP account provide a better explanation of published research on cognitive tasks than do other current theories of cognition in autism?	Chapter 2: Narrative literature review of four cognitive theories
Psychometrics generally assumes that quantitative variations in test scores reflect quantitative variations in cognitive ability among individuals, often neglecting the possibility that qualitative strategy variations, such as between visual and verbal strategies, can also play a role in the score achieved by an individual.	To what extent can a purely visual strategy, implemented as a computational model, be successful on the RPM tests?	Chapter 3: Constructive proof by building ASTI model
	How can this model be used to classify problems on the RPM according to their information processing demands?	Chapter 4: Ablation experiments using ASTI model
	To what extent can errors made on the RPM serve as behavioral markers to indicate the use of a visual versus verbal strategy?	Chapter 5: Observational study of errors made on the RPM
AI problem-solving systems built to date have focused almost exclusively on using propositional representations, even for the representation of visual information, but these accounts do not model the uses of mental imagery observed in human problem-solving.	To what extent can a purely visual strategy, implemented as a computational model, be successful on the RPM tests?	Chapter 3: Constructive proof by building ASTI model
	How do changes in the underlying representational commitments of the model affect its behavior?	Chapter 3: Experiments on ASTI model representations

2 VISUAL PROBLEM SOLVING IN AUTISM

In this chapter, I explore what I call the Thinking in Pictures (TiP) hypothesis about cognition in autism. This hypothesis is named after the well-known autobiography of Temple Grandin (2006), a woman with autism who believes that she "thinks in pictures." Grandin often describes how her visual thinking style provides benefits in some areas, such as in drafting and visualizing complex machinery for her work in engineering design, but also creates difficulties in other areas, such as in forming categories or understanding abstract concepts in her day-to-day life. Numerous other individuals on the autism spectrum have also posited that they tend to use visual mental representations instead of verbal ones (e.g. Hurlburt et al. 1994), though this phenomenon has not been documented in a systematic way.

In particular, I address two research questions:

1) To what extent is published research on cognitive tasks consistent with the TiP account, i.e. individuals with autism exhibit a bias towards using visual mental representations and away from using verbal ones?
2) To what extent does the TiP account provide a better explanation of published research on cognitive tasks than do other current theories of cognition in autism?

I first describe my predictions and methods, and I then present the results of my research on both questions, followed by a discussion of my claims and areas for future work.

2.1 Predictions of the TiP Hypothesis

A simplistic consideration of the TiP hypothesis might lead to predictions that individuals with autism will show good performance on visual tasks and poor performance on verbal tasks. In fact, general evidence suggesting a visual/verbal disparity among individuals on the autism spectrum can be found in studies of cognitive profiles, or patterns of verbal (V) versus nonverbal (NV) intelligence as measured by standardized IQ tests. Some studies have noted a V < NV (lower verbal than nonverbal IQ) pattern among individuals on the autism spectrum (Lincoln et al. 1988), though such findings have not been universal (Klin et al. 1995; Siegel et al. 1996). Joseph et al. (2002) found that, while children with autism were generally more likely to have a V-NV discrepancy in either direction than were TD children, children with autism having a V < NV pattern of abilities showed greater social impairment than the other children with autism, irrespective of absolute levels of verbal or general ability. The distinctiveness of the V < NV profile, and also its association with variables of diagnostic interest, led the authors to conjecture that such a profile might indicate "an etiologically significant subtype of autism" reflecting fundamental changes in cognition and neuroanatomy, rather than just the selective sparing of certain nonverbal abilities.

However, looking purely at existing measures of visual and verbal ability does not take into account the possibility of alternate strategies. In particular, psychology has classified cognitive tasks as visual or verbal according to how they are typically solved, without accounting for the alternate classification of how they might be solved using alternate strategies. Figure 1 illustrates the potential overlap between these two types of task classifications. The solid and dashed circles (**A** and **B**) represent tasks that **can** be solved visually or verbally, respectively, and their intersection *(***A** ∩ **B***)* represents tasks that can be solved either way. For example, matching one of two very similar shades of red to a target red patch can be solved using visual representations but not using verbal ones (at least not easily), and so this task lies inside solid circle **A** but outside dashed circle **B**. On the other hand, determining which of the words *shoe* or *now* rhymes with the word *too* can be solved using phonological verbal representations but not using visual ones, and so this task lies inside dashed circle **B** but outside solid circle **A**. Finally, deciding which of two red and green colored patches matches a target red patch can be solved using either visual or verbal representations (e.g. by matching on visual hue or on linguistic label), and so this task lies in the intersection **A** ∩ **B**.

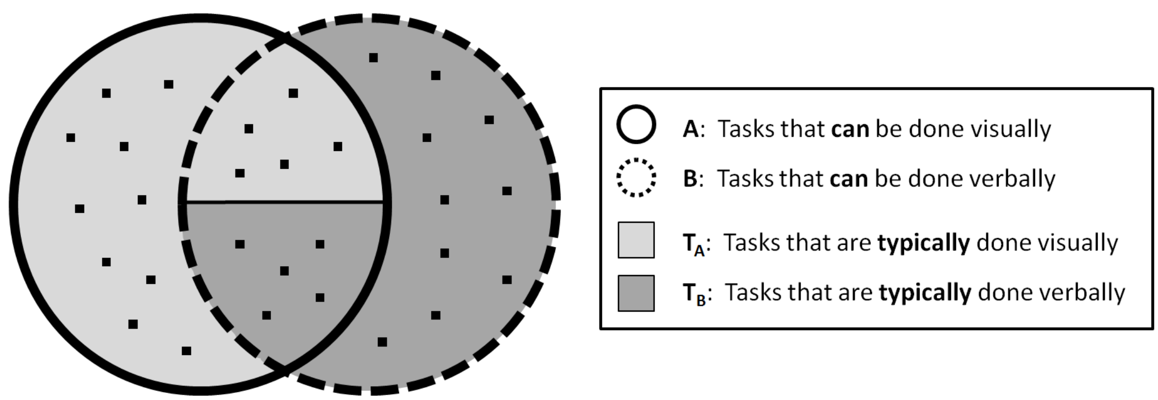


Figure 1. Task classifications according to how they can be solved (solid and dashed circles) and how they are typically solved (light and dark grey shadings).

The light grey and dark grey shaded regions ($\mathbf{T_A}$ and $\mathbf{T_B}$) represent tasks that are **typically** solved visually or verbally, respectively. The bulk of psychological evidence on how most humans solve cognitive tasks has given us $\mathbf{T_A}$ and $\mathbf{T_B}$, by definition. However, for a typically verbal task in $\mathbf{T_B}$, if that task happens to also be solvable visually (i.e. lies within **A** ∩ **B**), it is possible that an individual disinclined to use verbal representations can use a visual strategy to successfully solve that task, and vice versa. By making these distinctions, the performance of an individual on a given task can be evaluated independently of their strategy selection. Keeping this is mind, the TiP hypothesis can be used to make general predictions about the behavior of individuals with autism on three different types of tasks, as shown in Table 3.

Table 3. Behavioral predictions of the TiP hypothesis for strategy use and performance of typically developing (TD) individuals and individuals with autism (AUT).

Task type	**Task description**	**TD strategy**	**TD performance**	**AUT strategy**	**AUT performance**
exclusively in **B**	tasks that can only be done verbally	verbal	successful	visual	impaired
in $\mathbf{T_A}$	tasks typically done visually	visual	successful	visual	successful
in $\mathbf{T_B} \cap \mathbf{A}$	tasks typically done verbally that can be done visually	verbal	successful	visual	successful

The first prediction is, perhaps, the least useful for testing the TiP hypothesis, as impaired performance on verbal-only tasks is unlikely to inform us about what mental representations an individual who thinks in pictures is using; for instance, such individuals may not be engaging any task-relevant representations at all. Also, data from these tasks will not be very useful as a point of distinction between the TiP hypothesis and other cognitive-deficit accounts of autism.

Regarding the second prediction, that individuals with autism use visual strategies to solve tasks that are also typically solved visually, a conservative claim might be that the visual strategies used by the two groups are the same, and therefore no behavioral differences in either task performance or strategy selection ought to be observed. However, there is significant evidence for behavioral differences in autism on typically visual tasks, ranging from changes in low-level perception (e.g. Bertone et al. 2005) to superior performance on certain visual tasks like the Embedded Figures Task (e.g. Jolliffe & Baron-Cohen, 1997). One possible TiP explanation of these differences is that a bias towards using visual representations leads to a general "visual expertise" not shared by TD individuals. However, these findings can also been interpreted as indications of other forms of atypical cognitive processing, for example of greater detail-oriented processing (Happé & Frith, 2006) or superior low-level perceptual abilities (Mottron et al. 2006). If such processing differences are an integral aspect of autism, one important question for TiP will be how such differences might be related to a visual representation bias. In general, data from typically visual tasks are not necessarily the best test of the TiP hypothesis, as they alone cannot distinguish between the TiP account and other cognitive theories that posit superior or atypical aspects of cognitive processing in autism.

The third prediction, regarding tasks typically solved verbally that can also be solved visually, is the most useful for directly testing the TiP hypothesis. In particular, for a given task in this category, not only should there be evidence of successful performance by individuals with autism, but it should also be possible to measure secondary behavioral or neurological characteristics that point to a difference in the underlying strategy being used. This third TiP prediction provides the surest means of distinguishing TiP from other cognitive theories of autism, as (insofar as I have seen) no other cognitive account of autism explicitly posits visual/verbal strategy differences.

2.2 Methods

The first research question I address is this chapter is: to what extent is published research on cognitive tasks consistent with the TiP account, i.e. individuals with autism exhibit a bias towards using visual mental representations and away from using verbal ones? The specific predictions of the TiP hypothesis are:

1) On tasks that can only be done verbally, individuals with autism will show impaired performance.
2) On tasks that are typically done visually, individuals with autism will show intact performance.
3) On tasks that are typically done verbally but can be done visually, individuals with autism will show intact performance and will also exhibit secondary behavioral or neurological markers that point to the use of a visual strategy.

The first prediction is already consistent with general evidence for verbal impairments in autism—verbal impairment is, in fact, a definitional criterion of the disorder (*DSM-IV-TR*, 2000)—and so I did not address this prediction during the course of this research.

To test the second two predictions, I conducted a narrative literature review that sampled from the space of cognitive tasks found in the autism literature to provide a qualitative analysis of published data on each task. Tasks were selected according to two specific criteria. First, there had to be some evidence that the task fell into the relevant task category, i.e. for prediction #2, the task is typically done visually, and for prediction #3, the task is typically done verbally but at least some evidence exists to suggest that the task is amenable to successful solution using a visual strategy. Second, there had to be at least two published studies in the autism literature comparing performance on the task between individuals with autism and TD individuals.

The tasks that were selected according to these criteria are:

1) Tasks typically done verbally that can be done visually:
    a) The *n*-back task
    b) Serial recall
    c) Dual task studies
    d) Raven's progressive matrices
    e) Semantic processing
    f) False belief tasks
2) Tasks typically done visually:
    a) Visual search
    b) Spatial recall
    c) Visual recall

To find the studies of each task, potential articles were first located through searching online using Google Scholar or following reference trails in articles that I had already found. The time period was inclusive from the beginnings of cognitive research in autism until the time of the study in 2010. Studies considered for this review were restricted to those in which the diagnostic group had a diagnosis of "Autistic disorder." Studies in which this diagnosis was predominant, but a few individuals had other diagnoses on the autism spectrum, such as Asperger's Syndrome or Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) were included. Studies of participants only with diagnoses of Asperger's or PDD-NOS were excluded. Finally, studies were not screened on the basis of any other exclusionary criteria such as sample size or demographic characteristics.

The second research question addressed in this chapter is: to what extent does the TiP account provide a better explanation of published research on cognitive tasks than do other current theories of cognition in autism? In particular, for each existing cognitive theory of autism, I made the following two predictions:

1) The theory cannot explain empirical data showing a visual bias in autism for tasks that are typically done verbally but can be done visually.
2) For the major representative cognitive tasks that are currently cited in support of the theory, TiP provides at least as good an explanation for established findings.

To test these predictions, I performed another narrative literature review centered on each major cognitive theory in autism: Mindblindness, Executive Dysfunction, Weak Central Coherence, and Enhanced Perceptual Functioning. These theories are commonly cited as the main cognitive theories of autism. For each theory, I selected a small set of example tasks cited in the literature and provided a qualitative analysis of results and interpretations on these tasks. These selected tasks were restricted to those for which there was a published literature review in summary articles for each cognitive theory that examined the historical body of research on the task.

2.3 Review of Cognitive Tasks in Autism

2.3.1 The n-back task

In the *n*-back task (Kirchner, 1958), a subject is presented with a sequence of stimuli and asked whether the current stimulus matches that shown *n* steps ago. The variable *n* can take the value of one (respond "yes" to any succession of two identical stimuli), two (respond "yes" to any stimulus matching that presented two steps back), etc. Stimuli can vary in content and presentation, e.g. letters presented visually or auditorily, pictures, etc.

For TD individuals, the *n*-back task is thought to recruit verbal rehearsal processes in working memory (i.e. *phonological* verbal representations), among other executive resources (Smith & Jonides, 1999). Several published studies of the *n*-back task have not shown significant differences in accuracy or reaction time for individuals with autism relative to TD controls (see Table 4), which has led, in some cases, to the conclusion that verbal working memory is intact in autism (Williams et al. 2005).

However, recent fMRI studies have shown that, while behavioral measures on the *n*-back task may be similar, there can be significant differences in patterns of brain activation between individuals with autism and TD controls. In one study using stimuli of visually presented letters, the autism group showed less brain activation than controls in left prefrontal and parietal regions associated with verbal processing and greater activation in right hemisphere and posterior regions associated with visual processing (Koshino et al. 2005). In another study using stimuli of photographs of faces, a similar decrease in left prefrontal activation was found in the autism group (Koshino et al. 2008). Both of these studies suggest that individuals with autism may be using a visual strategy for the *n*-back task, whereas controls use at least a partially verbal strategy.

Table 4. Review of results from published studies of the n-back task in autism

Reference	N	Age	VIQ	PIQ	FSIQ	Task details	Results
Koshino et al. 2005	14	25.7	102.6	—	100.1	n = 0, 1, 2; visually presented letters.	NSGD on accuracy or RT. Significant group differences in patterns of brain activation.
Koshino et al. 2008	11	24.5 (10.2)	106.1 (14.1)	102.1 (13.8)	104.5 (13.1)	n = 0, 1, 2; grayscale face pictures.	NSGD on accuracy or RT. Significant group differences in patterns of brain activation.
Ozonoff & Strayer, 2001	25	12.94 (3.18)	94.6 (18.5)	99.3 (19.9)	96.3 (17.8)	n = 1, 2; visually presented colored shapes.	NSGD on accuracy or RT. RT tended to be correlated with VIQ.
Williams et al. 2005	31	26.58 (8.68)	111.10 (16.47)	103.13 (16.64)	108.65 (16.75)	n = 0, 1, 2; visually presented letters.	NSGD on accuracy or RT.
	24	11.75 (2.36)	112.50 (16.53)	106.38 (14.21)	109.67 (16.07)	n = 0, 1, 2; visually presented letters.	NSGD on accuracy or RT.

Note. N = number of participants; VIQ, PIQ, and FSIQ = verbal, performance, and full-scale IQ, respectively; NSGD = no significant group differences; RT = reaction time. Age and IQ values are shown as: *mean* (*standard deviation*) in years and scores, respectively. All subjects diagnosed with autism, all control groups TD, and NSGD in age or shown IQ measures.

2.3.2 Serial recall

In serial recall tasks, a subject is presented with a sequence of randomly ordered stimuli and then asked to reproduce the sequence in order, after a short delay. These tasks generally involve the visual or auditory presentation of letters, numbers, words, or pictures, after which the subject has to verbally repeat the sequence or point to items in the correct order.

For TD individuals, serial recall tasks are thought to recruit primarily verbal rehearsal processes in working memory (i.e. *phonological* verbal representations), for instance as evidenced by decreased memory spans for long words—the *word length effect*—or for phonologically similar items—the *phonological similarity effect* (Baddeley, 2003). These verbal effects are seen even with visually presented stimuli in TD children above seven years of age, suggesting that in later development, TD individuals tend to recode visual stimuli into a verbal form (Hitch et al. 1989). In younger TD children, there is evidence for visual (and not verbal) encoding of visual stimuli in the form of decreased memory spans for visually similar items—the *visual similarity effect* (Hitch et al. 1989b).

Several published studies on serial recall tasks show no significant group differences in overall performance between individuals with autism and controls (see Table 5). As with the *n*-back task, these data are often used to indicate intact verbal working memory in autism. For example, standardized tests such as the WISC and the WRAML use number and letter span subtests as components of verbal IQ, and individuals with autism have often shown peaks of ability on these particular subtests (Siegel et al. 1996). However, additional behavioral data, such as the presence or absence of the word length or similarity effects described in the previous paragraph, should be considered to determine what strategy an individual is actually using.

Two studies have examined the robustness of the word length effect in individuals with autism. Russell et al. (1996) found, for auditorily presented stimuli, no difference in word length effect in a verbal response condition between children with autism and TD controls as well as a group with moderate learning disabilities, but, oddly, the autism group's word length effect actually increased in a nonverbal (pointing) response condition. In contrast, Whitehouse et al. (2006) used visually presented stimuli with verbal responses and found a smaller word length effect in the autism group than in TD controls. Also, the word length effect increased in the autism group in an overt labeling condition, suggesting that the autism group may have relied to a lesser extent on verbal encoding than controls when not biased to do so by having to produce labels.

Williams et al. (2008) looked at a similar recall task with visually presented stimuli and verbal responses and measured the robustness of the phonological similarity and visual similarity effects in children with autism and in a control group with learning disabilities. They found no group differences in recall performance, but when subjects were divided by their verbal mental age (VMA), those with VMA over 7 years had better overall recall performance and a significant phonological similarity effect but no visual similarity effect, while subjects with VMA less than 7 years exhibited the opposite pattern. In other words, this study found VMA to better predict strategy use than did diagnostic group, and additional analyses found VMA to be a better predictor than cognitive profiles as well (Williams & Jarrold, 2010). While the authors of this study did not discount the significance of cognitive profile in predicting strategy use, they cautioned against treating it as the only variable of relevance, and they also pointed out the importance of looking at variables like VMA and cognitive profile, in addition to diagnostic group, in assessing results in experimental studies of autism. On both of these points, we wholeheartedly agree, and the question of how to experimentally identify and analyze data from subgroups within the ASD population is central to the continued understanding of how individuals with autism utilize various forms of mental representations.

In summary, many studies have reported individuals with autism achieving similar levels of performance on serial recall tasks as TD individuals, but at least some of these studies have found evidence of a visual strategy bias in autism.

Table 5. Review of results from published studies of serial recall tasks in autism

Reference	N	Age	VIQ	PIQ	FSIQ	Items	Presentation	Response	Results
Ameli et al. 1988	16[f]	22.7 (4.9)	81 (16)	90.6 (13.5)	83 (14)	digits	auditory	verbal	NSGD in digit span.
Bennetto et al. 1996	19[ac]	15.95 (3.3)	82.32 (15.2)	98.11 (15.9)	88.89 (11.1)	digits	auditory	verbal	NSGD in digit span.
Joseph et al. 2005	24[a]	8.9 (2.3)	94 (19)	99 (20)	96 (18)	words	auditory	pointing	NSGD in correct responses.
						pictures	visual	self-ordered pointing	NSGD in errors in nonverbal condition. AU had more errors in verbal than nonverbal condition and than controls.
Minshew et al.1992	15	21.13 (8.02)	98.53 (21.63)	92.87 (10.72)	95.73 (13.61)	digits	auditory	verbal	NSGD in digit span.
						digits	auditory	verbal	NSGD in correct responses.
Minshew et al. 1997	33	20.91 (9.69)	102.48 (16.35)	97.45 (11.19)	100.09 (12.96)	digits	auditory	verbal	NSGD in digit span.
						digits	auditory	verbal	NSGD in correct responses.
Minshew & Goldstein, 2001	52	22.33 (9.59)	94.96 (17.56)	91.52 (12.95)	92.88 (15.06)	letters	auditory	verbal	NSGD in correct sequences.
						words	auditory	verbal	AU impaired in number of correct sequences.
O'Connor & Hermelin, 1967	12[g]	11.8	---	---	---	words	auditory	verbal	NSGD in number recalled. Greater position and recency effects in AU than controls.
						pictures	visual	manual ordering	NSGD in number recalled.
Ozonoff & Strayer, 2001	25[e]	12.94 (3.18)	94.6 (18.5)	99.3 (19.9)	96.3 (17.8)	colored boxes	visual	self-ordered pointing	NSGD in number of perseverative errors, which tended to correlation with VIQ.

Table 5 continued. Review of results from published studies of serial recall tasks in autism

Russell et al. 1996	33[dh]	12.38 (2.95)	---	---	---	words	auditory	verbal, pointing	MLD had lower span than other groups. AU had greater nonverbal WLE than verbal WLE and than MLD group.
Whitehouse et al. 2006	23[i]	11	---	---	---	pictures	visual	verbal	NSGD in correct responses. AU had smaller WLE than controls and in silent vs. label condition.
Williams et al. 2005	24	11.75 (2.36)	112.50 (16.53)	106.38 (14.21)	109.67 (16.07)	digits, letters	auditory	verbal	NSGD in score. Correlated with FSIQ in AU but not in controls.
Williams et al. 2006	38	11.68 (2.46)	106.42 (15.97)	100.55 (14.19)	103.82 (14.29)	digits, letters	auditory	verbal	NSGD in score.
Williams et al. 2008	25[bc]	12.25 (3.08)	77.16 (15.25)	76.84 (20.27)	74.84 (15.99)	pictures	visual	verbal	NSGD in span. Subjects with high VMA had greater PSE than VSE; vice versa for low VMA.

Note. N = number of participants; VIQ, PIQ, and FSIQ = verbal, performance, and full-scale IQ, respectively; NSGD = no significant group differences; AU = autism group; MLD = mild learning disabilities group; WLE, PSE, and VSE = word length, phonological similarity, and visual similarity effects, respectively; VMA = verbal mental age. Age and IQ values are shown as: *mean* (*standard deviation*) in years and scores, respectively. All subjects diagnosed with autism, all control groups TD, and NSGD in age or shown IQ measures, except as noted.

[a] Subject group diagnosed as autism/PDD-NOS. [b] Subject group diagnosed as ASD. [c] Control group comprised of individuals with moderate learning difficulties (MLD). [d] Controls grouped as TD and MLD. [e] Controls grouped as TD and Tourette's Syndrome. [f] Subject group had slightly lower PIQ than controls, but NSGD in age. [g] NSGD in nonverbal ability according to Peabody Picture Vocabulary Test. [h] NSGD in VMA, but AU had higher age. [i] NSGD in verbal, nonverbal, and reading ability, but AU had higher age.

2.3.3 Dual task studies

Dual task studies aim to discern task strategy choices by looking at whether executing a simultaneous secondary task interferes with performance (Brooks, 1968). The basic assumption of the dual-task paradigm is that, because different cognitive modalities (e.g. visual versus verbal) draw upon separate and limited cognitive resources, performing two tasks simultaneously using the same modality will degrade performance more than performing two tasks that use different modalities (Jonides et al. 1996; Navon & Gopher, 1979). Whether a *primary task* uses a certain modality can be determined by finding out whether the simultaneous execution of a secondary task known to involve those resources affects performance (Baddeley & Hitch, 1974). Secondary tasks (a.k.a. *suppression tasks*) can be very simple, so there is little ambiguity about what cognitive resources are being used. Verbal or articulatory suppression (i.e. recruiting *phonological* verbal representations) often consists of repeating a word out loud. Visuospatial suppression can include holding an image in memory or performing a simple tapping or pointing task.

Dual task studies offer a good test of the TiP hypothesis, because their results can clearly indicate, for a particular individual or group, whether visual or verbal cognitive resources are necessary for some primary task. In particular, across a range of primary tasks typically done verbally (tasks for which controls show impairments under verbal but not visual suppression), the TiP hypothesis predicts that individuals with autism will show impairments under visual but not verbal suppression. Only a handful of dual task studies have been performed with individuals on the autism spectrum, and although none have had exactly this form, all have shown results generally consistent with the TiP hypothesis, though not necessarily interpreted as such.

García-Villamisar and Della Sala (2002) used a primary task of serial recall, with verbal recall of auditorily presented digits, and a secondary suppression task of visuomotor tracking, in which subjects had to manually mark a series of boxes on paper. No group differences were found for either task performed singly, but when performed together, the autism group showed a significant impairment on both tasks, while the control group showed no impairment. The authors read these results as marking a general deficit in simultaneous task performance in autism, but these data could also indicate that the group with autism was using a visual strategy for the digit span task, which, unlike the verbal strategy used by controls, was open to interference from the visual suppression task. Moreover, as discussed below, other dual task studies in autism have not found evidence of a general dual-tasking deficit.

Whitehouse et al. (2006) conducted a dual-task experiment in which the primary task was task-switching in written arithmetic, in which subjects had to alternately add and subtract pairs of numbers, and the secondary task was verbal suppression, with subjects repeating "Monday" out loud. No group differences were found in latency or accuracy in the single-task condition. However, the control group showed an increase in latency under articulatory suppression, matching previous studies on task switching in TD individuals (Baddeley et al. 2001; Emerson & Miyake, 2003), while the autism group did not. These results go against the idea of a general impairment in dual task performance in autism and also suggest that the autism group used a nonverbal (though not necessarily visual) task-switching strategy. Lidstone et al. (2009) re-analyzed these data divided by cognitive profile and found that the lack of a latency increase under articulatory suppression was limited to children with autism having a V < NV profile, irrespective of absolute levels of verbal ability. Controls with a V < NV profile did show impaired dual task performance under articulatory suppression, as did children with a V = NV profile in both groups. Wallace et al. (2009) looked at the Tower of London planning task as the primary task, with a secondary task of articulatory suppression, and similarly found that the control group showed a significant impairment in their primary task performance under articulatory suppression, whereas the autism group showed no such impairment.

Holland and Low (2010) repeated the task switching experiment of Whitehouse et al. (2006) but with an added visuospatial suppression task, with subjects tapping out a simple pattern on a set of blocks using their non-dominant hand. As in the study by Whitehouse et al. (2006), there were no significant group differences in latency or accuracy in the single-task condition. Dual task results showed that the autism group exhibited an increase in task-switching latency under visuospatial suppression but not under articulatory suppression, while the control group showed a similar latency increase under both suppression conditions. Similar dual-task results were obtained in a second experiment that looked at a Tower of Hanoi planning task. At first glance, these data seem to suggest that the autism group used visuospatial but not verbal resources for task-switching and planning, while controls used both visuospatial and verbal resources for both tasks. However, in the task-switching experiment, both groups also showed an increase in latency under visuospatial suppression for a baseline, non-task-switching version of the arithmetic task, suggesting that the visuospatial suppression task may have interfered with peripheral, non-task-switching demands of the primary task. For instance, the visuomotor demands of tapping blocks with the non-dominant hand while writing arithmetic answers with the dominant hand may have been in contention, in which case the visuospatial suppression task did not really target high-level task-switching resources.

While none of these dual-task studies taken singly provides a definitive test of the TiP hypothesis, together they are highly suggestive of individuals with autism using visual strategies for certain tasks that are typically done verbally.

2.3.4 Raven's Progressive Matrices

As mentioned in Chapter 1, Raven's Progressive Matrices (RPM) is a standardized intelligence test that consists of problems resembling geometric analogies, and published literature on the RPM suggests that TD individuals use a combination of visual and verbal strategies on this task. Recent research on the RPM has found an interesting discrepancy between the performance of TD individuals and individuals with autism. Whereas the RPM scores of TD individuals are usually strongly matched by their Wechsler IQ scores, individuals with autism have demonstrated RPM scores much higher than their Wechsler scores (Bölte et al. 2009; Dawson et al. 2007; Mottron, 2004). Individuals with Asperger's syndrome have shown a similar pattern (Hayashi et al. 2008).

One explanation for the RPM/Wechsler score discrepancies found in autism is that, while poor verbal abilities may in fact decrease performance on full-scale IQ tests, intact visual abilities might be recruited to successfully solve many RPM problems, which are after all visually presented and do not explicitly require any overt verbal abilities. Consistent with this explanation, Soulières et al. (2009) found, using fMRI, that individuals with autism had lower brain activation in prefrontal and parietal areas associated with language and working memory and higher activation in visual occipital areas than TD individuals did while solving the RPM. On a related but non-RPM set of matrix reasoning tasks, Sahyoun et al. (2009) found evidence through measures of response latency that individuals with autism exhibited a bias towards using visuospatial mediation, whereas TD individuals and individuals with Asperger's seemed able to use verbal mediation in solving the problems.

2.3.5 Semantic processing

Evidence from neuropsychology has suggested that visual and verbal semantic memory are somewhat dissociated, in that brain lesions can selectively impair the use of one or the other (Hart & Gordon, 1992). However, whether this dissociation reflects two separate, modality-specific semantic stores or a single store with multiple, modality-specific access schemes is unclear (Caramazza, 1996; Farah & McClelland, 1991). Either way, the TiP hypothesis predicts that individuals with autism have privileged or primary access to visual semantic information, whereas TD individuals are capable of accessing both visual and verbal semantics.

In one well-designed fMRI study, Kana et al. (2006) studied brain activation in individuals with autism and TD individuals while they answered true/false questions about high or low imagery sentences. High imagery sentences included statements like, "The number eight when rotated 90 degrees looks like a pair of eyeglasses," while low imagery sentences included statements like, "Addition, subtraction, and multiplication are all math skills." One way to conceptualize these two classes of stimuli is as follows:

(a) High imagery sentences require semantic understanding plus visual reasoning.

(b) Low imagery sentences require semantic understanding only.

The control group showed a significant difference between the high and low imagery conditions, with the high imagery condition eliciting more activity from temporal and parietal regions associated with mental imagery as well as from inferior frontal regions associated with verbal processing. This pattern fits the model that visual regions are used for visual reasoning, while verbal regions are used for lexical and semantic processing. (The baseline used for both conditions was a fixation task that involved no linguistic processing.) In contrast, the autism group showed similar activation in both conditions, with less activity in inferior frontal language regions than the control group in the high imagery condition, and greater activity in occipital and parietal visual regions in the low imagery condition. This pattern suggests that the individuals with autism may have used visual regions for both visual reasoning and semantic processing.

Many other studies have found significant differences in brain activity during semantic processing tasks between individuals with autism and TD controls, although the precise patterns of results have varied. Like the study by Kana et al. (2006), Gaffrey et al. (2007) found increased activation in posterior visual regions and decreased activation in frontal verbal regions for individuals with ASD during a task of determining whether a word belonged to certain semantic categories (tools, colors, and feelings), with a baseline perceptual processing task. However, Just et al. (2004), in a study of sentence comprehension with a fixation baseline, found reduced activity in visual, occipito-parietal regions in subjects with autism compared to TD controls, though the autism group did also show decreased activity in frontal language regions. Harris et al. (2006) found similar results of reduced frontal language region activation in an ASD group compared to TD controls during a word judging task with a perceptual processing baseline, and also found that the ASD group showed more similar activation in some language regions between the semantic and perceptual tasks than did the control group. In contrast, Knaus et al. (2008) used a response-naming task with a perceptual processing baseline and found that subjects with ASD had greater activation in frontal and temporal language areas than did TD controls.

One important factor in neuroimaging studies of semantic processing is the choice of a baseline task. For TD individuals, lexical-semantic tasks are often paired with perceptual processing tasks that use letter or word stimuli, in order to remove any perceptual components of the semantic understanding process. However, if a subject uses visual neural machinery to do semantic processing, then it is possible that subtracting the brain activation due to a perceptual processing task may remove semantic-related activation in visual regions as well.

In addition to these neuroimaging studies, several behavioral studies have also looked at semantic processing in individuals with autism. Kamio and Toichi (2000) used a word-completion task in which semantic priming was provided using either picture cues or word cues. TD controls performed similarly under both conditions, but the autism group performed much better with picture cues than word cues, suggesting that they were better able to retrieve verbal information through pictorial representations than through other verbal representations. Lopez and Leekam (2003) found that children with autism were as capable as TD controls of using visual semantic context to facilitate object identification; the same pattern was found for verbal semantic information, though ceiling effects were a possible confound in the verbal case.

In summary, while existing data are mixed, current modality-specific models of semantic memory (whether modality-specific in indexing alone or in storage as well) make semantic processing a good candidate for further testing of the TiP hypothesis.

2.3.6 False belief tasks

False belief tasks represent one experimental paradigm for testing theory of mind abilities, which center on the attribution of mentalistic or belief states to external entities. Theory of mind, in turn, represents one component of social cognition. False belief tasks comprise one domain that is widely found to be impaired among individuals on the autism spectrum (see review in Happé, 1995), and deficits in theory of mind (e.g. Mindblindness) and other aspects of social cognition have been suggested to be a central facet of autism (Baron-Cohen, 1995; Baron-Cohen & Belmonte, 2005).

One classic test of false belief understanding is the Sally-Anne task (Wimmer & Perner, 1983), in which the subject is shown a skit with two dolls, Sally and Anne. Sally places a marble into a basket and, after Anne leaves the room, moves the marble from the basket into a box. The subject is then asked where Anne will look for the marble when she returns. Responding correctly, that Anne will look in the basket, requires an understanding of Anne's *false belief* that the marble is still in the basket; Anne's belief is false in that it represents something that the subject watching the skit knows is not true.

Many interpretations of false belief task performance in autism posit that there is some fundamentally social deficit that leads to impaired theory of mind abilities (e.g. Baron-Cohen, 1995). One contrasting view is that false belief impairments in autism stem from a domain-general bias against using verbal representations, not from a domain-specific difference in social cognition. In particular, verbal mental age has been found to be strongly correlated with performance on false belief tasks in both individuals with autism and in TD controls (Happé, 1995; Yirmiya et al. 1998). While this pattern seems amenable to a straightforward TiP interpretation, it raises the question of precisely how verbal mental representations might be related to false belief tasks.

One possibility is that standard false belief tasks, which require explicit language comprehension and responding, overtax the weak language skills of individuals with autism. However, individuals with autism also show impairments on nonverbal analogues of false-belief tasks such as eye-tracking studies, making this explanation unlikely (Senju et al. 2009; Senju et al. 2010).

A second possibility is that *linguistic* verbal mental representations are required for developing concepts of false belief, on which both verbal and nonverbal versions of false-belief tasks rely (e.g. Fernyhough, 2008). However, two-year-old TD infants exhibit visual attentional patterns that seem to draw upon an understanding of false beliefs before significant linguistic abilities have developed (Southgate et al. 2007). While there is almost certainly a strong connection between linguistic representations and theory of mind abilities, these types of eye-tracking studies cast doubt on whether the relationship is strictly causal and sequential.

A third possibility, which we espouse, is that verbal representations are, after all, used to form false belief concepts, but where "verbal" in this case refers to *propositional* representations, not *linguistic* representations. Propositions can be thought of as the building blocks of a low-level representational system, where a single proposition takes the form of a related set of symbols that carries semantic meaning. Linguistic representations occur at a much higher level of abstraction than propositions and are explicitly tied to a particular language.

The idea of false belief impairments in autism having a low-level representational origin is not new; constructing false belief concepts has been described as requiring, for instance, the representation of complements (de Villiers & de Villiers, 2003; Hale & Tager-Flusberg, 2003) or meta-representation (Leslie, 1987). The gist of these arguments is that, in order to represent a false belief, an individual must have some mechanism for representing a statement as being held to be true in one context (e.g. as believed by an agent in a story), alongside the property of its being false in a different context (e.g. in the story itself). Recent modeling work in cognitive architectures has found that this type of information structure can be easily represented using propositions (Bello & Cassimatis, 2006).

From this perspective, individual performance on “mental” and “non-mental” versions of false belief tasks should be correlated. While for a time, several visual tasks such as the false photograph, false map, and false drawing tasks were thought to be appropriate non-mental analogues of false belief tasks (e.g. Leekam & Perner, 1991; Charman & Baron-Cohen, 1992; Leslie & Thaiss, 1992), Perner and Leekam (2008) have argued that these tasks do not tap the same representational structure as standard false belief tasks. Instead, they propose that the false sign (or false signal) task is the more appropriate non-mental analogue, and in support of their claim, correlated patterns of impairments have been observed in autism on the false signal task and standard false belief tasks (Bowler et al. 2005). These results support the view of false belief competency being more a function of domain-general *representational* ability than of domain-specific *social* ability.

Nevertheless, studies investigating performance across visual reasoning tasks, such as the false photograph and map tasks, and theory of mind reasoning tasks, such as standard false belief tasks, do reveal some very interesting patterns of results. Figure 2 gives a summary of results from studies that compared the performance of individuals with autism on visual false-item tasks with their performance on standard false belief tasks. Each data point represents the performance of a single group, as indexed in the legend, on the visual versus false belief tasks. Numbers within each data point refer to the experiment number in Table 6, which contains demographic and experimental design information. Note that this figure is intended as a qualitative illustration of trends in published data and not as a strict quantitative analysis.

These data, along with similar results on other visual tasks such as matching the state of a true model or photograph to a room (Charman & Baron-Cohen, 1995) and visual perspective taking (Reed & Peterson, 1990), suggest that many individuals with autism who show impaired performance on false belief tasks can exhibit intact or superior performance on these types of visual tasks. TD individuals, in contrast, seem equally predisposed towards having strengths on either one or the other (or both) of these task types, which we would expect if they emerge independently in normal development.

Finally, if false belief impairments in autism are due to deficits in underlying propositional representations, then false belief tasks may seem to fall under the first TiP prediction, regarding tasks only solvable verbally. However, there have been some recent attempts to help individuals with autism represent false belief concepts visually, for instance using thought bubbles or photograph-in-the-head analogies (McGregor et al. 1998a, 1998b; Swettenham et al. 1996; Wellman et al. 2002). These studies have generally shown positive results in teaching subjects to pass specific false belief tasks but less success in leading subjects to transfer their knowledge to new tasks.

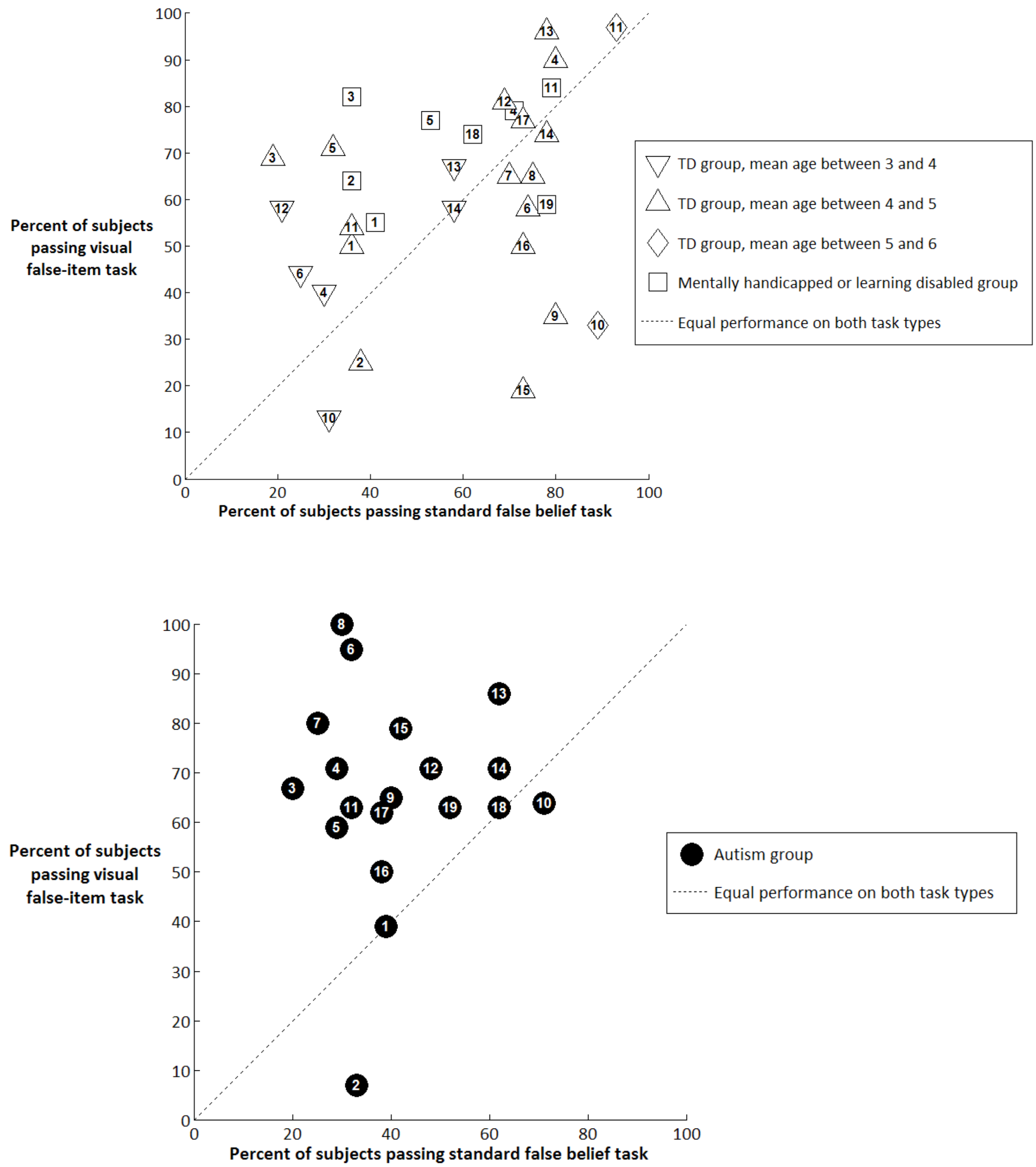


Figure 2. Summary of results from published studies of standard false belief versus visual false-item tasks in TD and learning disabled individuals (top) or individuals with autism (bottom).

