J Autism Dev Disord
DOI 10.1007/s10803-014-2188-5

ORIGINAL PAPER

# Visuo-Spatial Performance in Autism: A Meta-analysis

**Anne Muth · Johannes Hönekopp · Christine M. Falter**



**Abstract** Visuo-spatial skills are believed to be enhanced in autism spectrum disorders (ASDs). This meta-analysis tests the current state of evidence for Figure Disembedding, Block Design, Mental Rotation and Navon tasks in ASD and neurotypicals. Block Design ($d = 0.32$) and Figure Disembedding ($d = 0.26$) showed superior performance for ASD with large heterogeneity that is unaccounted for. No clear differences were found for Mental Rotation. ASD samples showed a stronger local processing preference for Navon tasks ($d = 0.35$); less clear evidence for performance differences of a similar magnitude emerged. We discuss the meta-analysis results together with other findings relating to visuo-spatial processing and three cognitive theories of ASD: Weak Central Coherence, Enhanced Perceptual Functioning and Extreme Male Brain theory.

**Keywords** Autism · Visuo-spatial · Figure Disembedding · Mental Rotation · Block Design · Navon Figures

A. Muth (✉)
School of Psychology, Bangor University, Adeilad Brigantia, Penrallt Road, Bangor LL57 2AS, UK
e-mail: A.Muth@Bangor.ac.uk

J. Hönekopp
Department of Psychology, Northumbria University, Newcastle upon Tyne, UK
e-mail: johannes.honekopp@unn.ac.uk

C. M. Falter
Department of Psychiatry, University Hospital Cologne, Cologne, Germany
e-mail: Christine.falter@cantab.net

## Introduction

The term autism spectrum disorder (ASD) is an umbrella term that encompasses Autistic disorder, Asperger's syndrome (AS), High-Functioning Autism (HFA) and pervasive developmental disorder not otherwise specified (PDD-NOS). As ASDs are often accompanied by learning and language impairments, it is most common for people with HFA to participate in research studies because their high-functioning level means that they are intellectually capable of meeting the needs of cognitive studies. Disorders of the autism spectrum are characterized by impairments in social interaction and communication, and restricted, repetitive and stereotyped patterns of behavior, interests and activities (Levy et al. 2009; APA 2000).

The question whether individuals with ASD have superior visuo-spatial skills compared to neurotypicals has inspired research for several decades and goes back to analyses of intelligence test performance of autistic children compared to typically developing (TD) control children that were matched for their overall IQ (Shah and Frith 1983). Furthermore, the assumption that visuo-spatial skills are indeed superior (Mitchell and Ropar 2004) has been used as a rationale for many studies and theories, yet findings reported in the literature are discrepant; while some claim performance in a certain task to be robustly superior in autism (e.g., Lee et al. 2007; Shah and Frith 1993), others disagree (e.g., Bölte et al. 2007; Manjaly et al. 2007; White and Saldana 2011).

Typical visuo-spatial tasks are Figure Disembedding, Block Design, Mental Rotation and Navon Figures. Originally developed by Witkin et al. (1971), the Figure Disembedding Test (also known as Embedded Figures Test, EFT) consists of cards depicting images made up of lines with embedded geometrical shapes, such as triangles, rectangles, or circles. A target shape is presented, which the participant is asked to locate as quickly as possible in the image. Adapted versions for preschool children and children are also available.

Block Design is a performance IQ subtest that is included in both the Wechsler Adult Intelligence Test (WAIS; Wechsler 1981) and the Wechsler Intelligence Test for children (WISC; Wechsler 1974). The participant receives a set number of blocks that have white, red, and both white and red sides. The task is to use these blocks to recreate a two-dimensional pattern that the participant is presented with on a card.

In the classic Mental Rotation task by Shepard and Metzler (1971) participants are presented with a two-dimensional picture representing a three-dimensional geometric object and have to decide whether it matches a rotated target or not. This task can also be used with objects and figures that are two-dimensional, such as geometric figures, drawings of objects or letters. This task requires spatial working memory because one has to hold the representation in mind and manipulate it.

Navon Figures (1977) were designed based on evidence that people do not process a picture all at once but extract information gradually, e.g., from coarse to fine detail, or as Navon puts it: as if "zoomed in on" (p. 354). To test this idea, he devised a task in which participants are shown large letters made up of the same or different smaller letters. The task allows identifying participants' processing style by testing whether the global level (large letter) or the local level (small letters) is attentionally salient for them. It has been found that TD people are faster and more accurate when responding to global forms (Wang et al. 2007).

As yet there is no single theory that can explain all of the phenomena seen in ASD, rather there are cognitive theories attempting to account for behavioral patterns within autism. Three influential theories have been evaluated frequently along with tests of visuo-spatial ability, namely the Weak Central Coherence (WCC; Happé and Frith 2006), Enhanced Perceptual Functioning (EPF; Mottron and Burack 2001; Mottron et al. 2006) and the Extreme Male Brain (EMB; Baron-Cohen 1999) theory, all of which will be summarised here and evaluated in the discussion.

The WCC account explains the perceptual profile of individuals with ASD by a tendency towards local processing and a weakness of global processing of information, based on the observation that children with ASD showed superior performance on Block Design and Figure Disembedding (Shah and Frith 1983, 1993). In short, this theory posits that individuals with ASD have an enhanced processing capacity for detail as opposed to a global processing style. Most TD individuals process perceptual stimuli with a global bias, looking at the whole, whereas people with autism appear to be biased towards a local processing style, which comes at the cost of often missing the gist. However, a local processing style is not detrimental per se, but rather depends on the requirements of a task at hand, i.e. this theory predicts an advantage for individuals with autism on tasks that require detail-focused processing, such as the Figure Disembedding and Block Design tasks. Support for this theory was provided in a review paper by Happé and Frith (2006) that includes studies in which individuals with ASD showed faster performance on the Block Design test compared to matched TD control groups. We will look at both these tasks and examine how well these predications hold up. Furthermore, we will also examine Mental Rotation, another visuo-spatial task but one that does not benefit from focusing on details but rather on comparing two complete figures. Thus, WCC theory would predict that people with autism would be disadvantaged in this task and perform worse than TD individuals. For Navon Figures, we should expect in the framework of WCC theory that any processing advantage of the global over the local stimulus dimension is stronger in TD people than in ASD affected people.

The model of Enhanced Perceptual Functioning (Mottron and Burack 2001; Mottron et al. 2006) partly agrees but also extends WCC. The authors suggest a complex relationship between local and global processing styles in ASD, with a preference, rather than a default option, for local processing. This preference is believed to enable people with ASD to switch to a global style if required by a task, in contrast to the global "deficit" proposed by WCC. Mottron et al. (2006) argue that this flexibility in processing results in superior performance for local tasks, and equivalent performance for global tasks compared to neurotypicals. Thus, for Navon Figures performance in ASD this model predicts a local preference but an equal ability when it comes to global processing requirements. According to Mottron et al. (2006) EPF predicts a "constant pattern of enhanced performance" (p. 30) in tasks in which local and global processing conflicts, such as Block Design and Figure Disembedding. More specifically, if the conflict is decreased, as is the case in pre-segmented Block Design tasks, the performance is predicted to be equivalent to that of TD individuals. We presume that EPF would predict no group difference for rotational aspects of Mental Rotation based on the assumption that Mental Rotation would be considered a task requiring global processing.

The Extreme Male Brain theory (Baron-Cohen 1999) is based on the idea that autism appears to be a testosterone driven "male condition". This is reflected in the fact that more males are affected by the condition than females (Levy et al. 2009). Further, digit ratio 2D:4D (i.e. index finger length divided by ring finger length) is believed to be a marker of prenatal testosterone exposure (Manning et al. 2013) and it is more male-typical in individuals with ASD than in TD individuals (Hönekopp 2012). Baron-Cohen (1999) lists evidence from research on tasks on which females are generally superior to males, and indeed these tasks are impaired in autism. It was argued that women appear to perform better than men on language tasks as well as tests of social judgment, and that they show more empathy and cooperation. Men, on the other hand, are supposed to be better at mathematical reasoning, "systemizing", the Embedded Figures task and Mental Rotation (Baron-Cohen 1999). In accordance with Baron-Cohen's argument, a meta-analysis examining spatial skills as reflected in the Mental Rotation task found an advantage for males (Linn and Petersen 1985). Yet more recent evidence indicates that these apparent sex differences are much smaller than previously thought and often non-significant (Hines 2004). For instance, a more recent fMRI study using Mental Rotation did not find any sex differences on accuracy or reaction times, but rather different analytic strategies (Hugdahl et al. 2006).

EMB theory predicts better performance for autistic individuals in Figure Disembedding, given the acclaimed sex difference in this task. Concerning Mental Rotation, predictions are not that straightforward however. In Mental Rotation tasks, accuracy and response time are commonly used as outcome measures. Typically, response time increases and accuracy decreases as the angle difference between picture and target increases. It is therefore sensible to regress, for each participant, time (and accuracy) on angle. The regression slope for time then represents rotation speed (with steeper slopes indicating slower speeds) whereas the intercepts reflects non-rotational aspects of the task, such as figure comparison, working memory, and decision making times (Falter et al. 2008). Falter and colleagues (Falter et al. 2008) have conducted a study with TD and ASD participants showing that overall the ASD group outperformed the control group in Mental Rotation. However, a closer look at task aspects revealed that the sex difference that is typically found in slopes (Falter et al. 2006) did not hold up for the ASD group that only showed superior performance in intercepts. Thus, although sex differences were found in slopes and EMB theory should therefore predict performance differences between ASD and TD participants for slopes, we already know that it is rather the non-rotational aspects of the task (intercepts), if any, in which participants with ASD excel.

Similarly, deriving predictions from EMB theory for Block Design and Figure Disembedding is not straightforward. Block Design is designed to be equally valid for females and males; hence sex differences should not exist. It is difficult to say what EMB theory predicts for Block Design performance; on the one hand the task is a visuo-spatial task for which male superiority is generally predicted by Baron-Cohen (1999), however, on the other hand it was designed to be impartial to sex differences. We therefore assume that no performance difference between ASD and TD is predicted for Block Design in the framework of EMB theory. Similarly, Navon Figure tests are visuo-spatial in nature, but given that we are not aware of any systematic investigations of whether or not Navon Figure performance is linked to sex or testosterone, no performance difference between ASD and TD should be predicted by EMB. We would like to note that although performance differences (or their absence) have repercussions for these theories; our aim here is not a general evaluation of these theories, which are broader in scope than the visual-spatial task performance that we focus on.

Studies on ASD performance in visuo-spatial skills tasks show a mixture of significant and non-significant results with inconsistent effect directions (see Tables 1, 2, 3). Here, we present the first meta-analysis of these data to shed light on the question whether visuo-spatial processing in ASD is superior, inferior, or equal to that in TD individuals. In particular, we examine the current state of knowledge by looking at the classic tests of visuo-spatial skills: Figure Disembedding, Block Design, Mental Rotation and Navon Figures. We then discuss implications of these findings for WCC, EPF and EMB. Finally, we look at other research findings in an attempt to understand our results better.

