Thinking in Pictures as a Cognitive Account of Autism

Maithilee Kunda and Ashok K. Goel
Design & Intelligence Lab, School of Interactive Computing
Georgia Institute of Technology
85 Fifth Street NW, Atlanta, GA 30332, USA

Abstract

We analyze the hypothesis that some individuals on the autism spectrum may use visual mental representations and processes to perform certain tasks that typically developing individuals perform verbally. We present a framework for interpreting empirical evidence related to this “Thinking in Pictures” hypothesis and then provide comprehensive reviews of data from several different cognitive tasks, including 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. We also discuss the relationships between the Thinking in Pictures hypothesis and other cognitive theories of autism including Mindblindness, Executive Dysfunction, Weak Central Coherence, and Enhanced Perceptual Functioning.

Key words: Autism; cognition; information processing; mental imagery; verbal representations; visual representations; visual reasoning.

Corresponding author: mkunda@gatech.edu
Introduction

Numerous individuals on the autism spectrum have posited that they tend to use visual mental representations instead of verbal ones (e.g. Hurlburt et al. 1994). In her well-known autobiographical book Thinking in Pictures, for example, Temple Grandin (2006) describes how her visual thinking style benefits her work in engineering design but also creates difficulties in understanding abstract concepts. Among cognitive theorists in the autism research community, this “Thinking in Pictures” idea seems to have received limited focused and sustained consideration. This relative lack of attention perhaps is due not only to the introspective nature of the above accounts but also because the hypothesis seems ill-defined.

The purpose of this article is to refine one formulation of the Thinking in Pictures (TiP) hypothesis about cognition in autism and examine existing empirical evidence relating to this hypothesis, expanding on our previous work (Kunda & Goel, 2008). Our formulation of this hypothesis has two main parts:

Assumption: Typically developing (TD) individuals are, in general, able to 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).

For the remainder of this paper, this (and only this) is what we mean by the TiP hypothesis.

Although for a time cognitive science debated whether visual mental representations even existed, the weight of evidence now seems to indicate that they do; they are usually described as being analogical (i.e. having some structural correspondence to what they represent) and closely tied to perceptual mechanisms (Kosslyn et al. 2006). In contrast, verbal mental representations are often described as being propositional (Pylyshyn, 2002). However, our reading of the literature on cognition in autism indicates that, like the literature on cognitive science in general, different interpretations of visual and verbal representations are often used in practice, usually on a task-by-task basis. In this paper, we will pin down precise meanings for visual and verbal in our discussions of individual tasks.

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.

A tendency to exhibit a V < NV profile is exactly what one might expect from an individual who thinks in pictures, with one important caveat: the standard tasks used to measure verbal and nonverbal abilities in IQ tests have been selected through extensive study of
neurotypical development and performance, and there is no guarantee that a test measuring a particular cognitive ability in TD individuals measures the same cognitive ability in individuals with autism, as there may be multiple different strategies that can be used to solve the same task. (We return to this point in the following section.)

Other behavioral data from autism suggesting an over-reliance on visual representations span many different cognitive and task domains (e.g. Heaton et al. 2008; Joseph et al. 2005; Whitehouse et al. 2006). On the neurobiological side, Mottron et al. (2006) reported that, across a variety of fMRI studies, individuals with autism tend to show increased brain activation in posterior, visual-perceptual brain regions and decreased activation in frontal brain regions often used for verbal processing.

For the remainder of this article, we first describe what sorts of predictions about behavior can be made using the TiP hypothesis, and then we give several examples of relevant empirical data from behavior and neurobiology. We conclude by discussing the relationship of the TiP hypothesis with several existing cognitive theories of autism, including Mindblindness, Executive Dysfunction, Weak Central Coherence, and Enhanced Perceptual Functioning.

**Effects of Thinking in Pictures on Behavior**

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. However, there are two different ways to classify tasks as *visual* or *verbal*: how a task can be solved (i.e. what sorts of mental representations and inferences are sufficient, but not necessary) and how tasks are typically solved (i.e. what sorts of mental representations and inferences do TD individuals generally use).

![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).](image)

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 $A \cap B$.

The light grey and dark grey shaded regions ($T_A$ and $T_B$) 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 $T_A$ and $T_B$, by definition, and it is tempting to treat these classifications as the final answer on whether a task is visual or verbal. However, for a typically verbal task in $T_B$, if that task happens to also be solvable visually (i.e. lies within $A \cap B$), it is possible that an individual disinclined to use verbal representations can use a compensatory visual strategy to successfully solve that task.

