



4

ANALOGICAL REASONING IN AUTISM

A systematic review and meta-analysis

*Kinga Morsanyi, Dušan Stamenković,
and Keith J. Holyoak*

Introduction

Analogical reasoning – the ability to find and exploit similarities based on relations among entities, rather than solely on the entities themselves – is a key mechanism underlying human intelligence and creativity (Gentner, 2010; Halford, Wilson, & Phillips, 2010; Holyoak, 2012). Among the species currently extant, the ability to formulate explicit relations and reason about them may be unique to *homo sapiens* (Penn, Holyoak, & Povinelli, 2008). Preschool children can already use simple analogies to solve problems (Brown, Kane, & Echols, 1986; Holyoak, Junn, & Billman, 1984); however, the ability continues to develop over childhood (e.g., Gentner & Toupin, 1986; Goswami, 1989), and is linked to improvements in executive functions (Richland & Burchinal, 2012) and to maturation of the prefrontal cortex (Knowlton, Morrison, Hummel, & Holyoak, 2012). For adults, advanced analogical reasoning impacts a broad range of human endeavors, including mathematics education (Richland, Zur, & Holyoak, 2007), engineering design (Chan & Schunn, 2015), and scientific discovery (Dunbar, 1995). Analogies are also ubiquitous in discourse, and can be used for various communicative purposes, such as to subtly express opinions or convey humor. Here is an example from Nasreddin, a Turkish philosopher of the 13th century:

Knowledge is like the carrot, few know by looking at the green top that the best part, the orange part, is there. Like the carrot, if you don't work for it, it will wither away and rot. And finally, like the carrot, there are a great many donkeys and jackasses that are associated with it.

Broadly speaking, reasoning by analogy involves finding coherent correspondences between disparate situations, focusing on *relations* between objects rather than

specific features of individual objects (for an overview see Holyoak & Thagard, 1995). Analogy seems to both require and promote cognitive flexibility, enabling transfer of knowledge and procedures between different contexts to solve novel problems (e.g., Gick & Holyoak, 1980, 1983). Furthermore, it has been argued that development of relational reasoning is closely linked to language, especially the acquisition of relational vocabulary (e.g., Gentner & Rattermann, 1991). Given these general characteristics, one might well expect to observe deficits in analogical reasoning among people with Autistic Spectrum Disorder (ASD), who are often characterized as lacking in cognitive flexibility, oriented toward detailed perceptual features, weak in central coherence, and impaired in language processing, including metaphor comprehension¹ (e.g., Frith, 2003; Kalandadze, Norbury, Nærland, & Næss, 2018; Pellicano, Maybery, Durkin, & Maley, 2006). It is therefore surprising that some empirical findings (e.g., Dawson, Soulières, Gernsbacher, & Mottron, 2007; Scott & Baron-Cohen, 1996) indicate that analogical reasoning may actually be spared in autism, or even constitute an area of relative cognitive strength.

If analogical reasoning is indeed spared in ASD despite apparent deficits in cognitive processes that have often been linked to this type of reasoning, then it may be useful to reconsider conceptions of the role played by analogy in normal cognitive development. ASD is characterized by a wide range of deficits in social communication (Jones, Gliga, Bedford, Charman, & Johnson, 2014). The status of analogical reasoning in individuals affected by ASD therefore has potential implications for design of interventions to compensate for deficits in social understanding associated with ASD. For individuals with ASD, spared analogical ability could provide a basis for therapeutic interventions (Green et al., 2017; McGregor, Whiten, & Blackburn, 1998; Swettenham, 1996).

By performing a systematic literature review and meta-analysis, the present study examines the accumulated evidence regarding analogy performance in the ASD population as it compares to that in typically developing controls. To set the stage, we will provide a brief overview of analogical reasoning, first in the typically developing population and then in the ASD population.

Analogical reasoning in the typically developing population

In general, analogical reasoning involves relating one situation (the *source* analog, usually familiar) to another (the *target* analog, usually more novel). Work in the psychometric tradition has focused on four-term or “proportional” analogies in the form $A:B:: C:?$. Often a list of options is provided from which the best completion for the missing D term is to be selected. Such problems can be stated verbally, as in “a toolbox for a hammer is like a lunchbox for a(n) ...” (response options: sandwich [correct], axe, picnic basket, dog). The list of response options typically includes distractor items that are conceptually related to the C term, but do not create a relational match with the $A:B$ pair (e.g., “picnic basket” in the above example). Problems can also be created using geometrical patterns or pictures, allowing distractors that are perceptually similar to the C term. In proportional analogy

problems, the source and target often play relatively symmetrical roles (although generally the second pair of terms is viewed as the target, because an inference is required to generate or evaluate the *D* term).

Spearman (1923, 1927) reviewed evidence that high correlations are found between performance in solving analogy problems and what he termed the *g* (general) factor in intelligence. Raven (1938) developed the Raven's Progressive Matrices test, which requires selection of a geometric figure to fill an empty cell in a two-dimensional matrix (typically 3×3 ; see Figure 4.1). Much like a geometric proportional analogy, Raven's matrices require participants to extract and apply information based on visuospatial relations. (See Carpenter, Just, & Shell, 1990, and Lovett & Forbus, 2017, for analyses of strategies for solving Raven's Progressive