## Methods

### Retrieval of Studies

We used the databases 'Web of Science', 'PubMed' and 'PsycInfo' by searching the literature using the following terms: ('autism' OR 'ASD') AND ('visuo-spatial' OR 'visuospatial' OR 'figure disembedding' OR 'embedd* figures' OR 'block design' OR 'mental rotation'). For the effect size analysis of Navon Figures we used the following search terms: ('autism OR 'ASD') AND ('Navon figure*' OR 'Navon task' OR 'hierarchisation' OR 'hierarchical letter' OR 'hierarchical stimuli' OR 'compound letter' OR 'free choice' OR 'divided attention' OR 'selected attention'). This procedure revealed 76 samples in line with our topic, made up of 24 samples for Block Design, 35 for Figure Disembedding, three for Mental Rotation slope analysis, eight for Mental Rotation intercept analysis and seven, four and nine for Navon Figures free choice, divided and selected attention respectively. Inclusion criteria were studies focusing on children, adolescents, and adults with ASD compared with at least one TD control group that participated in a task of visuo-spatial skills. Further, in order to complete the picture of performance in visuo-spatial skills other tests were included, such as mazes,

**Table 1** Primary studies comparing Block Design performance (M ± SD) in people with ASD (first line in each entry) and TD controls (second line)

| Source | N | Age | % male | IQ | Block Design | *d* |
|---|---|---|---|---|---|---|
| Pellicano (2006) | 40 | 5.6 ± 0.9 | 88 | 113.6[a] | 132.5 ± 17.25[g] | 1.75 |
| | 40 | 5.5 ± 1.0 | 78 | 112.5[a] | 105.25 ± 13.75[g] | |
| Ishida (2009) | 9 | 12.3 ± 2.0 | 100 | 98.2[b] | 13.6 ± 3.8[g] | 1.20 |
| | 9 | 11.6 ± 1.0 | 100 | 98.4[b] | 8.8 ± 4.2[g] | |
| Soulieres (2011) | 23 | 21.0 ± 5.8 | 87 | 101.9[b] | 13.7 ± 3.05[g,h] | 1.15 |
| | 14 | 19.4 ± 3.8 | 86 | 103.0[b] | 10.4 ± 2.5[g,h] | |
| Caron (2006) | 16 | 21.1 ± 6.3 | ? | 104.2[a] | 13.7 ± 3.5[g] | 1.12 |
| | 10 | 18.6 ± 3.5 | ? | 96.5[a] | 10.1 ± 2.7[g] | |
| Pring (2010) | 9 | 26.9 ± 6.3 | ? | 12.1[c] | 77 ± 44.0[i] | 1.09 |
| | 9 | 12.0 ± 2.0 | ? | 12.0[c] | 124 ± 42[i] | |
| Morgan (2003) | 19 | ≈4.5 ± 0.6 | ≈90 | 95.1[a] | 105.89 ± 20.45[g] | 0.76 |
| | 21 | 4.6 ± 0.4 | 76 | 104.6[a] | 92.43 ± 14.94[g] | |
| Shah (1993) | 10 | 18.5 ± 3.0 | 80 | 71.0[a] | 1.84 ± 0.22[j] | 0.63 |
| | 16 | 10.9 ± 0.3 | 94 | Matched[d] | 1.97 ± 0.20[j] | |
| Shah (1993) | 10 | 18.6 ± 1.7 | 90 | 96.7[a] | 1.65 ± 0.19[j] | 0.51 |
| | 17 | 16.0 ± 0.6 | 88 | 100.6[a] | 1.74 ± 0.17[j] | |
| Smalley (1990) | 9 | 20.3 | 100 | 95.1[b] | 11.8 ± 3.9[g] | 0.48 |
| | 9 | 24.3 | 100 | 100.3[b] | 10.1 ± 3.2[g] | |
| Planche (2011) | 15 | 8.5 | 93 | 109.1[a] | 12.5 ± 2.4 | 0.46 |
| | 15 | 9.0 | 80 | 107.1[a] | 11.5 ± 1.6 | |
| van Lang (2006) | 20 | ≈14.9 ± 2.2 | ≈77 | 57.0[a] | 5.4 ± 2.1[k] | 0.41 |
| | 17 | ≈14.5 ± 2.7 | ≈62 | 57.2[a] | 5.1 ± 2.1[k] | |
| | 20 | ≈14.9 ± 2.2 | ≈77 | 57.0[a] | 25.1 ± 12.6[i] | |
| | 17 | ≈14.5 ± 2.7 | ≈62 | 57.2[a] | 33.8 ± 12.7[i] | |
| Bölte (2011) | 46 | 14.1 ± 2.9 | 63 | 99.4[a] | 11.6 ± 4.9 | 0.26 |
| | 58 | 14.6 ± 4.8 | 40 | 103.3[a] | 10.6 ± 3.3 | |
| Bölte (2007) | 15 | 25.8 ± 7.7 | 100 | 100.1[a] | 10.9 ± 2.9[g] | 0.25 |
| | 15 | 27.0 ± 6.7 | 100 | 105.9[a] | 10.3 ± 1.8[g] | |
| Ropar (2001) | 11 | 11.1 ± 2.0 | 82 | 9.9[e] | 9.64 ± 3.93[g] | 0.16 |
| | 37 | 9.8 ± 1.4 | 51 | 9.7[e] | 9.05 ± 3.57[g] | |
| Ozonoff (1991) | 23 | 12.1 ± 3.2 | 91 | 98.4[a] | 10 ± 4.0[g] | 0.15 |
| | 20 | 12.4 ± 3.0 | 90 | 97.0[a] | 9.4 ± 4.07[g] | |
| Williams (2006) | 56 | 11.4 ± 2.2 | 82 | 102.1[a] | 11.75 ± 3.62[g] | 0.15 |
| | 56 | 11.8 ± 2.2 | 70 | 106.0[a] | 11.27 ± 2.92[g] | |
| Edgin and Pennington (2005) | 24 | 11.5 ± 2.3 | 47 | 104.4[f] | 12.08 ± 4.29[g] | −0.06 |
| | 34 | 12.0 ± 2.5 | 84 | 108.7[f] | 12.32 ± 4.06[g] | |
| Spek et al. (2011) | 83 | 39.2 ± 11.3 | 87 | 110.5[b] | 12.3 ± 3.6[g] | −0.18 |
| | 41 | 38.3 ± 9.7 | 73 | 114.2[b] | 12.9 ± 2.3[g] | |
| Drake (2013) | 15 | 9.1 ± 2.1 | 80 | 100.7[a] | 2.5 ± 0.8[k] | −0.31 |
| | 15 | 8.9 ± 1.8 | 80 | 107.7[a] | 2.9 ± 0.4[k] | |
| | 15 | 9.1 ± 2.1 | 80 | 100.7[a] | 2.7 ± 0.6[k] | |
| | 15 | 8.9 ± 1.8 | 80 | 107.7[a] | 2.6 ± 0.6[k] | |
| | 15 | 9.1 ± 2.1 | 80 | 100.7[a] | 2.5 ± 0.7[k] | |
| | 15 | 8.9 ± 1.8 | 80 | 107.7[a] | 2.5 ± 0.7[k] | |
| | 15 | 9.1 ± 2.1 | 80 | 100.7[a] | 1.7 ± 1.2[k] | |
| | 15 | 8.9 ± 1.8 | 80 | 107.7[a] | 2.3 ± 0.8[k] | |
| | 15 | 9.1 ± 2.1 | 80 | 100.7[a] | 2.5 ± 1.1[k] | |
| | 15 | 8.9 ± 1.8 | 80 | 107.7[a] | 2.8 ± 0.4[k] | |
| | 15 | 9.1 ± 2.1 | 80 | 100.7[a] | 1.6 ± 1.1[k] | |
| | 15 | 8.9 ± 1.8 | 80 | 107.7[a] | 2.1 ± 0.9[k] | |
| Pring et al. (1995) | 9 | 26.0 ± 5.7 | ? | 11.8[e] | 45 ± 24.0[i] | −0.39 |
| | 9 | 12.6 ± 1.8 | ? | 12.6[e] | 36.0 ± 22.0[i] | |

**Table 1** continued

| Source | N | Age | % male | IQ | Block Design | *d* |
|---|---|---|---|---|---|---|
| Bölte (2008) | 7 | 27.7 ± 7.8 | 100 | 98.4[a] | 10.8 ± 3.6[g] | −0.40 |
| | 7 | 25.3 ± 6.9 | 100 | 111.6[a] | 11.9 ± 1.5[g] | |
| Kaland (2007) | 13 | 15.8 ± 2.6 | 100 | 107.5[a] | 11.5 ± 7.0[i] | −0.40 |
| | 13 | 16.8 ± 2.4 | 100 | 108.1[a] | 9.3 ± 3.3[i] | |
| Burnette (2005) | 23 | 11.3 ± 1.6 | 83 | 110.1[a] | 12.52 ± 3.40[g] | −0.45 |
| | 20 | 11.0 ± 1.3 | 75 | 116.8[a] | 13.9 ± 2.57[g] | |
| Scheurich (2010) | 15 | 25.2 ± 8.5 | 100 | 88.5[a] | 9.9 ± 4.2[g] | −0.46 |
| | 16 | 26.6 ± 7.1 | 100 | 108.6[a] | 11.4 ± 2[g] | |

Entries are ordered by effect size (right-hand column)

[a] Non-verbal IQ

[b] Full-scale IQ

[c] Non-verbal mental age

[d] TD are matched on raw scores

[e] Verbal mental age

[f] Verbal IQ

[g] Scaled scores

[h] Personal communication

[i] Time to completion

[j] Log time to completion

[k] Accuracy

symmetry tasks, and visual search tasks, amongst others. However, the small number of studies for each of these latter tasks did not allow for meta-analyses.

## Analyses

The primary aim of meta-analyses is to estimate a population parameter, typically an effect size. Here, we used Cohen's *d* to quantify performance differences between ASD and TD. We always coded results in such a way that positive effect sizes indicated superior performance in ASD.

Naturally, a sample does not necessarily provide a truthful representation of the population; instead, some error attributable to sampling can be expected and $d_i$, the effect size observed in study *i*, will probably differ from the population effect size δ. As sampling error fluctuates randomly between studies, multiple studies can be expected to obtain different results (regarding the effect size of the performance difference between ASD and TD on a given task). The observed variability of results between studies can be compared against the variability that can be expected to arise from sampling error alone. In cases where the observed variability exceeds the expected variability, τ is the standard deviation for this excess variability (or *heterogeneity*). $\tau > 0$ suggests that the studies differ in one or more aspects that systematically affect the result (i.e. the effect size of the performance difference). This could be due to differences in measurement (e.g., what test was used or how it was scored), participant characteristics (e.g., age, IQ, sex), or the like. Consequently, the population effect size δ cannot be adequately described by a fixed number. Instead, random-effects meta-analysis, which we used here, considers δ as a random variable with a mean ($\mu_\delta$) and a standard deviation ($\sigma_\delta$) for which τ is the estimate. Obviously, even if δ is not a random variable but fixed, the observed variability between study effect sizes might, just by chance, be greater than expected. Therefore, it is sensible to test τ for statistical significance. This is done by test statistic *Q*, which follows a $\chi^2$-distribution with *k* (number of studies) degrees of freedom. Finally, meta-regression is a procedure to investigate if some known study characteristic (e.g., how a test is scored) moderates effect size. In other words, meta-regression tests if that study characteristic explains (a part of) the observed heterogeneity. We performed all analyses with Comprehensive Meta Analysis 2.2 (Borenstein et al. 2005).

*Ceteris paribus*, studies that find a large effect will have smaller *p* values than those that find a small effect. At the same time, non-significant results are less likely to be published than significant ones, a phenomenon known as publication bias. In combination, published results might therefore overestimate the magnitude of the effect of interest in the population. Here, we used funnel plots as a method to investigate potential publication bias (see Fig. 1 for an example). In funnel plots each study's effect size is plotted against its standard error. The latter is inversely related to the study's sample size because a study with