Table 6. References and experiment details from published studies of false belief-type tasks versus visual false-item tasks in autism.

Reference	N	Controls	Age	VMA	Type	Visual	Remarks	#
Bowler & Briskman, 2000	18	TD4, MH	11.3 (3.9)	4.6 (1.3)	location	false belief with photo cue		1
	15	TD4, MH	11.6 (3.7)	5.3 (2.3)	location	false belief with photo cue	between-groups design	2
					identity	false belief with photo cue	between-groups design	3
Charman & Baron-Cohen, 1992	17	TD3, TD4, MH	13.6 (3.9)	5.3 (2.0)	identity	false drawing		4
Charman & Lynggaard, 1998	21	TD, MH	10.9 (2.1)	4.8 (0.9)	identity	false belief with photo cue		5
Leekam & Perner, 1991	20	TD3, TD4	16.2 (2.8)	6.4 (1.8)	identity	false photograph	AU included PDD	6
Leslie & Thaiss, 1992	12	TD4	12	6.3	location	false photograph		7
					identity	false photograph		8
	18	TD4	11.4	6.7	location	false map		9
Parsons & Mitchell, 1999	15	TD3, TD5	12	7.5	location	false photograph	AU with one Asperger's subject	10
	19	TD4, TD5, MH	13.5	5.7	identity, location	false belief with thought bubble cue	two task pairs combined	11
Peterson & Siegal, 1998	21	TD3, TD4, deaf	10	9.1	location	false photograph		12
	21	TD3, TD4, deaf	9.6	--	identity	false photograph	same standard false belief task	13
					identity	false drawing		14
Peterson, 2003	14	TD4, deaf	9.7	--	identity	false belief about drawing		15
	8	TD4, deaf	8.3	--	identity	false belief via drawing	same standard false belief task	16
					identity	false belief about drawing		17

Table 6 continued. References and experiment details from published studies of false belief tasks vs. visual false-item tasks in autism.

Russell & Hill, 2001	27	MH	11.3 (3.5)	6.2 (1.8)	identity	false belief about drawing		18
					identity	false belief about drawing	standard false belief was a location task	19

Note. Results are illustrated in Figure 2 (# indicates data point label). N = number of participants; VMA = verbal mental age; TD3 = typically developing three- to four-year-olds; TD4 = typically developing four- to five-year-olds; TD5 = typically developing five- to six-year-olds; MH = mentally handicapped or learning disabled; AU = autism group. Age and VMA are shown as: *mean* (*standard deviation*) in years. Type indicates whether the false-belief (or other false-item) queried in each experiment had to do with the identity or location of stimuli. All subjects diagnosed with autism, except as noted in **Remarks**.

2.3.7 Visual search

One widely reported area of superior performance for individuals on the autism spectrum is visual search. For example, individuals on the spectrum have repeatedly demonstrated more accurate and/or more efficient performance on the Embedded Figures Task (EFT), in which a small figure must be located within a larger, more complex one (see review in Happé & Frith, 2006). Several recent papers have looked at classic target/distracter visual search tasks and have found similar patterns of superior performance by individuals on the autism spectrum, often through faster response latencies (see Table 7). Moreover, faster search performance in autism often grows more pronounced with more difficult search tasks, e.g. for conjunctive vs. feature search.

Studies of the EFT using fMRI have shown that individuals with autism tend to recruit more occipital visual processing brain regions for this task, whereas TD controls recruit more frontal and parietal working memory regions (Manjaly et al. 2007; Ring et al. 1999). However, looking at a target/distracter search task, Keehn et al. (2008) found increased activation in individuals on the autism spectrum compared to TD controls in both frontoparietal and occipital regions. This study also found that, while patterns of activation differed for controls between an easy feature search task and a more difficult one, no such differences were found for the autism group. In addition, significant group differences in eye-movement patterns (Keehn et al. 2009) and in sensitivity to task parameters (Baldassi et al. 2009) have been found on visual search tasks. These results are often explained by theories that posit processing strengths in autism, and in particular, some recent evidence suggests that enhanced low-level perceptual discrimination may contribute to faster search in autism (Joseph et al. 2009).

In general, many studies point to the existence of significant and widespread differences between individuals on the autism spectrum and TD individuals on visual search tasks and in overall patterns of visual attention, and these differences seem to developmentally precede many other cognitive processes (Brenner et al. 2007). Specific relationships between the TiP hypothesis and visual search and attention remain to be determined, especially in terms of development and basic perceptual processes.

Table 7. Review of results from published studies of target/distracter visual search tasks in autism.

Reference	N	Age	CPM	Search	Target	Distracters	Accuracy	RT
Jarrold et al. 2005	18[b]	12.42 (1.98)	21.56 (7.45)	F	red jumping clown	green skinny clown, red fat clown	---	AU faster.
				C	red jumping clown	green jumping clown, red skinny clown	---	AU faster.
Keehn et al. 2008	9[a]	15.1 (2.5)	---	F	upright *T*	right-pointing *T*	NSGD.	NSGD.
				F	upright *T*	right, left, bottom-pointing *T*	NSGD.	NSGD.
O'Riordan, 2000	11	9.2 (0.8)	27 (4)	C	red *X*	red *T*, green *X*	NSGD.	AU faster.
				C	red *X*	red letters and colored *X*	NSGD.	AU faster.
	12	9.4 (0.9)	29 (4)	C	red *X*	red *T*, green *X*	NSGD.	AU faster.
				C	green *T*	red *T*, green *X*	NSGD.	AU faster.
				C	red *X*, green *T*	red *T*, green *X*	NSGD.	AU faster.
O'Riordan, 2004	10	22.0 (3.6)	---	F	*N*	*P*, *Q*	NSGD.	NSGD.
				C	*R*	*P*, *Q*	NSGD.	AU faster.
				F	ellipse	circle	NSGD.	AU faster.
				C	red *X*	green *X*, red *C*	NSGD.	NSGD.
				C	red *F*	pink *F*, red *E*	AU more accurate.	NSGD.

Table 7 continued. Review of results from published studies of target/distracter visual search tasks in autism

O'Riordan & Plaisted, 2001	15	9.2 (1.1)	28 (3)	C	vertical red bar	vertical green bar, horizontal red bar	NSGD.	AU faster.
				C	vertical red bar	vertical green line, horizontal green bar, horizontal green line	NSGD.	NSGD.
				C	vertical red bar	vertical green bar, horizontal red bar, vertical red line	NSGD.	AU faster.
	13	9.0 (1.1)	28 (3)	C	red *X*	green *X*, red *C*	NSGD.	NSGD.
				C	red X	pink *X*, red *C*	NSGD.	AU faster.
				C	red *F*	green *F*, red *E*	NSGD.	AU faster.
				C	red *F*	pink *F*, red *E*	NSGD.	AU faster.
O'Riordan et al. 2001	12	8.4 (0.9)	26 (4)	F	red *S*	green *X*, red *T*	NSGD.	NSGD.
				C	red *X*	red *T*, green *X*	NSGD.	AU faster.
	12	8.8 (10.0)	28 (3)	F	tilted line	vertical lines	NSGD.	NSGD.
				F	vertical line	tilted lines	NSGD.	AU faster.
Plaisted et al. 1998	8[c]	8.8 (1.2)	---	F	red *S*	red *T*, green *X*	NSGD.	NSGD.
				C	red *X*	green *X*, red *T*	AU more accurate.	AU faster.

Note. N = number of participants; CPM = Raven's Colored Progressive Matrices; RT = reaction time; F = feature search; C = conjunctive search; AU = autism group; NSGD = no significant group differences. Age and CPM values are shown as: *mean* (*standard deviation*) in years and raw scores, respectively. All subjects diagnosed with autism, all control groups TD, and NSGD in age or cognitive measures, except as noted.

[a] Subject group diagnosed as ASD. [b] AU had higher chronological age. [c] Autism group had higher block design scores.

2.3.8 Spatial recall

Serial spatial recall tasks are a part of many psychometric tests, such as Finger Windows in the WRAML. These tasks involve the presentation of a sequence of spatial locations (e.g. holes on a card or blocks on a table), which the subject has to manually reproduce. Another spatial recall task uses self-ordered pointing, in which the subject must point to locations not previously selected. Both paradigms require the subject to reproduce a set or sequence of spatial locations. Individuals with autism often, but not always, show impaired performance on these tasks, and no studies of spatial recall were found in which the autism group showed superior performance (see Table 8).

Given that serial recall for items or objects appears to be unimpaired in autism, there appears to be a dissociation between how well individuals with autism remember visually discriminable items vs. visually indiscriminable spatial locations. Although these results seem to contradict the TiP hypothesis, one explanation could be that the visual representations used by individuals with autism do not, by themselves, represent spatial information adequately. In line with this idea, on tasks that combine visual and spatial information (i.e. recalling locations of visually discriminable stimuli), individuals with autism have shown intact performance (Ozonoff & Strayer, 2001; Williams et al. 2006).

Another possibility might be that spatial recall tasks actively recruit verbal working memory; correlations between spatial span and speech rate have been found in TD individuals, without similar correlations between spatial span and tapping or spatial movement rate (Chuah & Maybery, 1999; Smyth & Scholey, 1992, 1996). Studies have also found that articulatory suppression can interfere with spatial span tasks (Jones et al. 1995; Smyth et al. 1988; Smyth & Pelky, 1992).

Table 8. Review of results from published studies of spatial recall tasks in autism.

Reference	N	Age	VIQ	PIQ	FSIQ	Task	Details	Results
Caron et al. 2004	16[a]	17.6 (6.3)	102.2 (21.2)	112.3 (12.9)	107.7 (13.1)	Maze route learning	learn choice points in an actual maze	NSGD in accuracy or speed of going through maze.
Edgin & Pennington, 2005	24[a]	11.46 (2.32)	104.40 (20.24)	---	---	CANTAB Spatial Working Memory	touch one of identical boxes to find tokens (self-ordered pointing)	NSGD in errors or strategy. Both correlated with age and block design score in both groups.
Luna et al. 2002	11	32.3 (9.3)	106.7 (14.9)	96.5 (10.5)	102.7 (12.1)	Oculomotor Delayed Response	shift gaze to previous stimulus location	AU impaired in saccade accuracy compared to controls. Not impaired on baseline saccade task.
Minshew & Goldstein, 2001	52	22.33 (9.59)	94.96 (17.56)	91.52 (12.95)	92.88 (15.06)	Maze recall	learn choice points in hidden maze	AU impaired on complex mazes but not on simpler maze.
Minshew et al. 1992	15	21.13 (8.02)	98.53 (21.63)	92.87 (10.72)	95.73 (13.61)	Maze recall	learn choice points in hidden maze	NSGD.
Minshew et al. 1997	33	20.91 (9.69)	102.48 (16.35)	97.45 (11.19)	100.09 (12.96)	Maze recall	learn choice points in hidden maze	NSGD.
Minshew et al. 1999	26	20.2 (8.5)	98.5 (16.9)	90.1 (12.7)	94.0 (14.1)	Oculomotor Delayed Response	shift gaze to previous stimulus location	AU impaired in saccade accuracy compared to controls. Not impaired on baseline saccade task.
Morris et al. 1999	15[b]	29.5	99	100.1	---	Executive Golf Task	touch one of identical golf holes to find putts (self-ordered pointing)	AS impaired on within and between search errors.
Steele et al. 2007	29	14.83 (5.47)	107.52 (13.02)	106.21 (11.82)	107.76 (10.99)	CANTAB Spatial Working Memory	touch one of identical boxes to find tokens (self-ordered pointing)	AU showed greater errors and lower strategy score, both correlated with PIQ but not VIQ. TD scores not correlated with any IQ.
Verté et al. 2005	61[de]	9.1 (1.9)	---	---	99.2 (17.1)	Corsi block tapping	forward tapping recall of blocks	AU impaired.

Table 8 continued. Review of results from published studies of spatial recall tasks in autism.

Verté et al. 2006	50[f]	8.7 (1.9)	93.1 (18.0)	104.0 (17.8)	98.2 (17.3)	Corsi block tapping	forward tapping recall of blocks	AU impaired.
	37[b]	8.5 (2.1)	105.6 (20.0)	104.0 (15.9)	105.2 (16.3)	Corsi block tapping	forward tapping recall of blocks	AS impaired.
	25[cf]	8.5 (1.4)	93.3 (14.7)	105.8 (19.2)	98.3 (14.4)	Corsi block tapping	forward tapping recall of blocks	NSGD.
Williams et al. 2005	31	26.58 (8.68)	111.10 (16.47)	103.13 (16.64)	108.65 (16.75)	WMS-III spatial span	forward and backward tapping recall of blocks	AU impaired. Correlated with FSIQ in AU but not in controls.
	24	11.75 (2.36)	112.50 (16.53)	106.38 (14.21)	109.67 (16.07)	Finger Windows	poke sequence of holes in a card	AU impaired. No correlation with FSIQ in either group.
Williams et al. 2006	38	11.68 (2.46)	106.42 (15.97)	100.55 (14.19)	103.82 (14.29)	Finger Windows	poke sequence of holes in a card	AU impaired.

Note. N = number of participants; VIQ, PIQ, and FSIQ = verbal, performance, and full-scale IQ, respectively; NSGD = no significant group differences; AU = autism group; AS = Asperger's group. Age and IQ are shown as: *mean* (*standard deviation*) in years and scores, respectively. All subjects diagnosed with autism, all controls TD, and NSGD in age or shown IQ measures, except as noted.

[a] Subject group diagnosed as Autism/Asperger's. [b] Subject group diagnosed as Asperger's. [c] Subject group diagnosed as PDD-NOS. [d] Controls grouped as TD, Tourette's, and autism/Tourette's (comorbid). [e] Subject group had lower FSIQ. [f] Subject group had lower FSIQ and VIQ.

2.3.9 Visual recall

One form of visual recall test involves giving the subject an abstract design to draw from memory after an initial inspection. Two examples are the Benton Visual Retention Test and the Rey-Osterrieth Complex Figure Task. The Rey-Osterrieth task includes a copy condition to identify perceptual or motor impairments that could confound results.

Many studies of these tasks have revealed decreased performance in individuals with autism (see Table 9). Given the patterns of intact and even superior performance found in other visual domains, these visual recall data are rather puzzling. Moreover, both the Rey and Benton tests have been found, in TD individuals, to be correlated with the Block Design subtest of the Wechsler scales and not correlated with verbal measures (Mitrushina et al. 2005; Strauss et al. 2006), and Block Design has been commonly cited as an area of particular strength for individuals with autism (Siegel et al. 1996).

One explanation could be that perceptual and motor components of these drawing tasks cause difficulties for individuals with autism, rather than the memory requirements per se. Ropar and Mitchell (2001) examined this possibility by comparing differences in copy and recall scores across groups and found no differences between TD controls and subjects with autism or Asperger's. Alternately, individuals with autism could have difficulty on the spatial but not visual aspects of these tasks. Although the Rey-Osterrieth task is often described as a test of visual memory, the task contains both visual and spatial components that are somewhat dissociable (Breier et al. 1996).

As with spatial recall, data on visual recall in autism are mixed at best. It is unclear how these results fit the TiP hypothesis, and further study is needed of the cognitive processes that both individuals with autism and TD individuals recruit for these tasks.

Table 9. Review of results from published studies of visual recall tasks in autism.

Reference	N	Age	VIQ	PIQ	FSIQ	Task	Details	Results
Ambery et al. 2006	27[a]	37.6 (14.6)	106.1 (15.7)	103.7 (19.2)	---	Doors and People Test	draw shapes from memory, also recognize colored photos of doors	AU impaired on combined measure of visual recall and recognition.
Ameli et al. 1988	16[e]	22.7 (4.9)	81 (16)	90.6 (13.5)	83 (14)	BVRT	draw series of abstract figures from memory	AU impaired. Highly correlated with PIQ in AU but not in controls.
Gunter et al. 2002	8[a]	16.25 (10.19)	111.38 (15.22)	96.13 (13.13)	---	Rey-Osterrieth	draw abstract figure from memory	NSGD in copy/recall scores or score differences.
Minshew & Goldstein, 2001	52	22.33 (9.59)	94.96 (17.56)	91.52 (12.95)	92.88 (15.06)	Rey-Osterrieth	draw abstract figure from memory	AU impaired in number of elements correctly drawn for immediate and delayed recall.
Minshew et al. 1997	15	21.13 (8.02)	98.53 (21.63)	92.87 (10.72)	95.73 (13.61)	Rey-Osterrieth	draw abstract figure from memory	AU impaired in number of elements correctly drawn for delayed recall.
Minshew et al. 1992	33	20.91 (9.69)	102.48 (16.35)	97.45 (11.19)	100.09 (12.96)	Rey-Osterrieth	draw abstract figure from memory	AU impaired in number of elements correctly drawn for delayed recall.
Ropar & Mitchell, 2001	19[cf]	14.2 (2.4)	---	---	---	Rey-Osterrieth	draw abstract figure from memory	NSGD in copy/recall score differences or in strategy.
	11[acf]	11.8 (2.0)	---	---	---	Rey-Osterrieth	draw abstract figure from memory	NSGD in copy/recall score differences or in strategy.
Verté et al. 2005	61[dg]	9.1 (1.9)	---	---	99.2 (17.1)	BVRT	draw series of abstract figures from memory	AU impaired.

Table 9 continued. Review of results from published studies of visual recall tasks in autism.

Verté et al. 2006	50[h]	8.7 (1.9)	93.1 (18.0)	104.0 (17.8)	98.2 (17.3)	BVRT	draw series of abstract figures from memory	AU impaired.
	37[a]	8.5 (2.1)	105.6 (20.0)	104.0 (15.9)	105.2 (16.3)	BVRT	draw series of abstract figures from memory	AU impaired.
	25[bh]	8.5 (1.4)	93.3 (14.7)	105.8 (19.2)	98.3 (14.4)	BVRT	draw series of abstract figures from memory	AU impaired.
Williams et al. 2006	38	11.68 (2.46)	106.42 (15.97)	100.55 (14.19)	103.82 (14.29)	WRAML Design Memory	draw geometric shapes from memory	AU impaired.

Note. N = number of participants; VIQ, PIQ, and FSIQ = verbal, performance, and full-scale IQ, respectively; NSGD = no significant group differences; AU = autism group, BVRT = Benton Visual Retention Test. Age and IQ values are shown as: *mean* (*standard deviation*) in years and scores, respectively. All subjects diagnosed with autism, all control groups TD, and NSGD in age or shown IQ measures, except as noted.

[a] Subject group diagnosed as Asperger's. [b] Subject group diagnosed as PDD-NOS. [c] Controls grouped as 8-yr.-old TD, 11-yr.-old TD, and learning disabled. [d] Controls grouped as TD, Tourette's, and autism/Tourette's (comorbid). [e] Subject group had lower PIQ. [f] Subject group had higher age. [g] Subject group had lower FSIQ. [h] Subject group had lower FSIQ and VIQ.

2.4 Review of Cognitive Theories of Autism

There are four main cognitive theories of autism often cited in the literature: Mindblindness, Executive Dysfunction, Weak Central Coherence, and Enhanced Perceptual Functioning. A review of how research on Mindblindness, related to the TiP hypothesis was already presented in Section 2.3.6 on false belief tasks. This section reviews data from the three remaining theories.

The Executive Dysfunction (ED) theory posits impairments in a set of higher-level cognitive skills that underlie independent, goal-oriented behavior, such as planning, set-shifting, and generativity (Russell, 1997). However, evidence in support of the ED theory is consistent with the TiP hypothesis if the specific executive capacities found to be impaired in autism are those that cannot be performed using visual mental representations. For example, individuals with autism are often impaired on the Wisconsin Card Sorting Test (WCST), a test of set-shifting in which subjects must maintain knowledge of a sorting rule and then switch the rule as needed (see review in Hill, 2004). The WCST, however, has been found to rely heavily on language abilities and verbal working memory in TD individuals (Baldo et al. 2005). More generally, Russell et al. (1999) propose that individuals with autism may have trouble primarily with executive tasks that require the implicit verbal encoding of rules. However, despite these suggestive data, evaluating a potential link between executive functioning in autism and the TiP hypothesis will require re-examination of a wide range of tasks used to tap executive abilities to discern how they fit into the task decomposition presented earlier (i.e. can they be solved visually, verbally, or using either type of mental representation).

The Weak Central Coherence (WCC) theory suggests that individuals with autism may exhibit a bias towards local over global processing (Happé & Frith, 2006). Evidence for the WCC theory shows patterns of either poor performance in individuals with autism on tasks that are said to rely on global processing of stimuli, or intact or superior performance on tasks that are said to rely on local processing. However, at least some of the "local" tasks cited by the WCC theory are visual, e.g. embedded figures, block design, visual search, etc. Likewise, certain WCC "global" tasks are verbal, e.g. homograph pronunciation. For these tasks, the TiP hypothesis provides an explanation that is consistent with published data, although, as mentioned earlier, the TiP hypothesis does not currently provide a concrete explanation of autistic superiorities on certain tasks, beyond the speculation that a reliance on visual representations might lead to increased visual expertise. Moreover, the WCC literature has identified several non-visual local tasks that are also performed well by individuals with autism, such as pitch and melody perception (see review in Happé & Frith, 2006). The TiP hypothesis is, at present, silent about representational modalities other than visual or verbal, though these results question whether TiP should be extended to a more general perceptual/verbal distinction.

Along these lines, the Enhanced Perceptual Functioning (EPF) theory proposes that individuals with autism have enhanced low level perceptual processing across a variety of modalities, in contrast to cognitive processing that involves higher levels of neural integration (Mottron et al. 2006). For instance, several studies have found evidence of atypicalities, and often superiorities, in low-level visual perception in autism (e.g. Bertone et al. 2005; Vandenbroucke et al. 2008). In addition to low-level perceptual enhancements and atypicalities, Ropar and Mitchell (2002) have proposed that autistic perception can be characterized, at least in certain task domains, as being less influenced than in TD individuals by "top-down" cognitive processes that draw upon prior conceptual knowledge. Caron et al. (2006) suggest that a combination of locally oriented processing and enhanced perceptual processing leads to superiorities in autism on visual tasks, for the subgroup of individuals who share these two traits.

Unlike the TiP hypothesis, which at present focuses only on visual representations, EPF and other perceptual accounts of autism are stated broadly to encompass a variety of perceptual modalities. However, within consideration of the visual modality, there seems to be significant overlap between these accounts, especially in that both TiP and EPF propose "a successful, problem-solving use of perceptual [brain] areas" (Mottron et al. 2006). Also, inasmuch as working with verbal representations might fall under "high-level" cognition, additional overlaps between TiP and EPF are likely.

One major difference between the WCC and EPF theories and the TiP hypothesis is that WCC and EPF embody *process* accounts of cognition, equating various modalities—visual, auditory, etc.—within each of two distinct types of processing—local vs. global, or perceptual vs. high-level. TiP embodies a *content* account of cognition, equating various processing types—perception, working memory, long-term memory, etc.—within each of two distinct representational modalities—visual vs. verbal. Another difference is that WCC and EPF explicitly account for autistic superiorities on certain visual tasks, whereas TiP does not currently propose a concrete mechanism for superior performance, though several possibilities, such as increased visual expertise or the absence of a verbal bias, remain to be explored. It is plausible that these accounts are linked, both developmentally and cognitively, and the precise relationship between TiP and these theories remains to be determined.

2.5 Claims and Future Work

The first study in this chapter reviews empirical data from autism in several different task domains including an assessment of whether the data are consistent with my formulation of the Thinking in Pictures (TiP) hypothesis. As expected, the results of this analysis are mixed. Certain task domains offer evidence highly consistent with and well explained by the TiP hypothesis, including: (1) the n-back task, (2) serial recall, (3) dual tasking, (4) Raven's Progressive Matrices, (5) semantic processing, and (6) false belief tasks. Other domains, while not inconsistent with the TiP hypothesis, are not directly explained by it either, namely: (7) visual search. Finally, there are domains whose data seem to contradict the TiP hypothesis, which are: (8) spatial recall, and (9) visual recall.

The main finding of this study is that, across several domains, there is a significant amount of evidence that is highly consistent with the TiP hypothesis, which empirically substantiates the anecdotal evidence for visual thinking that has long been common in the autism community. This finding is even more interesting given that most of the studies reviewed did not explicitly use a visual/verbal hypothesis in the design or execution of their experiments. Of course, there are many experimental task paradigms that have not been addressed or have been only briefly touched upon, for instance block design, free recall, cued recall, visual or verbal recognition, executive functioning, etc. Future work investigating the TiP hypothesis should incorporate an increased breadth of tasks.

The second study presented in this chapter analyzes four different existing theories about cognition in autism to evaluate whether these theories can explain empirical findings of visual/verbal dichotomies in autism and whether TiP can provide explanations of data commonly cited in support of these theories. For each of the four theories—

Mindblindness, Executive Dysfunction, Weak Central Coherence, and Enhanced Perceptual Functioning— I find that no theory explicitly posits the differences in representational strategy use in autism that I observed during the first study, and TiP does explain many of the findings presented in support of each theory.

If certain individuals with autism do have a bias towards using visual mental representations, then several important questions remain to be answered about the TiP hypothesis. First, most of the studies that were reviewed in this chapter drew from only high-functioning participants with autism, i.e. IQ of a certain level. This was necessitated primarily by task demands. However, TiP may provide an account of cognition in lower-functioning individuals with autism as well, for instance in nonverbal individuals. How might TiP predictions be tested in this population without resorting to measures of performance on intellectually demanding cognitive tasks?

In addition, I have not explored the extent to which TiP might characterize cognition in other disorders on the autism spectrum, such as Asperger's Syndrome or PDD-NOS, or how it might be distributed among subsets within the diagnosis of autism itself. How might particular individuals or subgroups be identified for whom TiP applies?

Also, TiP may provide, on an individual or group basis, only partial accounts of cognitive bias, e.g. an individual might show a visual propensity for certain tasks or domains but a verbal propensity for others. What assessments could be designed to assess the breadth or selectivity of domains across which TiP does or does not apply? A related question is how TiP might play into savant-like abilities in autism.

One important feature of my characterization of the TiP hypothesis is that I specify both a bias towards visual representations as well as a simultaneous bias against verbal representations. Whether both of these biases are actually present, or if one of them causally or developmentally precedes the other, bears further investigation. It could be, for instance, that a bias against or difficulty with verbal representations occurs first in the developmental progression, and this causes the brain to compensate by recruiting visual brain areas in atypical ways; on the other hand, the main mechanism in development could be some predisposition towards using visual representations, which then leads to underutilization of verbal capabilities. In addition, for TD individuals, I do not examine whether, for tasks that can be solved either visually or verbally, there is any typical bias in one direction or another. Certainly, in many studies of individual differences in the TD population (e.g. MacLeod, Hunt, & Mathews, 1978), verbal thinkers seem to predominate, and so the TiP hypothesis could describe, in some sense, the absence of a typical verbal bias (e.g. Sahyoun et al., 2009). Furthermore, the TiP hypothesis presented in this chapter makes no claims as to whether a visual thinker in the ASD population and a visual thinker in the TD population exhibit similar forms of cognition or behavior, or indeed if they share any developmental trajectory at all. Given the continuity of the autism spectrum into ranges of typical behavior, however, it seems likely that visual thinkers in both populations may share aspects of neural development and recruitment.

Other important avenues for further inquiry include (1) the distinction, if any, between visual and spatial processing under the TiP account, as well as relationships with other types of perceptual processing, (2) the differences in predictions between TiP and the other cognitive theories of autism, and how they can be interpreted or reconciled, and (3) what role TiP might play in neurobiological and developmental accounts of autism.

3 VISUAL PROBLEM SOLVING ON THE RPM

As mentioned earlier, the Raven's Progressive Matrices (RPM) family of tests consists of geometric-analogy-like problems in which a matrix of figures is presented with one missing entry, and the correct missing entry must be selected from among a set of answer choices. Figure 3 shows examples of two-by-two (2x2) and three-by-three (3x3) matrix problems, respectively, that are similar to actual RPM problems.[3]

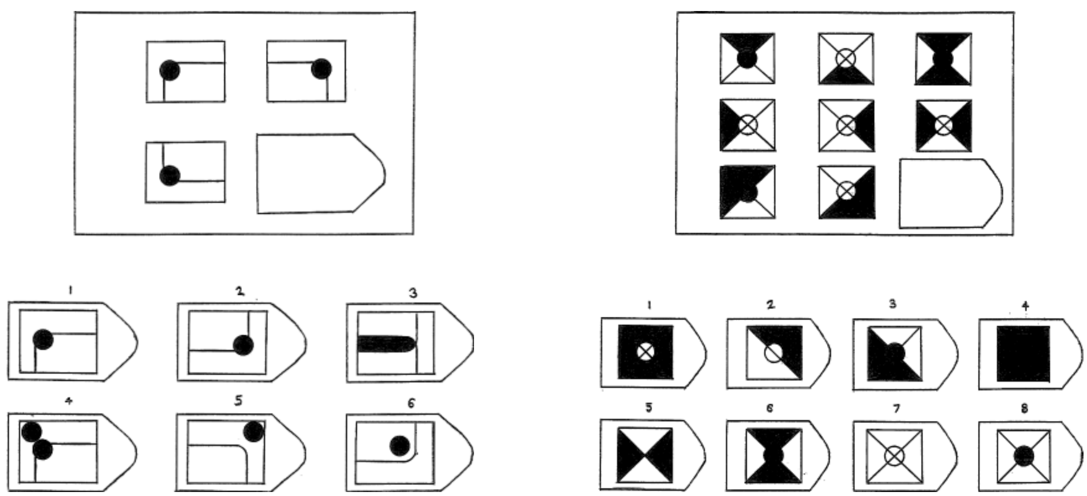


Figure 3. 2x2 (left) and 3x3 (right) example problem similar to those from the Raven's Progressive Matrices family of tests.

There are currently four different published versions of the RPM (Raven, Raven, & Court, 2003):

1) The original Standard Progressive Matrices (SPM).

2) The Colored Progressive Matrices (CPM), a simpler test than the SPM for use with children, the elderly, or other individuals falling into lower IQ brackets.

[3] To protect the confidentiality of the RPM, we present example problems that are similar, but not identical, to actual test problems.

3) The Advanced Progressive Matrices (APM), developed as a more difficult test to reduce the ceiling effects sometimes found with the SPM.

4) The Standard Progressive Matrices Plus (SPM+), a test that shares some items with the SPM but also contains more difficult items.

I use the term RPM to refer to the general family of tests, and SPM, CPM, APM, and SPM+ to refer to specific tests. The SPM+ is rarely used, compared to the prevalence of the SPM, CPM, and APM, and so I focus on the first three tests.

In this chapter, I address two primary research questions:

1) To what extent can a purely visual strategy, implemented as a computational model, be successful on the RPM tests?
2) How do changes in the underlying representational commitments of the model affect its behavior?

First, I motivate my investigation of visual problem solving on the RPM by reviewing evidence from the literature showing that humans do seem to use both visual and verbal strategies on the test, and that computational accounts to this date have not provided a model of problem solving on the RPM using visual representations. Then, I describe the ASTI model, which is a computational model that I have built which uses affine transformations and set operations, together with pixel-based representations of RPM problems, to generate solutions for the test. I present results from running the ASTI model against the CPM, the SPM, and the APM, and I also give results from the SPM for experiments in which the representational commitments of the ASTI model were varied. This chapter concludes with a discussion of my claims and areas for future work.

3.1 Visual Versus Verbal Strategy Use on the RPM

The RPM tests were originally designed by John Raven in the 1930s to measure only eductive ability, or the ability to extract and understand information from a complex situation, which is sometimes referred to as “fluid intelligence” (Raven et al., 2003). They were intended to be used together with the Mill Hill Vocabulary Scales, which measure reproductive ability, or the ability to recall previously learned information, sometimes called “crystallized intelligence.” Together, these two tests would provide a measure of Spearman’s general intelligence factor *g*, which Spearman had supposed could be decomposed into eductive and reproductive components (Spearman, 1923). However, over time, it was found that the RPM alone exhibited a very high level of correlation with other intelligence tests, leading the RPM to become widely considered one of the best single psychometric measures of *g* (Snow, Kyllonen, & Marshalek, 1984).

Using the RPM as a measure of general intelligence, though it consists only of problems in a single, visual format, stands in contrast to using broader IQ tests like the Wechsler scales, which are comprised of subtests that span several different verbal and nonverbal domains. In fact, Raven originally developed the RPM as an easy-to-administer, easy-to-score alternative to traditional multi-domain intelligence tests, which can take many hours to administer and yield complex, multi-dimensional subscores that must be combined to create a final IQ score (Raven et al., 2003).

The question of how people solve RPM problems was not addressed by John Raven during his original development of the test; the test was defined, and has later been refined, based on normative data and item analyses (Raven et al., 2003). Not until the emergence of computational and cognitive views of mental processes in the 1970s did researchers begin to take an in-depth look at problem solving on the RPM.

Hunt (1974) argued for a stronger consideration of strategy usage on psychometric tests as a window into the nature of individual differences in intelligence. He proposed the existence of two qualitatively different RPM problem-solving strategies which varied primarily in how problem inputs were represented—using modal iconic representations versus amodal propositional representations—and he described each of these strategies in terms of how they could be implemented in a computer program.

Despite the lack of a concrete implementation of his algorithms, Hunt made a significant contribution to the understanding of information processing on the RPM with his contention that qualitatively different strategies could exist and even potentially be equally successful. He emphasized the importance of taking possible variations in strategy into account when considering individual differences in overall performance: not only might individual performance differences stem from such strategy variations, but also equal levels of performance in two individuals could be masking variations in the underlying strategy. Such considerations, Hunt urged, should play a meaningful role in the interpretation of RPM results, rather than just treating an individual's score as an atomic measure of general intellectual ability. Interestingly, Spearman himself, though he generally approved of the RPM as a tool for measuring *g*, supposed that there might be two different ways to tackle RPM problems, which he termed "analytic" and "synthetic," and he also believed that only the analytic approach loaded heavily on *g* (cited in Hunt, 1974).

Despite Hunt's work on the existence and plausibility of multiple RPM strategies, computational studies of problem solving on the RPM have tended to embody only a single strategy (within any given information processing account), and these strategies have all relied on propositional forms of problem solving. As a result, the prevailing notion of how people solve RPM problems has been that they use verbal strategies.

However, Hunt's notion of dual strategies on the RPM has since been borne out across various psychological studies of human RPM performance. In particular, there is considerable evidence from both within-individual and between-individual studies that humans recruit both visual and verbal strategies on the RPM. The evidence from between-individual studies comes from research into autism and has already been summarized in Section 2.3.4 The remainder of this section is organized as follows:

1) Evidence from human studies of within-individual strategy differences across various RPM problems or test-taking conditions.
2) Computational accounts of problem solving on the RPM.
3) Operations of mental imagery in human cognition, including evidence for the kinds of affine and set transformations used by the ASTI model

3.1.1 Within-individual strategy differences on the RPM

Within-individual strategy differences have been studied as a function of problem type on the RPM tests, primarily through factor analyses of the SPM (Lynn, Allik, & Irwing, 2004; van der Ven & Ellis, 2000) and of the APM (Dillon, Pohlmann, & Lohman, 1981; Mackintosh & Bennett, 2005; Vigneau & Bors, 2008). These studies have identified multiple factors underlying RPM tests, indicating variations in the recruitment of particular cognitive mechanisms for different problems, and have often divided test problems into two primary categories: those solved using visuospatial or gestalt reasoning and those solved using verbal or analytic reasoning (though it should be pointed out that, while the factor loadings themselves are statistically determined, labels for the various factors appear to be based on the authors' own inspections of problem groupings by factor). Following the Gestalt/Analytic strategy divide proposed by Hunt (1974), Kirby and Lawson (1983) studied the performance effects of training students to use a particular strategy; part of this study involved developing a new series of test items on which the type of strategy being used led to a different selection of a "correct" answer choice, thus demonstrating the existence of strategy-linked answer types (in addition to strategy-linked problem types).

From neuroscience, one fMRI study of RPM performance (Prabhakaran, Smith, Desmond, Glover, & Gabrieli, 1997) found that patterns of brain activity differed significantly based on whether participants were solving "figural" versus "analytic" problems, using problem classifications derived from a computational study of the RPM (Carpenter, Just, & Shell, 1990). Figural problems induced brain activity primarily in spatial and object working memory regions, while analytic problems induced additional brain activity in verbal working memory and executive processing regions. Studies of patients with focal brain lesions have also found linkages between brain regions associated with specific types of visual or verbal processing and successful performance on figural versus analytic problems (Berker & Smith, 1988; Villardita, 1985).

Zaidel, Zaidel, & Sperry (1981) looked at the performance of two commissuorotomy and two hemispherectomy patients on the CPM and the SPM, to see if there were hemisphere effects on RPM performance. Overall, they found that RPM problems were equally solvable with either hemisphere, though the left hemisphere showed a slight dominance overall and on the more difficult problems within each set. They speculate on whether this dual hemispheric competence relates to Spearman's notion of two different RPM strategies, one analytic and one synthetic, which may correspond to left and right hemisphere solving, respectively.

DeShon, Chan, & Weissbein (1995) had participants complete the APM while simultaneously performing a "verbal overshadowing" protocol, in which they had to verbally describe their reasoning. The authors hypothesized that requiring overt verbal descriptions would bias participants towards using verbal instead of visual strategies and thereby impair performance on problems that would normally have been solved visually, which was borne out in the resulting data. These findings call into question the methodology of using verbal reporting protocols as a window into RPM problem solving, as the act of verbal reporting itself may cause shifts in an individual's strategy; this "verbal overshadowing" phenomenon has been observed in other problem domains as well (Schooler & Engstler-Schooler, 1990; Schooler, Ohlsson, & Brooks, 1993).

Jacobs and Vandeventer (1968) supposed that CPM problems require participants to both imagine what each answer choice looks like in the empty space in the matrix as well as perform induction to determine if an answer choice (thus imagined) is the correct one. They found that an experimental manipulation in which the cognitive loading of the first skill is removed, by having physical answer choices that could be moved by a participant into the matrix, made the task significantly easier for young children. This analysis assumes, to some extent, that participants are using an at least partially imagery-based strategy. It could be, however, that the overall strategy could change with the altered task; e.g., having a complete (but possibly incorrect) matrix would seem to suggest a Gestalt reasoning approach, whereas having to select an answer choice could admit

Gestalt approaches (with the imagination component) as well as rule-based, predictive approaches.

Babcock (1994) looked at how different hypothesized components of problem-solving on the RPM might be measurable by looking at correlations with other tests that load on these specific components. In particular, the author hypothesized that RPM problem solving requires rule identification, rule application, and rule coordination. (She hypothesized a fourth component, figure decomposition, but found in a pilot study that most variance on figure decomposition tasks was accounted for by processing speed, so omitted this fourth component from further analysis.) She then chose two tasks from psychology that exemplified each of these components: Shipley abstraction and letter sets for rule identification, geometric transformation and pattern transformation for rule application, and the calendar test and following directions for rule coordination.

Then, she analyzed results from these tests and other measures of processing speed and working memory, and also from the APM, from adults of varying ages to see which of these tests could account for the age-related variance on the APM. She found that working memory and processing speed accounted for much of the age-related variance on the APM, and neither rule identification nor coordination contributed substantial additional variance, but rule application did contribute, and the combination of rule application, working memory, and processing speed was sufficient to reduce the age-related variance on the APM to non-significance. This is particularly interesting given that the rule application task involved affine visual transformations very like operations performed by the ASTI model, which is described later in this chapter, in that it involved transforming given geometric patterns by a given geometric rule to choose the correct answer. Rule identification used two verbal tests that involve rule abstraction, and rule coordination involves following complex sets of rules, again that were given in verbal representational form. This could be considered evidence of participants on the APM using figural reasoning similar to that performed by the ASTI model, or it could just be that the rule application task was too similar to the APM, in comparison with the other verbal tasks, and this similarity presented a confound in the study results.

3.1.2 Computational models of the RPM

As described earlier, Hunt (1974) proposed two qualitatively different RPM strategies that varied primarily in how problem inputs were represented. Hunt's "Analytic" algorithm used amodal propositional representations and operations such as constancy and addition/subtraction. This algorithm proceeded by first abstracting features from each matrix entry and then iteratively applying operators to the entries within a row or column, or to an entire row or column, to generate partial answer predictions. If the predicted answer was found among the answer choices and was unique, then the algorithm halted. If either of these conditions were not met, then the algorithm iterated further to either predict a different answer or refine the current partial answer.

Hunt's "Gestalt" algorithm, akin to mental imagery, used modal iconic representations and perceptual operations like continuation and superposition. The algorithm successively applied various visual operations to entries from the matrix in order to obtain an answer that matched one given in the answer choices, using an answer-iteration procedure similar to that used in the Analytic algorithm.

While neither algorithm was actually implemented, Hunt performed a theoretical analysis of how each algorithm would fare on Set I from the APM, and he predicted that the Gestalt algorithm would successfully solve 6 of the 12 problems on Set I, whereas the Analytic algorithm would solve all 12 problems.