By making these distinctions, the performance of an individual on a given task (e.g. level of success) can be evaluated independently of their strategy selection (e.g. visual or verbal). Keeping this in mind, we now use the TiP hypothesis to make general predictions about the behavior of individuals with autism on three different types of tasks, as shown in Table 1.

**Table 1: General Behavioral Predictions for Autism from the TiP Hypothesis**

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Task type in Fig. 1</th>
<th>Task type description</th>
<th>TD strategy</th>
<th>TD performance</th>
<th>AU strategy</th>
<th>AU performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>exclusively in $B$</td>
<td>tasks that can only be done verbally</td>
<td>verbal</td>
<td>successful</td>
<td>visual</td>
<td>impaired</td>
</tr>
<tr>
<td>P2</td>
<td>in $T_A$</td>
<td>tasks typically done visually</td>
<td>visual</td>
<td>successful</td>
<td>visual</td>
<td>successful</td>
</tr>
<tr>
<td>P3</td>
<td>in $T_B \cap A$</td>
<td>tasks typically done verbally that can be done visually</td>
<td>verbal</td>
<td>successful</td>
<td>visual</td>
<td>successful</td>
</tr>
</tbody>
</table>

*Note. AU = Individuals with autism; TD = Typically developing individuals*

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 deficit accounts of autism. However, this TiP prediction is consistent with general evidence for verbal impairments in autism (*DSM-IV-TR*, 2000), though more precise relationships remain to be determined.

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 visual 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, it should be possible to design experiments that illuminate whether the
underlying representational strategy used by an individual is visual or verbal. Furthermore,
experiments testing this third TiP prediction will provide the surest means for distinguishing TiP
from other cognitive theories of autism, as (insofar as we have seen) no other cognitive account
explicitly posits visual/verbal representational differences.

In the following two sections, we review empirical data related to the third and second
TiP predictions, respectively. In particular, we look at:

- **Tasks typically done verbally that can be done visually:** (1) the *n*-back task, (2) serial
  recall, (3) dual task studies, (4) Raven’s Progressive Matrices, (5) semantic processing,
  and (6) false belief tasks.
- **Tasks typically done visually:** (7) visual search, (8) spatial recall, and (9) visual recall.

**TiP Prediction #3: Tasks Typically Done Verbally That Can Be Done Visually**

*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 the one 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 the one presented two steps back), and so on. Stimuli can vary as to their
content and presentation, such as 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,\(^1\) 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

---

\(^1\) See Appendix A.
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.

**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.\(^2\) 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

\(^2\) See Appendix B.
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 development of the TiP hypothesis.

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.

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), our 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.

*Raven’s Progressive Matrices*

Raven’s Progressive Matrices (RPM) is a standardized intelligence test that consists of problems resembling geometric analogies, in which a matrix of figures is presented with one entry missing and the correct missing entry must be selected from among a set of answer choices (Raven,
Although the test is only supposed to measure eductive ability, or the ability to extract and understand information from a complex situation, factor analyses have shown the RPM to be a good measure of Spearman’s g, and it is thus widely used as a general intelligence test (Raven et al. 2003). Using the RPM as a measure of general intelligence, though it consists only of problems in a single format, stands in contrast to using broader tests like the Wechsler scales, which contain subtests across several different domains.

Whereas the RPM scores of TD individuals are usually correlated with 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 have shown a similar pattern (Hayashi et al. 2008). One possible explanation for these results is that the RPM, in which both questions and answers are presented visually, might be amenable to solution using visual strategies. The Wechsler scales, on the other hand, are heavily verbal, and while individuals with autism often show good performance on certain subtests like Digit Span or Block Design, their performance on the other subtests can be much lower. In contrast, if TD individuals draw from a combination of visual and verbal strategies on both types of tests, then we would expect to see correlations in their scores between the two paradigms.

One widely cited computational modeling study proposes that TD individuals use a propositional, rule-based strategy to solve RPM problems (Carpenter et al. 1990). However, Hunt (1974) proposed the existence of two qualitatively distinct strategies: one visual, using perceptual operations like visual continuity and superposition, and one analytic, using formal operations based on logical rules. Several behavioral studies of TD individuals point to the possibility of distinct visual and verbal strategies being effective on the RPM (DeShon et al. 1995; Lynn et al. 2004; van der Ven & Ellis, 2000), and we are currently conducting computational studies to investigate whether visual-only algorithms can, in fact, successfully solve the RPM (Kunda et al. 2010a, 2010b).

Soulières et al. (2009) recently found, using fMRI, that individuals with autism had lower brain activation in verbal prefrontal and parietal areas and higher activation in visual occipital areas than TD controls while solving the RPM, consistent with the notion of a visual-strategy-bias in autism. On a related but non-RPM set of matrix reasoning tasks, Sahyoun et al. (2009) found evidence, through measures of response latency, of the autism group having a bias towards visuospatial mediation, whereas TD individuals and individuals with Asperger’s were able to use verbal mediation.