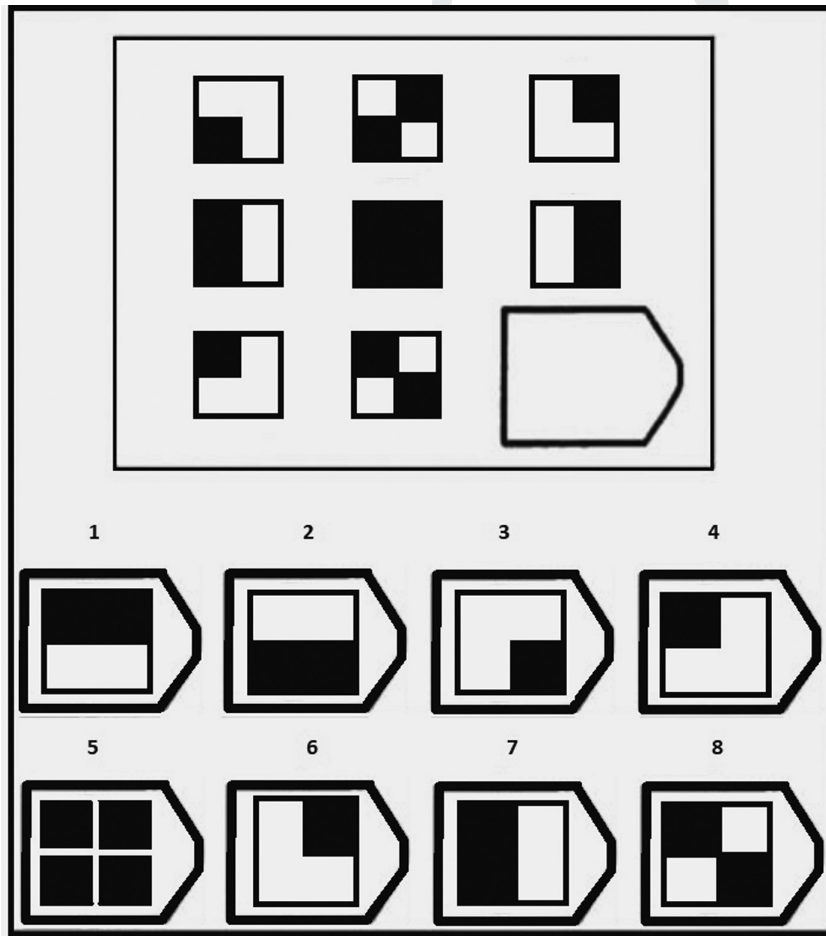


FIGURE 4.1 A matrix reasoning problem similar to Raven's Progressive Matrices. Option 3 is the correct response.

Matrices problems.) Scores on Raven's matrices tend to be highly correlated with performance on analogy tests (verbal as well as pictorial; Snow, Kyllonen, & Marshalek, 1984). Cattell (1971) distinguished between two components of *g*: *crystallized* intelligence, which depends on previously learned information or skills (often verbal in nature), and *fluid* intelligence, which involves reasoning with novel information. Cattell confirmed Spearman's (1946) observation that analogy tests and performance on Raven's matrices correlate highly with *g*, and clarified that these relational tests primarily measure fluid intelligence (although verbal analogies also depend on crystallized intelligence).

In addition to relatively formal analogy problems presented in the proportional format, psychological studies have often investigated more complex verbal and visual analogical reasoning in situations that require adapting a solution illustrated in the source analog to solve a novel target problem in a different domain. For example, Gick and Holyoak (1980) investigated transfer from a story about a general dividing his forces to capture a central fortress (source) to a medical problem (target) that could be solved by constructing an analogous "convergence" solution (apply converging weak rays to destroy an inoperable stomach tumor). In this sort of problem-solving paradigm, at least four major components of analogical processing can be distinguished: (1) initial *retrieval* of the source analog when the target is presented; (2) finding a *mapping*, or systematic correspondences between elements of the source and target; (3) making *inferences* to generate an analogous solution to the target; and (4) *generalizing* to create a more abstract schema that captures the commonalities between the two examples (Holyoak, 2012). An emerging schema can be refined if additional analogs are encountered (e.g., Catrambone & Holyoak, 1989).

In general, analogical mapping and inference are facilitated by overlap of properties relevant to goal attainment in the source analog (Holyoak & Koh, 1987; Keane, 1985; Keane, Ledgeway, & Duff, 1994). Early work by Gick and Holyoak (1980, 1983) revealed that when the source and target analogs are drawn from semantically disparate domains (e.g., a military story and a medical problem), a dissociation is observed between spontaneous analogical transfer (retrieval is relatively difficult) and transfer after a hint to make use of the source (mapping and inference are relatively easy). Spontaneous transfer is much more frequent when the source and target are drawn from similar domains (Keane, 1987). Spontaneous transfer can be elicited at most once in an experimental session (because noticing the relevance of a source analog to a target is likely to prompt the participant to use an analogy strategy to solve other target problems). This practical consideration (which conflicts with the need to administer multiple trials to achieve reliable estimates of analogy ability for an individual participant) has limited experimental research on spontaneous use of analogy (but see Kubricht, Lu, & Holyoak, 2017).

Many studies of analogical reasoning have used variations of a mapping task, which requires the participant to identify correspondences between specific objects in the source and target. Instructions sometimes specify that a relational match is to be sought, but may instead leave the match criterion relatively vague (then scoring how often a relational match is chosen; see Waltz, Lau, Grewal, & Holyoak, 2000).

Richland, Morrison, and Holyoak (2006) introduced a mapping task using “scene analogies” constructed using simple relations and objects familiar to young children. Examples of four sets of picture pairs are shown in Figure 4.2. On each trial in this task, participants are presented with two pictures (the source analog at the top, and the target analog at the bottom), with an arrow pointing at one item in the top picture (here, the child in a highchair). The participant is asked to indicate the item that “is the same part of the pattern” in the bottom picture as the item marked by the arrow in the top picture.

As summarized in the caption for Figure 4.2, these scene analogies vary in how many instances of the same critical relation (“feeds” in these examples) need to be considered (one or two), and whether or not the target picture includes a perceptual/semantic distractor. To generate the relational match (child maps to bird), the participant must avoid matches based on object features, when present (see Figure 4.2, Panels B and D). Studies using scene analogies have consistently shown that children select the correct relational match more often with increasing chronological age (over a range from preschool to early adolescence), especially in the more difficult conditions involving multiple relations and/or presence of a distractor (e.g., Richland et al., 2006; Simms, Frausel, & Richland, 2018). Studies using other analogy tests have confirmed the general increase in relational responding over a broad age range (e.g., Gentner & Rattermann, 1991; Whitaker, Vendetti, Wendelken, & Bunge, 2018).