All of the RPM models that have since been developed resemble Hunt's Analytic algorithm in that they rely on a conversion of problem inputs into amodal propositional representations. None of these models have adopted the approach suggested by Hunt in his Gestalt algorithm.

Model #1: Carpenter, Just, and Shell (1990) implemented a production system that took as input hand-coded propositional descriptions of problems from the APM. The system chose from a predefined set of rules over matrix elements in order to predict an answer for each problem. The predicted answer was compared to the answer choices in order to choose the best match. The predefined rules were generated by the authors from an *a priori* inspection of the APM and were validated in experimental studies by observing what "rules" participants used while taking the test, as evidenced by verbal reporting protocols. Differences between low- and high- scoring participants were modeled by developing two different versions of the production system; the more advanced system contained an increased vocabulary of rules and a goal monitor for setting and adjusting the high-level problem-solving process being used. Both systems were tested against 34 of the 48 problems from the APM. The basic system solved 23 of these 34 problems, while the more advanced system solved 32 of the 34 problems. This model also made predictions about the numbers of eye fixations that subjects might make on different types of problems. While the experimental results presented were consistent with the model, there is no direct relationship between the eye gaze data and the use of a purely propositional strategy; the data could be equally well fit by a model using iconic representations.

Model #2: Bringsjord and Schimanski (2003) used a theorem-prover to solve selected RPM problems stated in first-order logic, though no specific results were reported.

Model #3: Lovett, Forbus, and Usher (2010) combined automated sketch understanding with the structure-mapping technique for analogy to solve problems from the SPM. Input images from the test were first redrawn by hand in Powerpoint, and the resulting vector graphics objects were fed into the system. The system translated these inputs into amodal propositional descriptions using a procedure for automated sketch understanding. Then, a series of strategies based on the structure-mapping technique for analogy were applied to detect certain patterns of structural relationships between various elements in the matrix. Strategies focused either on differences between images in a row or column (Differences and Advanced Differences strategies) or on elements that are shared among images in a row/column (Literal or Advanced Literal strategies). The Advanced Differences strategy is different from the regular Differences strategy because objects are only mapped to other objects of the same shape, thus allowing for object additions or deletions but not object transformations. The Advanced Literal strategy is different from the regular Literal strategy because objects shared across matrix entries are removed, and spatial relations are also removed, causing objects to be matched independently to other objects (or segments of an object, in the case of a single object matrix entry.) These derived structural relationships were also used to refine object segmentation and groupings by revisiting the original vector-graphics-based representations and extracting modified propositional descriptions that allowed for improved structural matches. Finally, each answer choice was inserted into the matrix, and the answer providing the closest matching structural relationship within the matrix was selected. This system was tested against 48 of the 60 problems on the SPM and solved 44 of these 48 problems.

Model #4: The system of Cirillo and Ström (2010), like that of Lovett et al. (2010), took as inputs hand-drawn vector graphics representations of test problems and used an automated procedure to create hierarchical propositional representations of the problem information. Then, like the work of Carpenter et al. (1990), the system drew from a set of pre-defined patterns, derived by the authors from an *a priori* inspection of the SPM, to find the best-fit pattern for a given problem. The resulting pattern was used to predict an answer, though no explicit procedure was given for matching the predicted answer to one of the given answer choices. This system was tested against 36 of the 60 problems from the SPM and solved 28 of these 36 problems.

Model #5: Rasmussen and Eliasmith (2011) used a spiking neuron model to induce rules for solving RPM problems. Input images from the test were first hand-coded into vectors of propositional attribute-value pairs, and then the spiking neuron model was used to derive several individual transformations among these vectors and abstract over them to induce a general rule transformation for that particular problem. While the authors attested that this system could correctly solve RPM problems, they did not present any results regarding which specific tests or problems were addressed.

However, as they pointed out, their approach differed radically from earlier RPM models in that it explicitly accounted for the process of rule induction for a given problem, rather than relying on a predefined set of rules. They argued that rule induction is a fundamental component of problem solving on the RPM, as that is by definition the main characteristic of an eductive task; if the RPM were intended to measure an individual's pre-existing memory for a set of rules, then it would have to be classified as a test of reproductive ability instead of eductive ability. Furthermore, their model of rule induction naturally facilitates inter-problem learning and transfer, which undoubtedly plays a role in human problem solving on the RPM (Bors & Vigneau, 2001; Verguts & De Boeck, 2002) and for which none of the previous models readily account.

3.1.3 Operations of mental imagery

Hunt (1974) presented his intuition of the kinds of visual operations needed to solve RPM problems as resembling those of visual perception, specifically of continuation and superposition. More generally, these are consistent with a viewpoint of mental imagery as a form of analog mental model, in which the referents of imagery are modeled after rigid bodies in the world, and thus the operations of imagery are grounded in the affordances of manipulating (or observing manipulations of) such objects (Schwartz & Black, 1996). Mathematically, the manipulations of a 3D rigid body projected into a 2D plane correspond to affine transformations, and the combination of projections of multiple such bodies can be represented using operations on mathematical sets, in this case sets of points or elements in the 2D plane. In this section, I examine evidence relating to two classes of mental imagery operations: affine transformations and set operations.

The most well-known early experiments looking at functional properties of mental imagery are the mental rotation experiments of the early 1970s (e.g. Shepard & Metzler, 1971). Many variations on these experiments have been performed, with a recent plethora of studies that include neuroimaging measures (Zacks, 2008), but the basic findings have been that the reaction time taken by participants to mentally rotate an object is proportional to the angular degree through which the object must be rotated. This finding is replicated in situations when two stimuli are both visible for comparison (e.g. Shepard & Metzler, 1971) as well as when a single visible stimulus is compared to another representation from memory (Cooper & Shepard, 1973). More recently, certain individuals with autism have been found to be faster and more accurate on mental rotation tasks than typically developing individuals; the reaction times in the autism group in this study preserved the conventional pattern of increase with respect to angle of rotation (Soulières, Zeffiro, Girard, & Mottron, 2011).

Set operations in mental imagery can include manipulations such as union, intersection, and complement (see Section 3.2.4 for a detailed computation-based list). These manipulations correspond generally to the combination of elements of mental images in various ways, including: subtraction, in which participants were asked to subtract a visually presented shape from a remembered image in order to derive and identify a new mental image (Brandimonte, Hitch, & Bishop, 1992a), and combination, in which participants were asked to perform a similar identification after combining a new image element with a remembered image (Brandimonte et al, 1992c; Finke, Pinker, & Farah, 1989). Additional studies have found evidence, in some cases, of verbal coding interfering with this type of image transformation (Brandimonte, Hitch, & Bishop, 1992ab) and in other cases, of an interaction between seemingly pictorial depictive representations and non-pictorial descriptive representations (Hitch, Brandimonte, & Walker, 1995; Walker et al., 1997). In the paper discussed above looking at mental rotation in autism, another experiment was conducted to examine this type of image combination (Soulières et al., 2011). Participants first inspected and memorized an array of visually presented letters and numbers, and then were briefly shown a circular segment with a portion of one character inside it. Then, upon looking at the segmented circle alone, the task was to determine which segment would contain a greater visual proportion of the original character. This experiment can be thought of as requiring an operation akin to intersection, in visualizing which portion of the character falls into each segment of the circle, as if the two were overlaid, and also a comparison of visual similarity in terms of which character portion embodies a greater visual area.

Drawing upon this literature, the main types of transformations employed by the ASTI model, described in the following section, are affine transformations and set operations. The model's conceptualization of visual similarity, which might be considered a third class of visual mental transformations, relies on set-based notions of union and intersection (Tversky, 1977) and follows template-based theories of similarity in which image elements are compared according to the amount of visual overlap between them (Palmer, 1978).

3.2 Description of the ASTI Model

As described in Section 3.1.2, existing computational models have focused exclusively on propositional accounts of problem-solving on the RPM. This section describes a computational model for solving RPM problems, called the Affine and Set Tranformation Induction (ASTI) model, that, like Hunt's proposed Gestalt algorithm (1974), simulates modal reasoning by using iconic visual representations. In particular, this model uses pixel-based representations of problem inputs and reasons over these representations using affine transformations and set operations.

The ASTI model is one of a pair of visual RPM models under development by our research team. The other model, called the "fractal model," represents an effort led by fellow Ph.D. candidate Keith McGreggor (McGreggor et al., 2010). While this dissertation does not address the fractal model in depth, its development marks an important contribution to the general argument about visual RPM strategies that I present, in that the existence of two different visual computational models may yield additional insights as to the generalizability of each of our independent sets of results and analyses.

3.2.1 Inputs and outputs

The ASTI model uses representations consisting of two-dimensional arrays of grayscale pixels, with each pixel associated with a single intensity value. These pixel-based representations are iconic in that they preserve a spatial correspondence with the patterns of light and dark areas on the actual test problem inputs. They are modal in that they remain in the same pixel-based format that was generated when test problems were scanned using a digital scanner.

Specifically, the inputs to the ASTI model for a given problem are sets of images that represent individual matrix entries and answer choices as presented in the test booklet. For the 2x2 problem in Figure 3, the inputs to the ASTI model are the images shown in Figure 4, where $\mathbf{m_{ij}}$ refers to the entry at row **i** and column **j** of the matrix, and $\mathbf{a_1}$ through $\mathbf{a_n}$ represent the **n** answer choices given at the bottom of the problem.

The output of the ASTI model is a single number between **1** and **n**, denoting its chosen answer.

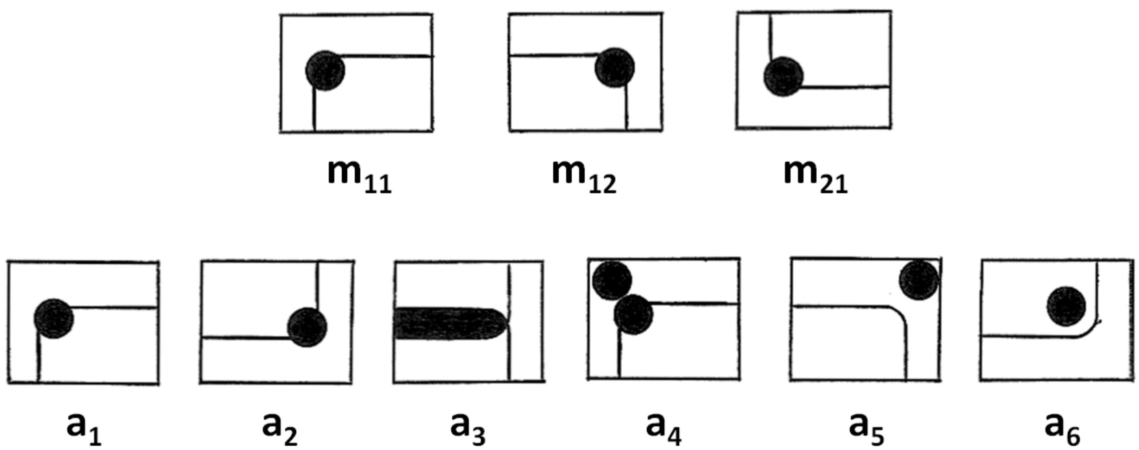


Figure 4. Imagistic representation of the RPM problem shown in Figure 3, fed as input into the ASTI model.

3.2.2 High-level approach

At a high level, the basic approach used by the ASTI model is to:

1) Inspect the matrix portion of an RPM problem to determine what relationship is present among the existing matrix entries.
2) Using this relationship, generate a predicted answer in the form of an image for what entry might occur in the empty spot in the matrix.
3) Compare the predicted answer to each given answer choice and select the choice that is most similar to the prediction.

The relationship that the ASTI model attempts to determine in Step 1 is an image transformation that best accounts for the variation among entries in any collinear set of entries in the matrix. In Step 2, the model applies this same transformation to whichever incomplete collinear set of entries is parallel to the first, in order to generate its predicted answer. In these two steps, the ASTI model is making two implicit assumptions about the structure of RPM problems: (1) entries in a single collinear set within the matrix are related according to some image transformation, and (2) parallel collinear image sets are analogous in that they share the same image transformation.

Schematic illustrations of which entries the ASTI model uses in Step 1 to induce row or column transformations are given in Figure 5 and Figure 6 for 2x2 and 3x3 matrices, respectively. A more detailed description of the image sets examined by the ASTI model is provided later in this section. These illustrations show which parallel incomplete rows or columns are used together with the induced transformation to generate the predicted answer in Step 2.

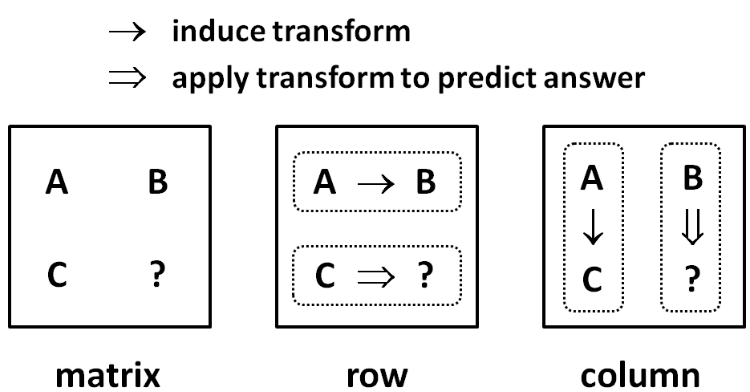


Figure 5. Schematic illustration of transformations considered by the ASTI model for a 2x2 RPM matrix.

For example, looking at a 2x2 matrix as shown in Figure 5, the model might induce a row transformation relating entries **A** and **B** and then apply this transformation to element **C** to predict an answer. Alternately, the model could try to induce a column transformation in the same manner, first relating entries **A** and **C** and then applying the induced transformation to entry **B**.

For 3x3 matrices, the set of possible transformations is much larger, as there are eight matrix entries to consider instead of just three. Beyond considering unary transformations as in the 2x2 case, i.e. transformations converting a single given image into a single transformed image, 3x3 matrices present the possibility of binary transformations, i.e. transformations converting two given images into a single transformed image. For a 3x3 matrix, looking at row transformations, the model might induce a unary row transformation between adjacent entries **A** and **B** or adjacent entries **B** and **C**, and then apply this transformation to entry **H** to predict an answer, as shown in the top "row" matrix in Figure 6. Or, the model might induce a binary row transformation relating all three entries **A**, **B**, and **C**, and then apply this transformation to entries **G** and **H**, as shown in the bottom "row" matrix in Figure 6. As with 2x2 matrices, all of these transformations for 3x3 matrices can be induced either across rows or along columns.

Therefore, for a given RPM problem, the ASTI model proceeds by first inducing all possible transformations for the matrix, using collinear sets of image entries. The transformation induction process is described in more detail below. Each induced transformation carries with it a measure of "fitness" that varies between 0.0 and 1.0 to indicate how well that particular transformation fits its associated row or column, where 0.0 indicates a poor fit and 1.0 indicates a perfect fit. The ASTI model selects that transformation and associated image set that has the highest measure of fitness, which completes Step 1.

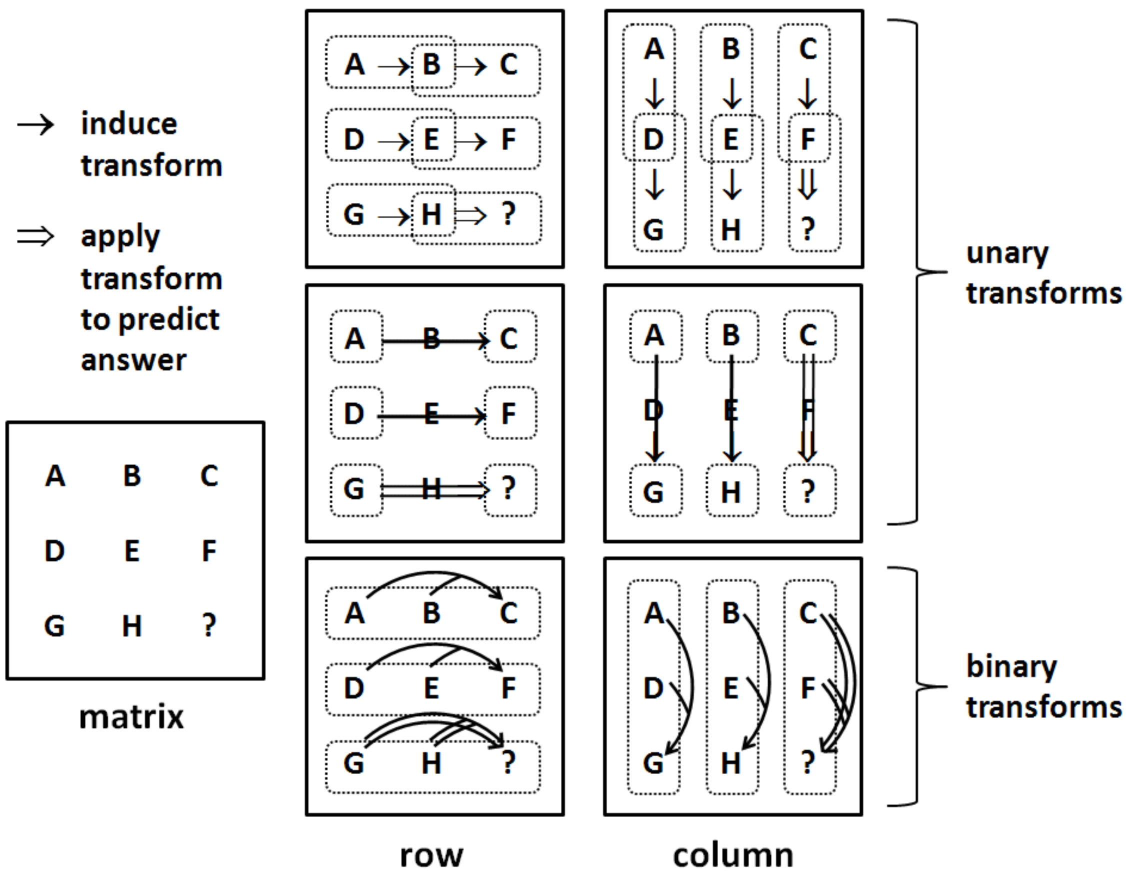


Figure 6. Schematic illustration of transformations considered by the ASTI model for a 3x3 RPM matrix.

In Step 2, the model applies this transformation to the appropriate incomplete parallel image set to predict an answer. Finally, in Step 3, the predicted answer is compared in turn to each given answer choice according to a similarity measure, which is also described below. The choice yielding the highest similarity value is chosen as the model's final answer.

3.2.3 Best-fit image transformations

This section describes the induction process for unary transformations (e.g. converting image **A** to image **B**), with the detailed algorithm given in Figure 7; induction of binary transformations (e.g. converting images **A** and **B** to image **C**) is a straightforward extension of this process.

To begin, suppose we have two images **A** and **B**. We wish to induce a transformation that represents the change from **A** to **B**. This process is akin to image registration, in which two images are aligned according to some criteria that ultimately enable a "best-fit" correspondence to be found. In image registration, a correspondence between two images is found by matching features between the images, and any remaining differences are modeled as a combination of various types of geometric deformations and/or color transformations (Zitová & Flusser, 2003). While image registration is typically performed on real-world images, this approach has been adapted for the ASTI model's transformation induction process, as it seems well able to capture differences between black-and-white line drawings of the type found in RPM problems.

In particular, the ASTI model defines a composite transformation between two images as a combination of two geometric transforms and one color-based transform:

1) A base affine transform **t** (e.g. rotation, reflection, etc.)
2) A translation **(x, y)**
3) A pixel-wise composition operation $\oplus$ (e.g. addition, subtraction) together with a composition operand **X**, which consists of another image

The ASTI model contains a finite set of base transforms which, for simplicity, are restricted to rectilinear rotations and reflections. Affine transformations such as shearing and scaling are not included, nor are other types of geometric image deformations.

Initialization

1 Read matrix entries into list of images M
2 Read answer choices into list of images A
3 For any two images a and b, define a similarity metric $S(a, b) \rightarrow z \in [0, 1]$
4 Define set of base transforms T
5 Define set of analogies $I_0 \rightarrow I_1$, where I_0 contains image sequences representing complete row, column, or diagonal lines in the matrix, and for each $i_0 \in I_0$, I_1 has the corresponding images i_1 representing the parallel partial line in the matrix

Transformation Induction

1 For each image sequence $i_0 \in I_0$, induce the best-fit composite transform t_C:
2   For each base transform $t \in T$:
3   Apply t to the first image(s) in i_0 to produce image i_t
4   Search all possible translation offsets (x, y) between i_0 and i_t to find the best match, as calculated by $S(i_{0(x,y)}, i_t)$
5   Select the best-fit base transform t_B as per S, as calculated above
6   t_C is then a composition of t_B and the translation offset (x, y)
7 Obtain a final transform t_F by selecting that t_C which produces the best average fit, across each subset of parallel $i_0 \in I_0$

Candidate Prediction and Answer Selection

1 Choose image sequence i_0 that results in the best-fit t_F, according to S as calculated in the previous step
2 Apply t_F to corresponding partial image sequence $i_1 \in I_1$ to produce candidate answer image i_C
3 For each answer choice $i_A \in A$, compute similarity $S(i_C, i_A)$
4 Select the best-fit answer choice iA as per S, as calculated above

Figure 7. Main algorithm for the ASTI model, including transformation induction.

To induce a composite transformation between two images, the ASTI model first uses a template-matching scheme to search across all possible base transforms and translations to find the combination of these two geometric transforms that results in the best correspondence between image **A** and image **B**. Then, given these particular geometric transforms, any remaining image discrepancies are accounted for by defining pixel differences between the two images as comprising the operand of an image composition operation, namely pixel-wise addition or subtraction. Which type of operation is selected depends on whether there are a greater number of pixels being added to or subtracted from image **A** to arrive at image **B**.

The combination of these three transforms—base transform, translation, and image composition—is defined to be the best-fit composite transformation between image **A** and image **B**. The degree of "fit" (i.e. the strength of the discovered correspondence) is defined as the similarity value found during the template-matching process

3.2.4 Base transforms

The base unary transforms (i.e. transforming image **A** into image **B**) used by the ASTI model during the induction of composite transformations are drawn from the set of image operations that fall under the category of affine transformations (hence the name of the model), and in particular are restricted to orthonormal transformations only (i.e. rotation and reflection, combined with translation). In addition to the fact that affine transformations are a well-defined and thoroughly-studied type of image operation, there is evidence that human visual processing can apply affine transformations like scanning (i.e. translation), zooming (i.e. scaling), and rotation to mental images, or operations that are computationally isomorphic (Kosslyn et al. 2006; Shepard & Metzler, 1971).

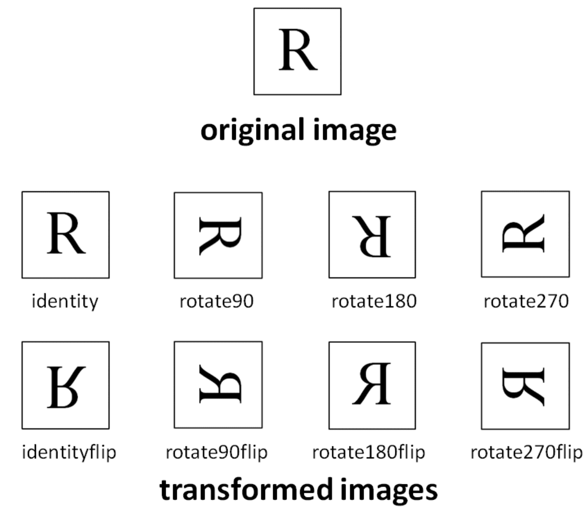
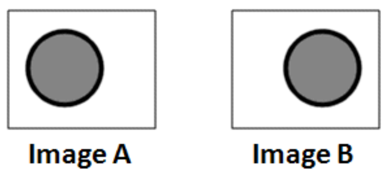


Figure 8. Eight base unary affine transforms used for 2x2 and 3x3 matrices.



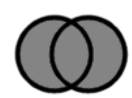
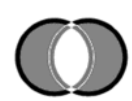
Transformation	Notation	Pixel operation on pixels $p_a \in A$ and $p_b \in B$	Resulting image
union	$A \cup B$	$\max(p_a, p_b)$	
intersection	$A \cap B$	$\min(p_a, p_b)$	
subtraction	$A - B$	$p_a - p_b$	
back-subtraction	$B - A$	$p_b - p_a$	
exclusive-or	$A \text{ xor } B$	$\max(p_a, p_b) - \min(p_a, p_b)$	

Figure 9. Five base binary set transforms used by the ASTI model for 3x3 matrices.

The ASTI model presently uses the eight base unary transforms shown in Figure 8, which comprise all possible rectilinear rotations and reflections. In addition, for 3x3 matrices, as mentioned earlier, the larger number of matrix entries introduces the possibility of using binary image transforms instead of unary ones (i.e. transforming images **A** and **B** into image **C**). The base binary image transforms used by the ASTI model are drawn from set composition operations to capture notions of image union, intersection, subtraction, etc., and are implemented at the pixel level as maximums, minimums, and differences of grayscale intensity values. Figure 9 illustrates the five base binary transforms used by the model.

To see why these five particular binary transforms were chosen, consider the full set of all binary transforms on two single black and white pixels (i.e. functions that map from two binary values onto one binary value). All possible values for these transforms are shown in Table 10.

Table 10. Definitions of pixel-wise operations for various pixel color representations.

A	**B**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

The two leftmost columns show the values of the two input pixels, labeled A and B; there are four possible combinations of input values that can occur, which are listed by row. The columns to the right show the sixteen different possible patterns of outputs that can occur for each pattern of inputs. All values are shown as if 0 is the value of a white pixel and 1 is the value of a black pixel. This table gives all possible combinations of inputs and outputs for a binary transform function on binary-valued pixels.

First consider the eight right-most columns that are shaded in brown. For all of these output functions, two white pixels (values of 0 and 0 for A and B) map to an output value of black (value of 1 for C). Since black is considered by the model to indicate figural pixels and white to indicate ground pixels, these functions would map the combination of two ground pixels onto a figural pixel, essentially "making up" pixel content where none previously existed. Operationally, this would turn all white regions shared between two images into black, and so the ASTI model omits these functions.

Next, look at the first green column output function. This function takes all possible inputs and maps them to white pixels (values of 0). Operationally, this function would turn all pairs of images into pure white images, and so the ASTI model omits this function as well.

Next, turn your attention to the output columns shaded in purple. Note that the first purple column is identical to the input values for A, and the second purple column is identical to the input values for B. Both of these output functions essentially ignore one of the inputs and instead map everything to the values of the other input. Operationally, given two images A and B, these functions would either map directly back to image A or back to image B, and so the ASTI model omits these functions.

The five remaining white output columns that are left correspond precisely to the five binary transforms that are implemented in the ASTI model. The first white column can be read as the intersection operation, mapping to black only when both input pixels are black. The second white column is subtraction, specifically A minus B. The third white column is also subtraction, specifically B minus A. The fourth white column is the XOR function, mapping to black only when one or the other (but not both) inputs are black. Finally, the fifth white column is the union function, mapping to black when either or both of the inputs are black.

3.2.5 Visual similarity

The same similarity measure is used by the ASTI model in Step 1, for template matching during the transformation induction process, and also in Step 3, to select the final answer choice based on the predicted answer image. This measure is adapted from Tversky's (1977) ratio model of similarity:

$$similarity(A, B) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)} \tag{1}$$

In this equation, f represents some function over features in each of the specified sets A and B. The constants α and β are used as weights for the non-intersecting portions of A and B. If α and β are both set to 1.0, this equation becomes:

$$similarity(A, B) = \frac{f(A \cap B)}{f(A \cup B)} \tag{2}$$

For calculating the similarity measure, each feature is defined to be a single pixel, and intersection, union, and subtraction operations are defined as the maximums, minimums, and differences, respectively, of the pixels' grayscale intensity values. The functions f are defined as summations of feature comparison values over the entire image.

The formulation of Tversky's ratio model used by the model makes one important assumption about pixels, which is that they can be treated as independent features within the pixel sets represented by images **A** and **B**. While this notion of pixel independence is a strong simplification, it matches the assumptions made by basic template theories of visual similarity that define similarity based purely on evaluations of the extent of overlapping figural units (Palmer, 1978), which in our case are individual pixels.

3.2.6 Consideration of image sets within matrix

Earlier, I stated that the ASTI model makes two assumptions about how entries in a problem matrix are related:

1) Collinear entries are related according to some image transformation.
2) Parallel pairs of collinear sets of entries are analogous in that they share the same (or a similar) image transformation.

These assumptions raise the question of how to select collinear entries and parallel pairs of collinear entries from a given problem matrix. In particular, we want the model to examine collinear sets of entries to discover an image transformation, and then find a parallel collinear set of entries which contains the empty space, in order to apply the previously discovered transformation to infer an answer.

In the 2x2 case, the situation is fairly simple. There are four entries, and thus we could consider collinear pairs of images or collinear triplets of images. Collinear pairs are possible in a 2x2 matrix, but collinear triplets are not, because the set of three elements **ABC** is not collinear in a 2x2 grid. So we restrict consideration to pairs only.

Combinatorially speaking, there are six ways to choose pairs of elements from this set: **AB**, **AC**, **BC**, **A?**, **B?**, and **C?**. Each of these image pairs represents a collinear set of elements in the matrix, because any two points on a grid are collinear. We need not consider the reverse pairs—e.g. **BA**, **CA**, and **CB**—because the base unary affine transforms, as a collective set of operations, are commutative over pairs of images:

$$Identity(A)=B \quad \Longrightarrow \quad Identity(B)=A$$

$$Rotate90(A)=B \quad \Longrightarrow \quad Rotate270(B)=A$$

$$Rotate180(A)=B \quad \Longrightarrow \quad Rotate180(B)=A$$

$$IdentityFlip(A)=B \quad \Longrightarrow \quad IdentityFlip(B)=A$$

$$Rotate90Flip(A)=B \quad \Longrightarrow \quad Rotate270Flip(B)=A$$

$$Rotate180Flip(A)=B \quad \Longrightarrow \quad Rotate180Flip(B)=A$$

Given the three complete pairs of collinear images **AB**, **AC**, and **BC**, the next step is to find corresponding parallel collinear pairs which contain the empty space **?** in the 2x2 matrix grid. For **AB** and **AC**, the collinear pairs are obvious: **C?** and **B?**, respectively, where **?** represents the empty space. However, it is not immediately clear what the parallel pair is for **BC**. To find this, we simply repeat the matrix entries as if they were part of a regular, infinite grid, as shown in Figure 10. **BC** is only parallel to **CB**, which does not help in solving the matrix problem. (Likewise for **A?** and **?A**.) Therefore, the analogies considered by the ASTI model for 2x2 matrices are: **A : B :: C : ?** and **A : C :: B : ?**. These correspond to rows and columns, respectively. **BC** and **A?**, which the model does not use, correspond to diagonals.

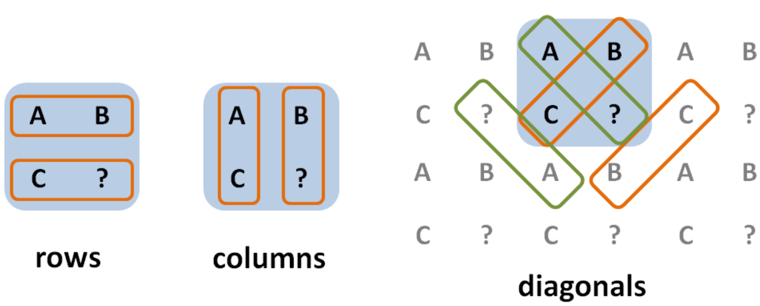


Figure 10. Collinear pairs of entries, and parallel sets of collinear pairs, in a 2x2 matrix.

For 3x3 matrices, the situation is more complex, though we can follow the same procedure to identify the analogies that the ASTI model will inspect. There are eight nonempty entries in a 3x3 matrix. However, only pairs of entries or triplets of entries can be collinear; any larger groupings will consist of non-collinear entries. To find the image pairs, we first note that combinatorially there are 36 ways to choose two entries from a set of nine. Further, as noted above, any two entries in a regular grid will be collinear, but we need not consider the reverse pairs, as the affine unary transforms are commutative. Figure 11 shows how these collinear pairs can be organized according to rows, columns, or diagonals. As 28 of these 36 pairs contain complete (i.e. non-empty) entries, there are 28 analogies that can be inspected by the model, given in Table 11.

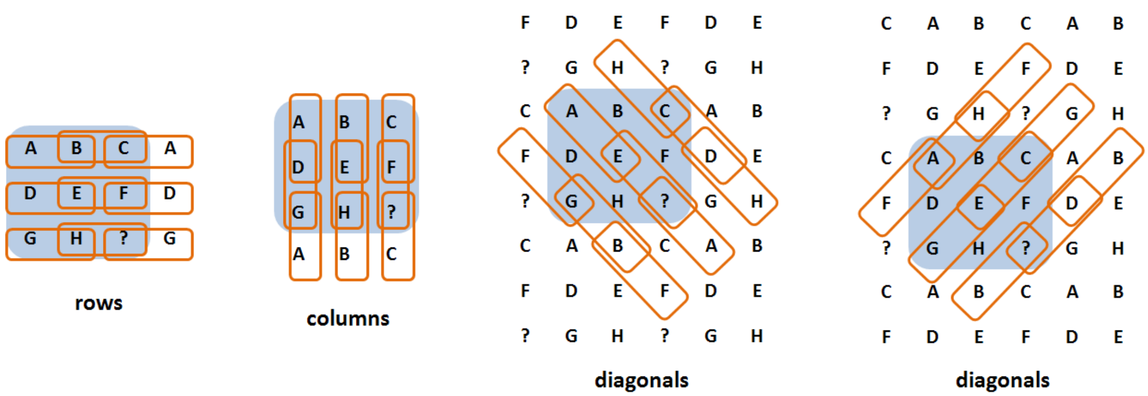


Figure 11. Collinear pairs of entries, and parallel sets of collinear pairs, in a 3x3 matrix.

Moving on to the case of image triplets within a 3x3 problem matrix, there are 84 ways to choose three entries from a set of nine. However, few of these will represent collinear triplets of entries in a 3x3 problem grid. In particular, we can count the collinear triplets by beginning again with the 36 possible collinear pairs. Of these pairs, one can form sets of three pairs which represent three collinear elements, e.g. **AB**, **AC**, and **BC**, which represent the triplet **ABC**. No other triplets in which any of these pairs participate will be collinear. Among the 36 possible collinear pairs, there are 36/3 = 12 such sets of three pairs representing a collinear triplet.

Note that for a given collinear triplet, e.g. **ABC**, there are actually six permutations that might be considered: **ABC**, **ACB**, **BAC**, **BCA**, **CAB**, and **CBA**. The five affine binary transforms are commutative with respect to the first two elements:

$$Union(A, B)=C \quad \Rightarrow \quad Union(B, A)=C$$

$$Intersection(A, B)=C \quad \Rightarrow \quad Intersection(B, A)=C$$

$$Subtraction(A, B)=C \quad \Rightarrow \quad Back\text{-}subtraction(B, A)=C$$

$$XOR(A, B)=C \quad \Rightarrow \quad XOR(B, A)=C$$

Thus, the triplet **ABC** is computationally equivalent to **BAC**, and likewise for **ACB** / **CAB** and **BCA** / **CBA**. However, examining the base binary transforms for triplet **ABC** will not be equivalent to examining the same transforms for triplets **ACB** or **BCA**; in particular, while the transposition of the first two images in a triplet do not matter, which entry takes the third place does matter. Furthermore, consider a triplet containing the empty entry in the matrix, call it **XY?**. As long as the empty entry is in the third place, then solving for it using a base binary transform is well-defined, as one just applies the binary transform to the first two known entries. However, if the empty entry is one of the first two places, e.g. **X?Y**, then solving for it will become ill-defined, as any number of different images may suffice to make the relationship true. Thus, the ASTI model restricts consideration of binary transforms over image triplets to only those triplets in which (or which are analogous to triplets in which) the empty entry takes the third place.

Graphically, these triplets are shown in Figure 12. There are a total of eight analogies that can be formed with these triplets, as shown in Table 11.

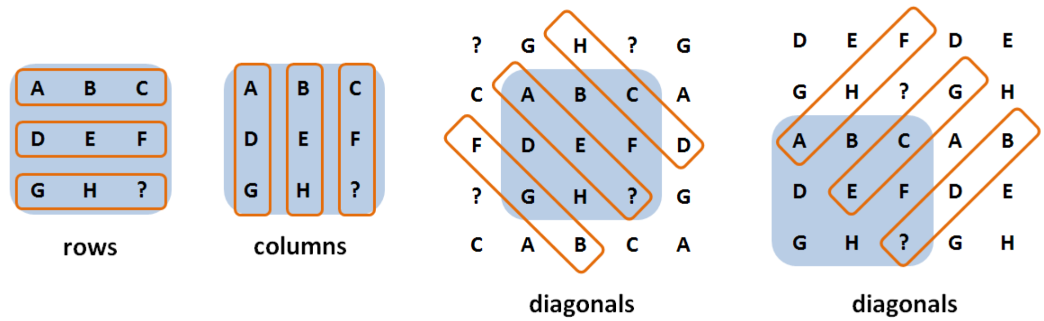


Figure 12. Collinear triplets and parallel sets of collinear triplets, in a 3x3 matrix.

Table 11. Listing of analogies in 3x3 matrix.

	Rows	Columns	Diagonals	
Pairs	A : B :: H : ? B : C :: H : ? D : E :: H : ? E : F :: H : ? G : H :: H : ? A : C :: G : ? D : F :: G : ?	A : D :: F : ? D : G :: F : ? B : E :: F : ? E : H :: F : ? C : F :: F : ? A : G :: C : ? B : H :: C : ?	F : G :: E : ? G : B :: E : ? H : C :: E : ? C : D :: E : ? A : E :: E : ? F : B :: A : ? H : D :: A : ?	F : H :: D : ? H : A :: D : ? G : C :: D : ? C : E :: D : ? B : D :: D : ? F : A :: B : ? G : E :: B : ?
Triplets	A : B : C :: G : H : ? D : E : F :: G : H : ?	A : D : G :: C : F : ? B : E : H :: C : F : ?	F : G : B :: A : E : ? H : C : D :: A : E : ?	F : H : A :: B : D : ? G : C : E :: B : D : ?

3.2.7 Generality of transform

There is much redundancy contained in 3x3 matrices; in particular, the same transform is repeated across each row or column of a matrix. One way for the ASTI model to take advantage of this redundancy is to look at best-fit image transformations across all row/column subsets, instead of just looking at the single best-fit image set. The basic affine configuration determines the best-fit image transformation for a matrix by searching among all possible base transforms and among all sets of entries listed in Figure 11 and Figure 12, and selecting the single set that yields the highest fitness value. For 3x3 problems, an alternate "aggregate" strategy was implemented that selects the best-fit image transformation by searching among the possible base transforms with fitness values *averaged* for sets of entries across all complete rows or columns. After the best-fit base transform has been chosen based on this aggregate fitness value, the single best-fit row or column is used together with the corresponding partial row/column, as in the standard model, to generate an answer prediction.

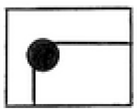
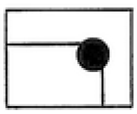
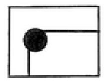
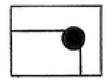
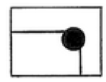
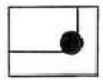
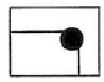
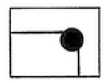
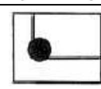
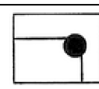
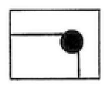
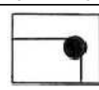
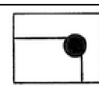
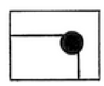
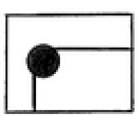
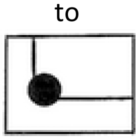
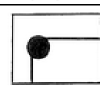
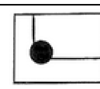
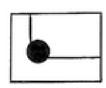
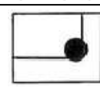
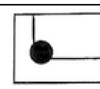
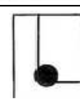
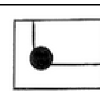
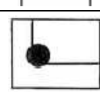
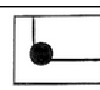
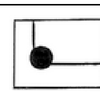
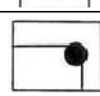
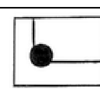
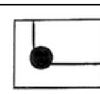
3.2.8 A detailed example of solving a 2x2 matrix problem

This section contains a detailed example of the operation of the ASTI model using the 2x2 sample RPM problem shown in Figure 3. The original problem image is first broken into the constituent matrix entry and answer images, as shown in Figure 4. Then, as shown in Figure 5, there are two possible combinations of entries that are used to induce transformations: the entries across the first row and the entries down the first column. The base transforms used in the induction process are the eight rotations/reflections shown in Figure 8.

For the top row and for the first column, the best-fit composite transformation $\mathbf{T_i}$ is calculated by the model according to the algorithm shown in Figure 7. The resulting similarity values from these calculations are given in Table 12. Once these similarity values have been calculated, the transformation yielding the highest similarity is chosen as the defining transformation for the matrix.