**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, under the TiP hypothesis, we predict 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.

**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). We investigate one contrasting view, namely 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.

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.

**TiP Prediction #2: Tasks Typically Done Visually**

Unlike the tasks discussed previously, which are typically done verbally but might be amenable to visual strategies, we now discuss empirical data related to the second TiP prediction, on tasks that are done visually both by TD individuals and by individuals with autism. As described earlier, while the TiP hypothesis provides a good explanation of why individuals with autism might show intact performance on these types of tasks, it does not provide a straightforward explanation of superior performance, and it is inconsistent with evidence of performance decrements. In this section, we discuss data that fall into both of these categories.

**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. Moreover, faster search performance in autism often grows more pronounced with more difficult search tasks, e.g. for conjunctive vs. feature search, etc.

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.

Spatial Recall

Serial spatial recall tasks are a part of many standardized intelligence 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 type of 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 types of tasks, and we found no study of spatial recall on which the autism group showed superior performance.

Given that serial recall for items or objects appears to be unimpaired in autism, as discussed earlier, there appears to be a dissociation between how well individuals with autism can 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 the 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 actually recruit verbal working memory; correlations between spatial span and speech rate have been found in TD individuals,

---

3 See Appendix C.
4 See Appendix D.
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).

*Visual Recall*

One paradigm for tests of visual recall 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 that helps to identify perceptual or motor impairments that could confound results.

Many studies of these types of tasks have revealed decreased performance in individuals with autism. 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 the Block Design subtest 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 are what 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 among experimental groups, instead of just looking at recall scores alone, and found no group 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 for individuals with autism are mixed at best. It is unclear how these results might be accounted for by the TiP hypothesis, and more detailed investigations are needed of what specific cognitive processes both individuals with autism and TD individuals recruit for these tasks.

*Discussion*

We have presented detailed reviews of empirical data on individuals with autism from several different task domains. For each task, we have attempted to give an objective assessment of whether the data are consistent with our formulation of a Thinking in Pictures (TiP) hypothesis about cognition in autism. As expected, the results of this analysis 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

---

5 See Appendix E.
search. Finally, there are task domains whose data seem to contradict the TiP hypothesis, which are: (8) spatial recall, and (9) visual recall.

Of course, there are many experimental task paradigms that we have not addressed or have only briefly touched upon, for instance free recall, cued recall, visual or verbal recognition, executive functioning, etc. However, the main point that we wish to convey is that, across several task domains, there is a significant amount of evidence that is highly consistent with the TiP hypothesis. This finding is even more interesting given that most of the studies we reviewed did not explicitly use a visual/verbal hypothesis in the design or execution of their experiments.

In the remainder of this section, we discuss the relationship of the TiP hypothesis to existing cognitive accounts of autism. We close with our final thoughts on TiP and important questions for further exploration.

**Theories of Cognition in Autism**

Several existing cognitive theories of autism aim to explain various aspects of autistic behavior. We briefly discuss three of these theories here—Executive Dysfunction, Weak Central Coherence, and Enhanced Perceptual Functioning—focusing on how the TiP hypothesis relates to them, on points of congruence as well as divergence. (Possible relationships of TiP to a fourth theory, Mindblindness, were covered in the section describing false belief tasks.)

The Executive Dysfunction (ED) theory posits that autism is characterized by impairments in a set of higher-level cognitive skills that underlie independent, goal-oriented behavior, such as planning, set-shifting, and generativity (Russell, 1997). We argue that 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 pieces of data, evaluating a potential link between executive functioning in autism and the TiP hypothesis will require a close 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). Much of the 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 at least these tasks, the TiP hypothesis can provide 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 our 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 raise the question of whether TiP can (or 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 representational modalities—visual, auditory, etc.—within each of two distinct types of processing—local vs. global, or perceptual vs. high-level. TiP, on the other hand, 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 more explicitly account for autistic superiorities on certain visual tasks, whereas TiP does not currently propose a concrete mechanism for this pattern of performance, though several possibilities, such as increased visual expertise, remain to be explored. It is plausible that these accounts are linked, both developmentally and cognitively, and the precise relationship between the TiP hypothesis and these theories is the subject of some of our current work.