Age-related changes in performance on analogy tests are closely linked to measures of executive functioning (Miyake et al., 2000), which encompass measures of working-memory capacity, inhibitory control, and cognitive flexibility (e.g., the ability to shift task set). Computational models of analogy imply that at least the first two of these three components are directly relevant to analogical reasoning: working memory is required to integrate multiple relations, and inhibitory control is required to select a relational match in the face of semantic and/or perceptual distractors (Hummel & Holyoak, 1997, 2003). A longitudinal study found that measures of executive functioning in preschool children predict their performance on a verbal analogy test administered at age 15 years (Richland & Burchinal, 2012; see also Simms et al., 2018). The predictive power of executive skills (especially inhibitory control) proved to be statistically separable from that of early vocabulary knowledge (which also had an impact on later analogy performance).

At the neural level, explicit analogical reasoning depends on the frontoparietal control network (for a review see Krawczyk, 2012). Frontal patients are seriously impaired in analogical tasks (e.g., Krawczyk et al., 2008; Morrison et al., 2004; Waltz et al., 1999). Neuroimaging studies with normal college students indicate that the rostrolateral prefrontal cortex (particularly on the left side; Bunge, Helskog, & Wendelken, 2009) is activated for analogies that involve multiple relations (Christoff et al., 2001) or that connect disparate semantic categories (Green, Kraemer, Fugelsang, Gray, & Dunbar, 2010). In addition, the inferior frontal gyrus is active whenever salient but irrelevant information must be inhibited (Cho et al., 2010; Whitaker et al., 2018).

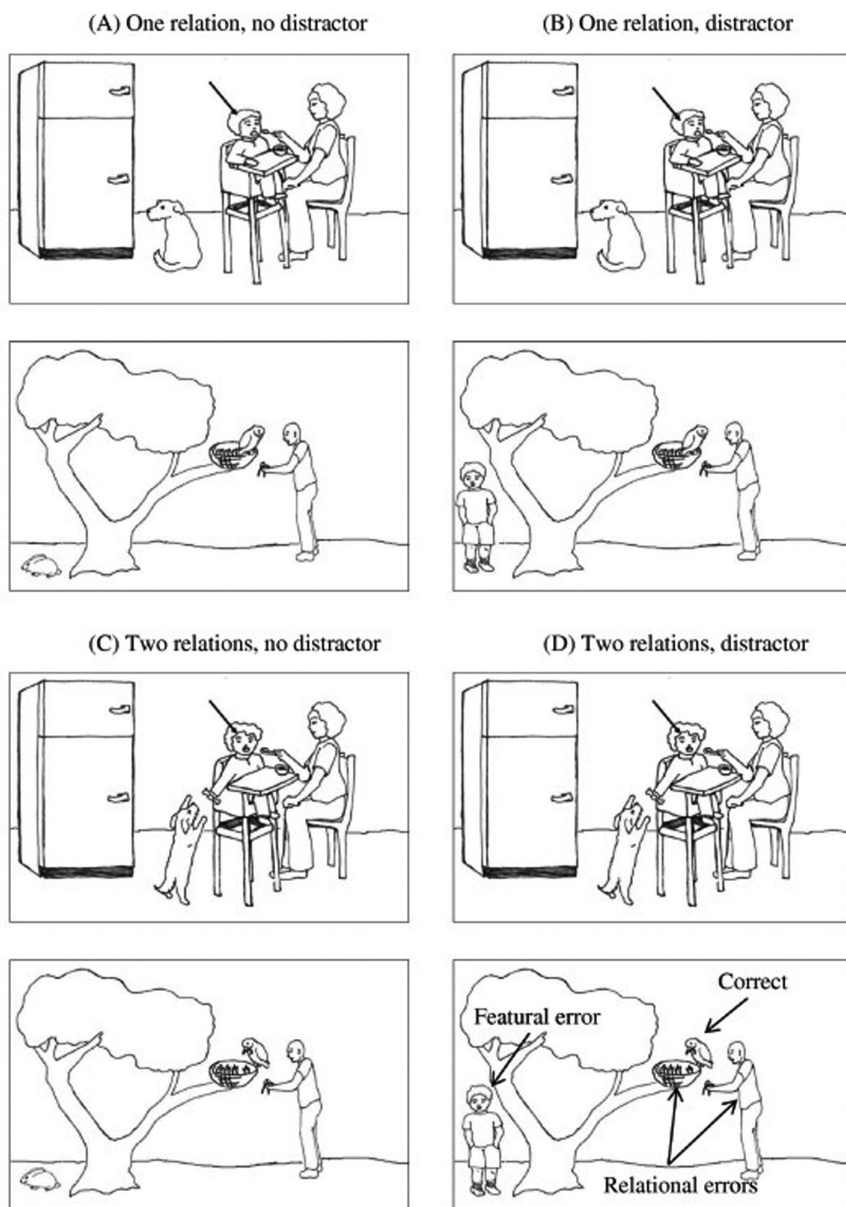


FIGURE 4.2 Examples of four conditions for Scene Analogy task problems, which vary relational complexity (A, B: one relation; C, D: two relations) and the presence of a competing object match distractor (A, C: no distractor; B, D: distractor). Children were asked to find the object in the target/bottom picture that corresponded with the object of interest in the source/top picture (identified here with an arrow). The coding of possible responses in the target (correct relational match, featural error, and relational errors) is labeled in Panel D, bottom.

Reprinted with permission from Simms et al. (2018)

Analogical reasoning in the ASD population

The nature of autistic intelligence has been the subject of long-standing interest, tracing back to the very first clinical description of autism. Kanner (1943) claimed that autistic individuals with profound intellectual disabilities may nonetheless maintain some “islets of abilities” (i.e., some special skills or talents with a relatively narrow scope). Indeed, this concept was central to Kanner’s (1943) diagnostic criteria for autism. Since then, a large number of studies have investigated analogical reasoning in autism, including various forms of nonverbal analogies (matrix analogies based on perceptual relations, pictorial analogies, and scene analogies), but with a notable lack of research involving verbal analogical reasoning. The age of participants in these studies typically ranges from early childhood to young adulthood, with some studies also including middle-aged participants. Although some of these studies included participants with intellectual impairments, most studies focused on high-functioning individuals.