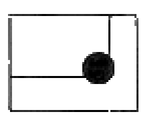
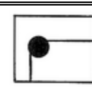
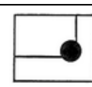
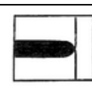
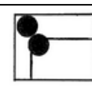
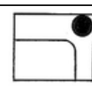
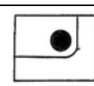
In this case, it is the rotate180-flip transform as applied to the images in the first row of the matrix, which yields a similarity value of 0.697. Then, for this particular transformation, the image composition operand is determined to be subtraction, as there are more pixels in **A** but not in **B** than vice versa, i.e. $\mathbf{\Sigma(A\text{-}B) > \Sigma(B\text{-}A)}$. In other words, the second image **B** roughly equals the first image **A** transformed and minus some pixels.

Table 12. Matrix similarity calculations for the example problem shown in Figure 3.

original images	base transform	first image transformed	second image	s	Σ(A-B)	Σ(B-A)
first row:  to 	identity			0.334	226.7	218.5
	rotate90			0.292	250.2	247.8
	rotate180			0.536	120.4	122.4
	rotate270			0.262	269.6	267.0
	identity-flip			0.318	235.9	229.4
	rotate90-flip			0.253	274.2	270.7
	rotate180-flip			0.697	59.5	58.7
	rotate270-flip			0.259	271.2	268.5
first column:  to 	identity			0.438	173.7	158.4
	rotate90			0.255	275.0	263.0
	rotate180			0.323	236.6	213.3
	rotate270			0.311	242.2	228.7
	identity-flip			0.608	104.6	86.3
	rotate90-flip			0.261	272.1	256.8
	rotate180-flip			0.289	254.9	234.1
	rotate270-flip			0.256	274.7	261.8

The predicted answer image is generated by taking the first image from the second row, applying the rotate180-flip transform, and subtracting the same pixels that represent the difference between the images in the top row. In this particular case, the first row images are fairly closely matched, and so the pixels that are subtracted are few in number but not zero, due to slight imperfections in the input images. Finally, the predicted answer image is compared to each of the answer choices, as shown in Table 13. The most similar answer choice is selected as the ASTI model's final answer, which is answer number 2, with a similarity value of 0.503.

Table 13. Answer similarity calculations for the example problem shown in Figure 3.

predicted answer image	**answer choice images**	**s**
		0.257
		0.503
		0.256
		0.211
		0.265
		0.277

3.2.9 Image processing to obtain inputs

To obtain inputs for the model in the form shown in Figure 4, I followed a standardized procedure, described in this section. First, each test booklet was cut at the spine, so that each problem was on a separate (but double-sided) sheet of paper. Each page was then scanned to create digital images of each problem. All scans were performed at 200 dpi on a grayscale setting (even the colored problems on the CPM). Each original scanned image was 1704 by 2197 pixels in size.

Then, each scanned image was cropped to remove the problem label at the top of each page. Also, any large dust specks that arose from the scanning process were manually removed, but only from the outer margins of each problem page (i.e. no pixels were altered within any of the actual problem content, either within the problem matrix box or within any answer choice box). These were the only manual corrections performed on the scanned images. The resulting images were then passed into an automated image processing system to carve up each problem into its constituent subimages.

One might assume that carving up each problem would be a simple matter of extracting subimages of particular sizes and positions across all problems; however, the RPM test booklets contain significant variation in sizing and positioning across problems, enough so that a simple one-size-fits-all solution was not sufficient to subdivide each problem. Therefore, automated methods were developed to extract the subimages, primarily by searching through the image to find the edges of the matrix and of each answer choice. Edges were determined using pixel intensity thresholds. For all of the image processing described here, the threshold for determining edges was set to 0.4 (where 0.0 is white and 1.0 is black). Thresholds for the CPM were set somewhat lower, ranging from 0.1 to 0.3, as some of the problems were very light when scanned.

First, each problem image was corrected for rotational misalignment. The image processing system searched from the top of each image to locate the upper edge of the box enclosing the problem matrix and performed a simple linear regression to determine the slope of this line. Then, the entire image was rotated to make this line horizontal.

Second, the position and dimensions of the matrix box were determined by searching for each of its four edges. For 2x2 matrices, the entire box was then divided into four quadrants, and the first three of these quadrants became the three matrix entry inputs into the ASTI model. For 3x3 matrices, once the edges of the matrix box were found, the position of the empty box within the matrix (in the lower-right-hand corner) was determined, again by searching to find its edges. The position of this empty box was used to define the size of each of the rows and columns in the 3x3 matrix. In particular, the bottom row and right-most column were defined to be aligned with the edges of the empty box. The top row and left-most column were defined to be the same sizes as the bottom row and right-most column, respectively.

One additional correction was performed by the image processing system to account for mis-aligned matrix entries in the original problems for 3x3 matrices. The height of the first two rows and the widths of the first two columns were automatically adjusted to avoid chopping off part of a figure in a matrix entry. In particular, if the divide between the first and second rows crossed any non-white pixels, its position was incrementally moved up or down (up to a maximum displacement) until it no longer crossed any non-white pixels. The divide between the first and second columns was automatically adjusted in the same fashion. Note that the resulting subimages for 3x3 matrix entries within a single problem were not necessarily the same size. Figure 13 illustrates results of the automated image processing system for 2x2 and 3x3 matrix problems.

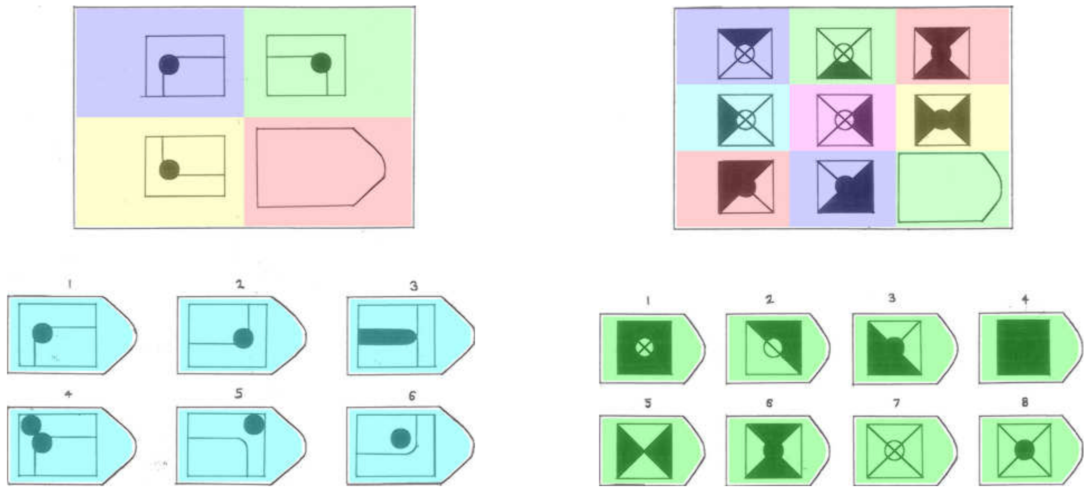


Figure 13. Automated segmentation of 2x2 and 3x3 matrix problems.

3.3 Results from the ASTI Model

In this section, I present results from running the ASTI model against the SPM, the CPM and the APM. When RPM results are typically analyzed, the total score from each test summarizes the overall level of performance. However, the raw numerical score may have little meaning to those unfamiliar with the test, and so the total score is typically compared to national age-group norms to determine a percentile ranking. As the ASTI model does not have an "age" with which to look up its percentile rank, it is instead assumed that the model is performing at the 50th percentile, and then an "average age" can be inferred at which human test-takers would show an equivalent level of performance. These comparisons with human norms are presented not to suggest that the problem solving processes used by humans of a certain age strictly resemble those of the ASTI model but merely to give readers unfamiliar with the RPM tests some indication of what level of ability is indicated by a particular score.

Because the ASTI model is not a process model of solving RPM problems, variables that have to do with the sequence or ordering of processing of inputs or of reasoning are not valid. Thus, measures like reaction time or attention (e.g. eye-tracking) are not valid for this model.

The Standard Progressive Matrices (SPM) consists of 60 problems divided into five sets of 12 problems each, labeled Sets A through E. The first 24 problems (Sets A-B) consist of 2x2 matrices. The Colored Progressive Matrices (CPM) consists of 36 problems divided into three sets of 12 problems each, labeled Sets A, AB, and B. Sets A and B are identical to Sets A and B from the SPM, except that problems are presented in color. Set AB is intended to be of intermediate difficulty between Sets A and B. All 36 of the problems on the CPM contain 2x2 matrices. The Advanced Progressive Matrices (APM) consists of 48 problems divided into two sets of 12 problems and 36 problems, respectively, labeled Sets I and II. All 48 problems on the APM consist of 3x3 matrices. The ASTI model was tested on all problems from all three tests.

Note that the ASTI model was developed primarily after my own inspection and completion of the SPM test. This test was used to provide insights into what kinds of problem solving capabilities the model might need to solve these types of matrix problems. However, I did not take the CPM or the APM prior to testing the model against these tests, and I had only minimal exposure to their problems.

3.3.1 Results and discussion for the SPM

The ASTI model correctly solves 50 of the 60 problems on the SPM. For typically developing children in the U.S., this total score corresponds to the 50th percentile for 17-year-olds, as shown in Figure 14 (Raven et al., 2003). A breakdown of this score across sets is given in Figure 15, along with the expected score composition across sets for participants achieving the same total score.

The ASTI model performs extremely well on the SPM, achieving a score near the upper end of the test's discriminable range. The scores across sets also resemble the expected score composition for human test-takers, deviating by no more than +/- 2 on any given set, indicating that the problems in Sets A and B which are easy for humans also seem to be easy for the ASTI model, while the model has trouble with more difficult problems found in Sets D and E.

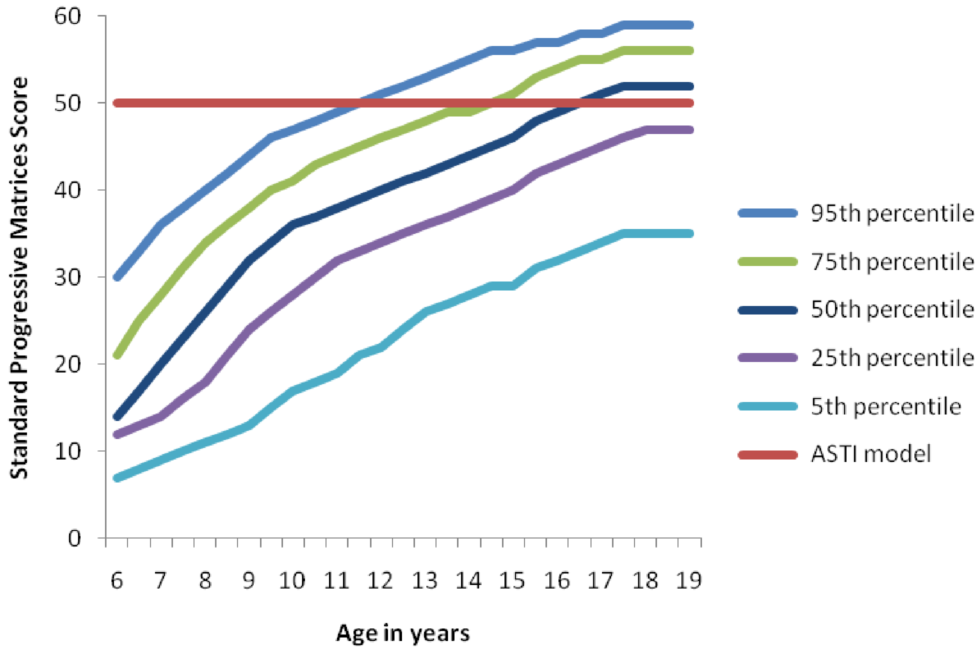


Figure 14. Score achieved by the ASTI model on the SPM, along with norms for typically developing children in the USA (Raven et al., 2003).

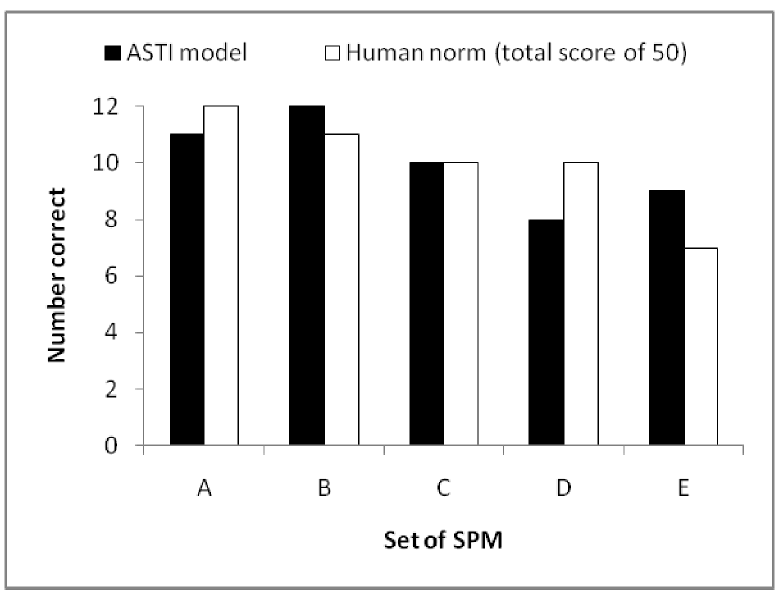


Figure 15. ASTI score broken down by Sets A through E on the SPM, with expected set score composition for same given total score (Raven, Raven, & Court, 2003).

Table 14 gives a comparison of the performance of the ASTI model with other computational models that have been tested against the SPM. The ASTI model achieves the highest overall score, but both propositional models perform better on the later sets than does the ASTI model. One reason for this might be that the ASTI model currently lacks the ability to perform segmentation of a single matrix entry into multiple entries that follow different transformation rules. For example, a problem might have three inner shapes that are permuted across rows and columns and three outer shapes that remain constant across rows, as illustrated in Figure 16. The ASTI model cannot currently account for these types of transformations, though there is no *a priori* reason why such transformations could not be implemented using iconic representations. Segmentation could be done by iteratively seeking transformations to successively explain differences between subsets of pixels in each matrix entry, until no pixels remain to be explained.

Table 14. Results from various computational models on the SPM, by set. Each set contains a total of twelve problems.

Model	Set A	Set B	Set C	Set D	Set E
Cirillo & Ström (2010)	n/a	n/a	8	10	10
Lovett et al. (2010)	n/a	44 total for sets B through E			
ASTI model	11	12	10	8	9

Segmentation of this type is likely necessary for any computational model to solve all of the problems on Set D. Both of the propositional models listed in Table 14 receive, as inputs, matrix entries already segmented into discrete shapes, as vector graphics, although the Lovett et al. (2010) system did have the ability to re-group and re-segment discrete shapes and edges within its vector graphics representations. The Carpenter et al. (1990) model (to be discussed later in Section 3.3.3) received hand-coded inputs already segmented into propositional features. A question for future work is how automated image processing techniques might be applied to perform image segmentation of RPM problems, and what background knowledge is needed regarding the identities of shapes and other visual entities in order to perform such segmentation.

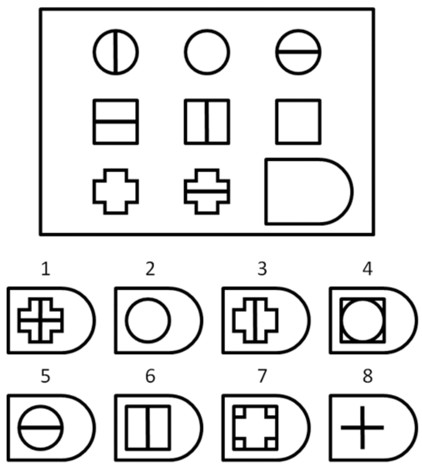


Figure 16. Example problem showing multiple elements in individual matrix entries.

One interesting aspect of the SPM results in Table 14 is that (as far as we can tell from published findings) only models using visual representations have ever attempted Set A of the SPM, which, according to human normative data, is purportedly the easiest set on the test The problems on Set A of the SPM (see Figure 17 for examples), are qualitatively different from the problems on Sets B through E in that they resemble pattern-completion problems more than geometric analogy problems. It may be that part of the reason that no propositional models have been tested against Set A is because these types of problems are very difficult to represent using propositions, especially within propositional schemes that focus on representing discrete shapes and attributes.

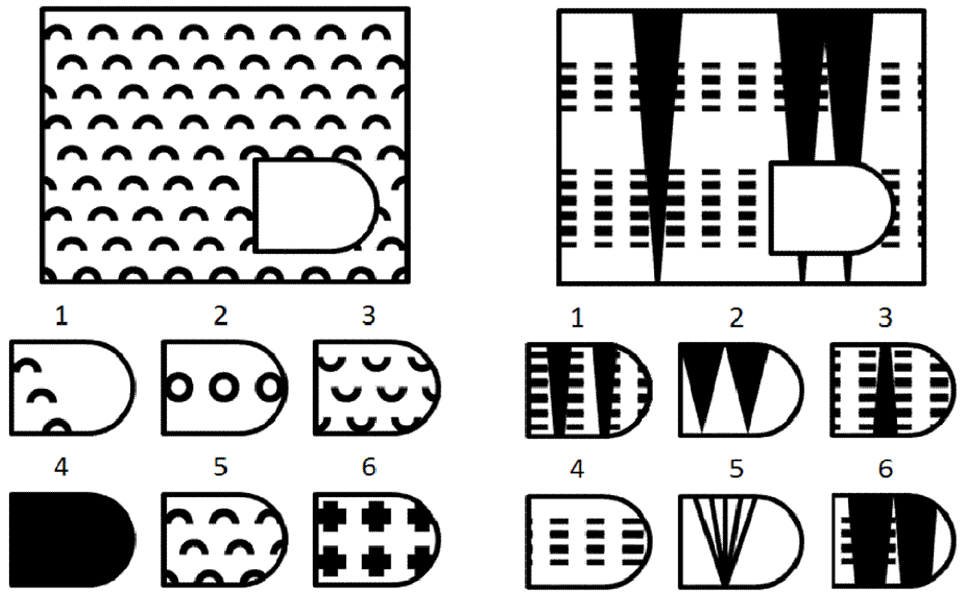


Figure 17. Two examples of the "pattern-completion" type of matrix problems found in Set A of the SPM and of the CPM, as well as in a few early problems on the APM.

Some of these types of problems could potentially be represented propositionally as textures, but this approach would be difficult for problems such as that shown on the right of Figure 17, in which no quadrant of the matrix contains a uniform texture. Furthermore, extracting propositional descriptions of texture directly from an image is in itself a difficult computational task. These problems might also be represented propositionally using a richer vocabulary that includes lower-level elements such as edges and lines (for example, as obtained by the edge segmentation process in the Lovett et al. 2010 model), but this approach greatly increases the computational complexity of the problem; instead of problems containing two or three or even ten elements per matrix entry, a single problem as shown in Figure 17 might have dozens or even hundreds of elements.

Such problems are very easy to represent using modal iconic representations of the type used by the ASTI model; the representation simply consists of the scanned images from the test. In fact, using pixel-based representations, none of the problems on the SPM are particularly more difficult to represent than any others, and this type of representation seems highly effective, as the ASTI model achieves a very high score on Set A.

Human factor-analytic studies of the SPM have typically classified these pattern completion problems as loading on a “gestalt” cognitive factor, in contrast to visuospatial or verbal factors. These data seem to suggest that pattern completion problems may be solved by humans using qualitatively different strategies than those used on the geometric analogy type of RPM problem. The ASTI model currently solves the problems in Set A using the same mechanisms used on later problems. In particular, the ASTI model looks at discrete transformations within the problem matrix, i.e. going from one image to another, which is akin to using a rule-based, albeit visual, approach (where the rules are conceptualized as image operations of affine and set transformations). A gestalt approach might differ by looking at the entire problem matrix as a whole, using principles of visual coherence such as symmetry and continuity.

3.3.2 Results and discussion for the CPM

The ASTI model correctly solves 35 of the 36 problems on the CPM. For typically developing children in the U.S., this total score corresponds to the 95th percentile for all children under 12 years of age, as shown in Figure 18 (Raven et al., 2003). A breakdown of this score across sets is given in Figure 19, along with the expected score composition across sets for participants achieving the same total score. Note that Sets A and B on the CPM are identical to Sets A and B on the SPM, except with colored diagrams instead of purely black and white ones.

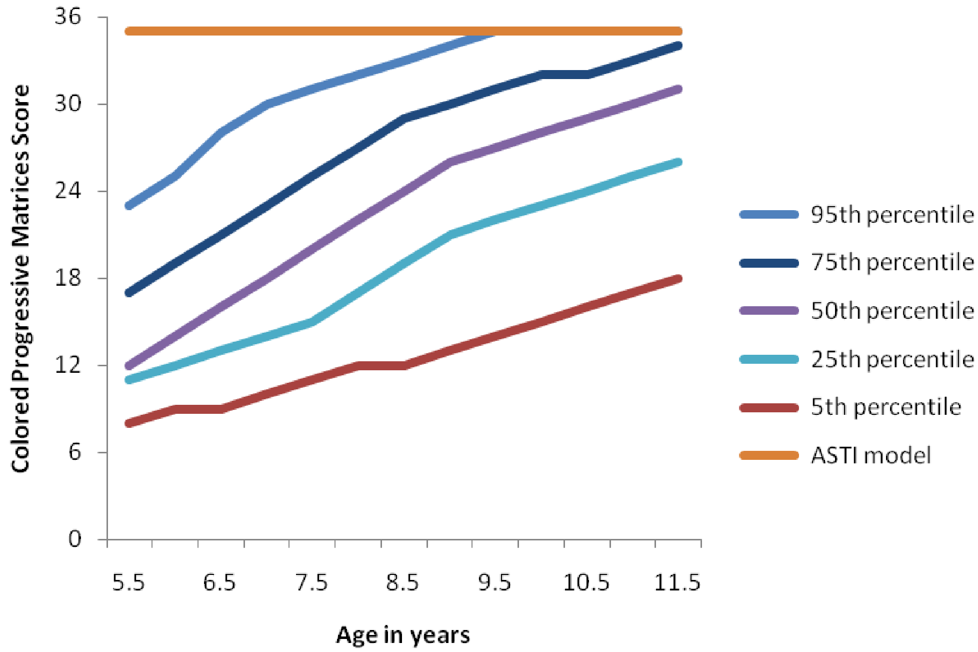


Figure 18. Score achieved by the ASTI model on the CPM, along with norms for typically developing children in the USA (Raven et al., 2003).

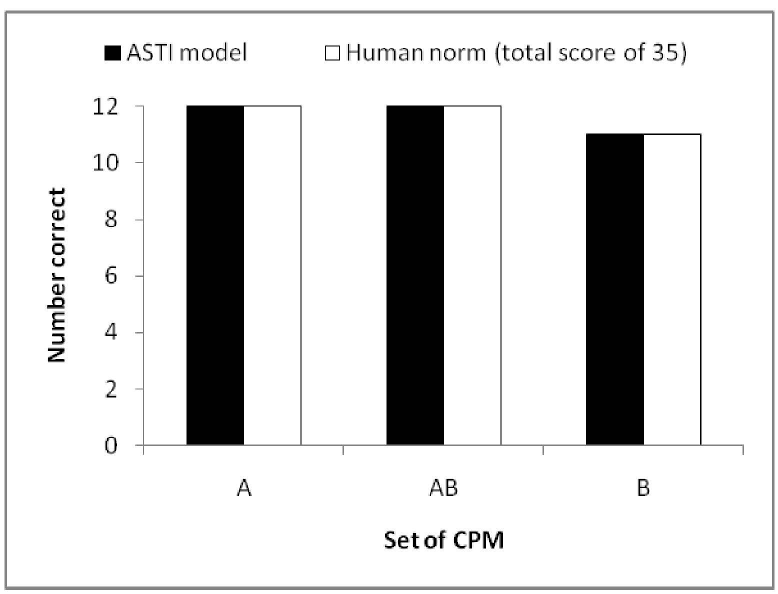


Figure 19. ASTI score broken down by Sets A through B on the CPM, with expected set score composition for same given total score (Raven, Raven, & Court, 2003).

The ASTI model performs extremely well on the CPM, achieving a score that is essentially at the ceiling of what the CPM can measure. The scores across sets exactly match the expected score composition for human test-takers. The single problem missed on the CPM was B9. This problem was actually solved correctly on the SPM, and the ASTI model correctly solves problem A12 on the CPM which was missed on the SPM. These two discrepancies are explained by the different pixel thresholds used in testing the ASTI model against the SPM and the CPM, which is discussed further in Section 3.4.1. No other computational model has ever been tested against the problems on the CPM.

3.3.3 Results and discussion for the APM

The ASTI model correctly solves 18 of the 48 problems on the APM. For typically developing adults in the U.S., this total score corresponds to around the 15th percentile for most adults, as shown in Figure 20 (Raven et al., 2003). Note that norms for the APM are given for adults instead of for children, and these data show a decline in scores with increasing age. A breakdown of this score across sets is given in Figure 21. The APM does not give data on the expected score composition across sets for all participants achieving the same total score, but for adults around 60 years of age, Figure 21 also shows their expected score composition alongside the affine results.

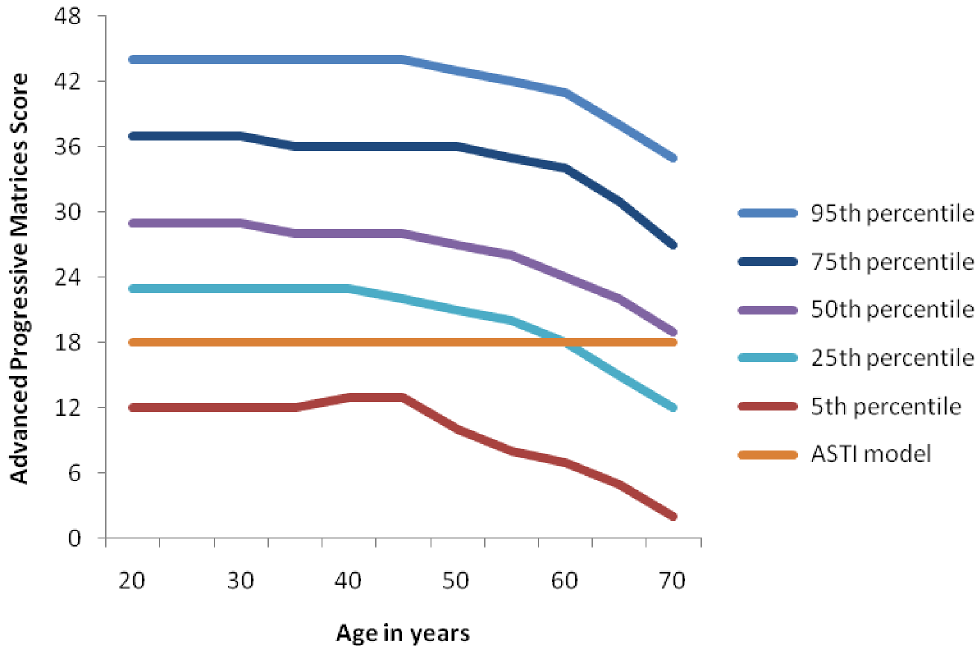


Figure 20. Score achieved by the ASTI model on the APM, along with norms for typically developing adults in the USA (Raven et al., 2003).

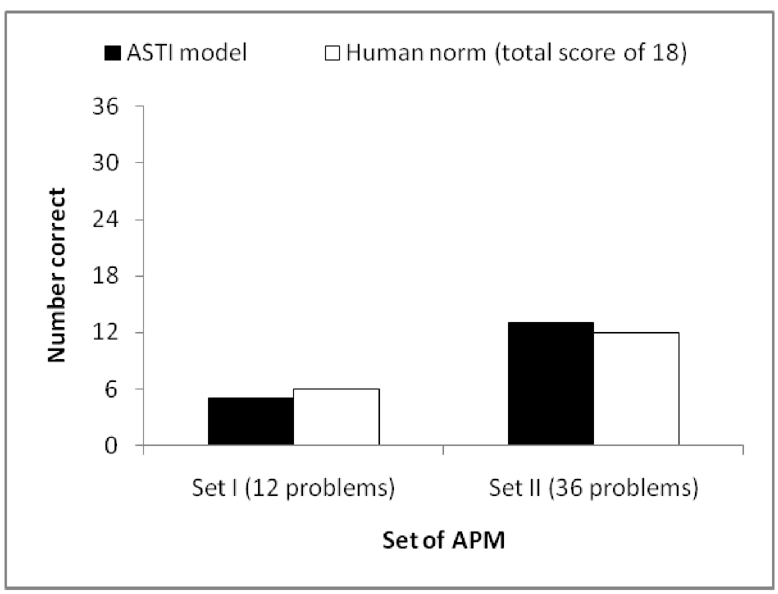


Figure 21. ASTI score broken down by Sets I and II on the APM, with expected set score composition for same given total score (Raven, Raven, & Court, 2003).

Carpenter et al. (1990) report results of running two versions of their production system model (FairRaven and BetterRaven) against a subset of APM problems. A comparison of their results with the results of the ASTI model is shown in Table 15.

Table 15. Results from various computational models on the APM given as total correct out of total number attempted, by set.

Model	Set I 12 problems	Set II 36 problems
Carpenter et al. (1990): FairRaven	7 out of 7	16 out of 27
Carpenter et al. (1990): BetterRaven	7 out of 7	25 out of 27
ASTI model	5 out of 12	13 out of 36

Both of the propositional models do much better than the ASTI model on the APM. As mentioned with regard to SPM results, part of the reason for this discrepancy may have to do with image segmentation. Both Carpenter models received as input hand-coded propositional feature vectors that already contained segmented descriptions of the problem content. It may be that adding a mechanism to perform visual segmentation to the ASTI model would be sufficient to boost its performance on the APM.

As seen in this table, both of the Carpenter et al (1990) models were tested on only 7 of the 12 problems in Set I, and 27 of the 36 problems in Set II. The reasons for this omission were described as stemming from the limitations of a digitized display system that was used for behavioral experiments conducted prior to the computational analysis, though there is no explicit mention of why these problems were omitted from the computational analysis as well. It may be that the same problem features that caused problems with the digital display also caused problems for the authors' hand-coding of propositional problem representations, though there is not enough information in the paper to know this for sure. The problems omitted were problems 1, 3, 4, 5, and 11 from Set I, and problems 2, 11, 15, 20, 21, 24, 25, 28, and 30 from Set II.

The first four problems in the APM resemble the pattern-completion problems in Set A of the SPM, as discussed in Section 3.3.1; of these four, three are omitted by the Carpenter et al. (1990) models. The fourth problem (problem 2 in Set I) contains discrete elements not unlike those in the geometric analogy type of problem, except with continuous lines added around the elements. Informal inspection might suggest that this particular problem can be solved even if its continuous, pattern-like content is ignored. So, as with the SPM, the results from Carpenter et al. (2010) suggest that propositional models have yet to attempt pattern-completion problems. Of the four pattern-completion problems on the APM, three are answered correctly by the ASTI model.

3.4 Altering Representations within the ASTI Model

In this section, I discuss the results of two experiments conducted with the ASTI model to investigate the effects of altering the low-level representational commitments made by the model, namely its representation of pixels and of visual similarity.

3.4.1 Representation of pixels

One tunable parameter in the ASTI model lies in how pixels are represented. In particular, pixels in the ASTI model are represented as binary black-and-white values, where only two values are possible: 0 (white) and 1 (black). While this approach reduces noise in the inputs, some fine-grained detail can be lost due to the radical shifting of raw pixel color values to extremes of black and white. Note that though an image in the test booklet may appear to be purely black and white, the test when scanned contains many grey-valued pixels, for instance on fine lines or at the edges of shapes.

Pixels in the ASTI model contain a threshold, expressed as a percentage value ranging from 0% (pure white) to 100% (pure black), above which pixels received as input from the original scanned images will be converted to black, and below which pixels will be converted to white. Changing this parameter essentially alters how the model performs figure-ground discrimination on the inputs that it receives. High values of the threshold are more resistant to grey-valued noisy pixels in the images, but lower values can capture more fine levels of detail.

The threshold for the initial run of the ASTI model against all three RPM tests described in the previous section was manually set to a threshold of 10%, based on visual inspection of sample test problem images. (For the CPM, to account for the variations in grey values among problems that were colored in different shades, the threshold was adaptively set by the ASTI model to either 25% or 62.5%.) In order to investigate the effects of varying this figure-ground separation threshold, I ran the ASTI model against the SPM using ten different threshold settings, distributed evenly from 5% up to 50%.

3.4.2 Results of pixel-representation experiment

Figure 22 shows a graph of the total scores achieved by the ASTI model on the SPM for the ten different assignments of the pixel threshold, ranging from 5% up to 50%. As this figure shows, the scores overall are very similar, with the lowest being 46 and the highest being 50. Thus, the ASTI model seems to be fairly robust to changes in the underlying pixel representation. The lower thresholds seem to yield slightly higher scores, which may indicate that the loss of visual information as the pixel threshold increases affects the performance of the ASTI model more than the addition of visual noise as the pixel threshold decreases.

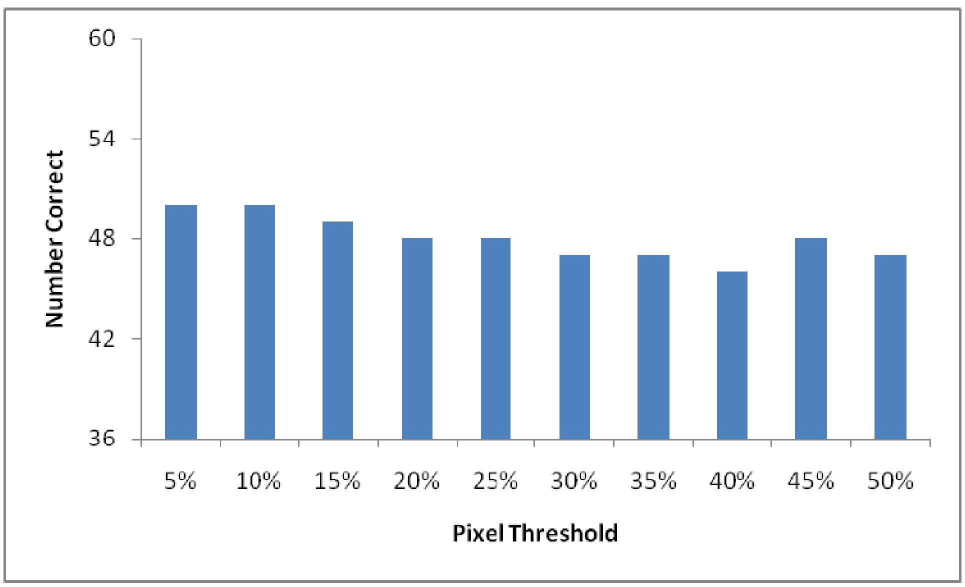


Figure 22. Total scores achieved by the ASTI model on the SPM for various assignments of the figure-ground separation pixel threshold value.

Figure 23 shows the scores for the same variations in pixel thresholds for each set on the SPM. The problems on Set A, which are densely packed with visual information, seem to suffer with low pixel threshold values, likely as a result of added noise that obscures the patterns and textures found in these problems. In contrast, the later problems, especially in Sets B, C and E, seem to show improved performance at lower thresholds, likely because these problems consist of fine line drawings for which higher thresholds may lose important visual details.

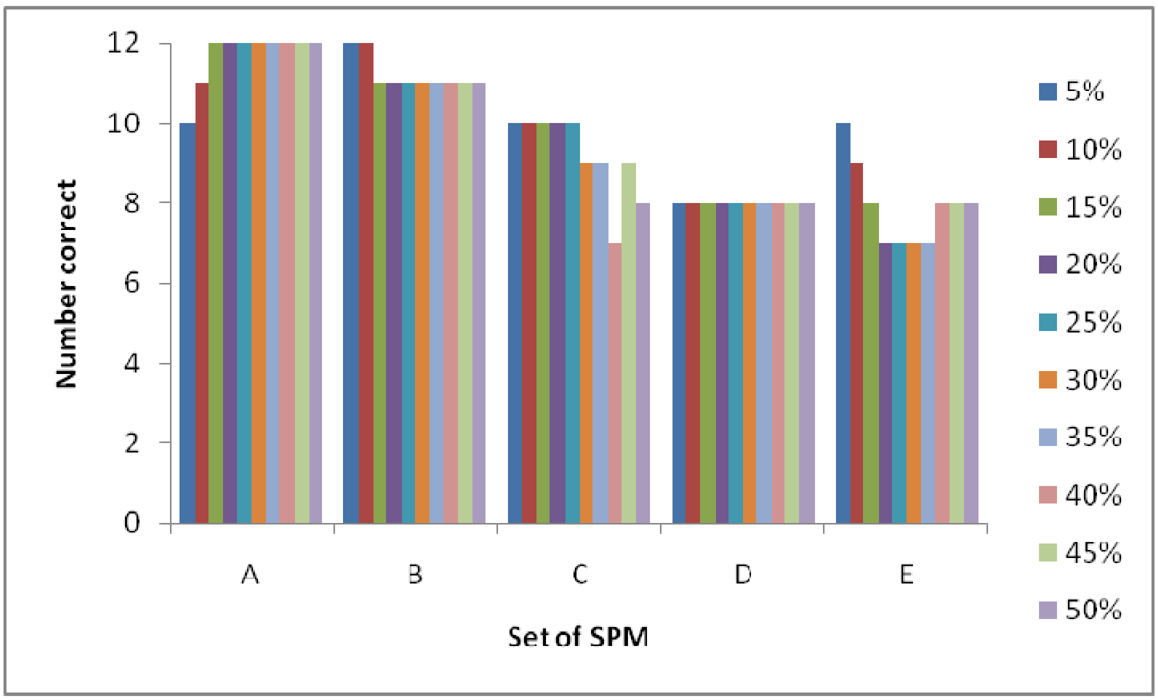


Figure 23. Scores achieved by the ASTI model for each set on the SPM for various assignments of the figure-ground separation pixel threshold value.

The highest performance achieved by any combination of thresholds is 52 total problems correct. One important area for future work will be in developing autonomous methods for a visual model like the ASTI model to tune its own pixel representation thresholds. This would essentially serve as a mechanism by which the model can identify salient features on its own and tune thresholds accordingly.

3.4.3 Representation of similarity

Calculating similarity is a central facet of the ASTI model and would be of any model using modal iconic representations. In order to investigate the effect of using different formulations of visual similarity, I implemented a sum-squared-differences (SSD) measure of similarity, in addition to the Tversky similarity measure originally used by the ASTI model, defined as:

$$similarity(A, B) = \frac{1}{1 + \sum(p_A - p_B)^2} \tag{3}$$

Note that the model takes the reciprocal of one plus the sum of squared differences between pixel intensities in order to convert the usual SSD measure of difference into one of similarity that varies between one (for identical images) to zero (for images with infinite differences in pixels), as shown in Figure 24.

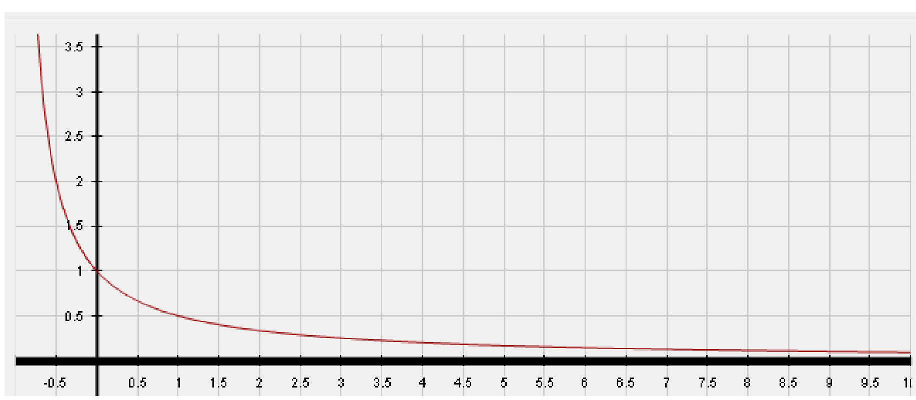


Figure 24. Graph showing range of values for sum-squared-difference similarity measure, where x-axis indicates magnitude of raw SSD calculation.

The absolute magnitude of the difference between the union and intersection of pixels approximates the result of the SSD measure (as the individual pixel values vary between zero and one). Thus, for the same magnitudes of SSD, we can approximate the Tversky similarity measure equivalents by setting SSD equal to the union of pixels minus the intersection, and then "solving" for the Tversky measure of union divided by intersection, as given in Equation (4):

$$SSD \approx union - intersection$$

$$union \approx SSD + intersection$$

$$Tversky = \frac{intersection}{union}$$

$$Tversky = \frac{intersection}{intersection + SSD} \tag{4}$$

Thus, for various values of the intersection between two images, the Tversky similarity measure will scale with the raw magnitudes of SSD very much like the SSD similarity measure defined above, as shown in Figure 25.

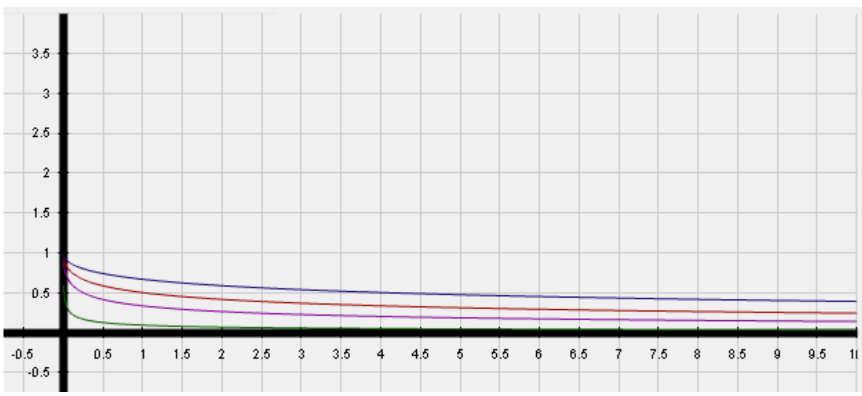


Figure 25. Graph showing how Tversky similarity measure changes for various values of intersection (curves) and magnitudes of SSD (x-axis).