**Final Thoughts**

Our results lead us to propose two main conclusions. First, given the existence of considerable evidence in line with the TiP hypothesis, the idea that certain individuals with autism may “think visually” should be taken seriously as a cognitive model and receive more focused and sustained attention in behavioral and neurobiological experiments. Second, and more generally, the interpretation of behavioral data from individuals with autism (or, indeed, with any form of atypical cognition) should be performed with care. Assumptions governing the relations between cognition and behavior that hold for TD individuals may not hold universally, and we
have presented several instances in which visual and verbal strategies seem to be recruited differently across experimental groups, despite often producing superficially similar behavior.

If a subset of individuals on the autism spectrum does have a bias towards using visual mental representations, then several important questions remain to be answered about the TiP hypothesis. How might this subset of individuals be identified, and how could experimental subgroups be appropriately defined to account for cognitive differences within the autism spectrum? Would these individuals display a $V < NV$ profile, and would such a profile be a necessary and/or sufficient marker of their cognitive style? What, if anything, might the TiP hypothesis tell us about individuals on the spectrum who showed $V = NV$ or $V > NV$ cognitive profiles? At present, we do not have answers to these questions.

Other important avenues for further inquiry include (1) the accuracy and interpretation of measures of visual and verbal IQ when potential differences in task strategies are taken into account, (2) the distinction, if any, between visual and spatial processing under the TiP account, as well as relationships with other types of perceptual processing, (3) how biases away from using verbal representations and towards using visual representations might be causally linked, and (4) what role TiP might play in neurobiological and developmental accounts of autism.

**Author Note**

Author information: Maithilee Kunda and Ashok K. Goel, Design & Intelligence Lab, School of Interactive Computing, Georgia Institute of Technology.

This research was supported in part by a grant (IIS Award 0534266) on “Multimodal Case-Based Reasoning in Modeling and Design” from the National Science Foundation, by a National Defense Science and Engineering Graduate Fellowship from the Office of Naval Research, and by a Graduate Research Fellowship from the National Science Foundation. The authors would like to thank A. Rozga for her careful reading of this manuscript, K. McGregor and E. Schumacher for many discussions on this topic, and G. Abowd, R. Arriaga, and the rest of the Autism Research Group at the Georgia Tech/Emory Health Systems Institute for their encouragement and support of this work.

Correspondence concerning this article should be addressed to Maithilee Kunda, School of Interactive Computing, Georgia Institute of Technology, 85 Fifth Street NW, Atlanta, GA 30332. Email: mkunda@gatech.edu
References


## Appendix A

### Review of Results from Published Studies of the n-Back Task in Autism

<table>
<thead>
<tr>
<th>Reference</th>
<th>N</th>
<th>Age</th>
<th>VIQ</th>
<th>PIQ</th>
<th>FSIQ</th>
<th>Task details</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koshino et al. 2005</td>
<td>14</td>
<td>25.7</td>
<td>102.6</td>
<td>—</td>
<td>100.1</td>
<td>$n = 0, 1, 2$; visually presented letters.</td>
<td>NSGD on accuracy or RT. Significant group differences in patterns of brain activation.</td>
</tr>
<tr>
<td>Koshino et al. 2008</td>
<td>11</td>
<td>24.5</td>
<td>106.1</td>
<td>102.1</td>
<td>104.5</td>
<td>$n = 0, 1, 2$; grayscale face pictures.</td>
<td>NSGD on accuracy or RT. Significant group differences in patterns of brain activation.</td>
</tr>
<tr>
<td>Ozonoff &amp; Strayer, 2001</td>
<td>25</td>
<td>12.94</td>
<td>94.6</td>
<td>99.3</td>
<td>96.3</td>
<td>$n = 1, 2$; visually presented colored shapes.</td>
<td>NSGD on accuracy or RT. RT tended to be correlated with VIQ.</td>
</tr>
<tr>
<td>Williams et al. 2005</td>
<td>31</td>
<td>26.58</td>
<td>111.10</td>
<td>103.13</td>
<td>108.65</td>
<td>$n = 0, 1, 2$; visually presented letters.</td>
<td>NSGD on accuracy or RT.</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>11.75</td>
<td>112.50</td>
<td>106.38</td>
<td>109.67</td>
<td>$n = 0, 1, 2$; visually presented letters.</td>
<td>NSGD on accuracy or RT.</td>
</tr>
</tbody>
</table>