**Table 2** Primary studies comparing embedded figure test performance (M ± SD) in people with ASD (first line in each entry) and controls (second line)

| Source | N | Age | % male | IQ | EFT performance | *d* |
|---|---|---|---|---|---|---|
| Pellicano (2006) | 40 | 5.6 ± 0.9 | 88 | 113.6[a] | 5.49 ± 3.2[f] | 2.36 |
| | 40 | 5.5 ± 1.0 | 78 | 112.5[a] | 13.82 ± 2.67[f] | |
| | 40 | 5.6 ± 0.9 | 88 | 113.6[a] | 5.68 ± 2.18[f] | |
| | 40 | 5.5 ± 1.0 | 78 | 112.5[a] | 9.89 ± 2.25[f] | |
| Pellicano (2005) | 20 | 9.5 ± 1.3 | 90 | | 23.60 ± 1.47[h,l] | 1.98[p] |
| | 20 | 9.8 ± 1.1 | 90 | Matched[a] | 22.95 ± 1.85[h,l] | |
| | 20 | 9.5 ± 1.3 | 90 | | | |
| | 20 | 9.8 ± 1.1 | 90 | Matched[a] | | |
| Brosnan (2012) | 13 | 13.9 ± 0.9 | 100 | 9.6[b] | 490 ± 309[g] | 1.88 |
| | 13 | 14.7 ± 0.8 | 100 | 9.5[b] | 8,776 ± 6,219[g] | |
| Shah (1983) | 20 | 13.3 ± 3.5 | 75 | | 20.55 ± 3.25[h] | 1.54 |
| | 20 | 9.3 ± 1.4 | 75 | Matched[a] | 15.70 ± 3.06[h] | |
| Jarrold (2005) | 18 | 12.4 ± 2.0 | | 103.1[c] | 14.41 ± 13.04[f] | 1.15 |
| | 18 | 6.5 ± 0.9 | | 101.7[c] | 28.56 ± 11.61[f] | |
| Koh (2012) | 11 | 11.2 ± 1.1 | 100 | 9.9[d] | 12.6 ± 1.3[h,k] | 0.88 |
| | 13 | 10.8 ± 1.5 | 100 | 9.7[d] | 11.2 ± 1.8[h,k] | |
| Ropar (2001) | 11 | 11.1 ± 2.0 | | 9.9[d] | 8.18 ± 4.19[h] | 0.68 |
| | 37 | 9.8 ± 1.4 | | 9.7[d] | 6.18 ± 2.78[h] | |
| | 11 | 11.1 ± 2.0 | | | 76.1 ± 57.1[f] | |
| | 37 | 9.8 ± 1.4 | | Matched[a] | 108.6 ± 41.8[f] | |
| Falter (2008) | 22 | 12.5 | 82 | Matched[a] | 72 ± 10[h] | 0.63 |
| | 22 | 12.6 | 51 | | 66 ± 7[h] | |
| | 22 | 12.5 | 82 | 100.9[a] | 15,543 ± 7,321[f] | |
| | 22 | 12.6 | 51 | 105.2[a] | 20,642 ± 10,467[f] | |
| Jolliffe (1997) | 34 | 29.2 ± 7.9 | ≈96 | 100.9[a] | 11.35 ± 0.99[h] | 0.61 |
| | 17 | 30.0 ± 9.1 | ≈97 | 105.2[a] | 10.76 ± 2.00[h] | |
| | 34 | 29.2 ± 7.9 | ≈96 | 107.5[a] | 30.74 ± 24.12[f] | |
| | 17 | 30.0 ± 9.1 | ≈97 | 108.1[a] | 52.63 ± 32.6[f] | |
| de Jonge (2006) | 18 | 18.9 ± 8.0 | 88 | 107.5[a] | 10.7 ± 1.3[h] | 0.57 |
| | 29 | 18.8 ± 7.7 | 88 | 108.1[a] | 9.8 ± 2.0[h] | |
| | 18 | 18.0 ± 8.0 | 88 | 95.1[a] | 25.9 ± 11.8[f] | |
| | 29 | 18.8 ± 7.7 | 88 | 104.6[a] | 36.0 ± 18.0[f] | |
| Morgan (2003) | 20 | ≈4.6 ± 0.5 | ≈90 | 95.1[a] | 17.75 ± 3.57[h] | 0.47 |
| | 21 | ≈4.6 ± 0.5 | 76 | 104.6[a] | 18.57 ± 2.91[h] | |
| | 20 | ≈4.6 ± 0.5 | ≈90 | 109[a] | 4.78 ± 2.04[f] | |
| | 21 | ≈4.6 ± 0.5 | 76 | 109[a] | 7.91 ± 2.54[f] | |
| Keehn (2009) | 12 | 12.9 ± 2.4 | | 109[a] | 2,041 ± 383[f,l] | 0.42 |
| | 11 | 13.9 ± 3.8 | | 109[a] | 2,408 ± 378[f,l] | |
| | 12 | 12.9 ± 2.4 | | 109[a] | 2,273 ± 450[f,l] | |
| | 11 | 13.9 ± 3.8 | | 109[a] | 2,990 ± 718[f,l] | |
| | 12 | 12.9 ± 2.4 | | 109[a] | 0.117 ± 0.115[i,l] | |
| | 11 | 13.9 ± 3.8 | | 109[a] | 0.077 ± 0.096[i,l] | |
| | 12 | 12.9 ± 2.4 | | 100.1[a] | 0.604 ± 0.168[i,l] | |
| | 11 | 13.9 ± 3.8 | | 105.9[a] | 0.577 ± 0.218[i,l] | |
| Bölte (2007) | 15 | 25.8 ± 7.7 | 100 | | 45.8 ± 17.6[f] | 0.37 |
| | 15 | 27.0 ± 6.7 | 100 | | 51.5 ± 12.6[f] | |
| van Lang (2006) | 22 | 14.9 ± 2.2 | 77 | 57.0[a] | 12.9 ± 5.7[h] | 0.35 |
| | 21 | 14.5 ± 2.7 | 57 | 57.2[a] | 10.9 ± 5.7[h] | |
| Edgin and Pennington (2005) | 24 | 11.5 ± 2.3 | 84[h] | 104.4[e] | 20.33 ± 3.77[h] | 0.25 |
| | 34 | 12.0 ± 2.5 | 47[h] | 108.7[e] | 20.96 ± 2.41[h] | |
| | 24 | 11.5 ± 2.3 | 84[h] | 104.4[e] | 13.89 ± 5.94[f] | |
| | 34 | 12.0 ± 2.5 | 47[h] | 108.7[e] | 20.64 ± 11.05[f] | |

**Table 2** continued

| Source | N | Age | % male | IQ | EFT performance | *d* |
|---|---|---|---|---|---|---|
| Koh (2012) | 13 | 10.3 ± 0.9 | 100 | 99.4[c] | 11.8 ± 1.8[h,k] | 0.23 |
| | 11 | 10.4 ± 1.2 | 100 | 104.5[c] | 11.2 ± 3.3[h,k] | |
| White (2011) | 45 | 9.3 ± 1.4 | 91 | 98.1[a] | 0.61 ± 0.21[h] | 0.07 |
| | 27 | 9.5 ± 1.4 | 78 | 103.[a] | 0.68 ± 0.21[h] | |
| | 45 | 9.3 ± 1.4 | 91 | 98.1[a] | 10.0 ± 3.8[f] | |
| | 27 | 9.5 ± 1.4 | 78 | 103.[a] | 11.7 ± 3.4[f] | |
| White (2011) | 62 | 10.6 ± 2.3 | | 97.3[a] | 0.61 ± 0.24[h] | 0.05 |
| | 50 | 10.5 ± 2.2 | | 96.4[a] | 0.62 ± 0.20[h] | |
| | 62 | 10.6 ± 2.3 | | 97.3[a] | 7.63 ± 3.80[f] | |
| | 50 | 10.5 ± 2.2 | | 96.4[a] | 8.30 ± 3.16[f] | |
| Schlooz (2006) | 12 | 10.4 | 100 | 103.9[a] | 80.1 ± 18.3[h] | 0.05 |
| | 12 | 10.5 | 100 | 107.0[a] | 81.0 ± 18.6[h] | |
| | 12 | 10.4 | 100 | 103.9[a] | 5.35 ± 1.95[f] | |
| | 12 | 10.5 | 100 | 107.0[a] | 5.68 ± 2.62[f] | |
| South (2007) | 19 | 14.0 ± 2.7 | 74 | 110.6[a] | ≈39.7 ± 26.9[f,k] | −0.05 |
| | 15 | ≈14.1 ± 2.9 | ≈61 | ≈113.3[a] | ≈38.3 ± 23.5[f,k] | |
| Chen (2008) | 14 | 9.1 ± 1.0 | 100 | 96.4[c] | 20.1 ± 2.81[h] | −0.11 |
| | 14 | 9.0 ± 0.9 | 100 | Matched[c] | 20.4 ± 2.47[h] | |
| Taylor et al. (2013) | 31 | 9.0 ± 2.1 | 81 | | 12.8 ± 5.3[f] | −0.15 |
| | 61 | 8.8 ± 1.7 | 54 | | 11.9[f,n] | |
| | 31 | 9.0 ± 2.1 | 81 | | 23.9 ± 1.1[h] | |
| | 61 | 8.8 ± 1.7 | 54 | | 23.8[h,n] | |
| | 31 | 9.0 ± 2.1 | 81 | | 10.6 ± 9.5[o] | |
| | 61 | 8.8 ± 1.7 | 54 | | 7.6[o,n] | |
| Kahland (2007) | 13 | 15.8 ± 2.6 | 100 | 107.5[a] | 39.4 ± 22.6[f] | −0.20 |
| | 13 | 15.8 ± 2.4 | 100 | 108.1[a] | 34.8 ± 23.2[f] | |
| Nydén (2011) | 10 | | | | 8.9 ± 4.2[h] | −0.24 |
| | 28 | 14–64 | | | 9.5 ± 1.5[h] | |
| Manjaly (2007) | 12 | 14.4 ± 2.7 | | 110.1[c] | 85.6 ± 8.7[h] | −0.27 |
| | 12 | 14.3 ± 2.7 | | 109.3[c] | 89.9 ± 5.5[h] | |
| | 12 | 14.4 ± 2.7 | | 110.1[c] | 1,825 ± 288[f] | |
| | 12 | 14.3 ± 2.7 | | 109.3[c] | 1,834 ± 194[f] | |
| Lee (2007) | 13 | 10.1 ± 1.6 | 85 | 109.3[c] | 61.83 ± 14.75[h,m] | −0.27 |
| | 13 | 10.6 ± 1.3 | 77 | 115.1[c] | 65.02 ± 11.66[h,m] | |
| | 13 | 10.1 ± 1.6 | 85 | 109.3[c] | 7.41 ± 6.52[f,m] | |
| | 13 | 10.6 ± 1.3 | 77 | 115.1[c] | 5.80 ± 3.43[f,m] | |
| Spencer (2012) | 38 | 14.6 ± 1.7 | 89 | 107.1[c] | 74.1 ± 18.3[h] | −0.31 |
| | 80 | 15.2 ± 1.9 | 40 | 111.3[c] | 80.7 ± 11.8[h] | |
| | 38 | 14.6 ± 1.7 | 89 | 107.1[c] | 2,549 ± 484[f] | |
| | 80 | 15.2 ± 1.9 | 40 | 111.3[c] | 2,491 ± 365[f] | |
| Malisza (2011) | 8 | 11.7 ± 1.4 | 75 | 84.4[a] | 80.48 ± 13.73[h] | −0.34 |
| | 9 | 12.1 ± 1.9 | 78 | | 92.41 ± 9.21[h] | |
| | 8 | 11.7 ± 1.4 | 75 | 84.4[a] | 1,212 ± 156[f] | |
| | 9 | 12.1 ± 1.9 | 78 | | 1,250 ± 146[f] | |
| Hodgson (2011) | 13 | 7.6 ± 1.8 | 100 | 5.7[b] | 11.4 ± 5.0[h] | −0.34 |
| | 28 | 7.8 ± 1.3 | 100 | 7.2[b] | 13.4 ± 6.1[h] | |
| | 13 | 7.6 ± 1.8 | 100 | 5.7[b] | 29.0 ± 7.8[f] | |
| | 28 | 7.8 ± 1.3 | 100 | 7.2[b] | 24.4 ± 9.2[f] | |
| Scheurich (2010) | 15 | 25.2 ± 8.5 | 100 | 88.5[a] | 75.7 ± 34.5[f] | −0.37 |
| | 16 | 26.6 ± 7.1 | 100 | 108.6[a] | 47.9 ± 16.0[f] | |
| Drake (2013) | 15 | 9.1 ± 2.1 | 80 | 100.7[a] | 5.1 ± 5.0[h] | −0.41 |
| | 15 | 8.9 ± 1.8 | 80 | 107.7[a] | 7.3 ± 5.8[h] | |

**Table 2** continued

| Source | N | Age | % male | IQ | EFT performance | *d* |
|---|---|---|---|---|---|---|
| Ozonoff (1991) | 23 | 12.1 ± 3.2 | 91 | 98.4[a] | 16.35 ± 5.77[h] | −0.44 |
| | 20 | 12.4 ± 3.0 | 90 | 97.0[a] | 18.75 ± 5.18[h] | |
| Damarla (2010) | 13 | 19 ± 5.5 | 85 | 109.5[a] | 76.0 ± 21.6[h] | −0.48 |
| | 13 | 22.1 ± 4.3 | 100 | 110.5[a] | 15.4 ± 84.6[h] | |
| | 13 | 19 ± 5.5 | 85 | 109.5[a] | 5,847 ± 1,520[f] | |
| | 13 | 22.1 ± 4.3 | 100 | 110.5[a] | 5,816 ± 835[f] | |
| Brian (1996) | 15 | 19.5 ± 5.6 | 87 | 93.6[a] | 383.7 ± 333.6[f] | −0.51 |
| | 15 | 11.8 ± 3.6 | 87 | Matched[a] | 257.0 ± 112.3[f] | |
| | 15 | 19.5 ± 5.6 | 87 | 93.6[a] | 220.4 ± 178.1[f] | |
| | 15 | 11.8 ± 3.6 | 87 | Matched[a] | 205.4 ± 114.9[f] | |
| | 15 | 19.5 ± 5.6 | 87 | 93.6[a] | 620.8 ± 309.9[f] | |
| | 15 | 11.8 ± 3.6 | 87 | Matched[a] | 554.6 ± 334.1[f] | |
| Burnette (2005) | 23 | 11.3 ± 1.6 | 87 | 110.1[a] | 25.26 ± 4.96[j] | −0.78 |
| | 20 | 11.0 ± 1.3 | 87 | 116.8[a] | 28.70 ± 3.61[j] | |

Entries are ordered by effect size (right-hand column)

[a] Non-verbal IQ

[b] Non-verbal mental age

[c] Full-scale IQ

[d] Verbal mental age

[e] Verbal IQ

[f] Time

[g] Negative efficiency (RT divided by proportion correct)

[h] Accuracy

[i] Errors

[j] Scores reflect time and accuracy (higher scores indicate better performance)

[k] Estimated from figure

[l] Personal communication

[m] 2AFC format

[n] Prediction for ASD group based on TD results

[o] False alarms

[p] Computed from test statistic

large *N* can be expected to yield a more precise estimate of δ than a study with small *N*. Results from studies with small standard errors should scatter narrowly around the population effect size, and the degree of scatter should increase as the standard error increases. Therefore, the distribution of study effect sizes should roughly look like an inverted funnel. With increasing standard errors, studies need to find ever larger effects in order to be statistically significant (cf. grey area in Fig. 1). Consequently, publication bias will result in an underrepresentation of results in the area of non-significance. This will break the symmetry of the funnel, especially at its base where studies with small effect sizes are then strongly underrepresented (e.g., Palmer 2000).