To study the effect of altering the ASTI model's representation of similarity, I tested the model using the SSD measure of similarity on the SPM using a pixel threshold of 0.1.

3.4.4 Results of similarity-representation experiment

Figure 26 shows a graph of the total scores achieved by the ASTI model on the SPM for the Tversky and SSD similarity metrics. As this figure shows, the SSD configuration does not perform very well, correctly answering only 32 problems as compared to 50 problems for the Tversky configuration. Thus, the ASTI model seems to be fairly sensitive to changes in the underlying similarity representation. Figure 27 shows the scores for both similarity metrics for each set on the SPM.

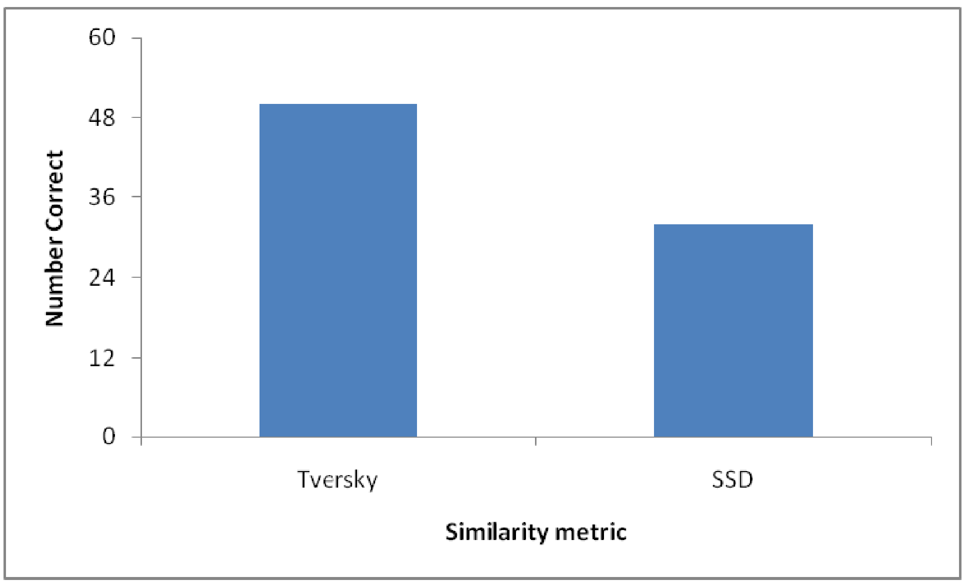


Figure 26. Total scores achieved by the ASTI model on the SPM for the Tversky and SSD similarity metrics.

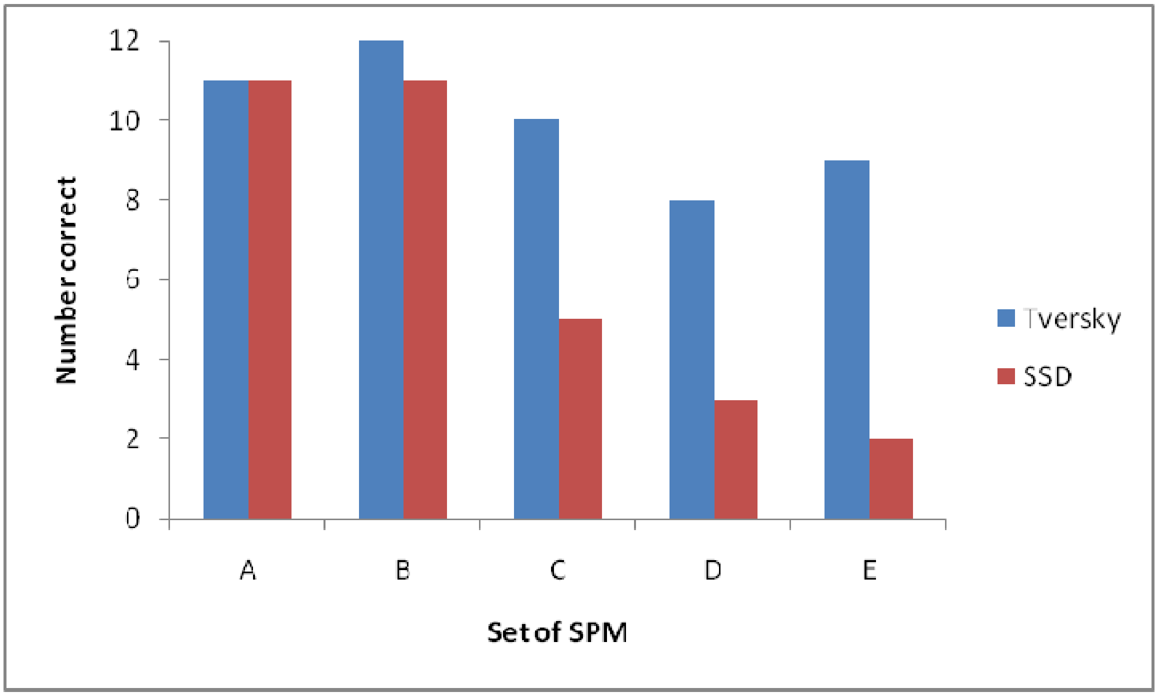
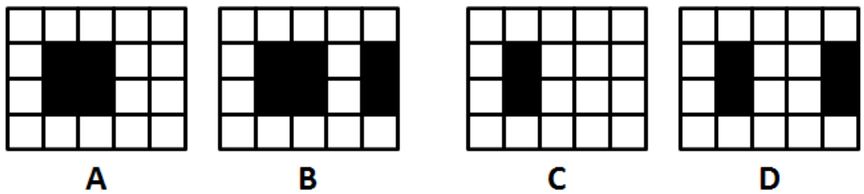


Figure 27. Scores achieved by the ASTI model for each set on the SPM for the Tversky and SSD similarity metrics.

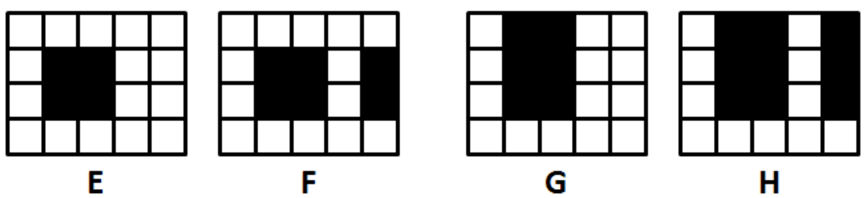
Why does changing the similarity measure affect the performance of the ASTI model? A closer look at each similarity measure suggests one possibility. Consider the image pairs shown in Figure 28. For pairs **AB** and **CD**, the number of pixels that are different between the images within each pair is the same—two pixels—but the amount of common pixel content that is shared is different—four pixels in pair **AB** and only two pixels in pair **CD**. The Tversky measure, given in Eq. (2), privileges matches that share more pixel content, and so images **AB** yield a higher similarity value than do images **CD**.

In contrast, the SSD similarity measure, given in Eq. (3), effectively ignores any pixel content that is shared; similarity is calculated only as a function of pixels that are different. Thus, the SSD measure yields identical similarity values for image pairs **AB** and **CD**, because within each image pair, there are two mismatched pixels. The opposite pattern can also occur: for image pairs **EF** and **GH**, the Tversky measure yields identical similarity values, but the SSD measure prefers pair **EF**, because **EF** has only two mismatched pixels, whereas **GH** has three mismatched pixels.

In summary, the Tversky measure considers both shared image content as well as content that differs between two images. This appears to be an important component of similarity for solving visual analogy problems like those found on the RPM.



Tversky $f(A \cap B) / f(A \cup B)$	4 / 6 = 0.667	2 / 4 = 0.500
SSD $1/(1+\Sigma(p_A-p_B)^2)$	1 / (1+2) = 0.333	1 / (1+2) = 0.333



Tversky $f(A \cap B) / f(A \cup B)$	4 / 6 = 0.667	6 / 9 = 0.667
SSD $1/(1+\Sigma(p_A-p_B)^2)$	1 / (1+2) = 0.333	1 / (1+3) = 0.250

Figure 28. Illustration of differences between Tversky (1977) and SSD similarity measures, as applied to pixel-based images.

3.5 A Note on Inputs

In this section, I discuss the rationale for choosing to work with raw, scanned images from the test, instead of using artificially generated, "clean" input images. The ASTI model takes as inputs scanned images from the actual RPM test booklets. Some preprocessing is done on these images; they are manually rotated to correct for rotational misalignments during the scanning process, and then they are automatically sliced into constituent images for each matrix entry and answer choice. Even after these preprocessing steps, however, the images fed as inputs into the ASTI model are still very noisy; they contain numerous pixel-level artifacts and misalignments from the scanning process, and in addition, the figures in the RPM test booklets are not (at a fine level of detail) as precise as they might appear to the human eye. For example, part of a matrix element that appears to be symmetric about its horizontal axis can be measured and found to be off by a half a millimeter, which sounds tiny until one considers that that entire segment of the figure is only seven millimeters wide. Entries that are clearly meant to appear identical across multiple matrix entries are not exact duplicates of one another; this becomes especially apparent in figures that incorporate textures such as stripes.

For these reasons, the similarity values calculated by the ASTI model are often much lower than one might expect. In the example problem discussed in Section 3.2.8, even though the predicted answer looks very like one of the given answer choices, the calculated similarity between the two images is only 0.503.[4]

One might ask, why not just create "clean" computerized input images to eliminate the imprecision found in scans of the SPM test booklet? There are three reasons for

[4] This example problem was hand-drawn using rulers, stencils, and ink, in order to emulate the level of imprecision found in the actual RPM test booklets.

choosing to work with the original scanned images. The first reason is simple: given that humans use paper copies of the test, one might say that the model should try to tackle inputs that are as close as possible to the originals. Humans do not receive the benefit of having "cleaned up" versions of RPM problems, and so neither should a computer model.

A second reason has to do with model robustness when faced with low-level representational irregularities. Part of the power of amodal propositional representations comes from their ability to abstract away from the raw pixel level, and, for example, call two squares "identical" despite slight mismatches in size or alignment. However, methods using modal iconic representations can also achieve similar levels of robustness using calculations of visual similarity at the pixel level, whether or not the inputs have been "cleaned up." The field of image processing regularly deals with noisy, imperfect images, and the ASTI model strives to maintain some of that realism and take the actual RPM test problems as they come.

The third and most important reason for choosing not to redraw RPM problems is that there is a strong methodological argument against it (whether they are redrawn as vector graphics or even just as more precise raster images). As an example, consider redrawing the shapes shown in Figure 29. At first glance, these images might appear to be identical, and it would be tempting to create the first circle with stripes and copy it in order to create the second. However, closer inspection will reveal that, although the high-level texture might be described in the same way, at a low level, the images are drastically different—the calculated similarity between these two images is a mere 0.253! While the outer circular outlines are alike, the inner "textured" portion of each circle is almost exactly a negative image of the other.

Figure 29. Illustration of “same” texture with wildly different pixel-level properties.

As another example, specific to redrawing problems as more precise raster images, consider the problem shown in Figure 30, which was drawn using vector graphics in PowerPoint and then exported as a raster image. Looking just at the top row of matrix entries, and using the set of eight affine base transformations shown in Figure 8, it becomes apparent that the top-row image transition could equally well be described as a “rotate180flip” transformation (i.e. a reflection about the vertical axis) or as a “rotate270” transformation (i.e. a one-quarter counter-clockwise rotation). It follows that the model *ought* to compute that either of these transformations is equally well-suited, and choose one according to whatever tie-breaker is in place.

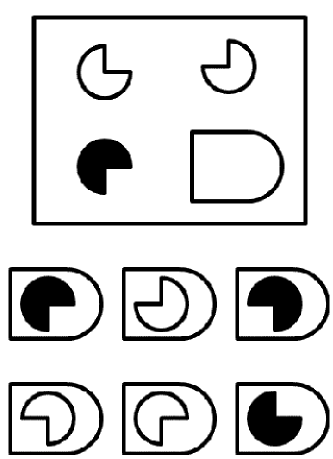


Figure 30. 2x2 example problem to illustrate impacts of “clean” input images.

However, the actual output of the model depends, in fact, on how the problem was originally created using vector graphics, even after the images have been rasterized. In particular, when recreating this problem using vector graphics in Powerpoint, the original, top-left image in the matrix was used to construct two different versions of the top-right image. For the first version (the “rotated version”), the top-left vector graphic was rotated 90 degrees to the left. For the second version (the “reflected version”), the top-left vector graphic was reflected about its vertical axis. Then, all of these vector images were rasterized to create input images to feed into the ASTI model.

Results from calculating similarity values over all base affine transformations for these two versions of the top-row images are shown in Table 16. For the rotated version, the rotate transformation is found to yield the highest image similarity. In contrast, for the reflected version, the rotate180-flip (i.e. reflection) transformation is found to yield the highest similarity. While the slight differences present in the final rasterized images would likely not influence the behavior of a human taking the test, these differences represent enough of a bias that they can completely change the output of a model that uses pixel-based representations.

As these examples show, when redrawing RPM problems, the specific choices by which “clean” images are created can have a non-trivial impact on the visual information contained in the problem and thus can significantly alter the output of a computer model. Redrawing could also introduce bias if the drafter has foreknowledge of the computer model to be tested against the problems, as they may consciously or unconsciously redraw problems with the problem-solving algorithm in mind. Lovett et al. (2010) notes that for their experiments, one of the SPM test problems was redrawn using a grey line instead of the original dotted line “for simplicity.” While humans solving this problem would likely not be much affected by such a change, it does raise questions of when such simplifications are appropriate and when they might, in fact, be materially changing the substance of a problem for a computational system. For all of these reasons, the ASTI model deliberately uses images scanned directly from the printed RPM test booklets.

Table 16. Similarity calculations for example problem shown in Figure 30. Underlined values indicate highest values of similarity for each calculation.

base transform	original images	s	original images	s
identity	**rotated**	0.456	**reflected**	0.439
rotate90	**version:**	0.347	**version:**	0.325
rotate180		0.449		0.431
rotate270		**0.884**		0.818
identity-flip	to	0.341	to	0.340
rotate90-flip		0.458		0.433
rotate180-flip		0.881		**0.825**
rotate270-flip		0.452		0.419

3.6 Claims and Future Work

In this chapter, I show that, for many RPM problems, it is not necessary to extract amodal symbols in order to arrive at the correct answer, and iconic visual representations often constitute a sufficient form of representation to solve these problems. The ASTI model is intended to serve as a complementary account to existing propositional models, which together may provide an integrated, dual-process account of human problem solving on the RPM. I conclude this chapter with a few remarks about this work.

First, my aim is not to show that the ASTI model is "better" or "worse" than previous computational models, but rather to explore to what extent a particular set of iconic representations and mechanisms can succeed on a body of RPM problems, just as previous computational models have explored to what extent particular propositional accounts can be successful. I discuss results from the ASTI model in comparison with other models in order to evaluate how the representational commitments made by such models affect their performance on various subsets of RPM problems.

Second, the ASTI model demonstrates only one possible instantiation of the use of modal iconic representations for RPM problem solving. The spectrum of possible iconic representations ranges from the type of low-level, pixel-based representation used by the ASTI model to more complex representations explicitly containing edges, lines, shapes, topological information, etc. One question for further exploration is how models that use other types of iconic representations might perform on the RPM.

Third, while the ASTI model does not seek to provide an account of or model all of the microstructures and processes of human visual cortical processing, the operations it uses (affine transformations and set operations) are mathematically grounded for general forms of imagery or visualization and are based upon evidence from studies of mental imagery. Both affine transformations and set operations can be formally defined as general types of transformations over any two-dimensional plane figures, whether pixels, edges, shapes, or otherwise. These types of operations, or operations that are computationally isomorphic, have been found to play a role in mental imagery tasks ranging from mental rotation (Shepard & Metzler, 1971) and scanning (Kosslyn, Ball, & Reiser, 1978) to image addition and subtraction (Brandimonte, Hitch, & Bishop, 1992ab).

Fourth, while the ASTI model was designed to use forms of inference similar to those evidenced by studies of mental imagery in humans, not all elements of the model are intended to be interpretations of human cognitive processing on the RPM. The primary intent of the model is to evaluate whether the content of the proposed knowledge representation is sufficient for solving RPM problems, using forms of inference that are cognitively plausible, even though certain aspects of the overall process may not be. Thus, the ASTI model represents a *content* model rather than a *process* model of how humans might solve RPM problems using iconic visual representations.

Future work on the ASTI model will include implementing additional forms of visual reasoning, such as scaling, image segmentation, and gestalt perception. In addition, much work remains in comparing the processing performed by the model with the mental imagery operations that humans apply when visually solving RPM problems. Part of this research will include investigating how the basic mechanisms used by the ASTI model, namely affine transformations, set operations, and similarity-based matching, might be implemented in a neural computational architecture, using detailed knowledge of the processing and neuronal structures contained in human visual processing brain areas.

4 CLASSIFICATION OF RPM PROBLEMS

One reason for forming problem type classifications for a standardized test such as the RPM is so that such classifications can be used in further computational, behavioral, and/or neuroimaging studies to better determine what cognitive processes people use or need to use to solve the test. This is an argument from cognitive science, in that problem classifications can be used to better understand human cognition.

Another reason has to do with understanding the psychometric properties of such a test, because if problems can be classified as being of different types, then finer-grained observations about an individual's cognition can be made by looking at their performance not just on the test as a whole but on individual problem classes. This is an argument from practicality, in situations involving the use of psychometric tests, as knowledge of problem classes can better inform interpretations of an individual's test results.

Developing such a classification scheme can also help to relate performance results across psychometric tests. There is currently a notion of very coarse problem classification in psychometrics along the dimensions of verbal and non-verbal (e.g. the Wechsler scales divide subtests into whether they contribute to "verbal IQ" or "performance IQ"). However, it may be that two different tests actually contain problems of the same type, or a single test might contain problems of multiple types. Classifying problems apart from just how they appear in a standardized test could affect how psychometric tests are studied. This is an argument from psychometrics, to improve the way in which tests are created, evaluated, and understood.

The specific research question that I address in this chapter is:

1) How can the ASTI model be used to classify problems on the RPM according

to their information processing demands?

I begin by summarizing existing approaches to problem classifications of the RPM tests. I then describe a method for using systematic ablations of the ASTI model to obtain problem classifications according to the information processing demands of particular problems. Then, using experimental data obtained from the ASTI model, I present a new classification of problems for the CPM and the SPM. The APM is not addressed in this section as the ASTI model does not do well enough on it to provide classifications for many of the APM problems. I conclude the chapter with a discussion of claims and future work.

4.1 Existing Approaches

Since John Raven's creation of the Progressive Matrices family of tests, most RPM analyses and refinements have been performed based on normative data from large samples of TD individuals (Raven et al., 2003). While there have been many efforts to identify problem types on RPM tests, most of these efforts have either used qualitative, introspective methods or have been based purely on human performance data, without a deep consideration of the information processing demands of individual problems.

Jacobs and Vandeventer (1972) collected 1335 matrix-type figural reasoning items from 22 standardized tests, and educed (using an informal inspection of the items) a set of 12 relations that they believed covered the majority of test items. These twelve relations are: identity, shape, shading, size, movement in a plane, flip-over, reversal, added element, addition, unique addition, number series, and elements of a set. Some inter-rater verification was performed to show that these relations could be used to describe a sample of the matrix problems, though not always uniquely. Then, they assessed how well Sets B through E of the SPM covered pairs of these relations (presumably to account for variation in rows and columns), and found that only 20 of 66 possible relation pairs are covered. It is noteworthy that they found all 48 problems readily describable in terms of pairs (or trios) of these 12 relations.

Mulholland, Pellegrino, and Glaser (1980) constructed a set of geometric analogy problems in which an analogy is presented and the participant must answer whether the analogy is true or false. Each problem varied systematically according to the number of elements and number and types of transformations (and the ways in which the analogy was altered to create incorrect analogies). They constructed a processing model in which elements and transformations are processed serially and independently, and with additive reaction times. The reaction time and error data that they collected supported their model to some extent, though they found significant interactions when the number of elements and transformations both increased, among other trends. They surmise that certain levels of problem complexity introduce significant working memory demands that cause increases in latency and error rates. While their data did support their model in the sense that number of elements and transformations do systematically affect reaction time, there may be other factors at play as well. One difference between their task and RPM is that, since their task was only to evaluate a single complete analogy, the prediction form of high-level strategy wasn't really warranted by the task; it was more of a compare-and-test task. Also, they stated that their model assumed that information was stored propositionally in mental representations, but it is not clear that any part of their model really relies on this assumption.

Horner and Nailling (1980) adapted a listing of problem types from Corman & Budoff (1973) and present a listing of the problem type for each problem in the CPM. In a study of left-, right-, and non-brain-damaged patients, they found that each group showed a similar pattern of accuracy across the four problem types, though absolute levels of accuracy differed somewhat.

Dillon, Pohlmann, & Lohman (1981) performed a factor analysis of Set II of the APM in which they accounted for problems being of different difficulty levels. Not accounting for problem difficulty could lead to confounds, because problems tapping into the same cognitive ability may be of different difficulties, and problems of the same difficulty may tap into different cognitive abilities, but difficulty levels may mask these other differences in a regular factor analysis. The authors also observed, in a brief review of previous factor-analytic studies of the RPM, that studies in which the RPM is examined along with other tests generally yield results that the RPM loads mainly on a single *g* factor, whereas studies in which the RPM is examined alone often yield results that the RPM has multiple factors. With regard to the APM, their analysis seemed to support a two-factor solution, in which they labeled their factors as Pattern Addition/Subtraction (Factor I) and Pattern Progression (Factor II). Many of the problems showed loadings on both factors, but the authors suggest subsets of problems that could serve as "pure" measures of each factor:

1) Factor I, Pattern Addition/Subtraction: 7, 9, 10, 11, 16, 21, 28, and 35
2) Factor II, Pattern Completion: 2, 3, 4, 5, 17, 26, 36.

However, they do not address apparent differences in difficulty levels between the two sets of problems that they suggest.

Stone and Day (1981) describe a study very similar to the reaction time study of Mulholland et al. (1980). They studied participants at three different ages (fifth grade, eighth grade, and college), and they used artificial matrix problems in which the number of elements and transformations was systematically varied. Their matrix problems were 3x3 problems, with the ninth element filled in either correctly or incorrectly, and the participant's task was to indicate whether the matrix was correct or incorrect. However, they instructed the participants to only look at the rows of the matrix, as column transformations were not implemented in the artificial matrix problems. So, like Mulholland et al. (1980), the task different from the regular RPM in that no processes of answer prediction or selection are required, and the analogies are all unidimensional. In

Mulholland et al. (1980), the analogies were strictly of the form A : B :: C : D, whereas in this paper, they were strictly of the form A : B : C :: D : E : F :: G : H : I. They found results broadly consistent with Mulholland et al. (1980), in that both increased elements and increased transformations increased reaction time in an additive fashion, but when both were increased simultaneously, the increases in reaction time were more than additive. In addition, the authors found that reaction time decreased with age, as did the non-additive increases with elements and transformations (i.e. the younger participants showed greater absolute increases in RT for the combination of more elements and transformations than did the older participants, though the increases were proportionally about the same). Little analysis was done of true versus false RTs; it appears in this study that the RTs were fairly similar across the two conditions in certain groups/manipulations, which goes against Mulholland et al. (1980)'s contention that false items involve self-terminating processing and should thus take less time overall.

Kirby and Lawson (1983) adopted the approach from Hunt (1974) in classifying RPM problems as either analytic or gestalt, in terms of what strategy might be more effective. They gave children one of four types of training in a strategy: strong or weak, gestalt or analytic. The weak training consisted of giving the participants a series of problems from the CPM and SPM that the authors evaluated to draw upon that particular strategy. The strong training consisted of giving the participants a series of problems that the authors evaluated to be strategy-neutral, and then they verbally described either a gestalt (i.e. pattern completion) or analytic (i.e. rule-based) strategy.

1) Weak gestalt: A1, A1, B2, A7, A8, A11, AB4, AB6, B4, and B5.
2) Weak analytic: A1, A2, B2, B3, B6, B7, B8, B9, B10, B11.

3) Strong: A1, A2, B2, AB7, AB9, AB12, B7, B8, C1, C4.

They developed a new set of "ambiguous" problems for a post-test in which at least one of the answers would follow from a gestalt strategy, and at least one of the answers would follow from an analytic strategy. They then scored how many analytic vs. gestalt answers each participant chose. They also administered Set I from the APM, of which they considered problems 1-6 as gestalt and 7-12 as analytic, following Hunt (1974). With the ambiguous figures, they found that strong analytic training did decrease the number of gestalt responses and increase the number of analytic responses, though the gestalt responses seemed preferred by respondents overall, and increased with age. For the APM, they found that gestalt problems were equally solvable in all groups (which we would expect, since Hunt classified these as gestalt or analytic), but analytic problems were solved with more accuracy by both weak and strong analytic training groups. They conclude with supposition that perhaps the RPM is a measure of strategy monitoring and selection, given that early items tend to bias the test-taker towards a gestalt strategy, while later items require an analytic approach. It would follow, then, that good psychometric tests are precisely those which are strategy ambiguous.

Smilansky (1984) classified RPM problems according to difficulty by assigning them each a score from 1 to 6, based on the number of elements within a matrix as well as their relationships, which he classified as representing either design patterns of linear, random, or complex relationships or arithmetic patterns of addition, subtraction, or complex relationships. He found good inter-rater reliability on scoring problems from the SPM as well as decent correlations for invented problems between the difficulty rating and reaction time of a group of students solving the problems.

Bethell-Fox, Lohman, and Snow (1984) performed a study similar to that of Mulholland et al. (1980) in that they gave participants geometric analogy problems with varying parameters and then measured their accuracy, reaction time, and eye-movements. The problems varied in number of elements, number of transformations, type of transformations (figural versus spatial), number of alternatives, and difficulty of distracters. They also included “ambiguous” items for which the expected answer choice was not present, but a similar one was, which was supposed to be the correct answer. One new finding they presented was that number of response alternatives (two versus four) greatly increased the difficulty/latency of problems, especially in cases with more elements. Also, spatial transformations (e.g. rotation, reflection) increased difficulty/latency more than figural transformations, which they supposed was because spatial transformations involved imagery operations that were not needed for figural transformations. They propose a componential model of performing these analogies with steps like encoding, inference, mapping, etc., and were able to fit the model to their data.

Their other significant finding was a distinction between strategies of constructive matching, where the answer choice is predicted and then compared to response alternatives, and response elimination, in which the response choices are inspected and eliminated in turn. They found, especially using the eye-movement data, that all subjects appeared to use constructive matching on easier problems, but that lower ability subjects switched to response elimination for more difficult problems. Higher ability subjects stuck with constructive matching, but just spent more time in constructing their answer and comparing it to the response choices. Evidence came from whether subjects looked back at the analogy before or after looking at the answer choices, and also at how many answer choices subjects considered before providing their answer. They also discuss another experimental manipulation in which they presented subjects with the first part of the analogy only, without answer choices, before presenting the full problem. In this manipulation, eye-tracking data suggested that constructive matching was preferred.

Green and Kluever (1991) presented a factor analysis of the CPM that found evidence of three factors, with most CPM items loading on a single factor, but certain items loading on multiple factors or only weakly on any of the factors:

1) Visual closure and pattern completion, visual orientation, discrimination
2) Visual analogies, particularly foreground/background discrimination, line/density discrimination
3) Perceptual matching

Green and Kluever (1992) tried to identify a system for predicting RPM item difficulty: They looked at parameters of the matrix (orientation, symmetry, progression, dimensions, curvature, number, density, color) as well as parameters of the answer choices (number of options, progression, rotation, reflection, directions, number of elements, reversal), and coded problems from the CPM and SPM using this scheme. They were mainly interested in modeling item difficulty, so they used these parameters as variables in a model that they based on the CPM and tested against problems from the SPM. The four parameters that seemed to contribute most were number of distinct options (trivially, as two very easy problems have repeated answer choices), reflection/rotation of options, number of dimensions/features (i.e. transformations) in the matrix, and number of directions of options.

DeShon, Chan, & Weissbein (1995) observed that while the APM is presented visually, it can be solved either using a visuospatial strategy that operates on visual representations of the problem or using a verbal-analytic strategy that operates on propositional representations of the problem. They surmised that certain problems on the APM may be more amenable to one or the other of these types of strategies, following Hunt (1974). Further, they hypothesized that concurrent verbalization would disrupt visuospatial problem solving, and thus impair performance only on those problems that were visuospatial to begin with. They developed a set of visuospatial rules and a set of verbal-analytic rules, following Hunt (1974) and Carpenter et al. (1990), as well as drawing on introspection, in the form of concurrent and retrospective reports of problem-solving on the APM from pilot studies. Using these rules, they classified all of the problems on the APM as "most likely" to be either 1) visual, 2) analytic, 3) either, or 4) both, using three independent coders for all 36 problems (in Set II, I assume). They then administered the APM to three groups of participants, with one group solving the test as usual, one group solving the test on the computer, and the third group solving the test on the computer with concurrent verbalization. Overall, the group with concurrent verbalization showed significantly lower accuracy than the other two groups, indicating that overt verbalization actually impairs performance on the APM. Furthermore, 7 of the 12 visual problems showed a significant decrement with verbalization, and the other 5 problems also showed some decrement, whereas none of the 9 verbal problems showed any decrement with verbalization. Response times showed a similar pattern, with the non-verbalization group having equivalent response times for visual and verbal items, but the verbalization group having (a) greater response times overall than the other group, and (b) greater response times for visual than for verbal items. Interestingly, as the authors observe, verbalization has been showed to increase performance on other tasks like the Tower of Hanoi, which Carpenter et al. (1990) compared to the APM in terms of requiring similar cognitive loadings on working memory. However, this study shows that even at a coarse level of analysis, verbalization does impair performance on the APM but improves it on the Tower of Hanoi, which might suggest that the two tasks tap into very different cognitive processes. They conclude by observing that actual strategy choice depends not only on the problem characteristics but also on the abilities of the problem solver.

Van der Ven and Ellis (2000) performed a Rasch analysis of the SPM by set. They found that sets B and E loaded on two factors each, and to some extent set C. They looked at the most frequent incorrect answer choice to hypothesize what the factors are, and they concluded that Set B requires Gestalt continuation for the early problems and analogical reasoning for the latter problems. Set C requires analogical reasoning and lack of resistance to perceptual distracters. Set E requires analogical reasoning and coping (or constructing an answer choice from the two adjacent entries in the matrix).

Matzen et al. (2010) constructed a set of artificial SPM-like items by using a fixed bank of shapes, features, transformations, and directions. They used six shapes (oval, rectangle, diamond, triangle, trapezoid, T), with each shape varying according to size, fill pattern, orientation, and numerosity. Transformations were of two types: object transformation (changes in shape, shading, orientation, size, and number) and logical transformations (AND, OR, and XOR). Directions were either rows, columns, diagonal (both ways), or outward from the top left corner of the matrix. Incorrect answer choices were created by systematically varying elements of the correct answer and of other matrix entries or answer choices. The authors hypothesized that three-relation problems would be more difficult than two-relation problems, and two-relation problems would be more difficult than one-relation problems. They also hypothesized that within three-relation problems, problems in which more relations were either diagonal or outward would be more difficult. Both of these difficulty hypotheses were borne out in participant data. In addition, outward transformations were more difficult than the other directions, though this result primarily applied to problems with shading changes as the salient transformation. The artificial object transformation problems seemed of equal difficulty to similar SPM problems, but the artificial logical problems were more difficult than similar SPM problems, probably because several of the SPM logical problems are very easy. They establish that direction of relation is a significant contributor to item difficulty, and they observe that direction can be equated with the Carpenter et al. (1990) "distribution" rules.

4.2 Problem Classification Using Model Ablation

One advantage of classifying problems using a computational model is that results are strictly quantitative; human coding or introspection necessarily includes qualitative influences. However, previous approaches to using a computational model to classify RPM problems—namely, the taxonomy developed by Carpenter et al. (1990) based on what rules their model used to solve particular problems—have shown sufficiency of particular reasoning mechanisms, but not necessity. In particular, just because a model can successfully solve a problem using a set of mechanisms does not directly inform us about which of those mechanisms were actually necessary.

In this section, I present a new approach for classification of RPM problems that relies on systematic ablations of the ASTI model to discover which, out of the set of mechanisms implemented in the model, are necessary for solving particular problems. I define the rules for this classification as follows:

1) To deem a particular set of mechanisms as sufficient for solving a particular problem, the model must be able to successfully answer the problem using these mechanisms.
2) To deem a particular set of mechanisms as necessary for solving a particular problem, removing any one mechanism must cause the model to fail to solve the problem.

The ASTI model has many orthogonal collections of mechanisms that can be ablated. I consider the types of base transformation used by the model as well as the image sets that are taken into account. These parameters are listed in Table 17.

Table 17. Configurations of ASTI model used for ablation experiments for 2x2 matrices.

Type	Image sets	Base transforms
2x2 matrices	1. Rows 2. Columns	1. Identity 2. Rotation/reflection 3. Addition/subtraction
3x3 matrices	1. Rows 2. Columns 3. Diagonals	1. Identity 2. Rotation/reflection 3. Addition/subtraction 4. Composition

4.2.1 Problem classification for the CPM

Table 18 shows results on the CPM from ablated configurations of the ASTI model, grouped by type of base transform and whether rows or columns were considered. Only one problem on the CPM, problem B9, was not solved by any configuration of the ASTI model that was tested and is thus unclassifiable under the current scheme.

Most of the problems on the CPM appear to be agnostic with respect to the use of rows or columns. Most of the problems on the CPM also appear to be solvable using either basic matching (i.e. the identity transform) or exclusively using one or the other of rotate/flip transforms or add/subtract transforms.

Table 18. Classification of problems from the CPM.

Set	#	Identity	Rotate/ flip	Add/ subtract	Rows	Columns
A	1	✓	-	-	✓	✓
	2	✓	-	-	✓	✓
	3	✓	-	-	✓	✓
	4	✓	-	-	✓	✓
	5	✓	-	-	✓	✓
	6	✓	-	-	✓	✓
	7	✓	-	-	✓	✓
	8	-	-	✓	✓	✓
	9	✓	-	-	✓	✓
	10	✓	-	-	✓	✓
	11	-	-	✓	✓	✓
	12	✓	-	-	✓	✓
AB	1	✓	-	-	✓	✓
	2	✓	-	-	✓	✓
	3	✓	-	-	✓	✓
	4	-	✓	-	✓	✓
	5	-	✓	-	✓	✓
	6	-	✓	-	✓	✓
	7	-	✓	-	✓	✓
	8	-	✓	-	-	✓
	9	✓	-	-	✓	✓
	10	-	✓	-	✓	✓
	11	-	✓	-	✓	✓
	12	-	-	✓	✓	-
B	1	✓	-	-	✓	✓
	2	✓	-	-	✓	✓
	3	✓	-	-	✓	✓
	4	-	✓	-	✓	✓
	5	-	✓	-	✓	✓
	6	-	✓	-	-	✓
	7	-	✓	-	✓	-
	8	-	-	✓	✓	✓
	9	-	-	-	-	-
	10	-	-	✓	-	✓
	11	-	-	✓	✓	✓
	12	-	-	✓	✓	✓

4.2.2 Problem classification for the SPM

Table 19 shows results for 2x2 SPM problems using a pixel threshold of 0.1. Notice that all 24 problems are solved correctly by some combination of transform and image set, though, as illustrated by the results presented in Section 3.3.1, the best single configuration only solves 23 of the 24 problems, due to interactions between various components within a single configuration.

Table 19. Classification of 2x2 problems from the SPM.

Set	#	Identity	Rotate/ flip	Add/ subtract	Rows	Columns
A	1	✓	-	-	✓	✓
	2	✓	-	-	✓	✓
	3	✓	-	-	✓	✓
	4	✓	-	-	✓	✓
	5	✓	-	-	✓	✓
	6	✓	-	-	✓	✓
	7	✓	-	-	✓	✓
	8	✓	-	-	✓	✓
	9	✓	-	-	✓	✓
	10	✓	-	-	✓	✓
	11	-	-	✓	✓	-
	12	✓	-	-	✓	✓
B	1	✓	-	-	✓	✓
	2	✓	-	-	✓	✓
	3	✓	-	-	✓	✓
	4	-	✓	-	✓	✓
	5	-	✓	-	✓	✓
	6	-	✓	-	-	✓
	7	-	✓	-	✓	-
	8	-	-	✓	✓	✓
	9	-	✓	✓	✓	-
	10	-	-	✓	-	✓
	11	-	-	✓	✓	-
	12	-	-	✓	✓	✓

There are small differences in these results compared to corresponding CPM problems, most likely due to slight differences in the pixel thresholds used for each of these tests. Like the CPM, most 2x2 SPM problems are not affected by the use of rows or columns, and most 2x2 SPM problems also appear to be solvable using either basic matching (i.e. the identity transform) or exclusively using one or the other of rotate/flip transforms or add/subtract transforms.

Table 20 shows results for 3x3 SPM problems also using a pixel threshold of 0.1. 30 of the 36 problems are solved by some configuration of the model at this threshold; problems C2, C9, D6, D9, D12, and E7. There are many more successful combinations of image sets and transforms with 3x3 problems than there are with 2x2 problems. Far fewer of the 3x3 problems are solvable using identity-based matching. Addition and subtraction seem to be predominant in Set C, and composition seems to be predominant in Set E, but Set D uses a mixture of all of these types of base transforms. In addition, Set C seems to be solvable using either rows or columns, while Set E seems to favor row-based transformations. Set D is clearly dominated by diagonal transformations.

Table 20. Classification of 3x3 problems from the SPM.

Set	#	ID	RF	AS	COMP	Rows	Cols	Diags
C	1	✓	-	-	-	✓	-	-
	2	-	-	-	-	-	-	-
	3	✓	-	-	-	✓	✓	-
	4	-	-	✓	-	✓	✓	-
	5	-	-	✓	-	✓	✓	-
	6	-	-	✓	-	✓	✓	-
	7	-	✓	-	✓	✓	✓	✓
	8	-	✓	✓	-	✓	✓	✓
	9	-	-	-	-	-	-	-
	10	-	-	✓	-	✓	✓	-
	11	-	✓	✓	-	✓	✓	✓
	12	-	-	✓	-	✓	✓	-
D	1	✓	-	-	-	✓	-	-
	2	✓	-	-	-	-	-	✓
	3	✓	-	-	-	-	-	✓
	4	-	-	✓	-	✓	-	-
	5	-	-	✓	✓	✓	✓	-
	6	-	-	-	-	-	-	-
	7	-	-	-	✓	-	-	✓
	8	-	-	✓	-	-	-	✓
	9	-	-	-	-	-	-	-
	10	✓	-	-	-	-	-	✓
	11	-	-	-	✓	-	-	✓
	12	-	-	-	-	-	-	-
E	1	-	-	✓	✓	✓	✓	-
	2	-	-	✓	✓	✓	✓	-
	3	✓	-	-	-	✓	-	-
	4	-	-	✓	✓	✓	✓	-
	5	✓	-	-	-	-	✓	-
	6	-	-	-	✓	✓	-	-
	7	-	-	-	-	-	-	-
	8	✓	-	-	-	✓	-	-
	9	-	-	✓	✓	✓	✓	✓
	10	-	-	-	✓	✓	-	-
	11	-	-	-	✓	✓	✓	-
	12	✓	-	-	-	-	-	✓

4.3 Claims and Future Work

In this section, I conducted computational experiments to test how systematic ablations of the ASTI model, removing various functionalities in different combinations, affect performance on individual RPM problems. Using this approach, I was able to define problem classifications based on the visual reasoning mechanisms necessary to solve particular problems, which provides a problem classification scheme based on the information processing demands of each problem.

My findings, apart from the problem classifications themselves, indicate that there do seem to be distinct problem types present on both the CPM and the SPM, which can be identified through computational modeling as I have shown. In addition, this work demonstrates that previous approaches that define large classes of RPM problems as “visual” or “verbal” are somewhat misleading, in that the majority of RPM problems can be solved using particular visual mechanisms.

There are many important areas for future work, including providing a model-based classification of problems on the APM, if future versions of the ASTI model are able to successfully address more problems. Additions to the ASTI model, such as image segmentation or gestalt reasoning, provide additional dimensions along which RPM problems might be classified. Furthermore, there is significant potential for using these problem classifications to provide more detailed studies of human behavior on the RPM. For example, one could imagine studying the neural activation of individuals with autism as they solve problems in the various classes identified in this section, to better understand the neural substrates that underlie various types of visual operations.

5 ERROR PATTERNS ON THE RPM

This chapter investigates one potential behavioral marker for the RPM that may or may not provide an indicator of whether a test-taker is using a visual or verbal problem-solving strategy. As I have shown in Chapter 3, visual strategies can be as successful as verbal strategies on many RPM problems, and so overall levels of accuracy cannot serve as this type of behavioral marker. However, beyond just the total score, an individual's responses on the test can actually furnish additional information based on the errors that they make. In particular, for two individuals who happen to achieve the same total score, their particular choice of distracters for problems answered incorrectly may differ, for instance if their preferred problem-solving strategies lead them down different paths of reasoning.