Early researchers investigating autism (e.g., Bartak, Rutter, & Cox, 1975; Lockyer & Rutter, 1970; Wurst, 1976) described a characteristic pattern in the performance of autistic individuals on the Wechsler intelligence scales, whereby they performed relatively well on subtests that assess visuospatial skills (in particular, the block design and object assembly subtests).² These findings, together with early clinical reports indicating that autistic individuals have outstanding skills in visual search (e.g., Wing & Wing, 1976), formed the basis for a cognitive theory of autism: the weak central coherence theory (WCC; Frith, 1989; Frith & Happé, 1994). Central coherence is the tendency to process incoming information in its context to extract higher-level meaning, at the expense of processing and remembering exact details. A central tenet of the original WCC theory was that autistic people show a bias toward featural, piecemeal processing, at the expense of integrating information into meaningful wholes. This proposal was later modified to suggest that although autistic people might show outstanding skills in processing exact details, their preference for featural processing does *not* necessarily come at the expense of global processing (e.g., Happé, 1999; Happé & Frith, 2006). Under this revised interpretation, WCC is viewed as a strength and a special ability, rather than a form of impairment.

The uneven profile of autistic individuals on the Wechsler intelligence scales has received continuing empirical support since the early 1970s. Stevenson and Gernsbacher (2013) presented an overview of 38 previously published studies that reported the subtest scores of autistic participants on the Wechsler scale. The typical finding in these studies (consistent with earlier reports) was that autistic individuals scored highest on the block design subtest (and lowest on the comprehension subtest) of the Wechsler scale.

Based on these findings regarding the autistic cognitive profile, it might be expected that autistic people would be particularly good at cognitive tasks that require processing of precise and exact details, but they might struggle at tasks that require a focus on higher-level relations, of which analogical reasoning would appear to offer a prime example. However, initial investigations of analogical

reasoning in ASD yielded mixed evidence. Scott and Baron-Cohen (1996) found no impairments in analogical reasoning in autistic children as compared to typically developing controls matched on mental age, as well as compared to a group of chronological- and mental-age-matched children with intellectual disabilities. By contrast, Reed (1996) reported that autistic children performed worse on two out of four analogical reasoning tasks as compared to typically developing controls matched on verbal mental age. Later investigations of analogical reasoning continued to yield mixed findings, some reporting impaired analogical reasoning in autism (Tzuriel & Groman, 2017), and some finding no group difference (e.g., Green et al., 2017; Morsanyi & Holyoak, 2010). Both Green et al. (2017) and Morsanyi and Holyoak (2010) reported that autistic and typically developing control participants did not differ in their sensitivity to the presence of distractors. In addition to perceptual distractors (see the examples above related to the scene analogies), these studies also included semantic distractors (see the example above in our description of proportional analogies). Moreover, the latter studies found that autistic participants were able to reason successfully not only with formal analogies, but also with problems based on thematic materials (i.e., materials that necessitated the processing of semantic meaning and context), even when the tasks included social content and multiple relations. However, none of these analogy tasks involved materials that clearly required understanding of the emotions or intentions of other people.

Dawson et al. (2007) found additional evidence for an uneven autistic cognitive profile by assessing performance on both the Wechsler intelligence scale and Raven's Progressive Matrices. These investigators compared the IQ profiles of children and adults with autism to those of control groups drawn from the typically developing population. Their study yielded two important findings. First, as previously reported, the Wechsler IQ profile for autistic participants was uneven, with some characteristic peaks and troughs in performance. Specifically, the full-scale Wechsler IQ of the sample of autistic children was at the 26th percentile, but their scores on the verbal comprehension subtest were even lower (around the 10th percentile³), whereas their scores on the block design subtest were much higher (around the 60th percentile). In addition to their "bumpy" Wechsler IQ profile, and despite their low full-scale IQ as measured by the Wechsler test, the autistic children achieved scores on Raven's matrices in the normal range (at the 56th percentile). Dawson et al.'s study also included a comparison group of typically developing children. As can be expected in typical populations, these children obtained approximately the same IQ scores regardless of the type of IQ measure that was used (i.e., verbal, nonverbal, and full-scale IQ on the Wechsler scale, as well as full-scale IQ assessed by Raven's matrices). However, a problem with this comparison sample was that the typically developing participants were not matched to the children with autism either on Wechsler or Raven IQ (in fact, the autistic group scored higher on both of these IQ measures), making it difficult to interpret the group differences in IQ profiles.

In general, for the interpretation of group differences on any cognitive task, it is necessary to match the samples on some basic characteristics. For example, in the case of performance on Raven's matrices, we would expect that nonverbal Wechsler intelligence will be a relevant factor, as well as the age of participants (especially in the case of developmental samples). If groups are matched on these variables, it is possible to argue more strongly that any differences found in matrix reasoning reflect genuine group differences and not just the fact that we are comparing participants from different age groups or with different levels of cognitive ability.

Dawson et al. (2007) also reported results for a group of autistic adults. These adults with autism scored at the same level on Raven's matrices as the control participants (at the 83th percentile), but had significantly lower Wechsler IQ (at the 50th percentile) than controls. This finding with adults replicated the general pattern found with children: a discrepancy between Wechsler and Raven IQ scores for the autistic sample, in contrast to a more even profile for the control group. Similar findings were later reported by Soulières, Dawson, Gernsbacher, and Mottron (2011) for children and adults diagnosed with Asperger's syndrome. Unfortunately, the latter study also did not match participants on their IQ, with the participants diagnosed with Asperger's scoring significantly lower than the typically developing sample on both the Wechsler scale and Raven's matrices. Overall, the most striking pattern observed in the studies by Dawson et al. (2007) and Soulières et al. (2011) was that autistic individuals showed a much larger difference than typically developing controls between their full-scale IQ as measured by the Wechsler scale versus Raven's matrices (although there was a tendency for autistic individuals to score lower on the verbal than on the nonverbal subtests of the Wechsler scale).