Finally, Navon Figure data require some additional comments. Three paradigms can be differentiated: In the *free choice* paradigm, participants indicate whether they spontaneously perceive the figure at the global or the local level (e.g., by naming the number they see, Wang et al. 2007). We used each participant's number of global preferences as the dependent variable and computed Cohen's *d* from that. We coded results such that positive effect sizes indicate stronger global preferences in the TD group, which is in line with expectations (Wang et al. 2007).

In the *divided attention* paradigm, participants indicate for each stimulus whether or not it contains a particular target; in the critical trials that we analyse here, the target appears either at the local or the global level (e.g., Johnson et al. 2010). In the *selective attention* paradigm, participants are asked to respond either to the global or the local stimulus level (e.g., by indicating via a button press which of two letters is shown at that level, Plaisted et al. 1999). In both paradigms, any global advantage is indicated by the performance difference between trials for which the global versus the local stimulus level is relevant for task performance. Of interest here is whether this global advantage (or disadvantage) differs between the ASD and TD group, which is reflected in the mixed interaction effect of group

**Table 3** Primary studies comparing Mental Rotation performance (M ± SD) in people with ASD (first line in each entry) and TD controls (second line)

| Task | DV | Group | Slope | Intercept | $d_{Slope}$ | $d_{Intercept}$ |
|---|---|---|---|---|---|---|
| Beacher et al. (2012)[a] | | | | | | |
| Letter | RT | ASD | | 1,607 ± 543 | | −0.20 |
| | | TD | | 1,507 ± 477 | | |
| Letter | Accuracy | ASD | | 0.695 ± 0.075 | | −0.34 |
| | | TD | | 0.717 ± 0.057 | | |
| Conson et al. (2013)[b,i] | | | | | | |
| Letter | RT | ASD | 0.65 ± 1.08 | 1,657 ± 202 | 0.00 | −0.46 |
| | | TD | 0.65 ± 1.45 | 1,569 ± 178 | | |
| Letter | Accuracy | ASD | 0.000 ± 0.000 | 0.931 ± 0.077 | 0.00 | 0.17 |
| | | TD | 0.000 ± 0.001 | 0.943 ± 0.066 | | |
| Hand | RT | ASD | 2.00 ± 1.77 | 2,348 ± 369 | −0.38 | 0.41 |
| | | TD | 2.64 ± 1.59 | 2,158 ± 548 | | |
| Hand | Accuracy | ASD | 0.000 ± 0.000 | 0.904 ± 0.109 | 0.03 | −0.32 |
| | | TD | 0.000 ± 0.000 | 0.933 ± 0.070 | | |
| Falter et al. (2008)[c] | | | | | | |
| 3D | RT | ASD | 15.95 ± 17.15 | 1,762 ± 431 | −0.52 | 0.81 |
| | | TD | 8.89 ± 8.88 | 2,208 ± 646 | | |
| 3D | Accuracy | ASD | −0.218 ± 0.245 | 0.956 ± 0.068 | 0.05 | 0.15 |
| | | TD | −0.232 ± 0.273 | 0.946 ± 0.070 | | |
| Hamilton et al. (2009)[d] | | | | | | |
| Object | RT | | | | | 0.73[k] |
| McGrath et al. (2012)[e] | | | | | | |
| 3D | RT | ASD | | 3,870 ± 1,340[j] | | 0.18 |
| | | TD | | 4,130 ± 1,520[j] | | |
| 3D | Accuracy | ASD | | 0.900 ± 0.156[j] | | 0.08 |
| | | TD | | 0.888 ± 0.155[j] | | |
| Nakano et al. (2012)[f] | | | | | | |
| 3D | Accuracy | ASD | | 0.57[i] | | 0.30[k] |
| | | TD | | 0.52[i] | | |
| Silk et al. (2006)[g] | | | | | | |
| 3D | RT | ASD | | 5,510 ± 1,055 | | 1.07 |
| | | TD | | 6,480 ± 770 | | |
| 3D | Accuracy | ASD | | 0.49 ± 0.15 | | −0.13 |
| | | TD | | 0.51 ± 0.16 | | |
| Soulieres et al. (2011)[h,i] | | | | | | |
| 3D | RT | ASD | 14.52 ± 9.36 | 3,891 ± 1,238 | −0.30 | 0.36 |
| | | TD | 11.57 ± 10.59 | 4,417 ± 1,769 | | |
| Shape | RT | ASD | 10.75 ± 5.08 | 2,885 ± 1,323 | −0.08 | 0.05 |
| | | TD | 10.28 ± 6.80 | 2,950 ± 1,305 | | |
| Hand | RT | ASD | 7.23 ± 6.00 | 1,902 ± 733 | 0.13 | 0.16 |
| | | TD | 8.13 ± 7.93 | 2,015 ± 640 | | |
| Letter | RT | ASD | 5.86 ± 4.46 | 1,583 ± 708 | −0.49 | 0.33 |
| | | TD | 3.83 ± 3.57 | 1,794 ± 500 | | |
| 3D | Accuracy | ASD | −0.077 ± 0.070 | 0.939 ± 0.080 | 0.13 | −0.79 |
| | | TD | −0.088 ± 0.099 | 0.855 ± 0.139 | | |
| Shape | Accuracy | ASD | −0.013 ± 0.054 | 0.961 ± 0.062 | −0.13 | −0.44 |
| | | TD | −0.004 ± 0.088 | 0.924 ± 0.114 | | |
| Hand | Accuracy | ASD | −0.014 ± 0.057 | 0.957 ± 0.078 | −0.35 | −0.61 |
| | | TD | 0.006 ± 0.059 | 0.904 ± 0.101 | | |

**Table 3** continued

| Task | DV | Group | Slope | Intercept | $d_{Slope}$ | $d_{Intercept}$ |
|---|---|---|---|---|---|---|
| Letter | Accuracy | ASD | −0.042 ± 0.076 | 0.978 ± 0.053 | −0.80 | −0.82 |
| | | TD | 0.026 ± 0.098 | 0.902 ± 0.136 | | |

Entries are ordered by effect size (right-hand column)

[a] $N_{ASD}$ = 26 (58 % male), $N_{TD}$ = 32 (50 % male); age: 32.8 (ASD) and 30.4 (TD); National Adult Reading Test score: 30.3 (ASD) and 35.2 (TD)

[b] $N_{ASD}$ = 24 (88 % male), $N_{TD}$ = 24 (83 % male); age: 13.4 (ASD) and 13.3 (TD); full scale IQ: 100.0 (ASD) and 99.4 (TD)

[c] $N_{ASD}$ = 19 (95 % male), $N_{TD}$ = 19 (95 % male); age: 12.9 (ASD) and 13.0 (TD); Raven score: 43 (ASD) and 45 (TD)

[d] $N_{ASD}$ = 23 (83 % male), $N_{TD}$ = 23 (≈63 % male); age: 8.0 (ASD) and 4.2 (TD); verbal mental age: 4.4 (ASD), 4.8 (TD)

[e] $N_{ASD}$ = 22 (100 % male), $N_{TD}$ = 22 (100 % male); age: 17.6 (ASD) and 17.5 (TD); performance IQ: 124.4 (ASD), 120.8 (TD)

[f] $N_{ASD}$ = 14 (71 % male), $N_{TD}$ = 20 (75 % male); age: 30.7 (ASD) and 27.6 (TD); performance IQ: 107 (ASD), matched (TD)

[g] $N_{ASD}$ = 7 (100 % male), $N_{TD}$ = 9 (100 % male); age: 14.7 (ASD) and 15.0 (TD); performance IQ: 118 (ASD), matched (TD)

[h] $N_{ASD}$ = 23 (87 % male), $N_{TD}$ = 14 (86 % male); age: 21.0 ± 5.8 (ASD) and 19.4 ± 3.8 (TD); performance IQ: 105.0 (ASD) and 99.1 (TD). Data stem from re-analysis of the original dataset

[i] Personal communication

[j] From figure

[k] Computed from *t* test statistics

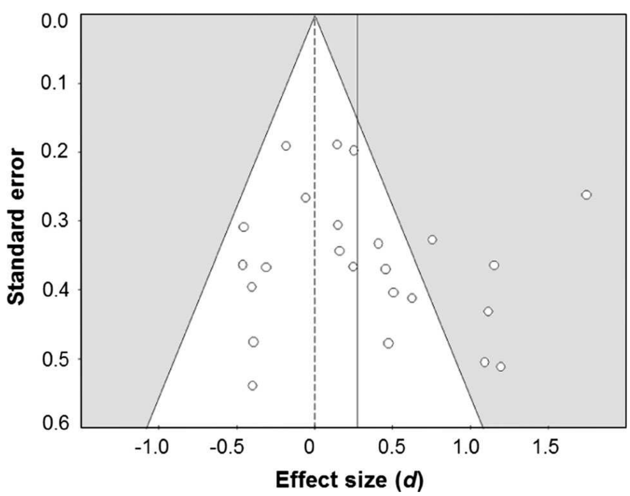



**Fig. 1** Funnel plot for 24 studies comparing Block Design performance between ASD and TD. *Note* Positive effect sizes indicate superior performance in ASD. Results outside the *white cone* are statistically significant for individual studies. The *solid vertical line* indicates the estimate for the population effect size (μδ)

(TD vs. ASD) × relevant stimulus level (global vs. local), where group is a between-subjects factor and target location is a within-subjects repeated-measures factor. We computed Cohen's *d* as the interaction effect in raw score units divided by the pooled standard deviation across the four conditions. In our analyses positive effect sizes indicate that the global advantage is stronger in the TD group than in the ASD group, which is in line with expectations (Wang et al. 2007). To the best of our knowledge, there is currently no method to determine the standard error for an effect size that stems from such a mixed interaction. In this case, we can therefore merely compile but not meta-analyse the relevant effect sizes.

In both, the divided attention and the selective attention paradigm, any advantage for the global stimulus dimension can either be expressed via reaction times or number of errors. Wherever possible, we present them separately as well as combined (see the three right-most columns in Table 4).

## Results

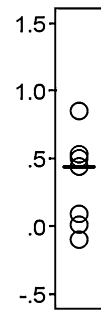
### Block Design

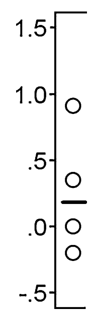
Twenty-four samples with altogether 520 ASD participants and 518 TD participants entered the analysis. Table 1 lists these studies and their results ordered by effect size; Fig. 1 shows the corresponding funnel plot. Most of the studies used scaled scores, which combine time and number of errors, as the dependent variable. Five studies used time to completion, instead, and one study used accuracy.