1) To what extent can errors made on the RPM serve as behavioral markers to indicate the use of a visual versus verbal strategy?

To answer this question, I have conducted an observational study of error patterns on the SPM using data drawn from three different populations: typically developing (TD) individuals, who likely use a combination of visual and verbal strategies, individuals with autism, who, as I have discussed in Chapter 2, may use predominantly visual strategies, and the ASTI model, which uses exclusively visual strategies.

I begin by describing existing approaches in the RPM literature for analyzing patterns and types of errors. Then, I present a new classification of conceptual error types represented by the distracters on the SPM. I then present a detailed method for the observation study of error patterns, followed by results. I conclude with a summary of my claims and areas for future work.

5.1 Existing Approaches for Examining RPM Errors

In this section, I examine existing approaches from the RPM literature for examining the number or types of errors that individuals make on the test.

Miller and Raven (1939) looked at the performance of two groups of children: one group of girls of unspecified school age, and another group of younger children between 5 ½ and 7 ½ years of age. Using variations of matrix problems, they established that there are at least two influences on which wrong answers participants choose, in terms of there being non-random effects on the distribution of answers that are chosen. One influence is the absolute position of the answer choice with respect to the matrix. When alternatives are all listed horizontally to the right of the matrix, the position effect is very marked, and participants tend to choose the left-most choices that are closer to the matrix. When alternatives are listed in rows underneath the matrix, position effects are less marked though still present, and participants tend to choose answers from the top row and those towards the middle-right of any particular row (i.e. closer to the empty space in the matrix). The other influence on answer choices is the conceptual type of error represented by the entry given in each respective answer choice, and in particular, for difficult problems, participants tend to make errors of repetition. These two influences are not independent, however; they do interact in complex ways. If a correct answer happens to be in the preferred position, it will be chosen more often than otherwise. Likewise, if an obviously implausible answer is put into this preferred position, participants will go on to examine more alternative choices, whereas if a “familiar” but still incorrect answer is in the preferred position, e.g. a repetition error, participants tend to stick with that answer.

Halstead (1943) compared results on the SPM from a clinical group of individuals diagnosed with neuroses with those of healthy controls. In order to examine group differences at a finer level of detail than overall score, Halstead created subgroups individually matched on raw scores. The groups did not differ on measures of "unevenness," i.e. score consistency across sets when compared to norms, or "reversals," i.e. scoring higher on a later set than on an earlier one. He also examined test variables as a function of age, time (for taking the test), attitude, etc. Finally, Halstead looked at the most frequent errors made by a very large number of control participants (n = 2790). He broadly classifies these errors according to conceptual type and observed that low ability participants tended to make "perceptual" errors like repetition, whereas high ability participants tended to make "inadequate reasoning" errors. He also observed that: "Mathematically minded subjects seem to do as well as any on the test, and indeed some items in Set E can only be solved logically. High scores have, however, been obtained by artistic people who have an eye for form (Gestalt), symmetry, etc." (p. 211).

Eysenck (1945) looked at the performance of elderly adults with senile dementia, compared to typical adults, on sets A and B of the CPM. She looked at errors in terms of the most frequent distracters chosen and found that in both groups, both the absolute position of distracters as well as distracters that repeat entries from the matrix influence the incorrect answer choices made by participants. In particular, distracters in positions 1, 2, and 6 were more frequently chosen than those in the other positions, but the only group difference was for position 2, which was chosen more frequently by the senile group. For both groups, matching the entry above the empty space accounted for a significant proportion of errors, and matching the entry to the left of the empty space accounted for a smaller, but still significant, proportion of errors.

Bromley (1953) wrote an interesting paper in which he studied performance on the SPM by a group of older individuals with various psychiatric disorders, whom he described as evincing "primitive" forms of thinking. The individuals were instructed to explain their reasoning as they took the test, and Bromley provided qualitative analyses of their responses. His main observation was that there seemed to be two ways to approach the SPM: one the intended way, with abstract, relational, analogical thinking, and another involving more global, holistic, concrete thinking (including mental imagery). He observed that the primitive thinking shown by the test participants fell more into the latter category and could explain many of the errors made on the test. With respect to errors, Bromley observed that the participants seemed susceptible to effects of both the absolute position of answer choices as well as other features of incorrect answer choices (e.g. repetition of a part of the matrix, etc.). He characterized the answer choice error types as: "part of the matrix, simple or distorted figure like the correct one, relatively unrelated figure, global figure, similar to part of the matrix reversed or distorted" (p. 384). The highest proportion of errors was for "part of the matrix" answer choices, followed rather distantly by "simple or distorted figure like the correct one" and "global figure." Bromley also listed the types of thinking that he supposed gave rise to errors on the test, and he emphasized that types of thinking differed significantly on an individual differences level. He also surmised that many of these forms of "primitive" thinking might have developed in an individual (and thus used on a test like the SPM) as a compensatory mechanism, to make up for difficulties with other forms of thinking (e.g. abstract, analogical, etc.).

1) Global responses are those that involve global/Gestalt solutions.
2) Concrete responses are those that fail to adequately abstract from the directly perceived features of the problem.
3) Mechanization of response involves the inability to switch set from initially successful strategies. Bromley points out that the test itself encourages this sort of mechanization, which echoes the strategy findings of Kirby and Lawson (1983), that early problems on the test influence the strategy chosen by participants for later problems. This seems very akin to perseveration.
4) Inability to explain refers to failures in verbalizing a strategy (for successful problems) or the difficulties with a particular problem (for unsuccessful problems).
5) Sensori-motor responses refer to the tendency of participants to point and trace their answer on the matrix and on the answer choices. Bromley observed that on occasion, participants would trace the correct answer but be unable to choose that answer choice.
6) Physiognomic responses.
7) Subjective responses occurred when participants seemed to think there was ambiguity in the answer choices, and the correct answer was a matter of personal preference.
8) Fluid responses were those in which participants seemed to use arbitrary selection criteria, including just picking the “odd man out” among the answer choices.
9) Avoidance of reality referred to participants who picked answer choices and

described how they should be different, or to participants who tried to evade the problem by trying to match the frame shapes instead of the entry content.

Forbes (1964), as part of an item analysis to revise the problems found on the APM, analyzed the types of errors made by participants as a function of their ability level (i.e. total APM score). He classified errors as being of four types: incomplete correlate, wrong principle, incomplete individuation, and repetition. His analysis looked at each third of the test with respect to a single ability level: low for the first third, average for the second third, and high for the third third. He noted that the incomplete correlate was the most frequent error type overall, but represented a smaller proportion of errors for the low ability group, for whom wrong principle was the most frequent error. Individuation and repetition errors were the least frequent in any group. He also looked at overall selection of answer choices as a function of position, and found that positions 6 and 7 tended to gain fewer responses than the others, and positions 1 and 4 were the most frequently chosen. He surmised that perhaps 1 is favored by typical scanning patterns, and 4 is closest to the empty space in the matrix.

Weatherick (1966) looked at the errors made by healthy adult subjects on the SPM to directly compare to Bromley's (1953) results with senile psychiatric patients. The subjects were overall high scoring, and he found "very close agreement between our sample of n = 236 and Bromley's sample of n = 35." As a result, Weatherick contends that the specific errors identified by Bromley do not indicate "primitive thought processes." Weatherick does observe that in instances where his control results do differ from Bromley's results, the senile patients tended to prefer (instead of the most frequent control error) a repetition error, of an answer that repeats a part of the matrix adjacent to the empty space.

Vejleskov (1968) looked at performance on the SPM among Danish children. He gave results on error frequencies for only a few problems, and observed that for these problems (all from Set B), girls tended to fail by choosing the distracter that was the same as the correct response except rotated or flipped.

Jacobs and Vandeventer (1970) looked at error patterns on the CPM. In particular, for a given 2x2 CPM problem, they classified the answer choices based on whether the answer choice followed a horizontal rule only, a vertical rule only, both (which would be the correct answer), or neither. They assumed that an incorrect answer choice was “superior” if it followed at least the horizontal or vertical rule (as opposed to answer choices that followed neither). They found that 18 of the 36 problems on the CPM contained both “superior” answer choices, and they restricted their analysis to these 18 problems. Then, for each participant, they calculated a proportion Ps that was the number of superior answer choices chosen divided by the total number of wrong answers. (Participants who answered fewer than five problems incorrectly were excluded.) Looking at data from American children in the first and third grades, Eskimo adults and young adults (from Canada), and Temne adults and young adults (from Sierra Leone), Ps appeared to be more strongly correlated with total number of correct answers in lower-ability groups of participants (i.e. those with lower average scores). In addition, Ps appeared to be higher in the more able groups (i.e. Ps for Eskimos was higher than Ps for Temne). One difficulty in this study is that Ps data from more able participants because less valid, because fewer errors have been made to contribute to the Ps score. In addition, Ps was defined solely based on answer choices that followed row or column rules in the matrix; it is a very strong assumption to say that these answer choices were “superior” to the other answer choices. A stronger methodology would require coding classes of distracters for all the answer choices, and then looking at types of errors of the various classes. It could be that the classes of distracters could still be ranked according to their “correctness” level, but that would depend on how the distracter classes were defined.

Carter (1970) supposed that solving SPM problems involved processes of induction as well as evaluating similarity (i.e. similarity of the induced answer to the given answer choices). He gave subjects five tests: regular problems from the SPM (induction + similarity), problems from the SPM in which the answer choices were omitted completely and the answer had to be described (pure induction), tests to rank the similarity of answer choices and matrix entries, according to shared features in a propositional encoding (pure similarity), and problems from other, non-visual tests of induction (pure induction). The similarity rankings of answer choices and matrix items by subjects might have given interesting insight into how they might be viewing the different distracters, but the study only scored them as correct or incorrect in their rankings. Further, the problems that they ranked were not the same as the ones in the first two tests, so it was not possible to see how their perceived rankings might have affected their actual performance on the problem. In fact, while the author designed the two ranking tasks to be different from inductive reasoning, it does seem as though evaluating similarities on a feature-by-feature basis would share a lot in common with solving a matrix task, even one without the answer choices, inasmuch as both tasks involve evaluating differences between entries in a systematic way.

Guttman (1974) looked at familial correlations in SPM scores among children and their parents. Guttman observed that for each item, two or three of the incorrect answer choices seemed to be chosen with greater frequency. However, she did not observe any inter-family differences in these frequency distributions.

Thissen (1976) characterized incorrect response choices on the SPM according to frequency: the first most-chosen, second most-chosen, and then all other incorrect answer choices. He used this information to calculate a latent trait model for each test item that gave probability of choosing a particular answer choice as a function of ability (the latent, unobserved trait). He did find that for different problems, the answer choices behaved differently for different levels of ability, but the analysis was purely done along this unidimensional notion of ability; no explanations were offered for why certain answer choices might be more or less chosen than others.

Horner and Nailling (1980) adapted a listing of error types from Raven (1965) and present a listing of the error type for each answer choice in the CPM. In a study of left-, right-, and non-brain-damaged patients, they found that each group showed nearly identical patterns of error types across the four error types. In particular, only one type of error, "repetition of a pattern," seemed to be made with significant frequency, other than the correct answer.

Kirby and Lawson (1983) developed a series of ambiguous items in which different answer choices were deemed correct depending on the strategy one were employing. This is one example of different strategies leading to different answer choices, though in this case both answers were deemed correct.

Vodegel Matzen, van der Molen, and Dudink (1994) looked at types of errors made on the SPM by typically developing children. They adopted error categories from the

APM to categorize SPM errors as: incomplete correlate, wrong principle, repetition, or additional elements, and they only analyzed sets C through E. Inter-rater reliability for coding the error types was only around 72%. They found that the incomplete correlate was most frequent overall, followed by wrong principle, repetition, and additional elements. When they divided the subjects by ability level, they found that low ability subjects made relatively more errors of the incomplete correlate and repetition types. The authors then devised an “experimental progressive matrices” test, in which all errors were of the “incomplete correlate” type, but they varied according to how many rules were omitted to generate that answer choice. The problems varied according to rule type and number, following Carpenter et al. (1990). The EPM was similar in difficulty to the SPM, and they found that most error choices (for any ability level) were made due to the omission of a single rule. Furthermore, the rules increased in difficulty (as measured by number of errors) in this order, for all ability levels: constant in a row, quantitative pairwise progression, distribution of three values, addition/subtraction, distribution of two values. This is the same rule ordering that was chosen by Carpenter et al. (1990) for inducing rules.

Van der Ven and Ellis (2000) looked at the most frequent incorrect answer choice for the SPM in sets B, C, and E, in order to determine what factors these problems might load upon. They identified different types of errors, including: “lack of completeness of analogical reasoning,” “freedom from perceptual distracters,” and “coping.” They also present data from sets C and E giving the frequencies of each answer choice for each problem, using their sample of several hundred Dutch schoolchildren.

Babcock (2002) classified each answer choice from the APM as being one of four different error types: incomplete correlate, wrong principle, confluence of ideas, or repetition. She studies the error responses made by adults of varying age and ability, according to whether their frequency of making a particular type of error was above or below chance levels. She found that adults of varying ages tended to make similar types of errors, but adults of high ability made different errors than those of low ability. In particular, high ability adults tended to make more incomplete correlate errors, and few errors of other types. Lower ability adults tended to make each type of error at chance levels. Also, she studied errors as a function of rule type, based on Carpenter et al. (1990), and found some differences between subjects of varying abilities.

Gunn & Jarrold (2004) looked at types of errors made by TD children, children with moderate learning disabilities (MLD), and children with Down syndrome (DS) on the CPM. They classified error choices as being of one of four types, following the CPM manual: difference, repetition of a figure, inadequate individuation, and incomplete correlates. They found that, even after controlling for total number of errors, the DS group made different types of errors than the other two groups. In particular, the DS group produced fewer repetition of a figure errors and more inadequate individuation errors and difference errors (which is choosing an unrelated answer choice). Furthermore, the pattern of errors produced by the DS group is similar to that shown by younger TD children, even in cases where the DS group shows better performance than younger TD children. The authors surmise that individuals with DS may have either difficulty in combining features to produce the target pattern, difficulty in visual discrimination, or less rigor in choosing their final response, in the case of incomplete or partial solutions.

Matzen et al. (2010) performed an analysis of errors on the SPM and on artificial SPM-like items. On the artificial items, errors were classified systematically according to how each distracter was related to content in the problem matrix. In particular, error types were classified as (for a single relation in the problem): match to diagonal, match to top left, match to adjacent, flanker, and unclassified. They were able to categorize some, but not all, SPM errors using the same scheme (i.e. the one-relation SPM problems but not the two-relation problems). For certain problems, they found that the error type seemed to have a relationship to the direction of the relation in the problem; for instance, problems that were diagonal in one direction tended to have more "match to adjacent" errors, whereas problems that were diagonal in the other direction tended to have more "flanker" errors. Though the authors do not draw this connection, it seems as though participants might have been distracted by Gestalt properties of the overall matrix in making such errors.

Fajgelj, Bala, and Katic (2010) as part of a factor analysis of the CPM, looked at types of errors made by their sample of Serbian children. They found that for younger children, more CPM problems had certain distracters that were chosen by significant portions of subjects (i.e. more than 20%). They observed that the most common distracters involved choosing the answer identical to the entry to the left of the empty space or above the empty space in the matrix. They also note, interestingly, that number 2 was chosen more frequently than other answer choices, and especially so for younger children, possibly because this choice is spatially closest to the empty spot in the matrix.

Facon and Nuchadee (2010) looked at relative item difficulties among items on the CPM between TD children, children with Down syndrome, and children with unspecified intellectual disability. They matched each group, participant-by-participant, on CPM raw score, to account for potential differences in overall ability level. They found no evidence of differential item functioning (i.e. particular items being more or less difficult for different individuals) among the groups. They did emphasize the importance of such analyses for various clinical groups because there is little evidence that psychometric tests involve the same cognitive processes for cognitively disparate groups of individuals.

Van Herwegen, Farran, and Annaz (2011) looked at error types on the CPM between TD children and individuals with Williams syndrome (WS). They classified errors on the CPM following the CPM manual into four categories: difference, inadequate individuation, repetition, and incomplete correlation. They looked at proportion of each error type out of total error for each participant. Participants were matched on CPM raw score, and the WS group had a much higher mean chronological age than did the TD group. Their results were very similar to those in Gunn and Jarrold (2004), in the proportions of each type of error made, on average, though they found no group differences in this study between the WS and TD individuals. They also studied developmental effects on error type, and again found similar results to Gunn and Jarrold (2004), in that the difference and inadequate individuation errors decreased and repetition errors increased; however, incomplete correlation errors did not increase with age. They also did an item analysis, following Facon and Nuchadee (2010), to look at whether items differed in difficulty between the two groups. Only 3 of the 36 items differed. They close with speculating that one might expect to see different patterns autism, since autism has perceptual atypicalities more so than WS and the RPM is a perceptual task.

5.2 Classification of Error Types on the SPM

The published manuals for both the CPM and the APM include small taxonomies of what types of conceptual errors are represented by the choice of a particular distracter from among the given answer choices. The CPM manual gives four broad categories of error types along with specific criteria that can be used to classify a particular distracter into one of these four categories, shown here in Table 21. The manual also gives a classification for each answer choice for all 36 problems on the CPM. The APM manual similarly lists four broad categories of error types, which are somewhat different in name from the four categories given for the CPM, and instead of specific classification criteria, gives broad descriptions of each type of error, shown here in Table 22. Then, for each of the 36 problems in Set II of the APM, the manual lists the top two most frequently chosen distracters along with their error type classifications. However, the manuals do not give details as to how these taxonomies were developed or how distracters were classified.

Table 21. Error type taxonomy and classification criteria from the CPM manual (Raven, Raven, & Court, 2003, p. 5).

Error type		Criteria
Difference	a	The piece has no figure of any kind on it
	b	The figure shown is quite irrelevant
Inadequate individuation	c	The figure is contaminated by irrelevancies or distortions
	d	It combines figures irrelevantly
	e	It is the whole or half the pattern to be completed
Repetition of the pattern	f	Above and to the left of the space to be filled
	g	Immediately above the space to be filled
	h	Immediately to the left of the space to be filled
Incomplete correlate	i	The figure is wrongly orientated [sic]
	j	It is incomplete, but correct as far as it goes

Table 22. Error type taxonomy and descriptions from the APM manual (Raven, Raven, & Court, 2003, p. 10).

Error type	Description
Incomplete solutions (Incomplete correlate)	These are errors arising from failure to grasp all the variables determining the nature of the correct option required to complete the problem pattern. Instead an option is chosen which is only partly correct.
Arbitrary lines of reasoning (Wrong principle)	The option chosen suggests that the person being tested is using a principle of reasoning qualitatively different from that demanded by the item.
Over-determined choices (Confluence of ideas)	These are errors involving failure to discriminate irrelevant qualities in the chosen option, and to select one which combines as many as possible of the individual characters shown in the matrix to be completed.
Repetitions	These involve selection of a [sic] option identical to one of the three options immediately adjacent to the space to be filled in the matrix.

The SPM manual does not contain a similar discussion of error types (Raven, Raven, & Court, 1998). Vodegel Matzen and colleagues (1994) attempted to use the APM classifications shown in Table 22 to categorize distracters for sets C through E of the SPM, but inter-rater reliability between two coders was found to be only around 70%, and the authors observed that classification of SPM distracters seemed “problematic,” as no explicit methodology for constructing distracters is apparent, either in the test itself or in the research literature on the SPM (Vodegel Matzen et al., 1994, p. 1).

For coding error types on the SPM, I first reconciled the two sets of error type categories given in the CPM and APM manuals which, although having different labels, seem to represent conceptually the same notions of error types. For each of these four error types, I give a preferred label, which best captures the intended conceptual underpinning of that error type, along with the other labels used to indicate the same type.

Incomplete correlate (IC) errors, or "incomplete solution" errors, are those in which the distracter is almost, but not quite, correct. For example, some IC distracters represent a rotation or reflection of the correct answer. Other IC distracters differ from the correct answer in a single feature dimension, e.g. they might have four elements instead of three, or straight elements instead of curvy ones, or have the correct shape but the wrong texture. Alternately, an IC distracter might be only missing an element from the correct answer. Oftentimes, an IC distracter might be correct in terms of a single row or column in the matrix, e.g. looking just at the right-most column or just at the bottom-most row, but when both rows and columns are taken into account, it no longer fits the matrix pattern. These kinds of errors are made when a test-taker more or less "gets" the problem, in terms of identifying and understanding the relevant matrix relationships, but then fails to fully account for all of the problem details when selecting an answer.

Repetition (R) errors are those in which the distracter is a copy of one of the matrix entries adjacent to the blank space. Choosing an R distracter may represent a sort of cognitive bias or fixation on the matrix, in which an answer is selected using perceptual matching between the answer choices and the matrix entries closest to the blank space. These entries may be privileged because of their proximity to the blank space, just as the answer choices in positions closest to the blank space tend to be chosen more frequently. Alternately, assuming a top-left to bottom-right visual scanning pattern, adjacent entries may be the last viewed before the test-taker moves on to look at the answer choices.

Difference (D) errors, or "over-determined choices" or "confluence of ideas" errors, are those in which the distracter is somehow qualitatively different in appearance from the other distracters. D distracters include those that are completely blank, as well as those that have extraneous shapes that are not found anywhere in the problem matrix. In addition, a D distracter is often the most complex-seeming answer choice, either combining all of the matrix entries together into a single agglomeration of matrix elements or taking some feature from the matrix and increasing its value until it surpasses all the other entries and answer choices. A D distracter might be chosen because it visually "pops" from among the other answer choices. Additionally, difference errors may be more common when a test-taker is adopting an answers-first strategy, i.e. response elimination, instead of a matrix-first strategy, i.e. constructive matching.

Wrong principle (WP) errors, or "arbitrary lines of reasoning" or "inadequate individuation" errors, are those in which the distracter is a copy or composition of elements from various matrix entries. A WP distracter might be chosen if the test-taker does not educe the correct relationship from the matrix entries and instead combines the entries according to some other rule or relationship to produce an answer choice.

Then, I developed criteria that mark a particular distracter as belonging to a particular error type. The difficulty in classification expressed by Vodegel Matzen and colleagues (1994) was perhaps partially due to the vague and qualitative error type descriptions taken from the APM manual. The approach in the CPM manual, with specific criteria for each error type, is much easier to adopt into a coding scheme, but its criteria do not cover all of the distracters present in the SPM. Thus, I surmised that developing a clear set of criteria would be important to establish a reliable coding of distracters on the SPM.

However, there is an additional difficulty in coding the SPM distracters, which is that often, it seems that the same distracter falls under multiple categories; it might be a repetition of an item in the matrix as well as an incomplete correlate of the correct answer. In particular, while trying to come up with a set of criteria satisfactory for coding the distracters on the SPM, I made the following key observation. The four error types listed above actually represent two orthogonal classifications of distracters:

I. Repetition, difference, and wrong principle errors all have to do with how a particular answer choice is related to information in the matrix and in the other answer choices, without any regard to the content of the correct answer choice. In particular, these errors assume that the test participant is attending to irrelevant or erroneous aspects of the problem, and that they are not able to discover even a partial solution to the problem. In particular, it may be possible for individuals who favor a particular error type to pick the correct answer by chance, even though they were simply doing it to repeat an entry or to choose the most different-looking answer choice.

II. Incomplete correlate errors, on the other hand, have to do with precisely how a particular distracter is related to the correct answer choice. These errors assume that the test participant correctly guesses some part of the solution, but does not quite attain the correct answer.

Therefore, I defined criteria for these four error types in two overlapping parts, which are shown in Table 23. Type I errors do not consider the correct answer; all distracters for each problem are coded according to criteria for repetition, difference, and wrong principle errors. Type II errors, in contrast, do consider the correct answer, but do not consider the matrix. Answer choices (with the exception of the correct answer) are coded according to how they are related to the correct answer, if at all; many distracters may not fit any criteria under the Type II error designation.

Table 23. Classification criteria for the SPM for Type I and Type II errors.

Error type	Code	Criteria
I: Repetition	R-Left	Repetition of matrix entry to left of blank space
	R-Top	Repetition of matrix entry above blank space
	R-Diag	Repetition of matrix entry to top-left of blank space
I: Difference	D-Blank	Filled completely white or black
	D-Union	Union of matrix entries or aspects of them, so that union has more components than any single matrix entry
	D-Plus	Maximizes some feature value or makes it more complex
	D-Diff	Differs qualitatively from matrix and other answers, or contains information not found anywhere in matrix
I: Wrong principle	WP-Copy	Copy of matrix entry not adjacent to blank space
	WP-Flip	Rotation/reflection of matrix entry
	WP-Matrix	Other transformations or combinations of matrix entries or aspects of them, including negative images
II: Incomplete correlate	IC-Neg	Negative (color-inversion) of correct answer
	IC-Fill	Change only in fill, texture, or style
	IC-Flip	Rotation/reflection of correct answer
	IC-Layout	Change only in spatial layout of elements
	IC-Scale	Change only in size or scale, in either or both dimensions (allowing for feature-wise scaling)
	IC-Num	Change only in number of discrete elements (allowing for slight changes in layout)
	IC-Inc	Incomplete, with missing element or portion

5.2.1 Coding method

The method I used for coding distracters on the SPM was as follows. First, from inspection of the SPM, I developed lists of criteria for Type I and Type II errors similar to those given in Table 23. Then, I wrote a small coding protocol to use for performing the distracter classification, which is given in its entirety in Appendix A. This protocol contained qualitative descriptions of each overall error type, an example problem illustrating the various types of criteria for both Type I and Type II errors, and finally, an instruction sheet with ordered codes and criteria to use for the classification, which proceeded in two parts, first for Type I errors and second for Type II errors. Type I error classification used a copy of the test booklet in which no answers had been marked. Type II error classification used another copy of the test booklet in which the correct answers had been marked, and additionally, the matrix portions of each problem had been cut off, so only the answer choices were used for the coding.

I used this protocol to perform a coding of all distracters on the SPM for both Type I and Type II errors. Then, an independent rater was given the same protocol. After a brief verbal discussion of the goals of the coding and the contents of the protocol, the second rater also coded all distracters for both Type I and Type II errors.

The initial agreement between the two raters was 82% for Type I errors and 82% for Type II errors. (The agreement in these percentages is purely coincidental; the two parts of the coding protocol had different numbers of items to code, different numbers of codes, and different counts of agreement between raters.) Kappa coefficients were calculated to test for independence between raters. The kappa values were 0.79 for Type I errors and 0.67 for Type II errors.

Then, I met with the second rater to discuss the items on which we disagreed. There were several systematic disagreements that were easily resolved by making the coding criteria more specific. For example, the D-Union criterion was modified to specify that this type of distracter had to have more elements in it than any entry in the matrix, which was not originally part of the criterion. All of these changes are incorporated into the final criteria listed in Table 23.

After the negotiation and criteria-revision phase, agreement between raters was recalculated. Post-negotiation agreement between the two raters was 95% for Type I errors and 98% for Type II errors. Any remaining disagreements were resolved based on consideration of the conceptual type of error intended to be captured.

Once the distracters had been assigned codes, the only step left in determining what error type a particular distracter choice represents was to resolve the priority ordering between Type I and Type II errors. For example, if a distracter is chosen that represents both a D-Diff error as well as an IC-Flip error, which code is assigned to determine the overall error type? Instead of trying to create a global ordering for all combinations of Type I and Type II error codes, I first observed which codes had been assigned in cases where both Type I and Type II errors were identified for the same distracter, and I resolved these conflicts only using a few simple rules, which are applied in order:

1) Type I repetition errors take precedence over any Type II error.
2) Any Type II error takes precedence over Type I WP-Matrix errors.
3) Type II IC-Flip errors take precedence over any Type I error.
4) Type I WP-Copy or WP-Flip errors take precedence over any Type II error.
5) Type I D-Plus or D-Diff errors take precedence over any Type II error.

5.2.2 Results

Figure 31 shows the overall proportions of each error type identified in the set of answer choices of the SPM. The proportions of each error type were found to be non-uniform, $\chi^2(4, N = 432) = 30.66$, $p < 0.001$. However, if the incomplete correlate errors are considered to represent a variation on choosing the correct answer, as shown in Figure 32, then the error types are distributed uniformly, $\chi^2(3, N = 432) = 2.57$, $p = 0.46$. This is interesting and suggests perhaps that Raven himself used some scheme in constructing distracters, and if so, the coding scheme that I have developed seems to match Raven's approach to a considerable extent. Figure 33 shows the distribution of error types across Sets A through E on the SPM. The distributions differ significantly across sets, $\chi^2(16, N = 432) = 70.88, p < 0.001$.

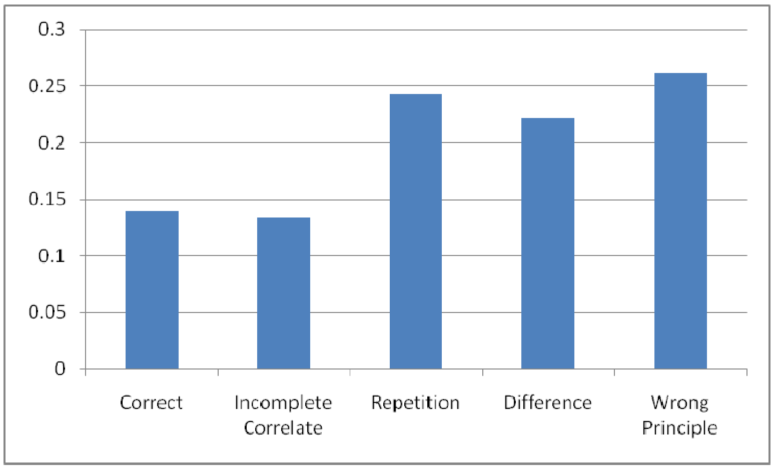


Figure 31. Proportion of error types found across all answer choices on the SPM.

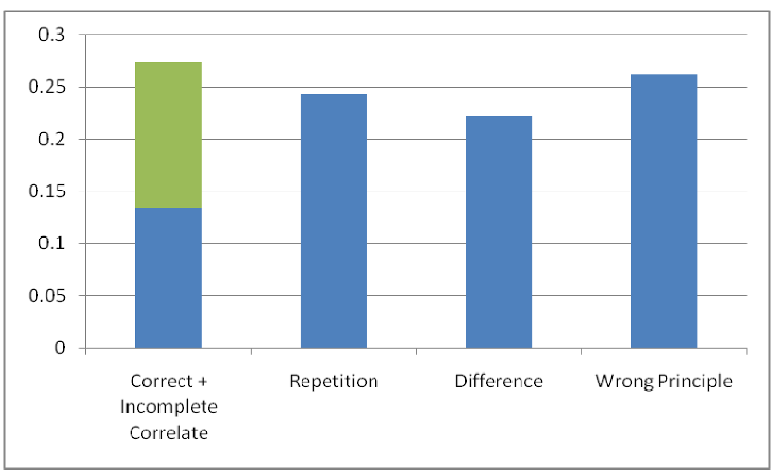


Figure 32. Proportion of error types found across all answer choices on the SPM after combining correct answers with incomplete correlates.

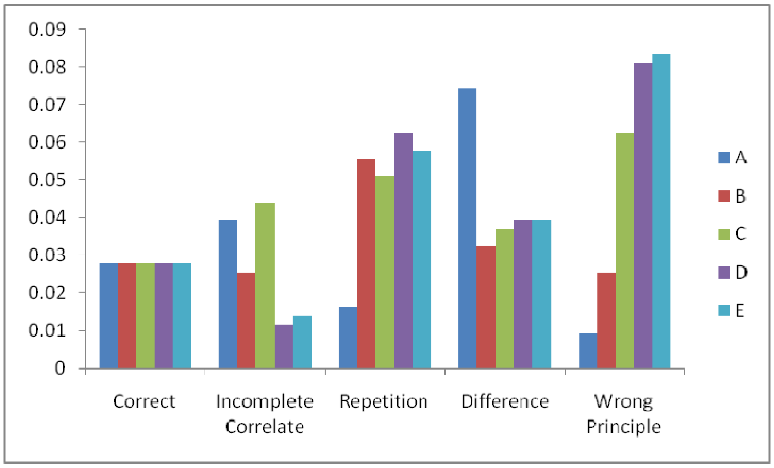


Figure 33. Proportion of error types found across each set on the SPM.

Figure 34 and Figure 35 show the distribution of error types across answer choices 1 through 6 for Sets A and B and across answer choices 1 though 8 for Sets C, D, and E. While the error frequencies were too small to perform a regular chi-square test of independence, data were analyzed using a simulated p-value, using the built-in functionality for this in the statistical software package R. The distributions are not significantly different across answer choice positions for either 2x2 matrices, $\chi^2(N = 144) = 6.71$, $p = 0.99$, or 3x3 matrices, $\chi^2(N = 288) = 16.46$, $p = 0.96$. The correct answers themselves are evenly distributed across answer choices, which was likely deliberately controlled by test creators, given that the effects of position had been observed to have a non-trivial effect on guessed answer choices.

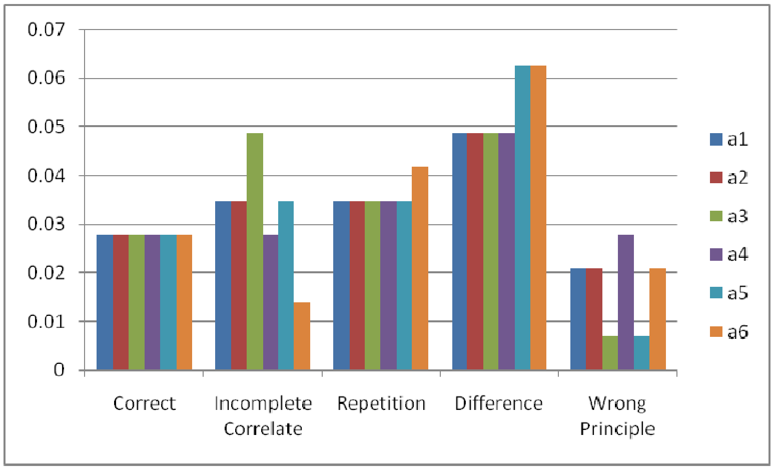


Figure 34. Proportion of error types found across each answer position for 2x2 matrices.

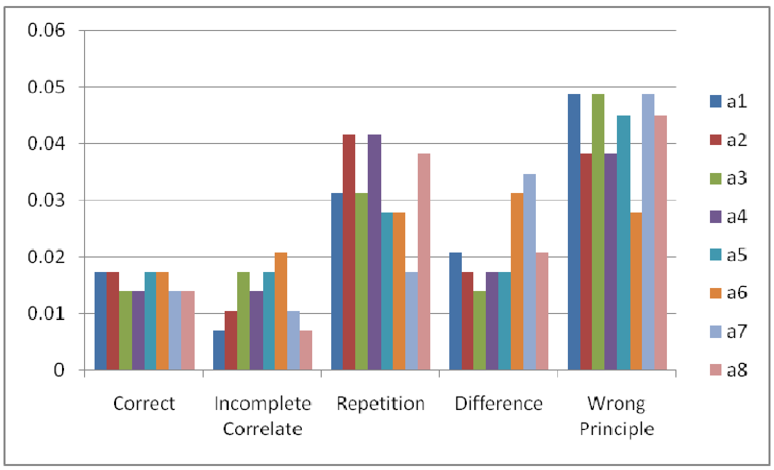


Figure 35. Proportion of error types found across each answer position for 3x3 matrices.

5.3 Methods

As can be seen from the literature, there are many different ways to examine errors made on the RPM, amidst general agreement that looking at errors in any of these ways can add valuable information about an individual's performance on the test. I focus on examining error choices by type, i.e., for problems that are not answered correctly, which conceptual types of distracters tend to be chosen?

This analysis is applied to existing human behavioral data, kindly made available by Dr. Isabelle Soulières at the University of Montreal. These data come from four groups: children and adults, either typically developing or with autism. These data are compared with each other as well as against computational data generated by the ASTI model.

5.3.1 Predictions

In this section, for each type of error discussed above, I make predictions about 1) whether the ASTI model is likely to make this type of error, and 2) whether human test-takers using either visual or verbal strategies are likely to make this type of error, i.e. does the probability of making this type of error differ based on strategy?

Incomplete correlate (IC): The ASTI model would likely make certain types of IC errors but not others. IC-Neg, IC-Fill, IC-Flip, IC-Layout, and IC-Scale errors all represent changes in the visual properties of the correct answer choice. Because the ASTI model represents visual properties explicitly (in the predicted answer image itself), it would be unlikely for the model to correctly reason about the problem far enough to get close to the correct answer but then make one of these types of errors. IC-Num and IC-Inc errors, on the other hand, have more to do with intrinsic content of the answer image, and it is possible that incorrect reasoning by the model might lead to making one of these two types of errors. If human test-takers use similar visual representations to the ASTI model, i.e. if these representations explicitly portray the visual properties of a shape in terms of its color, texture, orientation, layout, and size, then we would expect to see the same pattern: IC errors would only be made of the IC-Num and IC-Inc types. For human test-takers using verbal representations, on the other hand, number is likely one of the aspects of an answer choice that will be explicitly represented, and so we would not expect them to make IC-Num errors. However, it is plausible that they might make errors of the other types, i.e. IC-Neg, IC-Fill, IC-Flip, IC-Layout, IC-Scale, and IC-Inc.

Repetition (R): The ASTI model could very well make repetition errors, as it uses the matrix entries adjacent to the blank space to generate its predicted answer image. Thus, if the model reasons incorrectly about the matrix, it could easily select an answer choice due to the answer's similarity to an adjacent matrix entry. In humans, if repetition errors are made due to a cognitive fixation on the adjacent matrix entries, I would expect to see little difference in repetition errors between individuals who are solving the problem visually versus verbally. On the other hand, it may be that individuals who use visual representations are more likely to make repetition errors, if the visual priming/bias that occurs during inspection of the matrix is stronger for these individuals.

Difference (D): The ASTI model will likely not make difference errors, because the model reasons by trying to maximize similarity among matrix images, and difference errors represent answer choices that are very dissimilar from matrix entries. These answers would not likely be chosen by test-takers who are using a constructive matching strategy, because answer choices represent "difference" errors for the very reason that they would be unusual or complicated to construct using elements from the matrix. For this reason, I expect to see little change in difference errors between individuals solving the problem visually or verbally. On the other hand, as with repetition errors, it may be that individuals who use visual representations are more sensitive to attending to visually different stimuli, in which case their rate of difference errors might be greater.

Wrong principle (WP): The ASTI model could easily make wrong principle errors, as these errors might correspond to the ASTI model selecting an incorrect transform or image set from which to generate its predicted answer. For human test-takers, WP errors are likely to have equal frequencies of occurrence regardless of whether the test-taker is using visual or verbal representations. Differences in WP errors are probably greater among individuals of different ability levels, irrespective of what strategy they are using.

5.3.2 Overview of data

In this section, I give an overview of the quantity and type of human behavioral data that is available, along with summary statistics describing the demographics and the overall SPM performance levels shown by participants as well as the comparable SPM performance levels shown by the ASTI model.

The human behavioral data that was available for study consisted of data on SPM performance by children and adults who were either typically developing (TD) or who had been diagnosed with autism. Portions of these data were previously analyzed and published in other studies (Dawson et al., 2007; Soulieres et al., 2010). Participants included 106 TD individuals and 153 individuals with autism (AUT). Data were available for each participant giving which answer choice they chose for each of the 60 problems of the SPM, including a few instances in which an individual did not give an answer. Using age data, the participants were grouped into children and adults. I used a cutoff of 17 years for the maximum age of the children groups. One participant in the AUT group was excluded from all analyses, as he appeared to have adopted a strategy of answering "1" for more than half of the problems on the SPM.

Table 24. Breakdown of participant data by age (children versus adults) and group (typically developing versus autism versus the ASTI model).

	Children		Adults		ASTI model
	TD	**AUT**	**TD**	**AUT**	**ASTI**
N	54	108	52	44	96
SPM score: mean (SD)	42.61 (9.79)	37.43 (12.17)	50.69 (5.38)	48.43 (9.64)	32.57 (9.74)
Age in years: mean (SD)	11.96 (3.40)	11.02 (2.99)	22.98 (4.28)	26.80 (6.72)	n/a
Full scale IQ: mean (SD)	109.82 (10.35)	84.38 (20.03)	106.91 (11.76)	97.61 (16.40)	n/a

Note: Not all participants had FSIQ data available.

To obtain data samples from the ASTI model, I treated each configuration of the model described in Section 4.2 as an individual sample. In this way, I had 96 ASTI model "participants" who could then be compared with the TD and AUT human groups. Table 24 summarizes SPM, age, and full-scale IQ (FSIQ) information.

However, as noted in Section 5.1, if we wish to examine error patterns between groups, different overall levels of performance might be a confounding factor. One approach to address this confound is to select subgroups of participants who are individually matched on overall score. Then, any differences in error patterns will be due to group membership only, and not to potential group differences in ability.

This approach was applied to the child data along with data from the ASTI model. If multiple individuals had the same raw score, one was selected at random. Each pairing differed by at most 1 point. This resulted in groupings that share the same mean and very similar standard deviations. Data from this group matching are given in Table 25.

Table 25. Score-matched subgroups used for analysis of children data.