In summary, the empirical evidence regarding analogical/relational reasoning in ASD is mixed, with reports of impairment in autism, of no group differences, or even of outstanding performance for people with ASD. Here we present a systematic review of empirical investigations of analogical reasoning in autism, including studies that assessed performance on Raven's matrices. By extracting data from multiple studies and combining them in a meta-analysis, we aimed to obtain a clearer picture regarding the presence or absence of group differences in analogical reasoning. Given that analogical reasoning relies heavily on cognitive resources, we focused on studies in which the ASD and control samples were matched on chronological age and (verbal or full-scale) IQ. After excluding studies that did not match the groups, we found that all of the included studies used nonverbal analogy tasks. Thus, our review and meta-analysis does not cover verbal analogical reasoning in autism.

Meta-analyses can also be used to explore factors that might moderate group differences. For example, group differences might depend on the level of cognitive functioning of the participants or the type of analogy problems that are used in the study. In particular, the pattern of results might differ for Raven's matrices (which do not rely on factual knowledge), as compared to analogy problems with thematic content.

Method for systematic review and meta-analysis

Search strategy

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines in the design and reporting of our systematic review and meta-analysis (see more details in the Appendix).

Study inclusion criteria

We selected articles for the meta-analysis on the basis of the following predetermined criteria:

1. The paper reported the results of an original research study including an analogical reasoning task or Raven's Progressive Matrices, where scores on the relevant task were reported independently of other measures used in the study. Studies were excluded if insufficient data were available to calculate effect sizes and relevant data could not be obtained from the author(s).
2. Participants had to be diagnosed with ASD by experienced clinicians using the Diagnostic and Statistical Manual of Mental Disorders (DSM) or International Classification of Diseases (ICD) diagnostic criteria.
3. The study had to include a typically developing comparison group, matched to the ASD group on chronological age and either full-scale or verbal IQ.
4. When the study included the Raven test as an outcome measure, it was clear that the groups were not intentionally matched on Raven scores. (Please see Appendix for further details on the manuscript screening process.)

Coding

Study characteristics (title, authors, and publication year) were coded for descriptive purposes. For statistical analyses, we coded the number of ASD and typically developed participants, and the inferential statistics reported in the papers. The statistics of primary interest were means and standard deviations relating to performance of each group on the analogy task and/or Raven's matrices. For studies that included neuroimaging data, only the behavioral results were coded. When a study included multiple tasks relating to analogical reasoning, the results were combined into a single measure for the study, because computing effect sizes multiple times based on data from the same sample can distort the overall results (Borenstein, Hedges, Higgins, & Rothstein, 2009). When data from multiple subgroups were reported (i.e., high-functioning autism and also Asperger's syndrome), these were combined into a single score using the algorithm provided by the Comprehensive Meta-Analysis software. None of the studies included verbal analogy tests as a primary dependent measure; all tests were either Raven's matrices or some form of pictorial analogy test. Type of task (i.e., Raven's matrices

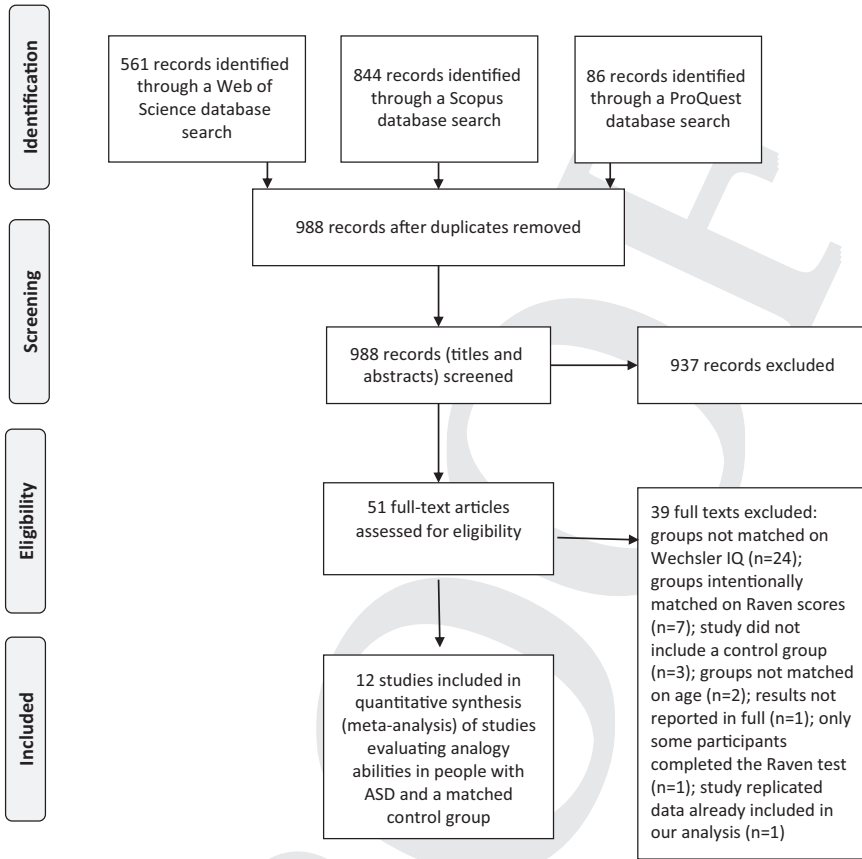


FIGURE 4.3 The PRISMA flow diagram describing the search process, the number of papers included in each step, and the reasons for exclusion

or pictorial analogies) was coded. We also coded the average age and the IQ of the participants in the ASD and typically developing samples, and the strategy used for matching the groups on intellectual ability (i.e., verbal or full-scale IQ). Matching strategy was not used as a variable in our analyses, because only three studies matched the samples on verbal ability, and the type of verbal ability measure used in these studies varied widely.