**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.
## Appendix B

**Review of Results from Published Studies of Serial Recall Tasks in Autism**

<table>
<thead>
<tr>
<th>Reference</th>
<th>N</th>
<th>Age</th>
<th>VIQ</th>
<th>PIQ</th>
<th>FSIQ</th>
<th>Items</th>
<th>Presentation</th>
<th>Response</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ameli et al. 1988</td>
<td>16</td>
<td>22.7</td>
<td>81 (16)</td>
<td>90.6</td>
<td>83 (14)</td>
<td>digits</td>
<td>auditory</td>
<td>verbal</td>
<td>NSGD in digit span.</td>
</tr>
<tr>
<td>Bennetto et al. 1996</td>
<td>19</td>
<td>15.95</td>
<td>82.32</td>
<td>98.11</td>
<td>88.89</td>
<td>digits</td>
<td>auditory</td>
<td>verbal</td>
<td>NSGD in digit span.</td>
</tr>
<tr>
<td>Joseph et al. 2005</td>
<td>24</td>
<td>8.9</td>
<td>94 (19)</td>
<td>99 (20)</td>
<td>96 (18)</td>
<td>words</td>
<td>auditory</td>
<td>pointing</td>
<td>NSGD in correct responses.</td>
</tr>
<tr>
<td>Minshew et al. 1992</td>
<td>15</td>
<td>21.13</td>
<td>98.53</td>
<td>92.87</td>
<td>95.73</td>
<td>digits</td>
<td>auditory</td>
<td>verbal</td>
<td>NSGD in digit span.</td>
</tr>
<tr>
<td>Minshew et al. 1997</td>
<td>33</td>
<td>20.91</td>
<td>102.48</td>
<td>97.45</td>
<td>100.09</td>
<td>digits</td>
<td>auditory</td>
<td>verbal</td>
<td>NSGD in digit span.</td>
</tr>
<tr>
<td>Minshew &amp; Goldstein, 2001</td>
<td>52</td>
<td>22.33</td>
<td>94.96</td>
<td>91.52</td>
<td>92.88</td>
<td>letters</td>
<td>auditory</td>
<td>verbal</td>
<td>NSGD in correct sequences.</td>
</tr>
<tr>
<td>O'Connor &amp; Hermelin, 1967</td>
<td>12</td>
<td>11.8</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>words</td>
<td>auditory</td>
<td>verbal</td>
<td>NSGD in number recalled. Greater position and recency effects in AU than controls.</td>
</tr>
<tr>
<td>Ozonoff &amp; Strayer, 2001</td>
<td>25</td>
<td>12.94</td>
<td>94.6</td>
<td>99.3</td>
<td>96.3</td>
<td>colored boxes</td>
<td>visual</td>
<td>self-ordered pointing</td>
<td>NSGD in number of perseverative errors, which tended to correlation with VIQ.</td>
</tr>
<tr>
<td>Study</td>
<td>Auditory Measures</td>
<td>Visual Measures</td>
<td>Verbal Measures</td>
<td>Visual Similarity</td>
<td>Phonological Similarity</td>
<td>Auditory Similarity</td>
<td>Age</td>
<td>IQ</td>
<td>Note</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>------------------------</td>
<td>---------------------</td>
<td>-----</td>
<td>------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Russell et al. 1996</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>12.38 (2.95)</td>
<td>---</td>
<td>MLD had lower span than other groups. AU had greater nonverbal WLE than verbal WLE and than MLD group.</td>
</tr>
<tr>
<td>Whitehouse et al. 2006</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>11 (---)</td>
<td>---</td>
<td>NSGD in correct responses. AU had smaller WLE than controls and in silent vs. label condition.</td>
</tr>
<tr>
<td>Williams et al. 2006</td>
<td>106.42 (15.97)</td>
<td>100.55 (14.19)</td>
<td>103.82 (14.29)</td>
<td>digits, letters</td>
<td>auditory</td>
<td>verbal</td>
<td>---</td>
<td>---</td>
<td>NSGD in score.</td>
</tr>
<tr>
<td>Williams et al. 2008</td>
<td>77.16 (15.25)</td>
<td>76.84 (20.27)</td>
<td>74.84 (15.99)</td>
<td>pictures</td>
<td>visual</td>
<td>verbal</td>
<td>---</td>
<td>---</td>
<td>NSGD in span. Subjects with high VMA had greater PSE than VSE; vice versa for low VMA.</td>
</tr>
</tbody>
</table>

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.
# Appendix C

Review of Results from Published Studies of Target/Distracter Visual Search Tasks in Autism