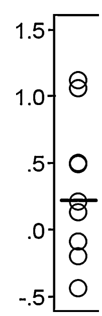
The population effect size $\mu_\delta$ was estimated as $d = 0.32$ ($Z = 2.53$, $p = .012$). When we re-run the analysis without the most extreme effect in order to test how much the estimate depended on a single finding, the estimate for $\mu_\delta$ did not change much, $d = 0.22$ ($Z = 2.24$, $p = .025$). Heterogeneity ($\sigma_\delta$) was estimated to be large ($\tau = 0.49$) and was significantly greater than zero ($Q_{23} = 77.3$, $p < .001$). Heterogeneity is also reflected in the large scatter of primary study results in the funnel plot (cf. Fig. 1). Year of publication, type of test scoring (accuracy, scaled scores, time to completion), and participants' age were tested in meta-regressions; neither of them turned out

**Table 4** Primary studies comparing performance with Navon stimuli in people with ASD and TD controls in three paradigms

| Source | N | Age (M ± SD) | % male | IQ | $d_{RT}$ | $d_{error}$ | $d_{overall}$ |
|---|---|---|---|---|---|---|---|
| Free choice | | | | | | | |
| Deruelle et al. (2006) | 13 | 9.1 ± 2.6 | 69 | MA = 7.8 | | | 0.85 |
| | 13 | 7.6 ± 3.3 | 69 | Matched | | | |
| Bölte et al. (2007) | 15 | 25.8 ± 7.7 | 100 | 100.1 | | | 0.53 |
| | 15 | 27.0 ± 6.7 | 100 | 105.9 | | | |
| Koldewyn et al. (2013) | 45 | 8.8 ± 1.6 | 80 | 109.4 | | | 0.50 |
| | 45 | 8.2 ± 1.7 | 80 | 111.7 | | | |
| Rondan and Deruelle (2007) | 26 | 26.2 ± 7.3 | 88 | 91.6 | | | 0.44 |
| | 26 | 27.7 ± 6.2 | 88 | ? | | | |
| Bernardino et al. (2012) | 20 | 12.1 ± 2.3 | 100 | 103.4 | | | 0.09 |
| | 20 | 15.7 ± 8.0 | 55 | 107.9 | | | |
| Wang et al. (2007) | 12 | ≈14.6 ± 4.0 | ≥83 | ≈67 | | | 0.01 |
| | 12 | ≈14.3 ± 3.5 | ≥83 | Matched | | | |
| Bernardino et al. (2012) | 19 | 13.3 ± 2.6 | 79 | 64.5 | | | −0.10 |
| | 20 | 15.4 ± 4.5 | 75 | 58.6 | | | |
| Divided attention | | | | | | | |
| Plaisted et al. (1999) | 17 | 10.3 ± 2.3 | ? | Normal | 0.77 | 1.04 | 0.91 |
| | 17 | 10.2 ± 2.0 | ? | Matched | | | |
| Katagiri et al. (2013) | 11 | 31.1 ± 6.1 | 27 | 105 | 0.58[a] | 0.12 | 0.35 |
| | 11 | 28.3 ± 5.4 | 27 | Normal | | | |
| Johnson et al. (2010) | 10 | 16.3 ± 2.3 | 60 | 112.4 | −0.06 | 0.06 | 0.00 |
| | 11 | 14.3 ± 2.6 | 45 | 114.6 | | | |
| Mottron et al. (1999) | 11 | 14.9 ± 3.5 | 91 | 110.9 | −0.50 | 0.11 | −0.20 |
| | 11 | 14.0 ± 2.5 | 91 | 116.1 | | | |
| Selective attention | | | | | | | |
| Behrmann et al. (2006) | 14 | 34.5 ± 10.9 | 86 | 104.1 | 1.12 | | 1.12 |
| | 27 | Matched | Matched | Matched | | | |
| Bernardino et al. (2012) | 19 | 13.3 ± 2.6 | 79 | 64.5 | | 1.06 | 1.06 |
| | 20 | 15.4 ± 4.5 | 75 | 58.6 | | | |
| Wang et al. (2007) | 14 | ≈14.6 ± 4.0 | ≥86 | ≈67 | 0.70 | 0.29 | 0.50 |
| | 14 | ≈14.3 ± 3.5 | ≥86 | Matched | | | |
| Bernardino et al. (2012) | 20 | 12.1 ± 2.3 | 100 | 103.4 | | 0.49 | 0.49 |
| | 20 | 15.7 ± 8.0 | 55 | 107.9 | | | |
| Koldewyn et al. (2013) | 45 | 8.8 ± 1.6 | 80 | 109.4 | 0.01 | 0.41[a] | 0.21 |
| | 45 | 8.2 ± 1.7 | 80 | 111.7 | | | |
| Rinehart et al. (2000) | 12 | 9.9 ± 3.0 | 92 | 93.7 | 0.13 | | 0.13 |
| | 12 | Matched | 92 | Matched | | | |
| Plaisted et al. (1999) | 17 | 10.3 ± 2.3 | ? | Normal | 0.11 | −0.28 | −0.09 |
| | 17 | 10.2 ± 2.0 | ? | Matched | | | |
| Rinehart et al. (2000) | 12 | 12.5 ± 3.8 | 83 | 102.8 | −0.20 | | −0.20 |
| | 12 | Matched | 83 | Matched | | | |
| Porter and Coltheart (2006) | 8 | 8.4 ± 1.5 | 50 | MA = 6.3 | −0.31 | −0.56 | −0.44 |
| | 27 | 16.7 ± 9.6 | ≈48 | MA = 14.7 | | | |




Entries are ordered by effect size (right-hand column)

For each study, the ASD sample is described in the top row and the TD sample in the bottom row. Positive effect sizes are in line with expectations and indicate that the global advantage (divided attention, selective attention) or the global preference (free choice) is stronger in the TD group than in the ASD group. For the divided attention and the selective attention paradigm, separate effect sizes are indicated for reaction time (RT) and errors as dependent variables if available. MA = mental age. The figures at the right illustrate the distribution of the overall effect size and their median (horizontal bar)

[a] Calculation partly based on personal communication

to moderate effect size ($p = .578$, $p = .754$, and $p = .123$, respectively).

For matching on IQ, non-verbal IQ should be most relevant for group comparisons on visuo-spatial skills. We therefore contrasted ASD and TD groups on non-verbal IQ wherever this information was available; otherwise we used full-scale or verbal IQ (cf. Table 1). Not all studies were equally successful in matching ASD and TD on IQ, which might contribute to the observed heterogeneity. We therefore considered IQ difference as a moderator. Relevant information was available for 20 studies, which showed a similar picture as the whole set of studies ($d = 0.31$, $\tau = 0.52$). Meta-regression showed that IQ difference (ASD–TD) significantly moderated the effect

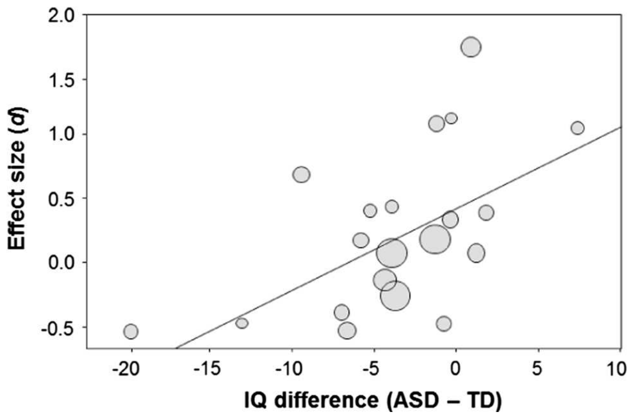



**Fig. 2** Block Design advantage in ASD individuals as a function of IQ difference in 20 studies. *Note* Circle sizes represent the weight of each study in the meta-regression; the regression line is based on a method-of-moments approach (Raudenbush 1994)

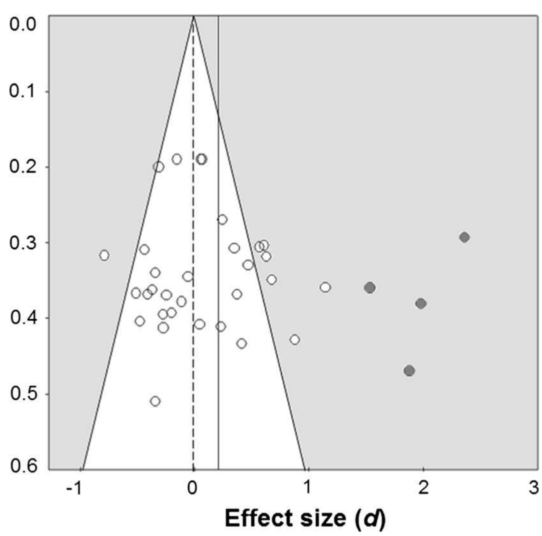



**Fig. 3** Funnel plot for 35 studies comparing Figure Disembedding Test performance between ASD and TD. *Note* Positive effect sizes indicate superior performance in ASD. Results outside the *white cone* are statistically significant for individual studies. The *solid vertical line* indicates the estimate for the population effect size (μδ)

size for BDT performance ($b = 0.059$, $Q_1 = 7.1$, $p = .008$). That is, for each IQ point an ASD group lags behind its TD control group, the ASD advantage in Block Design performance can be expected to be underestimated by $d = 0.059$ (see also Fig. 2). In this context it seems then relevant that, on average, IQ was 3.7 points lower in the ASD groups then in the TD groups (unweighted mean of 20 samples). In order to account for this effect, we used the result from the meta-regression ($b = 0.059$) to estimate for each of 20 relevant samples the effect size that had been observed if ASD and TD group had been matched perfectly on IQ. As expected, a meta-analysis on these corrected effect sizes resulted in a somewhat higher estimate for $\mu_\delta$, $d = 0.58$ ($Z = 3.80$, $p < .001$). However, heterogeneity was only marginally reduced ($\tau = 0.47$, $Q_{14} = 43.3$, $p < .001$), which suggests that precision of IQ matching explains only a small part of the heterogeneity observed in our initial analysis.

Returning to our initial analysis of all samples, it is interesting to note that the largest observed Block Design advantage ($d = 1.75$), which looks like an outlier in the funnel plot (cf. Fig. 1), appeared in one of the few samples in which IQ was higher in the ASD group than in the TD group (cf. Table 1; Fig. 2).

In the face of heterogeneity, formal tests of publication bias are often not very useful (Ioannidis and Trikalinos 2007). However, inspection of the funnel plot (cf. Fig. 1) showed no evidence for publication bias. As can be seen, results' scatter around the estimated population effect size was fairly symmetrical; more importantly, most of the primary study results were not statistically significant on their own. If there was strong publication bias in this area we would expect to see few statistically non-significant results in the published literature.

### Figure Disembedding

The results for all $k = 35$ samples (with 707 ASD participants and 803 TD participants) are shown in Table 2. As previously, studies are listed in decreasing order of effect size and positive effect sizes indicate superior performance in the ASD group.

The meta-analysis estimated the population effect size $\mu_\delta$ as $d = 0.26$ ($Z = 2.13$; $p = .033$); heterogeneity ($\sigma_\delta$) was very large ($Q_{34} = 167.2$; $p < .001$; $\tau = 0.63$), see also the funnel plot in Fig. 3. There was no indication of publication bias.

Without the four filled outliers (Brosnan et al. 2012; Pellicano et al. 2005, 2006; Shah and Frith 1983), all more than two standard deviations away from the mean of $d = 0.26$, heterogeneity was much lower but still substantial ($\tau = 0.29$, $p = .003$) and the estimate for $\mu_\delta$ shrank to $d = 0.05$ ($Z = 0.60$; $p = .550$).

We tried various variables as potential moderators in meta-regressions in order to account for the large variability in effect sizes across studies ($\tau = 0.63$). We did not find that type of dependent variable (time; time and accuracy; accuracy) had an influence on results ($p = .595$); the same was true for year of publication ($p = .191$) and participants' age ($p = .323$; 34 studies). For a sub-set of 23 studies, we could investigate whether the IQ difference between ASD and TD participants accounted for differences between studies. This resulted in a less steep slope than in the Block Design analysis ($b = .042$), which did not turn out to be statistically significant in Figure Disembedding ($p = .147$).