Meta-analytic procedures and analysis

All statistical analyses were conducted using the Comprehensive Meta-Analysis (CMA) software, version 3 (Biostat). A 95% confidence interval was computed for each effect size to indicate if it was statistically different from zero (i.e., if the confidence interval does not include zero, the effect is considered significant).

Results of meta-analysis

A total of 12 studies, involving 324 ASD and 335 control participants, were included in the analyses. The characteristics of the participants included in each study in the meta-analysis (number of participants in the ASD and control groups, and their mean chronological age and IQ) and the type of analogy task used in each study, are listed in Table 4.1. The studies are listed in the table in rank order of effect size (Hedge's g) for the group difference in analogical reasoning performance, from negative (i.e., control group performs better) to positive (the ASD group performs better). Overall, the studies included participants spanning a broad age range, from mid-childhood to middle age. Participants' mean IQ scores ranged from average to above average. Ten studies used Raven's matrices as an outcome measure, four studies used pictorial analogies, and two studies used both.⁴ The ASD and control samples were matched on full-scale Wechsler IQ in most studies, with the exception of the study of Tzuriel and Groman (2017), which matched the samples on verbal intelligence (the vocabulary subtest of the Wechsler Intelligence Scale for Children; WISC); Bodner, Williams, Engelhardt, and Minshew (2014), which matched the samples on Wechsler verbal IQ; and Terzi, Marinis, and Francis (2016), which matched the samples on the Peabody Picture Vocabulary Test.⁵

TABLE 4.1 Participant characteristics and type of analogy task in studies included in the meta-analysis

<i>Study</i>	<i>N</i>	<i>Mean age</i>	<i>Mean IQ</i>	<i>Task type</i>
	<i>ASD (TD)</i>	<i>ASD (TD)</i>	<i>ASD (TD)</i>	
Tzuriel & Groman (2017)*	32 (32)	9.33 (9.33)	102 (102)**	pictorial
Green et al. (2017)	41 (41)	10.98 (11.08)	115.32 (116.32)	pictorial
Sahyoun et al. (2009)	42 (21)	19.14 (18.43)	101.2 (103.7)	pictorial, RPM***
Hermann et al. (2013)	20 (20)	42.40 (35.10)	123.1 (123.1)	RPM
Bodner et al. (2014)*	37 (48)	11.85 (12.09)	107.08 (110.38)	RPM
Yamada et al. (2012)	25 (26)	30.70 (32.2)	106.9 (103.9)	RPM
Soulières et al. (2009)	15 (18)	22.40 (21.72)	100.87 (106.22)	RPM
Terzi et al. (2016)*	20 (20)	6.92 (6.58)	86 (89)**	RPM
Morsanyi & Holyoak (2010)	23 (49)	13.58 (13.14)	98.47 (99.97)	pictorial, RPM
Barbeau et al. (2013)	35 (30)	22.80 (20.3)	102.9 (105.0)	RPM
So et al. (2014)	17 (13)	8.82 (9.18)	94.52 (95.0)	RPM
Hayashi et al. (2008)	17 (17)	9.20 (9.50)	96.7 (99.8)	RPM

*Verbal IQ scores are reported for these studies

**Typically developing (TD) mean IQ not directly reported in paper, but ASD and TD groups were matched on verbal IQ (estimated from WISC-IV Vocabulary subtest)

***RPM stands for Raven's Progressive Matrices

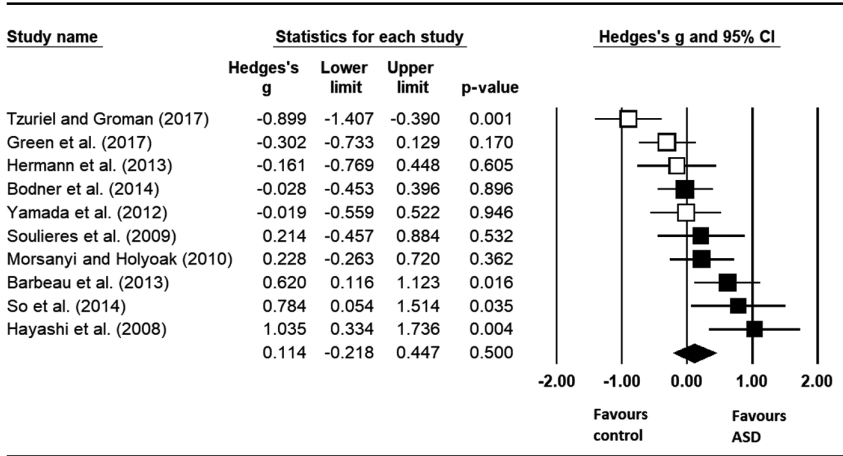


FIGURE 4.4 Hedges' *g* effect sizes with 95% confidence intervals for group differences in analogical reasoning between individuals with ASD and age- and Wechsler intelligence-matched controls
 The overall mean effect size is presented in the bottom line (and marked by \blacklozenge in the figure). Hollow squares are used to mark effect sizes for studies that used pictorial analogies.

Figure 4.4 presents the effect size of the group differences in analogical reasoning (Hedges' *g* with 95% CIs) between individuals with ASD and matched controls. Overall, the results showed no group differences in analogical reasoning ($g=.10, p=.503$). The heterogeneity between studies was significant ($Q(11)=38.49, p<.001, I^2=71.42$). Given that a considerable proportion of the variance in effect sizes appeared to be attributable to moderator variables, we conducted a meta-regression analysis with mean age and IQ of the ASD participants and type of analogy task (pictorial analogy vs. Raven's matrices) as predictor variables. The model was significant ($Q(3)=13.76, p=.003$), explaining 65% of the variance in effect sizes between studies. Type of analogy task was significantly related to the effect size of group differences ($p=.006$), with studies using Raven's matrices being more likely to yield trends favoring the ASD group.⁶ In contrast, neither chronological age ($p=.76$) nor IQ of the participants ($p=.21$) yielded a significant effect.