<table>
<thead>
<tr>
<th>Reference</th>
<th>N</th>
<th>Age</th>
<th>CPM</th>
<th>Search</th>
<th>Target</th>
<th>Distracters</th>
<th>Accuracy</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jarrold et al. 2005</td>
<td>18&lt;sup&gt;b&lt;/sup&gt;</td>
<td>12.42 (1.98)</td>
<td>21.56 (7.45)</td>
<td>F</td>
<td>red jumping clown</td>
<td>green skinny clown, red fat clown</td>
<td>---</td>
<td>AU faster.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>red jumping clown</td>
<td>green jumping clown, red skinny clown</td>
<td>---</td>
<td>AU faster.</td>
</tr>
<tr>
<td>Kechn et al. 2008</td>
<td>9&lt;sup&gt;a&lt;/sup&gt;</td>
<td>15.1 (2.5)</td>
<td>---</td>
<td>F</td>
<td>upright T</td>
<td>right-pointing T</td>
<td>NSGD.</td>
<td>NSGD.</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td>upright T</td>
<td>right, left, bottom-pointing T</td>
<td>NSGD.</td>
<td>NSGD.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>red X</td>
<td>red T, green X</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>red X</td>
<td>red letters and colored X</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td>O'Riordan, 2000</td>
<td>11</td>
<td>9.2 (0.8)</td>
<td>27 (4)</td>
<td>C</td>
<td>red X</td>
<td>red T, green X</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>red X</td>
<td>red T, green X</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>green T</td>
<td>red T, green X</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>red X, green T</td>
<td>red T, green X</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>P, Q</td>
<td>NSGD.</td>
<td>NSGD.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>P, Q</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td>O'Riordan, 2004</td>
<td>10</td>
<td>22.0 (3.6)</td>
<td>---</td>
<td>F</td>
<td>ellipse</td>
<td>circle</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>red X</td>
<td>green X, red C</td>
<td>NSGD.</td>
<td>NSGD.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>red F</td>
<td>pink F, red E</td>
<td>AU more accurate.</td>
<td>NSGD.</td>
</tr>
<tr>
<td>O'Riordan &amp; Plaisted, 2001</td>
<td>15</td>
<td>9.2 (1.1)</td>
<td>28 (3)</td>
<td>C</td>
<td>vertical red bar</td>
<td>vertical green bar, horizontal red bar</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>vertical red bar</td>
<td>vertical green line, horizontal green line</td>
<td>NSGD.</td>
<td>NSGD.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>vertical red bar</td>
<td>vertical green bar, horizontal green line</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>vertical red bar</td>
<td>vertical green bar, horizontal red bar</td>
<td>NSGD.</td>
<td>NSGD.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>vertical red bar</td>
<td>vertical green bar, horizontal red line</td>
<td>NSGD.</td>
<td>AU faster.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>red X</td>
<td>green X, red C</td>
<td>NSGD.</td>
<td>NSGD.</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Age Mean (SD)</td>
<td>CPM Mean (SD)</td>
<td>Feature Search</td>
<td>Search Mode</td>
<td>Group Differences</td>
<td>Reaction Time</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>O’Riordan et al. 2001</td>
<td>12</td>
<td>8.4 (0.9)</td>
<td>26 (4)</td>
<td>C red X, pink X, red C</td>
<td>NSGD</td>
<td>AU faster.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>8.8 (10.0)</td>
<td>F red S, green X, red S</td>
<td>NSGD</td>
<td>NSGD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>8.8 (10.0)</td>
<td>C red X, red T, green X</td>
<td>NSGD</td>
<td>AU faster.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>28 (3)</td>
<td>F tilted line, vertical lines</td>
<td>NSGD</td>
<td>NSGD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F vertical line, tilted lines</td>
<td>NSGD</td>
<td>AU faster.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plaisted et al. 1998</td>
<td>8</td>
<td>8.8 (1.2)</td>
<td>---</td>
<td>F red S, red T, green X</td>
<td>NSGD</td>
<td>NSGD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C red X, green X, red T</td>
<td>AU more accurate</td>
<td>AU faster.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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.

- Subject group diagnosed as ASD.
- AU had higher chronological age.
- Autism group had higher block design scores.
### Appendix D