	TD	**AUT**	**ASTI**
N	38	38	38
SPM score: mean (SD)	38.26 (8.07)	38.26 (8.09)	38.29 (8.07)
Age in years: mean (SD)	11.11 (3.30)	10.76 (2.71)	n/a
Full scale IQ: mean (SD)	106.08 (9.08)	88.83 (18.79)	n/a

5.4 Results

Comparisons of error types were conducted using three different analyses:

1) Between-groups analysis, using matched child subgroups
2) Between-groups analysis, using unmatched child groups
3) Between-groups analysis, using unmatched adult groups

5.4.1 Matched child subgroups

Figure 36 shows the proportion of total errors of each conceptual error type made by participants in the matched child subgroups. There is significant agreement in error type proportions between the TD and AUT groups, $\chi^2(N = 826) = 1.89$, $p = 0.60$, whereas the error type proportions made by the ASTI group differ significantly from each of the human groups, $\chi^2(N = 826) = 91.62$, $p < 0.001$ for TD, and $\chi^2(N = 826) = 98.69$, $p < 0.001$ for AUT. The relative scarcity of difference errors made by the ASTI model was one of the predictions made earlier, though the relative abundance of repetition errors had not been predicted for the ASTI model.

Figure 37 shows errors for the three groups using specific error codes individually, without aggregating into error type. Unlike the predicted outcomes, the proportions of individual IC errors do not differ significantly between the TD and AUT groups; again, across all error codes, performance between these two groups seems well matched. Also, the ASTI model does make some errors of the IC-Fill, IC-Flip, and IC-Layout types, though as predicted, more errors are made in the IC-Inc category. In addition, the ASTI model makes many more strict errors of repetition, including the WP-copy type of error, than do either of the human groups. The human groups, on the other hand, make significantly more WP-matrix errors. This may indicate that human participants have some notion that simply copying a matrix entry does not usually lead to the correct answer, whereas the ASTI model has no such proscription.

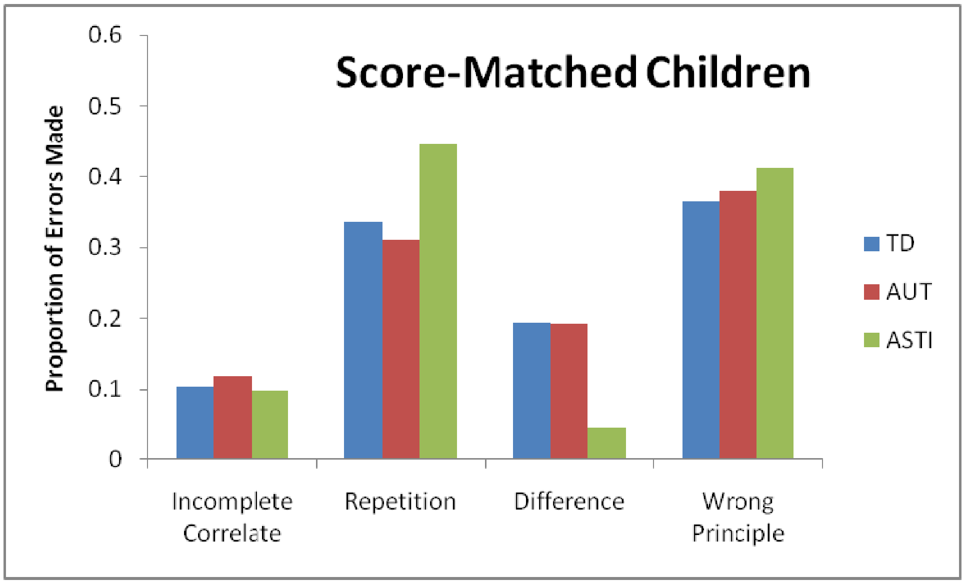


Figure 36. Error types made by the three score-matched subgroups of typically developing children (TD), children with autism (AUT), and the ASTI model (ASTI).

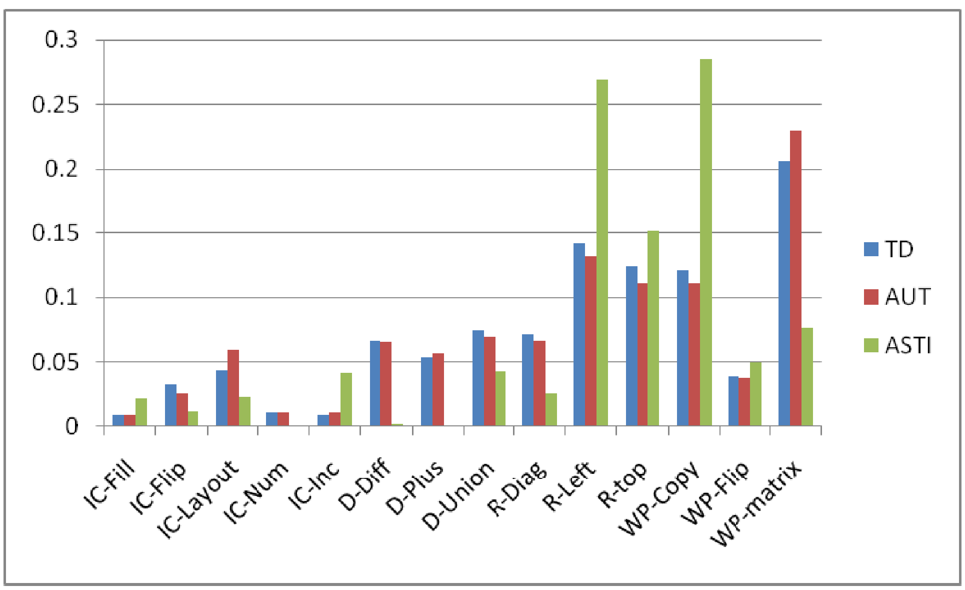


Figure 37. Specific errors made by the three score-matched subgroups of typically developing children (TD), children with autism (AUT), and the ASTI model (ASTI).

5.4.2 Unmatched groups

Figure 38 shows the proportion of total errors of each conceptual error type made by participants in the entire child groups, without matching subgroups based on total score (top) and the three adult groups (bottom). Both of these sets of results are similar to those found for the matched child subgroups described in the previous section.

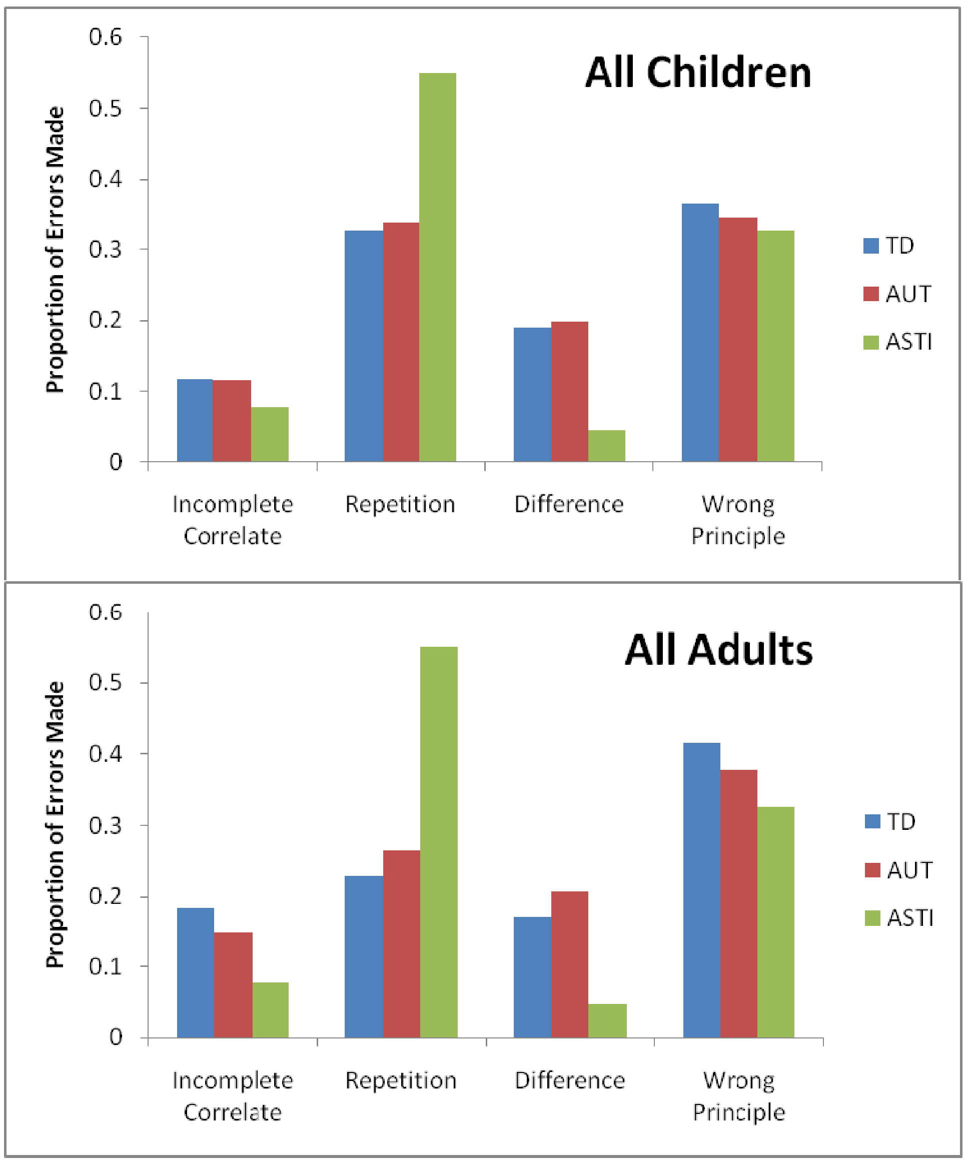


Figure 38. Error types made by children (top) and adults (bottom) in the three main groups: typically developing (TD), autism (AUT), and ASTI model (ASTI).

5.5 Claims and Future Work

Following the analyses presented in this chapter, it does not seem that error types alone provide a clear window into whether an individual is using a visual or verbal strategy on the SPM. One major contribution of this chapter is the classification of error types on the SPM. While earlier efforts to obtain such a classification were unsuccessful (Vodegel-Matzen, 1996), the current approach used a new set of more clearly defined classification criteria, as well as a new two-stage approach to classifying errors. Using this approach, an error classification for the SPM was obtained with 95% inter-rater reliability. This classification will have utility for other studies of human or machine performance on the SPM, and it adds significant information for the RPM family of tests, as both the CPM and the APM already had error classifications, but the SPM did not.

Future work on this topic should include addressing two important questions. First, what factors do affect an individual's particular choice of errors on a test like the SPM, if the visual/verbal nature of their problem-solving strategy does not seem to play a role? And second, what other behavioral markers may be useful to study the visual/verbal nature of an individual's strategy? Neuroimaging data from fMRI provides one avenue (e.g. Prabhakaran et al., 1996; Soulieres et al., 2011), as do eye-tracking studies or the use of verbal reporting protocols or qualitative observations (e.g. Bromley, 1957) to better understand an individual's problem-solving and decision-making processes.

6 CONCLUSION

This dissertation began with a discussion of problem solving, and in particular the types of knowledge representations and associated reasoning (i.e. cognitive strategies) that intelligent agents use to solve different kinds of problems. The ensuing work centers on two key insights about classifications of cognitive strategies. First, there is a difference between the way problem-solving tasks are *typically* or *readily* done and the way that they *can* be done. While a majority of research in psychology, psychometrics, and artificial intelligence has focused exclusively on the former, problems with this approach can arise when assumptions about typicality dominate, for instance by finding their way into studies of atypical cognition and problem solving under the radar, so to speak, without explicit justification, or by suppressing the study of alternate strategies.

The second insight has to do with what it means to "think visually," a concept inspired in this dissertation by the "visual thinking" accounts of many individuals on the autism spectrum. A simple definition might be that thinking visually means doing well on visual tasks and poorly on verbal tasks. However, the standard classifications of tasks as "visual" or "verbal" represent how tasks are *typically* done. Many of these tasks *can*, in fact, be done either visually or verbally, and so a more precise definition of thinking visually, from an information-processing perspective, is someone who 1) shows poor performance on tasks that can only be done verbally, 2) shows intact performance on tasks that are typically done visually, and 3) shows intact performance on tasks that are typically done verbally but can be done visually, by recruiting an atypical visual strategy.

Using this kind of task classification coupled with general behavioral predictions in order to describe a form of cognition is, to my knowledge, new, and explicitly accounts for the multiplicity of information processing strategies that can be successful in many problem domains as well as for the human brain's incredible capacity to adapt, creatively drawing upon areas of cognitive strength to compensate for areas that might be inaccessible or difficult to use. One way to conceptualize how this approach can characterize a particular form of cognition is to think in terms of a "cognitive phenotype," a notion recently coming into use to describe atypical forms of cognition, especially in neuropsychological disorders. Just as an organism's phenotype represents some presentation of physical characteristics that are tied to genetics and development, a cognitive phenotype represents a particular presentation of cognitive characteristics.

Autism seems unusual in the space of psychological disorders in the heterogeneity of cognitive phenotypes that are exhibited by affected individuals. While cognitive characterizations within autism seemed, for a time, to focus primarily on differences of *degree* (e.g. IQ, low- versus high-functioning, graded levels of language ability, etc.), it is becoming more apparent that there are also differences of *kind*, and these differences may represent distinct etiological subtypes within the autism spectrum (Charman et al., 2011). The notion of a "visual thinker" that I present in this dissertation attempts to describe one cognitive phenotype of autism that seems prevalent in introspective and anecdotal accounts. While I focus mainly on the occurrence of this cognitive phenotype within autism, it may have utility in describing individual differences in cognitive styles among typically developing individuals as well, for instance along the lines of Gardner's theories of multiple intelligences (Gardner, 1985).

My characterization of a "visual thinking" cognitive phenotype, i.e. the Thinking in Pictures (TiP) hypothesis, specifies a set of behavioral predictions that can be used to identify the occurrence of this phenotype in an individual. While the work presented in Chapter 2 of this dissertation examines these behavioral predictions in light of existing empirical studies, these predictions could also be used to design new experiments that explicitly test for the presence of this phenotype. This kind of assessment could yield numerous benefits for both the study of autism and for individuals on the autism spectrum, including 1) identifying a potential autism subtype defined by the presence of this cognitive phenotype, 2) tailoring the design of interventions to specifically recruit visual strategies, and 3) on an individual basis, informing the selection of certain interventions over others.

However, there are many aspects of the Thinking in Pictures hypothesis that remain important open questions. Clearly, reliance on a particular form of mental representation is only one aspect of cognition; studies of autism have also examined differences ranging from perception and sensory sensitivities to language and metacognition, and with the advancement of neuroimaging technologies, these various cognitive differences are becoming more and more linkable to specific differences in neural development and activation. If it is the case that there is a distinct visual thinking cognitive phenotype within autism, then it must ultimately be situated within an etiological framework that includes genetics, neurobiology, and development, as well as these other aspects of cognition and behavior.

In Chapter 3, I present a computational model, the ASTI model, primarily as a proof of concept of how one particular visual strategy can solve a large proportion of problems from the Raven's Progressive Matrices (RPM) series of intelligence tests. This model is fairly unique among computational models of the RPM, and indeed among AI models of high-level problem solving in general, in its use of iconic visual representations instead of propositional representations. Perhaps the biggest question surrounding the ASTI model is to what extent it serves as a model of human cognition, including 1) human cognition in general, 2) autistic cognition in particular, and 3) the neural bases of cognition.

In terms of general human cognition, the ASTI model is clearly a simplified model on many fronts. Mainly, the ASTI model is a content model of solving RPM problems, not a process model, and so it does not include many procedural aspects of problem solving, such as attention, learning, iteration, or cognitive control, especially in consideration of taking the test as a whole and not just addressing individual problems. On the content side, it provides a rather coarse-grained model of the kinds of operations that might be performed in mental imagery (or operations that are isomorphic in some sense); given the closed-world nature of individual RPM problems, and the visual simplicity of problem inputs (e.g. black-and-white shape and line drawings), this level of granularity appears to be appropriate, with the exceptions of operations of image scaling, filling, and segmentation that I mention in Chapter 3. In terms of the specific reasoning content generated for each RPM problem, one limitation of the ASTI model is that it uses a purely feed-forward approach, predicting a single concrete answer before inspecting any of the answer choices. This strategy of *constructive matching* is observed in humans but exists alongside the strategy of *response elimination*, which involves inspecting the answer choices up front, in conjunction with inspection of the matrix (Bethell-Fox et al., 1983). These two strategies, though ostensibly about the problem-solving process, also involve content as well, as they affect which aspects of the problem content (matrix vs. answer choices) are brought into play during various components of problem solving.

In terms of autistic cognition, the ASTI model attempts to approximate the way in which individuals with autism who have a visual thinking cognitive phenotype might solve RPM problems, respecting the above mentioned limitations of the model, in the sense that these individuals would be expected to use a purely visual strategy on the test, whereas typically developing individuals would be expected to use a combination of visual and verbal strategies, akin perhaps to a combination of the ASTI model with a propositional RPM model. One important question, however, is whether the kind of visual strategy used by individuals with autism who are visual thinkers would be the same as a typical visual strategy. A recent paper looking at mental image comparisons and mental rotation has found superior performance in individuals with autism (Soulières, Zeffiro, Girard, & Mottron, 2011). Given the depth of cognitive differences in domains like language and perception that exist between individuals with autism and typically developing individuals, it would seem surprising if there were not also qualitative differences in visual processing as well, though what form these differences might take remains to be seen.

In terms of its neural plausibility, the ASTI model in its current implementation makes no attempt to emulate aspects of neural computation. However, there are several avenues that could be explored in this direction. In terms of its perception and internal representation, the ASTI model uses pixel-based images, which are essentially two-dimensional intensity maps of rectilinearly arrayed points. We know that at least a portion of the primary visual cortex is dedicated to hierarchically processing visual inputs into spatial features such as edges, corners, and points, and the two-dimensional nature of this information is preserved according to the retinopically mapped structure of neurons.

If neural computation during mental imagery operates on this information in a way that preserves these two characteristics (Slotnick, Thompson, and Kosslyn, 2005), then these computations may, like those of the ASTI model, be described as combinations of affine and set transformations upon this information. Clearly, both from an efficiency standpoint as well as from what we know of neural networks, such computations are likely to be implemented in the brain in a massively parallel fashion. Again, this is an example in which the ASTI model as currently implemented is best described as a content model rather than a process model; it may be able to perform a similar or isomorphic operation, but currently uses only serial processing. The ASTI model could also be augmented to include features more similar to the edges and lines detected by the brain; computer vision offers many potential approaches in this direction.

Finally, in Chapter 5, I compare conceptual types of errors made on the RPM between typically developing individuals, individuals with autism, and various configurations of the ASTI model. A striking finding is how similar the error patterns made by the two human groups are, especially given the differences observed in behavioral measures like comparative IQ performance and reaction time, as well as patterns of brain activation (Dawson et al., 2007; Soulieres et al., 2010). One possible explanation is that whatever altered types of strategy or ability are causing the other observed differences, these alterations do not have any effect on the types of conceptual errors that are made; for instance, error patterns may depend more on overall ability using any type of representation, rather than being independent of ability within a particular representational paradigm. On the other hand, it may be that the error types themselves are not represented at a fine enough level of resolution or are capturing the wrong level of abstraction needed to see the effects of strategy differences. Additionally, the hypothesis that TD individuals and individuals with autism would exhibit different error patterns across the entire test assumes that strategy differences can be found on all, or at least most, problems. Behavioral studies of the RPM have long observed that different problems seem to elicit different problem-solving strategies, and so there may be subclasses of problems on which error patterns differences are more pronounced.

In sampling the types of errors made by the ASTI model, I chose to use an approach that does not explicitly model the decision-making process of the system choosing from among the given set of answer choices. Instead, in keeping with conceptualizing the ASTI model as a content model instead of a process model, I examined what kinds of errors would be made across the space of variations in visual strategy that can be represented by systematic ablations of the ASTI model. While the overall pattern of errors of the ASTI model is significantly different from both of the human group patterns, there are at least some similarities, and the differences seem to be explainable by particular features of the model. In particular, there seem to be three types of errors in which we see the largest differences between the model and human groups: 1) difference errors are rarely made by the model, which is likely due to the fact that the model uses a purely feedforward approach to inspecting answer choices and does not compute the salience of answer choices, 2) repetition errors are made much more often by the model than by the human groups, which is perhaps because the model always uses single transformations of elements taken directly from the matrix in order to construct a predicted answer, and 3) the model is far less likely to make an error that represents a more complex transformation of a matrix entry, perhaps for the same reason. Expanding the ASTI model to include a response elimination approach that takes into account the saliency of particular answer choices, in addition to the constructive matching approach that is currently used by the model, may increase the fidelity with which the model exhibits human-like error patterns.

In summary, this dissertation began by examining a new hypothesis about visual thinking in autism, namely that certain individuals with autism may have a bias towards using visual instead of verbal mental representations. I found evidence across several task domains indicating that this type of visual bias may be present in many individuals with autism, and that current cognitive theories of autism do not explicitly account for these findings. To show the feasibility of this hypothesis in one domain, the Raven's Progressive Matrices test, I constructed the ASTI model, which uses purely visual representations to solve many RPM problems. I tested the ASTI model against all three of the widely used Standard, Colored, and Advanced Progressive Matrices tests and found that the ASTI model correctly solves a considerable proportion of problems on these tests, with ceiling or near-ceiling performance on the Standard and Colored tests; future work will include incorporating mechanisms for image segmentation into the model, which is likely needed for improved performance on the Advanced test. I also conducted ablation experiments with the model to derive a new data-based classification of problem types on the Raven's Standard and Colored Progressive Matrices tests. Using this model, I then made predictions about the types of errors that might be made by individuals with autism versus by typically developing individuals. This analysis revealed first, that looking at error types may need to be done at a finer-grained level of analysis to reveal the differences, if any, in errors made by the two human groups.

Second, this analysis uncovered some limitations of the ASTI model, namely that its reliance on constructive matching and its lack of measures of attention and salience seem to cause an unusual pattern of errors relative to human performance.

This dissertation incorporates several theoretical and technical firsts, including:

1) the first systematic examination of a visual thinking bias in individuals with autism across multiple task domains
2) the first computational model that uses purely visual representations to solve problems from the Raven's Progressive Matrices tests
3) the first model of any kind to be tested against the entirety of all three of the Standard, Colored, and Advanced Progressive Matrices tests
4) the first RPM problem classification based on model ablation experiments
5) the first successful qualitative classification of conceptual error types on the Standard Progressive Matrices test

The main contribution of this dissertation lies in its integrated, interdisciplinary investigation into the role that visual mental representations can play in high-level problem solving. As discussed above, this work has significant implications for the neuropsychological study of autism, for the design and interpretation of psychometric intelligence tests, for the construction of AI computational models that reason visually instead of propositionally, and finally, for the development of information processing theories of human cognition, specifically in the areas of mental imagery and multimodal processing.

APPENDIX A: PROTOCOL FOR CODING SPM ERROR TYPES

This appendix contains a copy of the protocol provided to the second independent coder to classify distracters on the SPM according to error type (see Section 5.2).

(Protocol Page 1) Overview: Error type classification on the SPM:

There are four basic conceptual types of errors on the Raven's Standard Progressive Matrices Test: 1) incomplete correlate, 2) repetition, 3) difference, and 4) wrong principle.

Incomplete correlate (IC) errors are those in which the distracter is almost, but not quite, correct. For example, some IC distracters represent a rotation or reflection of the correct answer. Other IC distracters differ from the correct answer in a single feature dimension, e.g. they might have four elements instead of three, or straight elements instead of curvy ones, or have the correct shape but the wrong texture. Alternately, an IC distracter might be only missing an element from the correct answer. Oftentimes, an IC distracter might be correct in terms of a single row or column in the matrix, e.g. looking just at the right-most column or just at the bottom-most row, but when both rows and columns are taken into account, it no longer fits the matrix pattern. These kinds of errors are made when a test-taker more or less "gets" the problem, in terms of identifying and understanding the relevant matrix relationships, but then fails to fully account for all of the problem details when selecting an answer.

Repetition (RP) errors are those in which the distracter is a copy of one of the matrix entries that is adjacent to the blank space. Choosing an RP distracter may represent a sort of cognitive bias or fixation on the matrix entries, in which an answer is selected based on simple perceptual matching between the answer choices and the matrix entries closest to the blank space. These entries may be privileged because of their proximity to the blank space. Alternately, assuming a top-left to bottom-right visual scanning pattern, these adjacent entries may be the last viewed before the test-taker moves on to look at the answer choices, assuming a sequential inspection of the problem in a matrix-first, answers-second ordering.

Difference (DF) errors are those in which the distracter is somehow qualitatively different in appearance from the other distracters. DF distracters include those that are completely blank, as well as those that have extraneous shapes that are not found anywhere in the problem matrix. In addition, a DF distracter is often the most complex-seeming answer choice, either combining all of the matrix entries together into a single agglomeration of matrix elements or taking some feature from the matrix and increasing its value until it surpasses all the other entries and answer choices. A DF distracter might be chosen because it visually "pops" from among the other answer choices.

Wrong principle (WP) errors are those in which the distracter is a copy of or composition of various elements from various matrix entries (with the exception of copies of adjacent entries, which would still fall under the "repetition" error type). A WP distracter might be chosen if the test-taker does not successively educe the correct relationship from the matrix entries and instead combines the entries according to some other rule or relationship to produce an answer choice.

(Protocol Page 2) The following table gives specific criteria that can be used to distinguish among the various conceptual types of errors found on the SPM.

Error type taxonomy and classification criteria for the SPM

Error type		Criteria
Incomplete correlate	1	Negative (color-inversion) of correct answer
	2	Change only in fill, texture, or style
	3	Rotation/reflection of correct answer
	4	Change only in spatial layout of elements
	5	Change only in size or scale
	6	Change only in number of discrete elements
	7	Incomplete, with missing element or portion
Repetition	8	Repetition of matrix entry to left of blank space
	9	Repetition of matrix entry above blank space
	10	Repetition of matrix entry to top-left of blank space
Difference	11	Filled completely white or black
	12	Union or agglomeration of all or most matrix entries
	13	Maximizes some feature value
	14	Differs qualitatively from matrix and other answers, or contains information not found anywhere in matrix
Wrong principle	15	Repetition of matrix entry not adjacent to blank space
	16	Rotation/reflection of matrix entry
	17	Transformation/combination of matrix entries

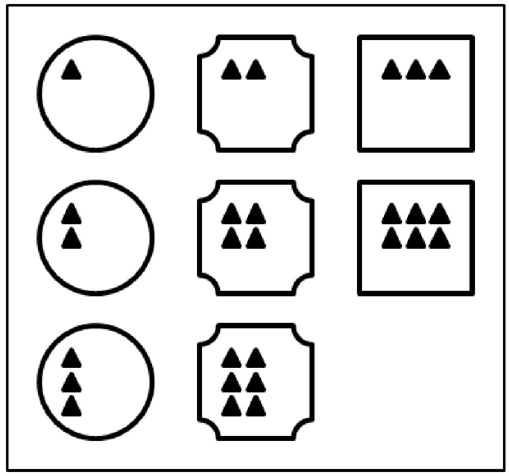
Note that these error type criteria can be broadly divided into two categories:

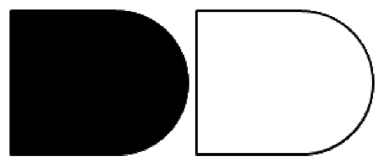
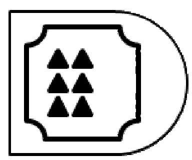
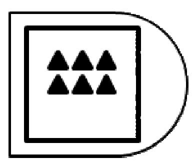
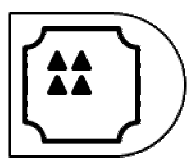
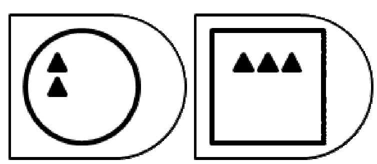
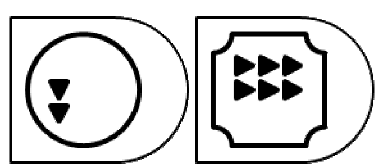
1) If an individual does not know or guess the correct answer, even partially, then they may make repetition, difference, or wrong principle errors. The distracters that represent these error types can be identified based on how the distracter is related to information in the matrix.
2) If an individual does partially guess the correct answer, then they may make incomplete correlate errors. The distracters that represent these error types can be identified based on how the distracter is related to the correct answer choice.

Therefore, the scheme for coding error types on the SPM actually has two parts:

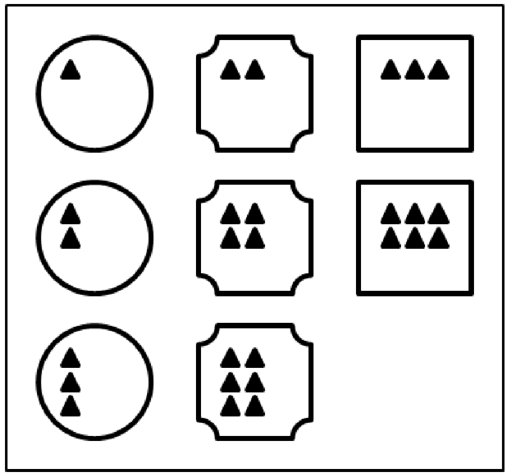
1) Without consideration of the correct answer, first code each answer choice in terms of criteria #8-17, which represent how each answer is related to information in the matrix.
2) Then, without consideration of the matrix, using only knowledge of the correct answer, code each answer choice in terms of criteria #1-7, which represent how each answer is related to the correct answer.

(Protocol Page 3) Here is an example problem, along with examples of the kinds of distracters that fall under each error type criterion.



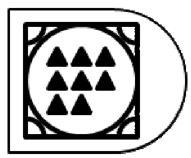
F	Mark any answer choice that is filled completely white or completely black.	
L	Mark any answer choice that is a repetition of the matrix entry directly to the left of the blank space.	
T	Mark any answer choice that is a repetition of the matrix entry directly to the top of the blank space.	
D	Mark any answer choice that is a repetition of the matrix entry directly to the diagonal top-left of the blank space.	
C	Mark any answer choice that is a copy of any matrix entry that is not directly adjacent to the blank space.	
↻	Mark any answer choice that is a rotation or reflection of any matrix entry.	

(Protocol Page 4)



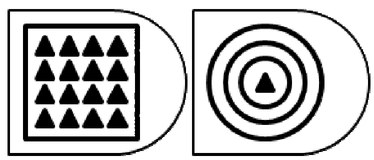
U

Mark any answer choice that is a union or agglomeration of all or most of the matrix entries (or aspects of them).



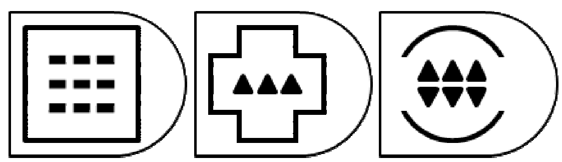
+

Mark any answer choice in which some particular feature found in the matrix is maximized or made more complex.



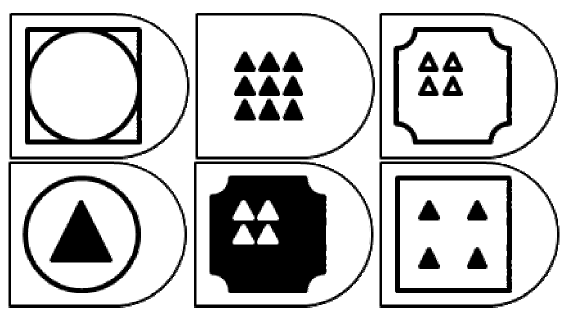
X

Mark any answer choice that contains new content not found in the matrix or other answer choices or is otherwise qualitatively different from the other answer choices.

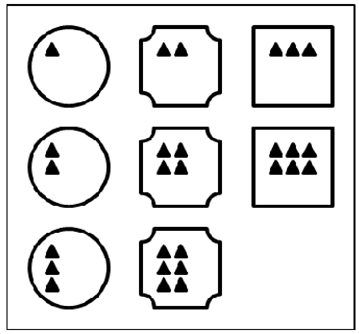


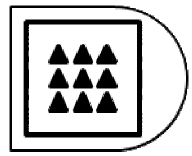
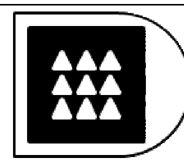
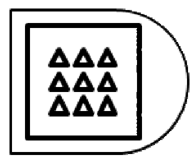
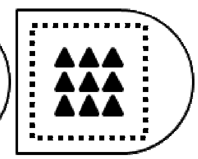
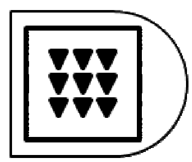
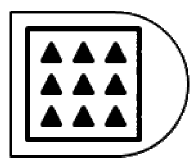
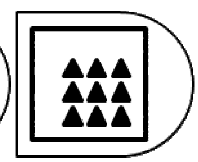
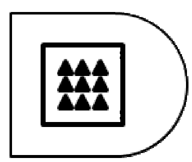
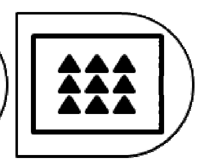
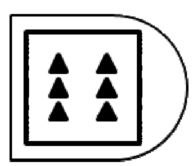
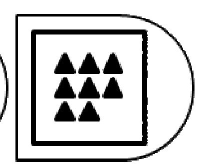
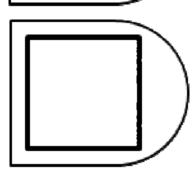
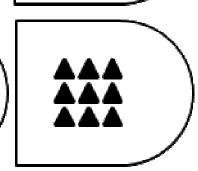
M

Mark any answer choice that represents any other transformation or combination of matrix entries.



(Protocol Page 5)



④	Mark the correct answers by circling the number choice.	
N	Mark any answer choice that is a negative (color-inverted) image of the correct answer.	
F	Mark any answer choice that is the same as the correct answer except with a change only in fill, texture, or style.	 
↻	Mark any answer choice that is a rotation or reflection of the correct answer.	
L	Mark any answer choice that is the same as the correct answer except with a change only in spatial layout of elements.	 
S	Mark any answer choice that is the same as the correct answer except with a change only in size or scale.	 
#	Mark any answer choice that is the same as the correct answer except with a change only in number of elements.	 
—	Mark any answer choice that is the same as the correct answer but with a missing element or portion.	 

(Protocol Page 6) Part 1: How answer choices are related to content in the problem matrix:

Instructions: Part 1 uses Test Booklet A, which contains the complete matrix and answers for each problem.

First, using the six codes found in Step 1 in the table below, go through each problem in the test and mark any answers that fit these six criteria. If more than one criterion fits a particular answer choice, just mark the first one that applies using the order specified in the table below.

Then, go through the test once more, this time marking answers that fit the codes and criteria listed in Step 2. If an answer choice has already been marked during Step 1, skip it. At the end, each answer choice should have exactly one code assigned to it.

Step	Code	Criteria
1)	F	Mark any answer choice that is filled completely white or completely black.
	L	Mark any answer choice that is a repetition of the matrix entry directly to the left of the blank space.
	T	Mark any answer choice that is a repetition of the matrix entry directly to the top of the blank space.
	D	Mark any answer choice that is a repetition of the matrix entry directly to the diagonal top-left of the blank space.
	C	Mark any answer choice that is a copy of any matrix entry that is not directly adjacent to the blank space.
	↻	Mark any answer choice that is a rotation or reflection of any matrix entry.
2)	U	Mark any answer choice that is a union or agglomeration of all or most of the matrix entries (or aspects of them).
	+	Mark any answer choice in which some particular feature found in the matrix is maximized or made more complex.
	X	Mark any answer choice that contains new content not found in the matrix or other answer choices or is otherwise qualitatively different from the other answer choices.
	M	Mark any answer choice that represents any other transformation or combination of matrix entries.

(Protocol Page 7) Part 2: How answer choices are related to correct answer:

Instructions: Part 1 uses Test Booklet B, which contains only the answers for each problem.

Step 0 has already been completed; the correct answers have been marked by circling the number of the appropriate choice.

Using the seven codes found in Step 1 in the table below, go through each problem in the test and, for each answer choice <u>other</u> than the correct one, mark any answers that fit these seven criteria. If more than one criterion fits a particular answer choice, just mark the first one that applies using the order specified in the table below. Not all answer choices need to be marked; if an answer choice fits none of these seven criteria, then just leave it blank. At the end, each answer choice (excluding the correct answer) should have zero or one codes assigned to it.

Step	Code	Task
0)	④	Mark the correct answers by circling the number choice.
1)	N	Mark any answer choice that is a negative (color-inverted) image of the correct answer.
	F	Mark any answer choice that is the same as the correct answer except with a change only in fill, texture, or style.
	↻	Mark any answer choice that is a rotation or reflection of the correct answer.
	L	Mark any answer choice that is the same as the correct answer except with a change only in spatial layout of elements.
	S	Mark any answer choice that is the same as the correct answer except with a change only in size or scale.
	#	Mark any answer choice that is the same as the correct answer except with a change only in number of discrete elements.
	I	Mark any answer choice that is the same as the correct answer but is incomplete, with a missing element or portion.

REFERENCES

Ambery, F., Russell, A., Perry, K., Morris, R., & Murphy, D. (2006). Neuropsychological functioning in adults with Asperger syndrome. *Autism*, 10, 551-564.

Ameli, R., Courchesne, E., Lincoln, A., Kaufman, A., & Grillon, C. (1988). Visual memory processes in high-functioning individuals with autism. *Journal of Autism and Developmental Disorders*, 18, 601-615.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders (Revised 4th Ed.)*. Washington, DC: American Psychiatric Association.

Babcock, R. (1994). Analysis of adult age differences on the Raven's Advanced Progressive Matrices test. *Psychology and Aging*, 9 (2), 303-314.

Babcock, R. L. (2002). Analysis of age differences in types of errors on the Raven's Advanced Progressive Matrices. *Intelligence*, 30(6), 485-503.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829-839.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, 130, 641-657.

Baldassi, S., Pei, F., Megna, N., Recupero, G., Viespoli, M., Igliozzi, R., Tancredi, R., et al. (2009). Search superiority in autism within, but not outside the crowding regime. *Vision Research*, 49, 2151-2156.

Baldo, J., Dronkers, N., Wilkins, D., Ludy, C., Raskin, P., & Kim, J. (2005). Is problem solving dependent on language? *Brain and Language*, 92, 240-250.

Barbey, A. K., & Barsalou, L. W. (2009). Reasoning and problem solving: Models. *Encyclopedia of Neuroscience*, Vol. 8, pp. 35-43.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S., & Belmonte, M. (2005). Autism: A window onto the development of the social and the analytic brain. *Annual Review of Neuroscience*, 28, 109-126.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577-660.

Bello, P., & Cassimatis, N. (2006). Developmental accounts of theory-of-mind acquisition: Achieving clarity via computational cognitive modeling. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, pp. 1014-1019.

Bennetto, L., Pennington, B., & Rogers, S. (1996). Intact and impaired memory functions in autism. *Child Development*, 67, 1816-1835.

Berker, E., & Smith, A. (1988). Diaschisis, site, time and other factors in Raven performances of adults with focal cerebral lesions. *The International Journal of Neuroscience*, 38(3-4), 267-285.

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. *Brain*, 128, 2430-2441.

Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye-movement analysis of geometric analogy performance. *Intelligence*, 8, 205-238.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15 (11), 527-536.

Bölte, S., Dziobek, I., & Poustka, F. (2009). Brief report: The level and nature of autistic intelligence revisited. *Journal of Autism and Developmental Disorders*, 39 (4), 678-682.

Bors, D. A., & Vigneau, F. (2001). The effect of practice on Raven's Advanced Progressive Matrices. *Learning and Individual Differences*, 13(4), 291-312.

Bowler, D. M., & Briskman, J. A. (2000). Photographic cues do not always facilitate performance on false belief tasks in children with autism. *Journal of Autism and Developmental Disorders*, 30(4), 305-316.

Bowler, D., Briskman, J., Gurvidi, N., & Fornells-Ambrojo, M. (2005). Understanding the mind or predicting signal-dependent action? Performance of children with and without autism on analogues of the false-belief task. *Journal of Cognition and Development*, 6, 259-283.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992a). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 157.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992b). Verbal recoding of visual stimuli impairs mental image transformations. *Memory & Cognition*, *20*(4), 449.

Brandimonte, M.A., Hitch, G.J., & Bishop, D.V.M. (1992c). Manipulation of visual mental images in children and adults. *Journal of Experimental Child Psychology*, 53(3), 300-312.

Breier, J., Plenger, P., Castillo, R., Fuchs, K., Wheless, J., Thomas, A., et al. (1996). Effects of temporal lobe epilepsy on spatial and figural aspects of memory for a complex geometric figure. *Journal of the International Neuropsychological Society*, 2, 535-540.

Brenner, L., Turner, K., & Müller, R. (2007). Eye movement and visual search: Are there elementary abnormalities in autism? *Journal of Autism and Developmental Disorders*, 37, 1289-1309.

Bringsjord, S., & Schimanski, B. (2003). What is artificial intelligence? Psychometric AI as an answer. In *Proceedings from IJCAI* (Vol. 18, pp. 887–893).

Bromley, D. B. (1953). Primitive forms of response to the matrices test. *Journal of Mental Sciences*, 99, 374-393.

Brooks, L. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 22, 349-368.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1), 139-159.