Given that the type of analogy task had a significant influence on the size of group differences, with Raven's matrices tending to yield an advantage for the ASD groups, we performed an additional analysis using the ten out of 12 studies listed in Table 4.1 that included Raven's matrices as an outcome measure.⁷ The results, shown in Figure 4.5, revealed a significant, medium-sized group difference in performance on the Raven's matrices ($g=.37, p=.001$), favoring the ASD group.

We also intended to run a meta-regression to investigate the effects of chronological age and Wechsler IQ of the participants on group differences. However, it was not possible to perform this analysis, due to an overfitting problem (most likely

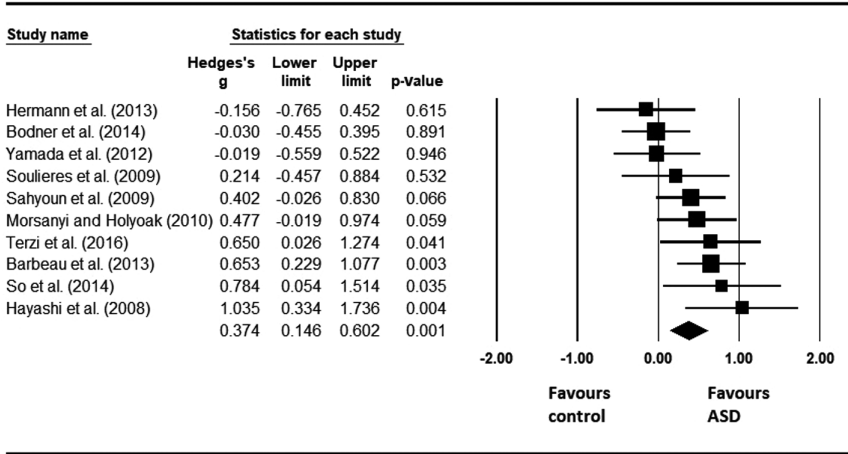


FIGURE 4.5 Hedges' *g* effect sizes with 95% confidence intervals for group differences in performance on Raven's matrices between individuals with ASD and age- and Wechsler intelligence-matched typically developing controls. The overall mean effect size is presented in the bottom line (and marked by \blacklozenge in the figure).

because of the relatively small number of studies). Nevertheless, when we created a regression plot to investigate the association between the effect size of group differences and the average IQ of ASD participants within each study (Figure 4.6), we found an association between the two (i.e., a strong and significant negative correlation: $r(8) = -.76, p = .011$). That is, the effect size of group differences was close to zero in the case of participants with an IQ of around 100 or higher, but the effect size increased among participants with lower IQ levels. This pattern suggests that the group difference on Raven's matrices (i.e., relatively superior performance in the ASD group compared to age- and IQ-matched controls) may be restricted to participants with Wechsler IQs below 100).

Figure 4.7 presents a funnel plot to assess the possible impact of publication bias. This analysis was based on the 12 studies included in the meta-analysis presented in Figure 4.4. The funnel plot is organized with a measure of standard error for each study on the vertical axis and effect size on the horizontal axis (Borenstein et al., 2009). Studies with smaller standard errors appear toward the top of the graph and tend to cluster around the mean effect size; those with larger standard errors appear toward the bottom, and tend to be more dispersed. In the absence of publication bias, studies are expected to be symmetrically distributed on each side of the overall mean effect size. If publication bias is present, a higher concentration of studies is expected on one side of the mean toward the bottom of the plot. We used the fail-Safe *N* statistic (Rosenthal, 1979) to statistically test for publication bias. This analysis revealed no evidence of such a bias ($p = .283$). However, this result should be treated with caution due to the relatively small number of studies included in the analysis (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006).

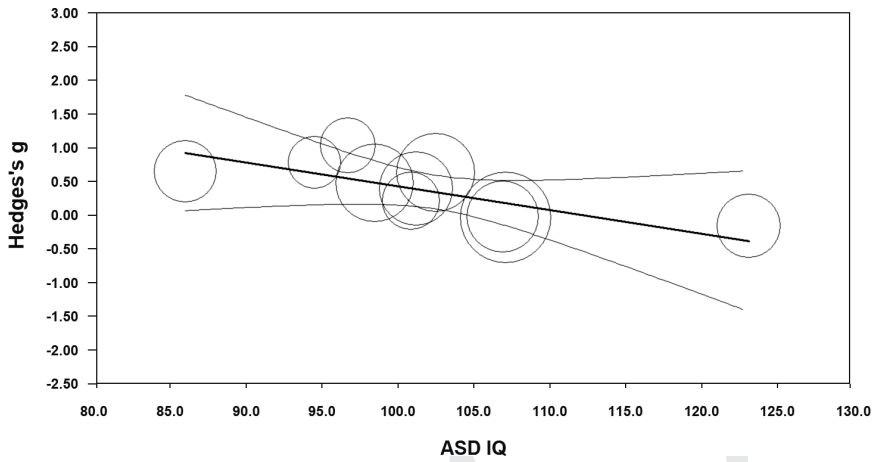


FIGURE 4.6 Plot presenting the regression of Hedges' *g* effect sizes for group differences in Raven scores in each study on the average IQ of the ASD participants in the study