**Review of Results from Published Studies of Spatial Recall Tasks in Autism**

<table>
<thead>
<tr>
<th>Reference</th>
<th>N</th>
<th>Age</th>
<th>VIQ</th>
<th>PIQ</th>
<th>FSIQ</th>
<th>Task</th>
<th>Details</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caron et al., 2004</td>
<td>16</td>
<td>17.6 (6.3)</td>
<td>102.2 (21.2)</td>
<td>112.3 (12.9)</td>
<td>107.7 (13.1)</td>
<td>Maze route learning</td>
<td>learn choice points in an actual maze</td>
<td>NSGD in accuracy or speed of going through maze.</td>
</tr>
<tr>
<td>Edgin &amp; Pennington, 2005</td>
<td>24</td>
<td>11.46 (2.32)</td>
<td>104.40 (20.24)</td>
<td>---</td>
<td>---</td>
<td>CANTAB Spatial Working Memory</td>
<td>touch one of identical boxes to find tokens (self-ordered pointing)</td>
<td>NSGD in errors or strategy. Both correlated with age and block design score in both groups.</td>
</tr>
<tr>
<td>Luna et al., 2002</td>
<td>11</td>
<td>32.3 (9.3)</td>
<td>106.7 (14.9)</td>
<td>96.5 (10.5)</td>
<td>102.7 (12.1)</td>
<td>Oculomotor Delayed Response</td>
<td>shift gaze to previous stimulus location</td>
<td>AU impaired in saccade accuracy compared to controls. Not impaired on baseline saccade task.</td>
</tr>
<tr>
<td>Minshew &amp; Goldstein, 2001</td>
<td>52</td>
<td>22.33 (9.59)</td>
<td>94.96 (17.56)</td>
<td>91.52 (12.95)</td>
<td>92.88 (15.06)</td>
<td>Maze recall</td>
<td>learn choice points in hidden maze</td>
<td>AU impaired on complex mazes but not on simpler maze.</td>
</tr>
<tr>
<td>Minshew et al., 1992</td>
<td>15</td>
<td>21.13 (8.02)</td>
<td>98.53 (21.63)</td>
<td>92.87 (10.72)</td>
<td>95.73 (13.61)</td>
<td>Maze recall</td>
<td>learn choice points in hidden maze</td>
<td>NSGD.</td>
</tr>
<tr>
<td>Minshew et al., 1997</td>
<td>33</td>
<td>20.91 (9.69)</td>
<td>102.48 (16.35)</td>
<td>97.45 (11.19)</td>
<td>100.09 (12.96)</td>
<td>Maze recall</td>
<td>learn choice points in hidden maze</td>
<td>NSGD.</td>
</tr>
<tr>
<td>Minshew et al., 1999</td>
<td>26</td>
<td>20.2 (8.5)</td>
<td>98.5 (16.9)</td>
<td>90.1 (12.7)</td>
<td>94.0 (14.1)</td>
<td>Oculomotor Delayed Response</td>
<td>shift gaze to previous stimulus location</td>
<td>AU impaired in saccade accuracy compared to controls. Not impaired on baseline saccade task.</td>
</tr>
<tr>
<td>Morris et al., 1999</td>
<td>15</td>
<td>29.5</td>
<td>99</td>
<td>100.1</td>
<td>---</td>
<td>Executive Golf Task</td>
<td>touch one of identical golf holes to find putts (self-ordered pointing)</td>
<td>AS impaired on within and between search errors.</td>
</tr>
<tr>
<td>Steele et al., 2007</td>
<td>29</td>
<td>14.83 (5.47)</td>
<td>107.52 (13.02)</td>
<td>106.21 (11.82)</td>
<td>107.76 (10.99)</td>
<td>CANTAB Spatial Working Memory</td>
<td>touch one of identical boxes to find tokens (self-ordered pointing)</td>
<td>AU showed greater errors and lower strategy score, both correlated with PIQ but not VIQ. TD scores not correlated with any IQ.</td>
</tr>
<tr>
<td>Verté et al., 2005</td>
<td>61</td>
<td>9.1 (1.9)</td>
<td>---</td>
<td>---</td>
<td>99.2 (17.1)</td>
<td>Corsi block tapping</td>
<td>forward tapping recall of blocks</td>
<td>AU impaired.</td>
</tr>
<tr>
<td>Verté et al., 2006</td>
<td>50</td>
<td>8.7 (1.9)</td>
<td>93.1 (18.0)</td>
<td>104.0 (17.8)</td>
<td>98.2 (17.3)</td>
<td>Corsi block tapping</td>
<td>forward tapping recall of blocks</td>
<td>AU impaired.</td>
</tr>
<tr>
<td>37&lt;sup&gt;b&lt;/sup&gt;</td>
<td>8.5 (2.1)</td>
<td>105.6 (20.0)</td>
<td>104.0 (15.9)</td>
<td>105.2 (16.3)</td>
<td>Corsi block tapping</td>
<td>forward tapping recall of blocks</td>
<td>AS impaired.</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>25&lt;sup&gt;cf&lt;/sup&gt;</td>
<td>8.5 (1.4)</td>
<td>93.3 (14.7)</td>
<td>105.8 (19.2)</td>
<td>98.3 (14.4)</td>
<td>Corsi block tapping</td>
<td>forward tapping recall of blocks</td>
<td>NSGD.</td>
<td></td>
</tr>
</tbody>
</table>

William 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. |

William 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.