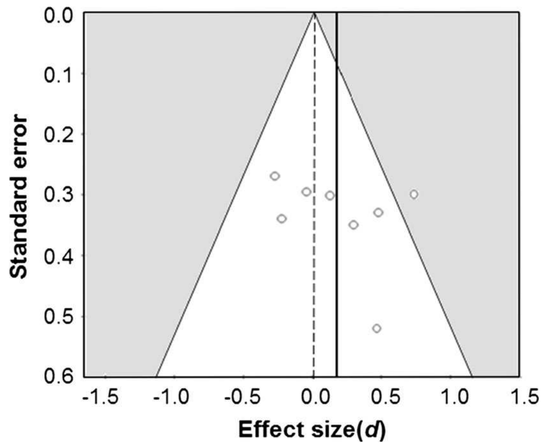



**Fig. 4** Funnel plot for eight studies comparing Mental Rotation performance (intercepts) between ASD and TD. *Note* Positive effect sizes indicate superior performance in ASD. Results outside the white cone are statistically significant for individual studies. The *solid vertical line* indicates the estimate for the population effect size (μδ)

### Mental Rotation

Given the different skills required for this task that can be divided into rotational (slopes) and non-rotational (intercepts) performance aspects, we present separate meta-analyses on regression slopes and intercepts.

Regarding regression slopes, three samples with altogether 66 ASD participants and 57 TD participants entered the analysis. The results for two samples (Conson et al. 2013; Soulieres et al. 2011) depended on a reanalysis of the raw data, which is presented in Table 3. In the Conson et al. (2013) study, participants solved two different tasks. Across both, the effect size for the performance difference between ASD and TD participants was $d = -0.19$ for RT slopes and $d = 0.02$ for accuracy slopes. As previously, positive effect sizes indicate superior performance in the ASD group. Across both measures, the performance difference was $d = -0.09$ ($SE = 0.29$). In the Soulieres et al. (2011) study, each participant solved four mental rotation tests. The effect size for the performance difference between ASD and TD participants across all four tests was $d = -0.19$ for RT slopes and $d = -0.29$ for accuracy slopes. Across both measures, the performance difference was $d = -0.23$ ($SE = 0.34$). For the other sample (Falter et al. 2008), the performance difference between ASD and TD participants across measures (RT and accuracy) was $d = -0.24$ ($SE = 0.33$). The meta-analysis of $k = 3$ samples estimated the population effect size $\mu_\delta$ of the combined performance measure (RT slope and accuracy slope) as $d = -0.18$ ($Z = -0.99$; $p = .323$). Heterogeneity was estimated as zero, which makes a confidence interval on $\mu_\delta$ particularly meaningful; the 95 % CI was [−0.54, 0.18].

Regarding regression intercepts eight samples with altogether 158 ASD participants and 163 TD participants entered the analysis. Detailed descriptions are presented in Table 3, the funnel plot is shown in Fig. 4. For the Conson et al. (2013) study, results across subtests were $d = -0.03$ for RT and $d = -0.08$ for accuracy; for the Soulieres et al. (2011) sample, results across four subtests were $d = 0.23$ for RT and $d = -0.67$ for accuracy. This pattern, where ASD participants showed a stronger relative performance for RT than for accuracy, also emerged in all four other relevant samples, as can be seen from Table 3 (Beacher et al. 2012; Falter et al. 2008; McGrath et al. 2012; Silk et al. 2006). This might reflect that speed-accuracy tradeoff differs between ASD and TD participants. For one sample (Hamilton et al. 2009), only RT data were available, for another sample (Nakano et al. 2012) only accuracy data. Thus, none of the two outcome measures (RT and accuracy) was overrepresented in the analysis, which averaged across both whenever RT and accuracy results were available. For $k = 8$ samples, the population effect size $\mu_\delta$ was estimated as $d = 0.16$ ($Z = 1.21$; $p = .227$). The estimate for heterogeneity ($\sigma_\delta$) was moderate ($\tau = 0.19$) but statistically non-significant ($Q_7 = 9.4$, $p = .226$), which may reflect low power due to small sample size ($k = 8$). The funnel plot does not suggest publication bias because there is no overrepresentation of significant studies with small sample sizes. Because of the small number of samples we did not try to identify moderators.

### Navon Figures

#### *Free Choice*

The results for all $k = 7$ free choice samples (with 150 ASD participants and 151 TD participants) are shown in the top panel of Table 4. As previously, studies are listed in decreasing order of effect size; positive effect sizes indicate that preference for the local stimulus level was greater in the ASD group than in the TD group. The distribution of effect sizes is also illustrated in the top figure at the right of Table 4.

The meta-analysis estimated the population effect size $\mu_\delta$ as $d = 0.35$ ($Z = 3.01$; $p = .003$); heterogeneity ($\sigma_\delta$) was absent ($Q_6 = 5.9$; $p = .430$; $\tau = 0.00$). In light of the latter, a confidence interval for $\mu_\delta$ becomes particularly informative; the 95 % CI obtained was [0.12, 0.57]. Following Cohen's (1988) popular convention of effect sizes, the combined data then suggest that, compared to TD, ASD affected people show a small to medium preference for the local level of Navon stimuli.

Due to the small number of studies we did not look for publication bias. However, five out of the seven primary effect sizes were not statistically significant on their own; thus, it appears unlikely that publication bias is an issue in this domain.

### *Divided Attention*

Fewer data were available for the divided attention paradigm. The results for the $k = 4$ samples (with 49 ASD participants and 50 TD participants) are shown in the middle panel of Table 4. Positive effect sizes indicate that the global advantage is stronger in the TD than in the ASD group. The overall effect sizes (reaction time and errors combined where possible) ranged from $d = -0.20$ to $d = 0.90$, with a median of $d = 0.18$. Again, the distribution of effect sizes can be seen in the middle figure to the right of Table 4. A similar picture emerged for reaction times only (median $d = 0.26$) and number of errors only (median $d = 0.12$). Thus, the limited evidence that is available tentatively suggests that any advantage for the global stimulus level might be slightly smaller in ASD affected people than in TD.

### *Selective Attention*

The results for the $k = 9$ selective attention samples (with 161 ASD participants and 194 TD participants) are shown in the bottom panel of Table 4. Again, positive effect sizes indicate that the global advantage is stronger in the TD than in the ASD group. For one of the studies (Porter and Coltheart 2006), matching on age and intellectual abilities was poor. However, the comparison between the ASD and TD group is not about absolute performance but instead about global relative to local performance within the same group. From this viewpoint, the lack of matching should not be overly problematic. However, this study might reflect age effects instead of ASD effects. We therefore present analyses with and without this sample included.

The overall effect sizes (reaction time and errors combined where possible) showed a fairly wide spread that ranged from $d = -0.44$ to $d = 1.12$, with a median of $d = 0.21$ ($d = 0.35$ without the Porter and Coltheart 2006, sample); their distribution is depicted in the bottom figure to the right of Table 4. A one-sample $t$ test against zero resulted in a large $p$ value ($t_8 = 1.7$, $p = .309$; without the Porter et al. sample: $t_7 = 2.3$, $p = .053$). When we only look at reaction times, the median effect size was $d = 0.11$ ($d = 0.12$ without the Porter et al. sample); when we only look at number of errors, the median effect size was $d = 0.01$ ($d = 0.41$ without the Porter et al. sample). Overall then, there is some tentative evidence for a stronger global advantage in TD as compared to ASD groups; however any such difference appears to be small and it did not show reliably across studies.

## Discussion

The question of visuo-spatial superiority in ASD compared to TD individuals is both complex and not easy to answer on the basis of research done so far. First, we review the results of the meta-analyses, and then we examine how the theories discussed in the introduction hold up against these results. Next, we look at findings from other tasks and revisit the theories we introduced earlier. Finally, we examine the issue of heterogeneity.

A meta-analysis of 35 studies comparing the Figure Disembedding performance of participants with ASD compared to TD found that participants with ASD are, on average, superior at this task. However, that difference is small and heterogeneity was enormous. A closer look at the funnel plot reveals that four studies (Brosnan et al. 2012; Pellicano et al. 2005, 2006; Shah and Frith 1983) are outliers. If we disregard them, the overall effect vanishes and heterogeneity is greatly reduced.

The effect size for the 24 studies of Block Design was small, and subject to substantial heterogeneity. While it is unknown what accounts for the heterogeneity we know that it is not due to the way the tests were scored, the year of publication or the participants' age.

Mental Rotation results are divided into two aspects: the intercept (non-rotational aspects of the task such as the speed with which one mentally compares the objects, decision making and response variables) and the slope (rotational aspects of the task, i.e. the speed with which one mentally rotates) (Falter et al. 2008). This meta-analysis found no consistent superiority for ASD on intercepts, which was carried out on eight studies. An analysis of *slopes*, for which we had three samples, even suggested a slight inferiority for the ASD group. However, given the small number of studies there is no strong evidence for the lack of an effect.

Finally, findings for Navon Figures suggest that the differences between TD and ASD are not as profound as assumed in the literature. The results of seven studies using the free choice paradigm point towards a stronger local preference in ASD affected people as compared to TD people, as hypothesized. Nonetheless, this difference in preference was small. For both the divided and selective attention paradigms, with four and nine samples respectively, there was tentative evidence that the relative advantage of global over local tasks is indeed stronger in TD people than in ASD affected people. However, any such differences were very small.

A good theory of autism should not only cover the deficits seen in the disorder but also be able to explain intact and superior abilities. How do the cognitive theories hold up with regards to the present meta-analysis results?

WCC theory presumes superior performance on Figure Disembedding and Block Design, however we only found unequivocal evidence for superior performance for Block Design. For Mental Rotation it predicts no difference or an inferior performance, which receives some support by our findings, but one must keep in mind that this result is based on only a small number of studies. Concerning Navon Figures, WCC assumes a stronger local preference by individuals with ASD in the free choice task and a weaker global-over-local advantage in the divided and selective attention tasks. While our analyses support the former, the support for the latter is also tentative.

Like WCC, EPF predicts superior performance for Block Design and Figure Disembedding, yet clear superiority was only found for Block Design. We presumed an equal performance prediction for Mental Rotation, which is supported by our data. For free-choice Navon Figures a local preference in ASD was predicted and our meta-analysis did find this to a greater extent than in TD, but the effect size was small. Our finding, that there is only a small difference (if at all present) between the two groups for the divided and selected attention paradigms, supports the idea of a flexible processing style. Neither WCC, nor EPF is able to deal with the large heterogeneity we observed.

The predications made by EMB theory on the basis of which tasks have shown sex differences in previous studies, are not supported by our findings: while we found a superior performance pattern for Block Design, EMB should predict no difference due to the lack of sex difference in this IQ subtest. Furthermore, the predicted superiority was not found for Mental Rotation and it received only equivocal support for Figure Disembedding. Since no known sex difference exists for Navon Figures, EMB would not predict a group difference, which our data agrees with. In sum, no group difference predictions made by EMB theory are supported by the results of the current meta-analysis. It appears that predictions made in the framework of EMB theory are grounded on assumptions of sex differences in visuo-spatial tasks, for which only weak support exists. As a result, EMB theory can only make predictions about a very confined number of perceptual and cognitive tasks, for which clear sex differences exist, and its relevance for explaining the autistic cognitive profile is therefore limited.

All things considered, we would like to remark that even though the data presented here are partially at odds with the presented theories, the latter are wider in scope and not solely about performance differences on the tasks we studied in this meta-analysis.

Apart from the classic tests of visuo-spatial ability there are also other tests that have been used to assess visuo-spatial skills. For instance, tasks involving mazes (Caron et al. 2004; Edgin and Pennington 2005; Pellicano et al. 2006) test visuo-spatial skills in a more realistic and life-like setting. All three studies using mazes showed typical performance for individuals with ASD, which speaks against a generalized visuo-spatial superiority in ASD across different types of tasks.

Simmons et al. (2009) argue that both Figure Disembedding and Block Design tasks may be regarded as visual search tasks. Plaisted et al. (1998) explicitly investigated visual search performance and found enhanced performance in ASD concluding that superior visual search skills might explain the superior performance on Figure Disembedding tasks. However, Shah and Frith (1983) investigated qualitatively which search strategies were employed by their participants during the performance on the children's version of the Figures Disembedding task (CEFT) and concluded that children with ASD rarely used a visual search strategy. Instead, the correct shape seemed to "pop out", resulting in an immediate response most of the time.

Simmons et al. (2009) summarized the results from several studies, and found that participants with ASD demonstrated superior performance on visual search tasks even when the task was made more complex such as in multiple conjunction searches. Yet the finding that participants do not seem to rely on a search strategy during the Figure Disembedding leads one to believe that superior visual search performance does not account alone for enhanced Figure Disembedding performance. Alternatively, Brosnan et al. (2012) suggested that the enhanced performance on Figure Disembedding seen in ASD is due to increased visual acuity.