Caramazza, A. (1996). Pictures, words, and the brain. *Nature*, 383, 216-217.

Caron, M., Mottron, L., Berthiaume, C., & Dawson, M. (2006). Cognitive mechanisms, specificity and neural underpinnings of visuospatial peaks in autism. *Brain*, 129, 1789-1802.

Caron, M., Mottron, L., Rainville, C., & Chouinard, S. (2004). Do high functioning persons with autism present superior spatial abilities? *Neuropsychologia*, 42, 467-481.

Carpenter, P.A., Just, M.A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97 (3), 404-431.

Carter, D. E. (1970). *An Exploratory Investigation of the Processes Involved in the Completion of Raven's Progressive Matrices*. M. A. Thesis, Department of Psychology and Sociology, University of Canterbury.

Casakin, H., & Goldschmidt, G. (1999). Expertise and the use of visual analogy: Implications for design education. *Design Studies*, *20*, 153–175.

Chalmers, D., French, R., & Hofstadter, D. (1992). High-level perception, representation and analogy: A critique of artificial intelligence methodology. *Journal of Experimental and Theoretical Artificial Intelligence*, *4*, 185-211.

Chandrasekaran, B. (2011). When is a bunch of marks on paper a diagram? Diagrams as homomorphic representations. *Semiotica*, *186* (1), 69-87.

Chandrasekaran, B., Banerjee, B., Kurup, U., & Lele, O. (2011). Augmenting cognitive architectures to support diagrammatic imagination. *Topics in Cognitive Science*, *3* (4), 760-777.

Charman, T., & Baron-Cohen, S. (1992). Understanding drawings and beliefs: A further test of the metarepresentation theory of autism: A research note. *Journal of Child Psychology and Psychiatry*, 33, 1105-1112.

Charman, T., & Baron-Cohen, S. (1995). Understanding photos, models, and beliefs: A test of the modularity thesis of theory of mind. *Cognitive Development*, 10(2), 287-298.

Charman, T., & Lynggaard, H. (1998). Does a photographic cue facilitate false belief performance in subjects with autism?. *Journal of Autism and Developmental Disorders*, 28(1), 33-42.

Charman, T., Jones, C. R. G., Pickles, A., Simonoff, E., Baird, G., & Happé, F. (2011). Defining the cognitive phenotype of autism. *Brain Research*, 1380, 10-21.

Chuah, Y., & Maybery, M. (1999). Verbal and spatial short-term memory: Common sources of developmental change? *Journal of Experimental Child Psychology*, 73, 7-44.

Cirillo, S., & Ström, V. (2010). *An Anthropomorphic Solver for Raven's Progressive Matrices* (No. 2010:096). Goteborg, Sweden: Chalmers University of Technology.

Clement, J. (2008). *Creative Model Construction in Scientists and Students: The Role of Imagery, Analogy, and Mental Simulation*. Dordrecht: Springer.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 75-176). New York: Academic Press.

Corman, L., & Budoff, M. (1974). Factor structures of Spanish-speaking and non-Spanish-speaking children on Raven's Progressive Matrices. *Educational and Psychological Measurement*, 34(4), 977-981.

Croft, D., & Thagard. P. (2002). Dynamic imagery: A computational model of motion and visual analogy. In Magnani, L., & Nersessian, N. (eds.), *Model-based reasoning: Science, technology, & values.* New York: Kluwer Academic/Plenum Publishers, pp. 259-274.

Davies, J., & Goel, A. (2001). Visual analogy in problem solving. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI-01),* 377-382.

Davies, J., & Goel, A. (2008). Visual re-representation in creative analogies. *Open AI Journal*, *2*, 11-20.

Davies, J., Goel, A., & Nersessian, N. (2009). A computational model of visual analogies in design. *Journal of Cognitive Systems Research*, 10, 204-215.

Davies, J., Goel, A., & Yaner, P. (2008). Proteus: Visuospatial analogy in problem solving. *Knowledge-Based Systems*, *21*(7): 636-654.

Davies, J., Nersessian, N., & Goel, A. (2005). Visual models in analogical problem solving. *Foundations of Science*, *10*(1), 133-152.

Dawson, M., Soulieres, I., Gernsbacher, M.A, & Mottron, L. (2007). The level and nature of autistic intelligence. *Psychological Science*, 18 (8), 657-662.

de Villiers, J., & de Villiers, P. (2003). Language for thought: Coming to understand false beliefs. In *Language in Mind: Advances in the Study of Language and Thought* (pp. 335–384). Cambridge, MA: MIT Press.

DeShon, R.P, Chan, D., & Weissbein, D.A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. *Intelligence*, 21 (2), 135-155.

Dillon, R. F., Pohlmann, J. T., & Lohman, D. F. (1981). A factor analysis of Raven's Advanced Progressive Matrices freed of difficulty factors. *Educational and Psychological Measurement*, 41, 1295–1302.

Edgin, J., & Pennington, B. (2005). Spatial cognition in autism spectrum disorders: Superior, impaired, or just intact? *Journal of Autism and Developmental Disorders*, 35, 729-745.

Emerson, M., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, 48, 148-168.

Evans, T. (1968). A heuristic program to solve geometric analogy problems. In Minsky, M. (ed.), *Semantic information processing*. Cambridge, MA: MIT Press.

Eysenck, M. D. (1945). A study of certain qualitative aspects of problem-solving in senile-dementia patients. *Journal of Mental Science*, 91, 337-345.

Facon, B., & Nuchadee, M. L. (2010). An item analysis of Raven's Colored Progressive Matrices among participants with Down syndrome. *Research in Developmental Disabilities*, 31, 243-249.

Fajgelj, S., Bala, G., & Katic, R. (2010). Latent structure of Raven's Colored Progressive Matrices. *Collegium Anthropologicum*, 34 (3), 1015-1026.

Fancher, R. E. (1985). *The Intelligence Men*. W. W. Norton & Company: New York.

Farah, M., & McClelland, J. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120, 339-357.

Fernyhough, C. (2008). Getting Vygotskian about theory of mind: Mediation, dialogue, and the development of social understanding. *Developmental Review*, 28, 225-262.

Finke, R. A., Pinker, S., & Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science*, 13(1), 51-78.

Forbes, A.R. (1964). An item analysis of the advanced matrices. *British Journal of Educational Psychology*, 34 (3), 223-236.

Gaffrey, M., Kleinhans, N., Haist, F., Akshoomoff, N., Campbell, A., Courchesne, E., et al. (2007). Atypical participation of visual cortex during word processing in autism: An fMRI study of semantic decision. *Neuropsychologia*, 45, 1672-1684.

García-Villamisar, D., & Della Sala, S. (2002). Dual-task performance in adults with autism. *Cognitive Neuropsychiatry*, 7, 63-74.

Gardner, H. E. (1985). *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books.

Glasgow, J., & Papadias, D. (1992). Computational imagery. *Cognitive Science*, 16(3), 355-394.

Goddard, H. H. (1922). *Human Efficiency and Levels of Intelligence*. Princeton University Press: Princeton, NJ.

Goel, A., & Bhatta, S. (2004). Use of design patterns in analogy-based design. *Advanced Engineering Informatics*, *18* (2), 85-94.

Goel, A. K., Bhatta, S., & Stroulia, E. (1997). Kritik: An early case-based design system. In Maher, M., & Pu, P., (eds.), *Issues and applications of case-based reasoning in design*. Mahwah, NJ: Erlbaum, pp. 87-132.

Goel, A. K., & Chandrasekaran, B. (1988). Integrating case-based and model-based reasoning for design problem solving. In *Proceedings of the AAAI-88 Workshop on AI in Design*. Minneapolis, MN, August, 1988.

Grandin, T. (2006). *Thinking in pictures, expanded edition: My life with autism*. New York, NY: Vintage Books.

Green, K. E., & Kleuver, R. C. (1991). Structural properties of Raven's Coloured Progressive Matrices for a sample of gifted children. *Perceptual and Motor Skills*, 1991, 72, 59-64.

Green, K. E., & Kluever, R. C. (1992). Components of item difficulty of Raven's

Matrices. *The Journal of General Psychology*, 119 (2), 189-199.

Griffith, T., Nersessian, N., & Goel, A. (2000). Function-follows-form transformations in scientific problem solving. In *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gunn, D. M., & Jarrold, C. (2004). Raven's matrices performance in Down syndrome: Evidence of unusual errors. *Research in Developmental Disabilities*, 25, 443-457.

Gunter, H., Ghaziuddin, M., & Ellis, H. (2002). Asperger syndrome: Tests of right hemisphere functioning and interhemispheric communication. *Journal of Autism and Developmental Disorders*, 32, 263-281.

Guttman, R. (1974). Genetic analysis of analytical spatial ability: Raven's Progressive Matrices. *Behavior Genetics*, 4 (3), 274-284.

Hale, C., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science*, 6, 346-359.

Halstead, H. (1943). An analysis of matrix test results. *Journal of Mental Science*, 89, 202-215.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66, 843-855.

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 36, 5-25.

Harris, G., Chabris, C., Clark, J., Urban, T., Aharon, I., Steele, S., et al. (2006). Brain activation during semantic processing in autism spectrum disorders via functional magnetic resonance imaging. *Brain and Cognition*, 61, 54-68.

Hart, J., & Gordon, B. (1992). Neural subsystems for object knowledge. *Nature*, 359, 60-64.

Hayashi, M., Kato, M., Igarashi, K., & Kashima, H. (2008). Superior fluid intelligence in children with Asperger's disorder. *Brain and Cognition*, 66, 306-310.

Heaton, P., Ludlow, A., & Roberson, D. (2008). When less is more: Poor discrimination but good colour memory in autism. *Research in Autism Spectrum Disorders*, 2, 147-156.

Hill, E. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review*, 24, 189-233.

Hitch, G. J., Brandimonte, M. A., & Walker, P. (1995). Two types of representation in visual memory: Evidence from the effects of stimulus contrast on image combination.

Memory & Cognition, 23(2), 147-154.

Hitch, G., Halliday, M., Dodd, A., & Littler, J. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, 7, 347-362.

Hitch, G., Woodin, M., & Baker, S. (1989b). Visual and phonological components of working memory in children. *Memory and Cognition*, 17, 175-185.

Hofstadter, D., ed. (1995). *Fluid concepts and creative analogies*. New York: Basic Books.

Holland, L., & Low, J. (2010). Do children with autism use inner speech and visuospatial resources for the service of executive control? Evidence from suppression in dual tasks. *British Journal of Developmental Psychology*, 28, 369-391.

Horner, J. & Nailling, K. (1980). Raven's Colored Progressive Matrices: Interpreting results through analysis of problem-type and error-type. *Clinical Aphasiology Conference*, BRK Publishers, Minneapolis, MN. 226-239.

Hunt, E. (1974). Quote the raven? Nevermore! In L.W. Gregg (Ed.), *Knowledge and Cognition* (pp. 129–158). Hillsdale, NJ: Erlbaum.

Hunt, E. (1980). Intelligence as an information processing concept. *British Journal of Psychology*, 71, 449-474.

Hurlburt, R., Happé, F., & Frith, U. (1994). Sampling the form of inner experience in three adults with Asperger syndrome. *Psychological Medicine*, 24, 385-396.

Indurkhya, B. (1998). On creation of features and change of representation. *Cognitive Studies*, *5*(2), 43-56.

Jacobs, P. I., & Vandeventer, M. (1968). Progressive matrices: An experimental, developmental, nonfactorial analysis. *Perceptual and Motor Skills*, 27, 759-766.

Jacobs, P. I., & Vandeventer, M. (1970). Information in wrong responses. *Psychological Reports*, 26, 311-315.

Jacobs, P. I., & Vandeventer, M. (1972). Evaluating the teaching of intelligence. *Educational and Psychological Measurement*, 32, 235-248.

Jarrold, C., Gilchrist, I., & Bender, A. (2005). Embedded figures detection in autism and typical development: Preliminary evidence of a double dissociation in relationships with visual search. *Developmental Science*, 8, 344-351.

Jolliffe, T., & Baron-Cohen, S. (1997). Are people with autism and Asperger syndrome faster than normal on the embedded figures test? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38, 527-534.

Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1008-1018.

Jonides, J., Reuter-Lorenz, P., Smith, E., Awh, E., Barnes, L., Drain, M., et al. (1996). Verbal and spatial working memory in humans. *Psychology of Learning and Motivation*, 35, 43–88.

Joseph, R., Keehn, B., Connolly, C., Wolfe, J., & Horowitz, T. (2009). Why is visual search superior in autism spectrum disorder? *Developmental Science*, 12, 1083-1096.

Joseph, R., Steele, S., Meyer, E., & Tager-Flusberg, H. (2005). Self-ordered pointing in children with autism: Failure to use verbal mediation in the service of working memory? *Neuropsychologia*, 43, 1400-1411.

Joseph, R., Tager-Flusberg, H., & Lord, C. (2002). Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 43, 807-821.

Just, M., Cherkassky, V., Keller, T., & Minshew, N. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity. *Brain*, 127, 1811-1821.

Kamio, Y., & Toichi, M. (2000). Dual access to semantics in autism: Is pictorial access superior to verbal access? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41, 859-867.

Kana, R., Keller, T., Cherkassky, V., Minshew, N., & Just, M. (2006). Sentence comprehension in autism: Thinking in pictures with decreased functional connectivity. *Brain*, 129, 2484-2493.

Keating, D. P. (1984). *The Emperor's New Clothes: The "new look" in intelligence research*. In R. J. Sternberg (Ed.), Advances in the Psychology of Human Intelligence, Vol. 2, pp. 1-46. Lawrence Erlbaum Associates: Hillsdale, NJ.

Keating, D. P., & Bobbitt, B. L. (1978). Individual and developmental differences in cognitive-processing components of mental ability. *Child Development*, 155-167.

Keehn, B., Brenner, L., Palmer, E., Lincoln, A., & Müller, R. (2008). Functional brain organization for visual search in ASD. *Journal of the International Neuropsychological Society*, 14, 990-1003.

Keehn, B., Brenner, L., Ramos, A., Lincoln, A., Marshall, S., & Müller, R. (2009). Brief report: Eye-movement patterns during an Embedded Figures test in children with ASD. *Journal of Autism and Developmental Disorders*, 39, 383-387.

Kirby, J. R. & Lawson, M. J. (1983). Effects of strategy training on Progressive Matrices performance. *Contemporary Educational Psychology*, 8, 127-140.

Kirchner, W. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55, 352-358.

Klin, A., Volkmar, F., Sparrow, S., Cicchetti, D., & Rourke, B. (1995). Validity and neuropsychological characterization of Asperger syndrome: Convergence with nonverbal learning disabilities syndrome. *Journal of Child Psychology and Psychiatry*, 36, 1127-1140.

Knaus, T., Silver, A., Lindgren, K., Hadjikhani, N., & Tager-Flusberg, H. (2008). fMRI activation during a language task in adolescents with ASD. *Journal of the International Neuropsychological Society*, 14, 967-979.

Kokinov, B. (1998). Analogy is like cognition: Dynamic, emergent, and context-sensitive. In Holyoak, K., Gentner, D., & Kokinov, B. (Eds.), *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences*. Sofia: New Bulgarian University Press.

Koshino, H., Carpenter, P., Minshew, N., Cherkassky, V., Keller, T., & Just, M. (2005). Functional connectivity in an fMRI working memory task in high-functioning autism. *NeuroImage*, 24, 810-821.

Koshino, H., Kana, R., Keller, T., Cherkassky, V., Minshew, N., & Just, M. (2008). fMRI investigation of working memory for faces in autism: Visual coding and underconnectivity with frontal areas. *Cerebral Cortex*, 18, 289-300.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, 4(1), 47.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York, NY: Oxford University Press.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65-100.

Leekam, S., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition*, 40, 203-218.

Leighton, J., & Sternberg, R. (2004). *The Nature of Reasoning*. Cambridge, UK: Cambridge University Press.

Leslie, A. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review*, 94, 412-426.

Leslie, A., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, 43, 225-251.

Leyton, M. (2001). *A generative theory of shape*. Berlin: Springer.

Lidstone, J., Fernyhough, C., Meins, E., & Whitehouse, A. (2009). Brief report: Inner speech impairment in children with autism is associated with greater nonverbal than verbal skills. *Journal of Autism and Developmental Disorders*, 39, 1222-1225.

Lincoln, A., Courchesne, E., Kilman, B., Elmasian, R., & Allen, M. (1988). A study of intellectual abilities in high-functioning people with autism. *Journal of Autism and Developmental Disorders*, 18, 505-524.

Lopez, B., & Leekam, S. (2003). Do children with autism fail to process information in context? *Journal of Child Psychology and Psychiatry*, 44, 285-300.

Lovett, A., Forbus, K., & Usher, J. (2010). A structure-mapping model of Raven's Progressive Matrices. *Proceedings from the 32nd Annual Conference of the Cognitive Science Society*.

Luna, B., Minshew, N., Garver, K., Lazar, N., Thulborn, K., Eddy, W., et al. (2002). Neocortical system abnormalities in autism: An fMRI study of spatial working memory. *Neurology*, 59, 834-840.

Lynn, R., Allik, J., & Irwing, P. (2004). Sex differences on three factors identified in Raven's Standard Progressive Matrices. *Intelligence*, 32 (4), 411-424.

Mackintosh, N., & Bennett, E. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence*, 33(6), 663-674.

MacLeod, C. M., Hunt, E. B., & Mathews, N. N. (1978). Individual differences in the verification of sentence-picture relationships. *Journal of Verbal Learning and Verbal Behavior*. 17, 493-507.

Manjaly, Z., Bruning, N., Neufang, S., Stephan, K., Brieber, S., Marshall, J., et al. (2007). Neurophysiological correlates of relatively enhanced local visual search in autistic adolescents. *NeuroImage*, 35, 283-291.

Markman, A. (1999). *Knowledge Representation*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

Matzen LE, Benz ZO, Dixon KR, Posey J, Kroger JK, & Speed AE. (2010). Recreating Raven's: software for systematically generating large numbers of Raven-like matrix problems with normed properties. *Behavior Research Methods*, 42 (2), 525-541.

McGreggor, K., Kunda, M., & Goel, A. K. (2010). A fractal approach towards visual analogy. In *Proceedings of the First International Conference on Computational Creativity*. Lisbon, Portugal.

McGregor, E., Whiten, A., & Blackburn, P. (1998a). Teaching theory of mind by highlighting intention and illustrating thoughts: A comparison of their effectiveness with 3-year-olds and autistic individuals. *British Journal of Developmental Psychology*, 16, 281–300.

McGregor, E., Whiten, A., & Blackburn, P. (1998b). Transfer of the picture-in-the-head analogy to natural contexts to aid false belief understanding in autism. *Autism*, 2, 367-387.

Miller, F. M., & Raven, J. C. (1939). The influence of positional factors on the choice of answers to perceptual intelligence tests. *British Journal of Medical Psychology*, 18, 35-39.

Minshew, N., & Goldstein, G. (2001). The pattern of intact and impaired memory functions in autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42, 1095-1101.

Minshew, N., Goldstein, G., & Siegel, D. (1997). Neuropsychologic functioning in autism: Profile of a complex information processing disorder. *Journal of the International Neuropsychological Society*, 3, 303-316.

Minshew, N., Goldstein, G., Muenz, L., & Payton, J. (1992). Neuropsychological functioning in nonmentally retarded autistic individuals. *Journal of Clinical and Experimental Neuropsychology*, 14, 749-61.

Minshew, N., Luna, B., & Sweeney, J. (1999). Oculomotor evidence for neocortical systems but not cerebellar dysfunction in autism. *Neurology*, 52, 917-922.

Minsky, M. (1967). *Computation: Finite and Infinite Machines*. Englewood Cliffs, New Jersey: Prentice Hall.

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, 55 (2), 195-215.

Mitrushina, M., Boone, K., & Razani, J. (2005). *Handbook of normative data for neuropsychological assessment*. New York: Oxford University Press.

Morris, R., Rowe, A., Fox, N., Feigenbaum, J., Miotto, E., & Howlin, P. (1999). Spatial working memory in Asperger's syndrome and in patients with focal frontal and temporal lobe lesions. *Brain & Cognition*, 41, 9-26.

Mottron, L. (2004). Matching strategies in cognitive research with individuals with high-functioning autism: Current practices, instrument biases, and recommendations. *Journal of Autism and Developmental Disorders*, 34, 19-27.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, 36, 27-43.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, 12, 252-284.

Narayanan, N. H., Glasgow, J., & Chandrasekaran, B. (1995). *Diagrammatic Reasoning: Cognitive and Computational Perspectives*. Cambridge, MA: MIT Press.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review*, 86, 214–255.

Nersessian, N. (2008). *Creating Scientific Concepts*. Cambridge, MA: MIT Press.

Newell, A. (1994). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. (1972) *Human Problem Solving*. Upper Saddle River, New Jersey: Prentice Hall.

Newell, A., & Simon, H. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19 (3), 113-126.

O'Connor, N., & Hermelin, B. (1967). Auditory and visual memory in autistic and normal children. *Journal of Intellectual Disability Research*, 11, 126-131.

Ojha, A., & Indurkhya, B. (2009). Perceptual vs. conceptual similarities and creation of new features in visual metaphor. In Kokinov, B., Holyoak, K., & Gentner, D. (Eds.), *New frontiers in analogy research*. Sofia: New Bulgarian University Press.

O'Riordan, M. (2000). Superior modulation of activation levels of stimulus representations does not underlie superior discrimination in autism. *Cognition*, 77, 81-96.

O'Riordan, M. (2004). Superior visual search in adults with autism. *Autism*, 8, 229-248.

O'Riordan, M., & Plaisted, K. (2001). Enhanced discrimination in autism. *Quarterly Journal of Experimental Psychology A*, 54, 961-979.

O'Riordan, M., Plaisted, K., Driver, J., & Baron-Cohen, S. (2001). Superior visual search in autism. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 719-730.

Ozonoff, S., & Strayer, D. (2001). Further evidence of intact working memory in autism. *Journal of Autism and Developmental Disorders*, 31, 257-263.

Palmer, S.E. (1978). Structural aspects of visual similarity. *Memory and Cognition*, 6(2), 91-97.

Parsons, S., & Mitchell, P. (1999). What children with autism understand about thoughts and thought bubbles. *Autism*, 3(1), 17-38.

Perner, J., & Leekam, S. (2008). The curious incident of the photo that was accused of being false: Issues of domain specificity in development, autism, and brain imaging. *Quarterly Journal of Experimental Psychology*, 61, 76-89.

Peterson, C. C. (2003). Drawing insight from pictures: The development of concepts of false drawing and false belief in children with deafness, normal hearing, and autism. *Child Development*, 73(5), 1442-1459.

Peterson, C. C., & Siegal, M. (1998). Changing focus on the representational mind: Concepts of false photos, false drawings and false beliefs in deaf, autistic and normal children. *British Journal of Developmental Psychology*, 16, 301-320.

Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998). Enhanced visual search for a conjunctive target in autism: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 39, 777-783.

Prabhakaran, V., Smith, J. A. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997). Neural Substrates of Fluid Reasoning: An fMRI Study of Neocortical Activation during Performance of the Raven's Progressive Matrices Test. *Cognitive Psychology*, 33(1), 43-63.

Quill, K. (1997). Instructional considerations for young children with autism: The rationale for visually cued instruction. *Journal of Autism and Developmental Disorders*, 27, 697-714.

Rasmussen, D., & Eliasmith, C. (2011). A neural model of rule generation in inductive reasoning. *Topics in Cognitive Science*, 3(1), 140-153.

Raven, J., Raven, J. C., & Court, J.H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment.

Reed, T., & Peterson, C. (1990). A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective taking. *Journal of Autism and Developmental Disorders*, 20(4), 555-567.

Ring, H., Baron-Cohen, S., Wheelwright, S., Williams, S., Brammer, M., Andrew, C., et al. (1999). Cerebral correlates of preserved cognitive skills in autism: A functional MRI study of embedded figures task performance. *Brain*, 122, 1305-1315.

Ropar, D., & Mitchell, P. (2001). Susceptibility to illusions and performance on visuospatial tasks in individuals with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42, 539-549.

Ropar, D., & Mitchell, P. (2002). Shape constancy in autism: The role of prior knowledge and perspective cues. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 43, 647-653.

Russell, J. (1997). *Autism as an executive disorder*. New York: Oxford University Press.

Russell, J., & Hill, E. L. (2001). Action-monitoring and intention reporting in children with autism. *Journal of Child Psychology and Psychiatry*, 42(03), 317-328.

Russell, J., Jarrold, C., & Henry, L. (1996). Working memory in children with autism and with moderate learning difficulties. *Journal of Child Psychology and Psychiatry*, 37, 673-686.

Russell, J., Jarrold, C., & Hood, B. (1999). Two intact executive capacities in children with autism: Implications for the core executive dysfunctions in the disorder. *Journal of Autism and Developmental Disorders*, 29, 103-112.

Sahyoun, C., Soulières, I., Belliveau, J., Mottron, L., & Mody, M. (2009). Cognitive differences in pictorial reasoning between high-functioning autism and Asperger's syndrome. *Journal of Autism and Developmental Disorders*, 39 (7), 1014-1023.

Schooler, J.W., & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 22(1), 36-71.

Schooler, J.W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, 122(2), 166-183.

Schwartz, D. L., & Black, J. B. (1996). Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology*, 30(2), 154-219.

Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., Osanai, H., et al. (2010). Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and Psychopathology*, 22, 353-360.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger Syndrome. *Science*, 325, 883-885.

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701-703.

Siegel, D., Minshew, N., & Goldstein, G. (1996). Wechsler IQ profiles in diagnosis of high-functioning autism. *Journal of Autism and Developmental Disorders*, 26, 389-406.

Slotnick, S., Thompson, W., & Kosslyn, S. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*, 15 (10), 1570-1583.

Smilansky, J. (1984). Problem solving and the quality of invention: An empirical investigation. *Journal of Educational Psychology*, 76 (3), 377-386.

Smith, E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes.

Science, 283, 1657-1661.

Smyth, M., & Pelky, P. (1992). Short-term retention of spatial information. *British Journal of Psychology*, 83, 359-374.

Smyth, M., & Scholey, K. (1992). Determining spatial span: The role of movement time and articulation rate. *Quarterly Journal of Experimental Psychology A*, 45, 479–501.

Smyth, M., & Scholey, K. (1996). The relationship between articulation time and memory performance in verbal and visuospatial tasks. *British Journal of Psychology*, 87, 179-191.

Smyth, M., Pearson, N., & Pendleton, L. (1988). Movement and working memory: Patterns and positions in space. *Quarterly Journal of Experimental Psychology A*, 40, 497-514.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. *Advances in the Psychology of Human Intelligence*: Volume 2, 47-103.

Soulières, I., Dawson, M., Samson, F., Barbeau, E.B., Sahyoun, C.P., Strangman, G.E., Zeffiro, T.A., & Mottron, L. (2009). Enhanced visual processing contributes to matrix reasoning in autism. *Human Brain Mapping*, 30 (12), 4082-4107.

Soulières, I., Zeffiro, T.A., Girard, M. L., & Mottron, L. (2011). Enhanced mental image mapping in autism. *Neuropsychologia*, 49 (5), 848-857.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18, 587-592.

Spearman, C. E. (1923). *The Nature of "Intelligence" and the Principles of Cognition*. Macmillan: London.

Stafford, B. (2001). *Visual analogy: Consciousness as the art of connecting.* Cambridge, MA: MIT Press.

Steele, S., Minshew, N., Luna, B., & Sweeney, J. (2007). Spatial working memory deficits in autism. *Journal of Autism and Developmental Disorders*, 37, 605-612.

Stone, B. & Day, M. C. (1981). A developmental study of the processes underlying solution of figural matrices. *Child Development*, 52, 359-362.

Strauss, E., Sherman, E., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

Swettenham, J., Baron-Cohen, S., Gomez, J., & Walsh, S. (1996). What's inside someone's head? Conceiving of the mind as a camera helps children with autism acquire an alternative to a theory of mind. *Cognitive Neuropsychiatry*, 1, 73-88.

Thissen, D. M. (1976). Information in wrong responses to the Raven's Progressive Matrices. *Journal of Educational Measurement*, 13 (3), 201-214.

Thomas, N. (2010). Mental imagery. *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/spr2013/entries/mental-imagery/>

Thompson-Schill, S. (2003). Neuroimaging studies of semantic memory: Inferring "how" from "where". *Neuropsychologia*, 41, 280-292.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327-352.

van der Ven, A.H.G.S., & Ellis, J.L. (2000). A Rasch analysis of Raven's standard progressive matrices. *Personality and Individual Differences*, 29 (1), 45-64.

Van Herwegen, J., Farran, E., & Annaz, D. (2011). Item and error analysis on Raven's Coloured Progressive Matrices in Williams Syndrome. *Research in Developmental Disabilities*, 32, 93-99.

Vandenbroucke, M., Scholte, H., van Engeland, H., Lamme, V., & Kemner, C. (2008). A neural substrate for atypical low-level visual processing in autism spectrum disorder. *Brain*, 131, 1013-1024.

Vejleskov (1968). An analysis of Raven matrices responses in fifth grade children. *Scandinavian Journal of Psychology*, 9, 177-186.

Verguts, T., & De Boeck, P. (2002). The induction of solution rules in Raven's Progressive Matrices Test. *European Journal of Cognitive Psychology*, 14(4), 521-547.

Verté, S., Geurts, H., Roeyers, H., Oosterlaan, J., & Sergeant, J. (2005). Executive functioning in children with autism and Tourette syndrome. *Development and Psychopathology*, 17, 415-445.

Verté, S., Geurts, H., Roeyers, H., Oosterlaan, J., & Sergeant, J. (2006). Executive functioning in children with an Autism Spectrum Disorder: Can we differentiate within the spectrum? *Journal of Autism and Developmental Disorders*, 36, 351-372.

Vigneau, F., & Bors, D. A. (2008). The quest for item types based on information processing: An analysis of Raven's Advanced Progressive Matrices, with a consideration of gender differences. *Intelligence*, 36(6), 702-710.

Villardita, C. (1985). Raven's colored Progressive Matrices and intellectual impairment in patients with focal brain damage. *Cortex*, 21(4), 627-634.

Vodegel Matzen, L. B. L., van der Molen, M. W., & Dudink, A. C. M. (1994). Error analysis of Raven test performance. *Personality and Individual Differences*, 16 (3), 433-445.

Walker, P., Hitch, G. J., Dewhurst, S. A., Whiteley, H. E., & Brandimonte, M. A. (1997). The representation of nonstructural information in visual memory: Evidence from image combination. *Memory & Cognition*, 25(4), 484-491.

Wallace, G., Silvers, J., Martin, A., & Kenworthy, L. (2009). Brief report: Further evidence for inner speech deficits in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39, 1735-1739.

Weatherick, N. E. (1966). The responses of normal adult subjects to the Matrices test. *British Journal of Psychology*, 57 (3-4), 297-300.

Wellman, H., Baron-Cohen, S., Caswell, R., Gomez, J., Swettenham, J., Toye, E., et al. (2002). Thought-bubbles help children with autism acquire an alternative to a theory of mind. *Autism*, 6, 343-363.

Whitehouse, A., Maybery, M., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry*, 47, 857-865.

Williams, D., & Jarrold, C. (2010). Brief report: Predicting inner speech use amongst children with autism spectrum disorder (ASD): The roles of verbal ability and cognitive profile. *Journal of Autism and Developmental Disorders*, 40, 907-913.

Williams, D., Goldstein, G., & Minshew, N. (2006). The profile of memory function in children with autism. *Neuropsychology*, 20, 21-29.

Williams, D., Goldstein, G., Carpenter, P., & Minshew, N. (2005). Verbal and spatial working memory in autism. *Journal of Autism and Developmental Disorders*, 35, 747-756.

Williams, D., Happe, F., & Jarrold, C. (2008). Intact inner speech use in autism spectrum disorder: Evidence from a short-term memory task. *Journal of Child Psychology and Psychiatry*, 49, 51-58.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128.

Yaner, P., & Goel, A. (2006). Visual analogy: Viewing retrieval and mapping as constraint satisfaction. *Journal of Applied Intelligence*, *25*(1), 91-105.

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin*, 124, 283-307.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of Cognitive Neuroscience*, 20(1), 1-19.

Zaidel, E., Zaidel, D. W., & Sperry, R. W. (1981). Left and right intelligence: Case studies of Raven's Progressive Matrices following brain bisection and hemidecortication. *Cortex*, *17*(2), 167-86.

Zitová, B., & Flusser, J. (2003). Image registration methods: A survey. *Image and Vision Computing*, 21, 977-1000.

CURRICULUM VITAE

Education

Georgia Institute of Technology, Atlanta, GA: Ph.D., Computer Science, to be awarded May 03, 2013.

Massachusetts Institute of Technology, Cambridge, MA: B. S., Mathematics with Computer Science, 2006.

Research Positions

Research Assistant, School of Interactive Computing, Georgia Tech: Design & Intelligence Lab, 2007 – 2012. BORG Lab, 2006 – 2007.

Consultant and Research Scientist (part-time), Aerotonomy, Inc.: Technical proposal development, 2004 – 2011. Image-aided sensing and guidance, 2007 – 2011.

Intern, Environmental Sciences Division, Oak Ridge National Laboratory: Ecosystem Simulation Science Group, 2004 – 2006.

Research Assistant, Computer Science and Artificial Intelligence Laboratory, MIT: Robust Robotics Group, 2004 – 2006.

Intern, Accurate Automation Corporation: UAV development, 2003.

Awards and Honors

Foley Scholar: GVU Center, Georgia Tech, 2010.

Anita Borg Memorial Scholar: Google, 2010.

PhD Student Research Excellence Commendation: School of Interactive Computing, Georgia Tech, 2008.

Graduate Research Fellowship Program: National Science Foundation, 2009 – 2011.

National Defense Science and Engineering Graduate Fellowship: Department of Defense, 2006 – 2009.

Undergraduate Research Opportunity Program Scholarship: Massachusetts Institute of Technology, 2006.

Higher Education Research Experience Fellowship: Oak Ridge National Lab, 2006.

Global Change Education Program Fellowship: Department of Energy, 2004 – 2006.

Research Experience for Undergraduates: National Science Foundation, 2003.

Teaching

Co-Instructor, School of Interactive Computing, Georgia Tech: Computational Creativity / Knowledge-Based Modeling and Design (undergraduate / graduate), 2013.

Teaching Assistant, School of Interactive Computing, Georgia Tech: Introduction to Cognitive Science (graduate), 2011.

Guest Lecturer, School of Interactive Computing, Georgia Tech: Knowledge-Based AI (graduate), 2010 – 2011. Intro to Cognitive Science (undergraduate), 2009 – 2012.

Instructor, Let's Get Ready, MIT: SAT Verbal Course (high school level), 2003 – 2004. Tutoring in Writing and AP Calculus (high school level), 2004.

Publications

Refereed Journal Papers

Kunda, M., McGreggor, K., & Goel, A. K. (2013). "A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations." *Cognitive Systems Research*, 22-23, pp. 47-66.

McGreggor, K., Kunda, M., & Goel, A. K. (2013). "Fractals and ravens." In preparation for submission to *Journal of Artificial Intelligence Research*.

Kunda, M., & Goel, A. K. (2011). "Thinking in Pictures as a cognitive account of autism." *Journal of Autism and Developmental Disorders*, 41 (9), pp. 1157-1177.

Neidhoefer, J., Gibson, C., Kunda, M., and Johnson, E. (2007). "Determinism and autonomy in the National Airspace System (NAS)." *Journal of Aerospace Computing, Information, and Communication*, 4 (11), pp. 1037-1045.

Refereed Conference Papers

Kunda, M., McGreggor, K., & Goel, A. K. (2012). "Reasoning on the Raven's Advanced Progressive Matrices test with iconic visual representations." *34th Annual Conference of the Cognitive Science Society*, Portland, OR, pp. 1828-1833.

Kunda, M., McGreggor, K., & Goel, A. K. (2011). "Two visual strategies for solving the Raven's Progressive Matrices intelligence test." *AAAI National Conference*, San

Francisco, CA.

McGreggor, K., Kunda, M., and Goel, A. K. (2011). "Fractal analogies: Preliminary results from the Raven's test of intelligence." *2nd International Conference on Computational Creativity*, Mexico City, Mexico.

Kunda, M., McGreggor, K., & Goel, A. K. (2010). "Taking a look (literally!) at the Raven's intelligence test: Two visual solution strategies." *32nd Annual Conference of the Cognitive Science Society*, Portland, OR, pp. 1691-1696.

McGreggor, K., Kunda, M., and Goel, A. K. (2010). "A fractal approach towards visual analogy." *1st International Conference on Computational Creativity*, Lisbon, Portugal.

Kunda, M., and Goel, A. K. (2008). "How Thinking in Pictures can explain many characteristic behaviors of autism." *7th IEEE International Conference on Development and Learning*, Monterey, CA, pp. 304-309.

Kunda, M., and Goel, A. K. (2008). "What can pictorial representations reveal about the cognitive characteristics of autism?" *5th International Conference on the Theory and Application of Diagrams*, Herrsching, Germany. *LNCS*, 5223, pp. 103-117.

Kunda, M., and Goel, A. K. (2008). "Thinking in Pictures: A fresh look at cognition in autism." *30th Annual Conference of the Cognitive Science Society*, Washington, DC, pp. 321-326.

Neidhoefer, J., Gibson, C., Kunda, M., and Johnson, E. (2007). "Determinism in autonomy for applications in the National Airspace System (NAS)." *American Institute for Aeronautics and Astronautics: Infotech@Aerospace*, Rohnert Park, CA.

Marquez, J., Cummings, M., Roy, N., Kunda, M., and Newman, D. (2005). "Collaborative human-computer decision support for planetary surface traversal." *American Institute for Aeronautics and Astronautics: Infotech@Aerospace*, Arlington, VA.

Refereed Workshops and Symposia

McGreggor, K., Kunda, M., and Goel, A. K. (2010). "A fractal analogy approach to Raven's test of intelligence." *AAAI Workshop on Visual Representations and Reasoning*, Atlanta, GA.

Kunda, M., McGreggor, K., and Goel, A. K. (2009). "Addressing the Raven's Progressive Matrices test of general intelligence." *AAAI Fall Symposium on Multi Representational Architectures for Human Level Intelligence*, Washington, D.C.

Conference Presentations

Kunda, M., Soulières, I., Mottron, L., and Goel, A. K. (2011). "Comparing patterns of errors on the Raven's Progressive Matrices test: Strategy differences among typically developing individuals, individuals with autism, and computational models." Poster presentation, *International Meeting for Autism Research (IMFAR)*, San Diego, CA.

Kunda, M., McGreggor, K., and Goel, A. K. (2010). "Can the Raven's Progressive Matrices intelligence test be solved by thinking in pictures?" Oral presentation, *International Meeting for Autism Research (IMFAR)*, Philadelphia, PA.

Kunda, M., and Goel, A. (2009). "Evidence for Thinking in Pictures as a cognitive account of autism." Poster presentation, *International Meeting for Autism Research (IMFAR)*, Chicago, IL.

Kunda, M., Marland, G., Canella, L., Schlamadinger, B., and Bird, N. (2005). "Impact of albedo change on carbon sequestration strategies." Oral presentation, *Third USDA Symposium on Greenhouse Gases & Carbon Sequestration in Agriculture and Forestry*, Baltimore, MD.

Kunda, M., Bird, N., Canella, L., Marland, G., and Schlamadinger, B. (2005). "Carbon management, earth surface albedo, and biomass fuels." Poster presentation, *14th European Biomass Conference and Exhibition: Biomass for Energy, Industry, and Climate Protection*, Paris, France.

Invited Talks

March 14, 2013: "Identifying the Visual Cognitive Phenotype in Autism." Meeting of the Atlanta Autism Consortium Research SIG. Atlanta, GA.

May 21, 2012: "Visual thinking in autism: Computational approaches for studying cognition." Instructional session for Georgia Leadership & Education in Neurodevelopmental and Related Disabilities (GaLEND) program. Atlanta, GA.

April 12, 2012: "Visual thinking and intelligence," as part of *An Illustrated Conversation: Visual Thinking in Autism, Art, and Creativity*. Neuro-Humanities Entanglement Conference and Neuro-Salon, Georgia Tech, Atlanta, GA.

September 15, 2010: "Can the Raven's Progressive Matrices test be solved visually?" Laboratoire de Neurosciences Cognitives des Troubles Envahissants du Developpement, University of Montreal. Montreal, Canada.

September 13, 2010: "Can the Raven's Progressive Matrices test be solved by thinking in pictures?" Yale Early Social Cognition Lab, Yale University. New Haven, CT.

September 25, 2009: "Visual thinking in autism: Design of behavioral experiments using the dual-task paradigm." Meeting of the Atlanta Autism Consortium. Atlanta, GA.

Professional Activities

Co-chair: AAAI Workshop on Visual Representations and Reasoning, 2010.

Organizer: Cognitive Science discussion group, College of Computing, Georgia Tech, 2009 – 2011.

Member

Association for the Advancement of Artificial Intelligence

Cognitive Science Society

International Society for Autism Research

Atlanta Autism Consortium

Reviewer

Cognitive Science Journal

Journal of Autism and Developmental Disorders

ACM Conference on Creativity and Cognition

International Conference on Computational Creativity

International Conference on the Theory and Application of Diagrams

National Conference of the Association for the Advancement of Artificial Intelligence (AAAI)

Annual Conference of the Cognitive Science Society