<sup>a</sup> Subject group diagnosed as Autism/Asperger.’s.  
<sup>b</sup> Subject group diagnosed as Asperger’s.  
<sup>c</sup> Subject group diagnosed as PDD-NOS.  
<sup>d</sup> Controls grouped as TD, Tourette’s, and autism/Tourette’s (comorbid).  
<sup>e</sup> Subject group had lower FSIQ.  
<sup>f</sup> Subject group had lower FSIQ and VIQ.
Appendix E

Review of Results from Published Studies of Visual Recall Tasks in Autism

<table>
<thead>
<tr>
<th>Reference</th>
<th>N</th>
<th>Age (mean (SD))</th>
<th>VIQ (mean (SD))</th>
<th>PIQ (mean (SD))</th>
<th>FSIQ (mean (SD))</th>
<th>Task</th>
<th>Details</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambery et al. 2006</td>
<td>27</td>
<td>37.6 (14.6)</td>
<td>106.1 (15.7)</td>
<td>103.7 (19.2)</td>
<td>--</td>
<td>Doors and People Test</td>
<td>draw shapes from memory, also recognize colored photos of doors</td>
<td>AU impaired on combined measure of visual recall and recognition.</td>
</tr>
<tr>
<td>Ameli et al. 1988</td>
<td>16</td>
<td>22.7 (4.9)</td>
<td>81 (16)</td>
<td>90.6 (13.5)</td>
<td>83 (14)</td>
<td>BVRT</td>
<td>draw series of abstract figures from memory</td>
<td>AU impaired. Highly correlated with PIQ in AU but not in controls.</td>
</tr>
<tr>
<td>Gunter et al. 2002</td>
<td>8</td>
<td>16.25 (10.19)</td>
<td>111.38 (15.22)</td>
<td>96.13 (13.13)</td>
<td>--</td>
<td>Rey-Osterrieth</td>
<td>draw abstract figure from memory</td>
<td>NSGD in copy/recall scores or score differences.</td>
</tr>
<tr>
<td>Minshew &amp; Goldstein, 2001</td>
<td>52</td>
<td>22.33 (9.59)</td>
<td>94.96 (17.56)</td>
<td>91.52 (12.95)</td>
<td>92.88 (15.06)</td>
<td>Rey-Osterrieth</td>
<td>draw abstract figure from memory</td>
<td>AU impaired in number of elements correctly drawn for immediate and delayed recall.</td>
</tr>
<tr>
<td>Minshew et al. 1997</td>
<td>15</td>
<td>21.13 (8.02)</td>
<td>98.53 (21.63)</td>
<td>92.87 (10.72)</td>
<td>95.73 (13.61)</td>
<td>Rey-Osterrieth</td>
<td>draw abstract figure from memory</td>
<td>AU impaired in number of elements correctly drawn for delayed recall.</td>
</tr>
<tr>
<td>Minshew et al. 1992</td>
<td>33</td>
<td>20.91 (9.69)</td>
<td>102.48 (16.35)</td>
<td>97.45 (11.19)</td>
<td>100.09 (12.96)</td>
<td>Rey-Osterrieth</td>
<td>draw abstract figure from memory</td>
<td>AU impaired in number of elements correctly drawn for delayed recall.</td>
</tr>
<tr>
<td>Ropar &amp; Mitchell, 2001</td>
<td>11</td>
<td>11.8 (2.0)</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>Rey-Osterrieth</td>
<td>draw abstract figure from memory</td>
<td>NSGD in copy/recall score differences or in strategy.</td>
</tr>
<tr>
<td>Vérté et al. 2005</td>
<td>61</td>
<td>9.1 (1.9)</td>
<td>---</td>
<td>---</td>
<td>99.2 (17.1)</td>
<td>BVRT</td>
<td>draw series of abstract figures from memory</td>
<td>AU impaired.</td>
</tr>
<tr>
<td>Vérté et al. 2006</td>
<td>50</td>
<td>8.7 (1.9)</td>
<td>93.1 (18.0)</td>
<td>104.0 (17.8)</td>
<td>98.2 (17.3)</td>
<td>BVRT</td>
<td>draw series of abstract figures from memory</td>
<td>AU impaired.</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>8.5 (2.1)</td>
<td>105.6 (20.0)</td>
<td>104.0 (15.9)</td>
<td>105.2 (16.3)</td>
<td>BVRT</td>
<td>draw series of abstract figures from memory</td>
<td>AU impaired.</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Age (mean, SD)</td>
<td>IQ (mean, SD)</td>
<td>Test</td>
<td>Task</td>
<td>Result</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>----------------</td>
<td>--------------</td>
<td>------</td>
<td>------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams et al. 2006</td>
<td>38</td>
<td>11.68 (2.46)</td>
<td>106.42 (15.97)</td>
<td>WRAML Design Memory</td>
<td>Draw geometric shapes from memory</td>
<td>AU impaired.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.