Evidence from neuroimaging studies points towards different patterns of processing of visuo-spatial information in ASD. Three recent studies (Damarla et al. 2010; Lee et al. 2007; Manjaly et al. 2007) made use of functional Magnetic Resonance Imaging (fMRI) during a Figure Disembedding task. While none of these studies found performance differences between ASD and TD groups, differential brain activation was observed. Lee et al.'s (2011) study revealed that while TD control group participants showed activation in the left frontal cortex, the ASD group did not recruit these frontal regions. Frontal brain activation is particularly interesting because it is thought to be involved in the execution of complex tasks (Silk et al. 2006). Damarla et al. (2010) also found decreased frontal activation in ASD participants compared to TD controls. These findings suggest that individuals with ASD did not require as many resources as TD individuals in order to achieve the same level of performance, which supports the idea of superiority of Figure Disembedding in individuals with ASD. Furthermore, the fMRI study conducted by Manjaly et al. (2007) utilized a different control task than the previous two studies, and concluded that the ASD group showed enhanced local processing in visual areas, a conclusion supported by Damarla et al. (2010). Silk et al. (2006) investigated activity levels in frontoparietal networks during Mental Rotation. Similar to the Figure Disembedding tasks mentioned above, no differences in performance were found, but differential brain activation revealed processing differences between the ASD and TD group. Again, individuals with ASD showed less recruitment of frontal areas while exhibiting intact task performance.

In sum, literature on fMRI techniques combined with visuo-spatial tasks showed that while performance in ASD remains intact, differences are evident in the recruitment of frontal brain areas. The main finding was that individuals with ASD did not rely as much as TD individuals on frontal areas but rather on visual areas to solve visuo-spatial tasks. Together with Simmons et al. (2012) paper it does appear that any ASD advantage found in previous studies in both Figure Disembedding (Lee et al. 2007) and Block Design (Shah and Frith 1993) might stem from enhanced visual processing skills, which may not be as helpful in Mental Rotation and may thus explain the lack of consistent superiority in this task.

Heterogeneity was large in Block Design and Figure Disembedding, suggesting that the included studies differed in at least one aspect that systematically affected the results. Sampling differences might be a common source of heterogeneity. Many of the reviewed studies used a mixed sample of individuals with autism, AS and HFA, a distinction no longer made (American Psychiatric Association 2013), and it may be that such differences contribute to the observed heterogeneity.

Barbeau et al. (2013) advise that perceptual peaks may disappear according to matching decisions. Matching is commonly done on the basis of age, gender, and IQ (performance IQ, verbal IQ, or full scale IQ). However, it is also possible not to match groups based on an IQ profile but rather compare them on a range of performance measures, thereby capturing the specific ASD profile without losing sight of the overall difference. We tested whether the differences in IQ affected the results; while this was the case the effect was not strong enough to cause the large amount of heterogeneity.

Edgin and Pennington (2005) suggested that age might play a role: they found a significant age by group interaction in their maze study. However, if age were a promising moderator we would see a systematic increase or decrease in the age-column from top to bottom in our tables, as they are ordered by effect size; but this is not the case. Therefore, we conclude that age does not affect the results.

The diverse ranges of tests that have been used to assess visuo-spatial skills introduce another possible source of heterogeneity. Each of them taps slightly different skills and cognitive processes. Moreover, tests may be administered using the classical paper-and-pencil format or a computerized version. Another variable that might be promising to test in the future is stimulus complexity. The Block Design task lends itself to a variety of manipulations, e.g., manipulating the angles of rotation or oblique lines, which can be used to create Block Design patterns of various difficulties, which was done by Shah and Frith (1993). It was found that the only variable that produced an advantage for the ASD group was segmentation (slicing the design into smaller parts), on which they were considerably faster than the TD control group, i.e. when the design was not pre-segmented the ASD had an advantage over the TD group. Caron et al. (2006) manipulated difficulty of the design with perceptual cohesiveness, task uncertainty and matrix size in order to better understand the segmentation advantage, which they interpreted as evidence for enhanced locally oriented processing. Indeed, the only advantage was found for locally oriented processing. Like in Mental Rotation, it appears that the difference does not lie in the ability to rotate but instead, we may conclude that any advantage that is found is due to a visual rather than a spatial processing advantage.

In sum, we considered age, IQ matching and sampling as factors that may affect the outcome of a study on visuo-spatial skills. The present results indicate that age did not influence the results. IQ matching only had a weak effect on them. Happé and Ronald (2008) suggest that the triad of impairments apparent in autism may in fact be due to a combination of genetic, neurological and behavioral factors rather than a single underlying cause. If this were the case we might be able to account for the large heterogeneity in terms of the huge variability within the autistic spectrum.

In contrast to what has been reported in the literature, this meta-analysis presents evidence that does not confirm that people affected by ASD generally show superior performance at visuo-spatial tasks: no support was found for Mental Rotation and only weak support for an ASD advantage in Navon Figures; although we did find the expected local preference, the expected global disadvantage was small at best. Thus, Block Design and Figure Disembedding were the only tasks on which we found superiority but with a small effect size and huge heterogeneity. Observations about enhanced vision in visual search tasks in ASD, along with evidence from fMRI studies seem to support the idea that any advantage might be due to differential processing of visual stimuli. Large heterogeneity was found across studies that could not be easily accounted for. Speculatively, the large variability in skills and cognitive profiles within the autistic spectrum might be responsible for this heterogeneity. Overall, this meta-analysis reveals that the assumption that individuals with ASD generally excel in visuo-spatial tasks is wrong.

**Acknowledgments** We would like to thank Frank Renkewitz for helpful discussions on analytical issues, and we are grateful to all researchers who supplied additional information about their studies on request.

## References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Text Revision. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.

Barbeau, E. B., Soulieres, I., Dawson, M., Zeffiro, T. A., & Mottron, L. (2013). The level and nature of autistic intelligence III: Inspection time. *Journal of Abnormal Psychology, 122*(1), 295–301. doi:10.1037/a0029984.

Baron-Cohen, S. (1999). The extreme male-brain theory of autism. In H. Tager-Flusberg & H. Tager-Flusberg (Eds.), *Neurodevelopmental disorders* (pp. 401–429). Cambridge, MA: The MIT Press.

Beacher, F. C., Radulescu, E., Minati, L., Baron-Cohen, S., Lombardo, M. V., Lai, M., & Critchley, H. D. (2012). Sex differences and autism: Brain function during verbal fluency and mental rotation. *Plos ONE, 7*(6). doi:10.1371/journal.pone.0038355.

Behrmann, M., Avidan, G., Leonard, G., Kimchi, R., Luna, B., Humphreys, K., et al. (2006). Configural processing in autism and its relationship to face processing. *Neuropsychologia, 44*(1), 110–129. doi:10.1016/j.neuropsychologia.2005.04.002.

Bernardino, I., Mouga, S., Almeida, J., van Asselen, M., Oliveira, G., & Castelo-Branco, M. (2012). A direct comparison of local–global integration in autism and other developmental disorders: Implications for the central coherence hypothesis. *PLoS ONE, 7*(6), e39351. doi:10.1371/journal.pone.0039351.

Bölte, S., Duketis, E., Poustka, F., & Holtmann, M. (2011). Sex differences in cognitive domains and their clinical correlates in higher-functioning autism spectrum disorders. *Autism, 15*(4), 497–511. doi:10.1177/1362361310391116.

Bölte, S., Holtmann, M., Poustka, F., Scheurich, A., & Schmidt, L. (2007). Gestalt perception and local–global processing in high-functioning autism. *Journal of Autism and Developmental Disorders, 37*(8), 1493–1504. doi:10.1007/s10803-006-0231-x.

Bölte, S., Hubl, D., Dierks, T., Holtmann, M., & Poustka, F. (2008). An fMRI-study of locally oriented perception in autism: Altered early visual processing of the block design test. *Journal of Neural Transmission (Vienna, Austria: 1996), 115*(3), 545–552.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta analysis (Version 2.2) [software]*. Englewood, NJ: Biostat.

Brian, J., & Bryson, S. (1996). Disembedding performance and recognition memory in autism/PDD. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 37*(7), 865–872. doi:10.1111/j.1469-7610.1996.tb01482.x.

Brosnan, M. J., Gwilliam, L. R., & Walker, I. (2012). Brief report: The relationship between visual acuity, the embedded figures test and systemizing in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*(11), 2491–2497.

Burnette, C., Mundy, P., Meyer, J., Sutton, S., Vaughan, A., & Charak, D. (2005). Weak central coherence and its relations to theory of mind and anxiety in autism. *Journal of Autism and Developmental Disorders, 35*(1), 63–73. doi:10.1007/s10803-004-1035-5.

Caron, M., Mottron, L., Berthiaume, C., & Dawson, M. (2006). Cognitive mechanisms, specificity and neural underpinnings of visuospatial peaks in autism. *Brain, 129*, 1789–1802. doi:10.1093/brain/awl072.

Caron, M., Mottron, L., Rainville, C., & Chouinard, S. (2004). Do high functioning persons with autism present superior spatial abilities? *Neuropsychologia, 42*(4), 467–481.

Chen, F., Lemonnier, E., Lazartigues, A., & Planche, P. (2008). Non-superior disembedding performance in children with high-functioning autism and its cognitive style account. *Research in Autism Spectrum Disorders, 2*(4), 739–752. doi:10.1016/j.rasd.2008.03.003.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Conson, M., Mazzarella, E., Frolli, A., Esposito, D., Marino, N., Trojano, L., et al. (2013). Motor imagery in Asperger syndrome: Testing action simulation by the hand laterality task. *PLoS ONE, 8*(7), e70734. doi:10.1371/journal.pone.0070734.

Damarla, S. R., Keller, T. A., Kana, R. K., Cherkassky, V. L., Williams, D. L., Minshew, N. J., et al. (2010). Cortical underconnectivity coupled with preserved visuospatial cognition in autism: Evidence from an fMRI study of an embedded figures task. *Autism Research: Official Journal of the International Society for Autism Research, 3*(5), 273–279.

de Jonge, M. V., Kemner, C., & van Engeland, H. (2006). Superior disembedding performance of high-functioning individuals with autism spectrum disorders and their parents: The need for subtle measures. *Journal of Autism and Developmental Disorders, 36*(5), 677–683.

Deruelle, C., Rondan, C., Gepner, B., & Fagot, J. (2006). Processing of compound visual stimuli by children with autism and asperger syndrome. *International Journal of Psychology, 41*(2), 97–106. doi:10.1080/00207590500184610.

Drake, J. E. (2013). Is superior local processing in the visuospatial domain a function of drawing talent rather than autism spectrum disorder? *Psychology of Aesthetics Creativity and the Arts, 7*(2), 203–209. doi:10.1037/a0030636.

Edgin, J. O., & Pennington, B. F. (2005). Spatial cognition in autism spectrum disorders: Superior, impaired, or just intact? *Journal of Autism and Developmental Disorders, 35*(6), 729–745.

Falter, C. M., Arroyo, M., & Davis, G. J. (2006). Testosterone: Activation or organization of spatial cognition? *Biological Psychology, 73*(2), 132–140. doi:10.1016/j.biopsycho.2006.01.011.

Falter, C. M., Plaisted, K. C., & Davis, G. (2008). Visuo-spatial processing in autism—Testing the predictions of extreme male brain theory. *Journal of Autism and Developmental Disorders, 38*(3), 507–515.

Hamilton, A. F., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition, 113*(1), 37–44.

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36*(1), 5–25.

Happé, F., & Ronald, A. (2008). The 'fractionable autism triad': A review of evidence from behavioural, genetic, cognitive and neural research. *Neuropsychology Review, 18*(4), 287–304. doi:10.1007/s11065-008-9076-8.

Hines, M. (2004). *Brain gender*. New York, NY: Oxford University Press.

Hodgson, E. S., & McGonigle-Chalmers, M. (2011). Partial occlusion depiction and its relationship with field independence in children with ASD. *Autism, 15*(4), 473–495. doi:10.1177/1362361310363279.

Hönekopp, J. (2012). Digit ratio 2D:4D in relation to autism spectrum disorders, empathizing, and systemizing: A quantitative review. *Autism Research, 5*, 221–230.

Hugdahl, K., Thomsen, T., & Ersland, L. (2006). Sex differences in visuo-spatial processing: An fMRI study of mental rotation. *Neuropsychologia, 44*(9), 1575–1583. doi:10.1016/j.neuropsychologia.2006.01.026.

Ioannidis, J. P. A., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: A large survey. *Canadian Medical Association Journal, 176*, 1091–1096. doi:10.1503/cmaj.060410.

Ishida, R., & Kamio, Y. (2009). Perceptual distortions of visual illusions in children with high-functioning autism spectrum disorder. *Psychologia, 52*, 175–187.

Jarrold, C., Gilchrist, I. D., & 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*(4), 344–351.

Johnson, S. A., Blaha, L. M., Houpt, J. W., & Townsend, J. T. (2010). Systems factorial technology provides new insights on global–local information processing in autism spectrum disorders. *Journal of Mathematical Psychology, 54*(1), 53–72. doi:10.1016/j.jmp.2009.06.006.

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, 38*(5), 527–534. doi:10.1111/j.1469-7610.1997.tb01539.x.

Kaland, N., Mortensen, E. L., & Smith, L. (2007). Disembedding performance in children and adolescents with Asperger syndrome or high-functioning autism. *Autism: The International Journal of Research and Practice, 11*(1), 81–92.

Katagiri, M., Kasai, T., Kamio, Y., & Murohashi, H. (2013). Individuals with Asperger's disorder exhibit difficulty in switching attention from a local level to a global level. *Journal of Autism and Developmental Disorders, 43*(2), 395–403. doi:10.1007/s10803-012-1578-9.

Keehn, B., Brenner, L. A., Ramos, A. I., Lincoln, A. J., Marshall, S. P., & Mueller, R. (2009). Brief report: Eye-movement patterns during an embedded figures test in children with ASD. *Journal of Autism and Developmental Disorders, 39*(2), 383–387. doi:10.1007/s10803-008-0608-0.

Koh, H., & Milne, E. (2012). Evidence for a cultural influence on field-independence in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 42*(2), 181–190. doi:10.1007/s10803-011-1232-y.

Koldewyn, K., Jiang, Y. V., Weigelt, S., & Kanwisher, N. (2013). Global/local processing in autism: Not a disability, but a disinclination. *Journal of Autism and Developmental Disorders, 43*(10), 2329–2340. doi:10.1007/s10803-013-1777-z.

Lee, P. S., Foss-Feig, J., Henderson, J. G., Kenworthy, L. E., Gilotty, L., Gaillard, W. D., et al. (2007). Atypical neural substrates of embedded figures task performance in children with autism spectrum disorder. *NeuroImage, 38*, 184–193.

Levy, S. E., Mandell, D. S., & Schultz, R. T. (2009). Autism. *The Lancet, 374*(9701), 1627–1638. doi:10.1016/S0140-6736(09)61376-3.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*(6), 1479–1498. doi:10.2307/1130467.

Malisza, K. L., Clancy, C., Shiloff, D., Foreman, D., Holden, J., Jones, C., et al. (2011). Functional evaluation of hidden figures object analysis in children with autistic disorder. *Journal of Autism and Developmental Disorders, 41*, 13–22. doi:10.1007/s10803-010-1013-z.

Manjaly, Z. M., Bruning, N., Neufang, S., Stephan, K. E., Brieber, S., Marshall, J. C., et al. (2007). Neurophysiological correlates of relatively enhanced local visual search in autistic adolescents. *NeuroImage, 35*(1), 283–291.

Manning, J. T., Kilduff, L. P., & Trivers, R. (2013). Digit ratio (2D:4D) in Klinefelter's syndrome. *Andrology, 1*, 94–99.

McGrath, J., Johnson, K., Ecker, C., O'Hanlon, E., Gill, M., Gallagher, L., et al. (2012). Atypical visuospatial processing in autism: Insights from functional connectivity analysis. *Autism Research: Official Journal of the International Society for Autism Research, 5*(5), 314–330.

Mitchell, P., & Ropar, D. (2004). Visuo-spatial abilities in autism: A review. *Infant and Child Development, 13*(3), 185–198. doi:10.1002/icd.348.

Morgan, B., Maybery, M., & Durkin, K. (2003). Weak central coherence, poor joint attention, and low verbal ability: Independent deficits in early autism. *Developmental Psychology, 39*(4), 646–656. doi:10.1037/0012-1649.39.4.646.

Mottron, L., & Burack, J. (2001). Enhanced perceptual functioning in the development of autism. In Charman. Burack & Zelazo. Yirmiya (Eds.), *The development of autism: Perspectives from theory and research* (pp. 131–148). Mahwah, NJ: Erlbaum.

Mottron, L., Burack, J., Stauder, J., & Robaey, P. (1999). Perceptual processing among high-functioning persons with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40*(2), 203–211. doi:10.1017/S0021963098003333.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders, 36*(1), 27–43.

Nakano, T., Kato, N., & Kitazawa, S. (2012). Superior haptic-to-visual shape matching in autism spectrum disorders. *Neuropsychologia, 50*(5), 696–703. doi:10.1016/j.neuropsychologia.2011.12.024.

Navon, D. (1977). Forest before trees—Precedence of global features in visual-perception. *Cognitive Psychology, 9*(3), 353–383. doi:10.1016/0010-0285(77)90012-3.

Nydén, A., Hagberg, B., Goussé, V., & Rastam, M. (2011). A cognitive endophenotype of autism in families with multiple incidence. *Research in Autism Spectrum Disorders, 5*(1), 191–200. doi:10.1016/j.rasd.2010.03.010.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32*(7), 1081–1105.

Palmer, A. R. (2000). Quasireplication and the contract of error: Lessons from sex ratios, heritabilities, and fluctuating asymmetry. *Annual Review of Ecology and Systematics, 31*, 441–480.

Pellicano, E., Gibson, L., Maybery, M., Durkin, K., & Badcock, D. R. (2005). Abnormal global processing along the dorsal visual pathway in autism: A possible mechanism for weak visuospatial coherence? *Neuropsychologia, 43*, 1044–1053. doi:10.1016/j.neuropsychologia.2004.10.003.

Pellicano, E., Maybery, M., Durkin, K., & Maley, A. (2006). Multiple cognitive capabilities/deficits in children with an autism spectrum disorder: 'Weak' central coherence and its relationship to theory of mind and executive control. *Development and Psychopathology, 18*(1), 77–98. doi:10.1017/S0954579406060056.

Plaisted, K. C., 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, 39*(5), 777–783.

Plaisted, K. C., Swettenham, J., & Rees, L. (1999). Children with autism show local precedence in a divided attention task and global precedence in a selective attention task. *Journal of Child Psychology and Psychiatry, 40*(5), 733–742. doi:10.1111/1469-7610.00489.

Planche, P., & Lemonnier, E. (2011). Le pic d'habileté aux tâches visuospatiales chez les autistes de haut niveau désigne-t-il vraiment un déficit du traitement global? (Does the islet of ability on visuospatial tasks in children with high-functioning autism really indicate a deficit in global processing?). *L'Encéphale, 37*, 10–17. doi:10.1016/j.encep.2010.03.007.

Porter, M. A., & Coltheart, M. (2006). Global and local processing in Williams syndrome, autism, and Down syndrome: Perception, attention, and construction. *Developmental Neuropsychology, 30*(3), 771–789. doi:10.1207/s15326942dn3003_1.

Pring, L., Hermelin, B., & Heavey, L. (1995). Savants, segments, art and autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 36*(6), 1065–1076. doi:10.1111/j.1469-7610.1995.tb01351.x.

Pring, L., Ryder, N., Crane, L., & Hermelin, B. (2010). Local and global processing in savant artists with autism. *Perception, 39*(8), 1094–1103.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–322). New York: Russel Sage Foundation.

Rinehart, N., Bradshaw, J., Moss, S., Brereton, A., & Tonge, B. (2000). A typical interference of local detail on global processing in high-functioning autism and Asperger's disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(6), 769–778. doi:10.1017/S002196309900596X.

Rondan, C., & Deruelle, C. (2007). Global and configural visual processing in adults with autism and Asperger syndrome. *Research in Developmental Disabilities, 28*(2), 197–206. doi:10.1016/j.ridd.2006.02.007.

Ropar, D., & Mitchell, P. (2001). Susceptibility to illusions and performance on visuospatial tasks in individuals with autism. *Journal of Child Psychology and Psychiatry, 42*(4), 539–549.

Scheurich, A., Fellgiebel, A., Müller, M. J., Poustka, F., & Bölte, S. (2010). Erfasst der FBT lokale visuelle Informationsverarbeitung bei Autismus-Spektrum-Störungen? *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 38*(2), 103–110. doi:10.1024/1422-4917/a000017.

Schlooz, W. M., Hulstijn, W., van den Broek, P. A., van der Pijll, A. M., Gabreëls, F., van der Gaag, R. J., et al. (2006). Fragmented visuospatial processing in children with pervasive developmental disorder. *Journal of Autism and Developmental Disorders, 36*(8), 1025–1037. doi:10.1007/s10803-006-0140.

Shah, A., & Frith, U. (1983). An islet of ability in autistic children: A research note. *Journal of Child Psychology and Psychiatry, 24*(4), 613–620.

Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the Block Design task? *Journal of Child Psychology and Psychiatry, 34*(8), 1351–1364.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Silk, T. J., Rinehart, N., Bradshaw, J. L., Tonge, B., Egan, G., & O'Boyle, M. (2006). Visuospatial processing and the function of prefrontal-parietal networks in autism spectrum disorders: A functional MRI study. *The American Journal of Psychiatry, 163*(8), 1440–1443.

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research, 49*(22), 2705–2739.

Smalley, S. L., & Asarnow, R. F. (1990). Brief report: Cognitive subclinical markers in autism. *Journal of Autism and Developmental Disorders, 20*(2), 271–278. doi:10.1007/BF02284724.

Soulieres, I., Zeffiro, T. A., Girard, M. L., & Mottron, L. (2011). Enhanced mental image mapping in autism. *Neuropsychologia, 49*(5), 848–857.

South, M., Ozonoff, S., & McMahon, W. M. (2007). The relationship between executive functioning, central coherence, and repetitive behaviors in the high-functioning autism spectrum. *Autism, 11*(5), 437–451. doi:10.1177/1362361307079606.

Spek, A. A., Scholte, E. M., & Van Berckelaer-Onnes, I. A. (2011). Local information processing in adults with high functioning autism and asperger syndrome: The usefulness of neuropsychological tests and self-reports. *Journal of Autism and Developmental Disorders, 41*(7), 859–869. doi:10.1007/s10803-010-1106-8.

Spencer, M. D., Holt, R. J., Chura, L. R., Calder, A. J., Suckling, J., Bullmore, E. T., et al. (2012). Atypical activation during the embedded figures task as a functional magnetic resonance imaging endophenotype of autism. *Brain, 135*, 3469–3480. doi:10.1093/brain/aws229.

Taylor, L. J., Maybery, M. T., Grayndler, L., & Whitehouse, A. J. O. (2013). Evidence for distinct cognitive profiles in autism spectrum disorders and specific language impairment. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-013-1847-2.

van Lang, D. J., Bouma, A., Sytema, S., Kraijer, D. W., & Minderaa, R. B. (2006). A comparison of central coherence skills between adolescents with an intellectual disability with and without comorbid autism spectrum disorder. *Research in Developmental Disabilities, 27*, 217–226.

Wang, L., Mottron, L., Peng, D., Berthiaume, C., & Dawson, M. (2007). Local bias and local-to-global interference without global deficit: A robust finding in autism under various conditions of attention, exposure time, and visual angle. *Cognitive Neuropsychology, 24*(5), 550–574. doi:10.1080/13546800701417096.

Wechsler, D. (1974). *Wechsler Intelligence Scale for children—Revised*. New York: The Psychological Corporation.

Wechsler, D. (1981). *Wechsler adult intelligence scales—Revised*. New York: The Psychological Corporation.

White, S. J., & Saldana, D. (2011). Performance of children with autism on the embedded figures test: A closer look at a popular task. *Journal of Autism and Developmental Disorders, 41*(11), 1565–1572.

Williams, D. L., Goldstein, G., & Minshew, N. J. (2006). Neuropsychologic functioning in children with autism: Further evidence for disordered complex information-processing. *Child Neuropsychology, 12*(4–5), 279–298. doi:10.1080/09297040600681190.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. (1971). *A manual for the embedded figures test*. California: Consulting Psychologists Press.