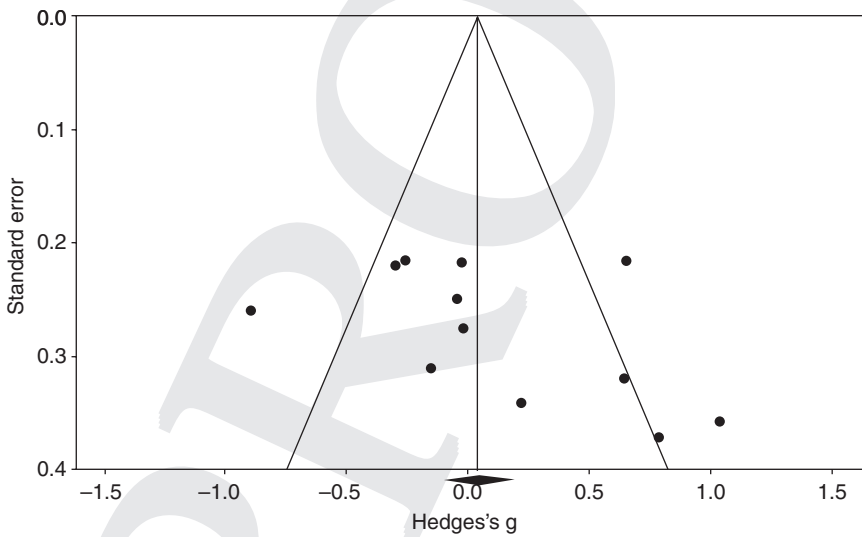


FIGURE 4.7 Funnel plot to assess the potential impact of publication bias

General discussion

Summary of meta-analysis findings

This chapter has presented a systematic review and meta-analysis of studies on analogical reasoning in autism and typically developing controls, including papers that investigated performance on the Raven's Progressive Matrices test and on various

types of analogy problems presented in pictorial formats. We investigated this topic because contradictory claims have been made in the literature about both the existence and the direction of group differences in analogical reasoning. We considered analogical reasoning and performance on Raven's matrices together, given that both types of task rely on reasoning with relations and performance on these tasks tends to be highly correlated (e.g., Snow et al., 1984).

The results of our literature search highlighted a number of issues regarding research into analogical reasoning in autism. Although many papers have investigated this issue, a large proportion of them had serious methodological limitations. In particular, many papers compared ASD participants to control participants who were not matched on chronological age and intelligence (in fact, we had to exclude about 75% of the papers investigating analogical reasoning in autism due to these issues). Matching on chronological age is especially important in the case of child samples, as analogical reasoning ability improves with age over childhood and adolescence (e.g., Richland et al., 2006; Whitaker et al., 2018). Matching on general intellectual functioning is also desirable in light of the high cognitive demands imposed by explicit analogical reasoning (Holyoak, 2012). In particular, given that ASD might be associated with deficits in executive functioning (see Hill, 2004; Rajendran & Mitchell, 2007 for reviews), as well as with impairments in language and communication skills (e.g., Brynskov, Krøjgaard, & Eigsti, 2016; Eigsti, de Marchena, Schuh, & Kelley, 2011), matching on chronological age only might result in samples with markedly different cognitive profiles, making it impossible to draw conclusions about the specific impact of an ASD diagnosis on analogical reasoning performance. We were able to identify 12 studies that met our inclusion criteria.

Overall, our meta-analysis found no reliable difference in analogical reasoning ability between ASD participants and carefully matched controls. Nevertheless, there was significant heterogeneity across the findings from the studies. The results of a meta-regression analysis indicated that the effect size of the difference between groups varied significantly between studies based on pictorial analogies as compared to studies that used Raven's matrices, with a relative ASD advantage in performance on Raven's matrices as compared to pictorial analogies. We also performed an additional meta-analysis including only those studies that used Raven's matrices as an outcome measure. This meta-analysis revealed a reliable ASD advantage, with an effect size that was small to medium. Moreover, this group difference was moderated by the Wechsler IQ of participants, with a larger group difference observed in those studies that tested participants with relatively low scores. This is a remarkable finding, particularly considering the relatively restricted range of participants' level of intelligence (average to above average) in these studies.

If a similar trend were observed among participants with an intellectual disability, the ASD advantage in performance on Raven's matrices might be even more pronounced. Indeed, some studies that did not meet all the inclusion criteria for our meta-analysis have found that ASD participants with relatively low Wechsler IQ perform in the normal range on Raven's matrices, or that there at least appears to be a significant discrepancy in their performance between the two IQ tests

(e.g., Chen, Planche, & Lemonnier, 2010; Dawson et al., 2007; Nader, Courchesne, Dawson, & Soulières, 2016; Soulières et al., 2011).

These variations in group differences as a function of type of test and IQ level cast a new light on some earlier findings regarding ASD. First, and remarkably, the effect of participants' Wechsler IQ level was very strongly related to group differences in performance on Raven's matrices, resolving the apparent contradiction in the literature regarding the presence/lack of group differences in performance on Raven's matrices. Second, these findings enrich our understanding of the nature of autistic intelligence. In particular, in line with a series of earlier studies, they provide further evidence that autism is characterized by a relative strength in certain aspects of intelligence (see e.g., Stevenson & Gernsbacher, 2013, for a review). Specifically, Raven's matrices are highly formal (i.e., knowledge-free) in nature, and depend on the ability to code perceptual relations among geometric patterns (similar to block design, a test that reveals an ability spared in ASD).⁸ In addition, the Raven test can be considered a gold standard measure of fluid intelligence (Mackintosh, 1998); hence good performance on the test (especially in the case of participants with low Wechsler IQ) has important implications for clinical and research purposes. For example, these findings can inform intervention approaches that build on existing strengths in autistic individuals, and can also contribute to discussions regarding the best procedures for matching ASD and control samples on intellectual functioning in research studies (e.g., Jarrold & Brock, 2004; Mottron, 2004).

Third, heterogeneity is a defining feature of autism, existing not only at the behavioral level, but also in terms of the genetic origins of ASD (e.g., Betancur, 2011). Past debate regarding cognitive functioning in ASD has centered on the question of whether high-functioning autism and Asperger's syndrome are characterized by distinguishable cognitive profiles (e.g., Ehlers et al., 1997; Ozonoff, South, & Miller, 2000). The IQ interaction identified by the present analysis can be interpreted in terms of the general picture of ASD being associated with an uneven IQ profile, which becomes more pronounced at lower levels of Wechsler intelligence. Specifically, for groups matched with normal to above normal IQ scores, both ASD and typically developing participants have relatively even IQ profiles, which include those abilities that support success on Raven's matrices. When groups are instead matched at lower IQ levels, the typically developing participants continue to exhibit a relatively even (though lower) profile across subtests. In contrast, lower-IQ ASD participants have an uneven profile, with deficits primarily on subtests sensitive to verbal comprehension and other skills that are *not* critical for solving Raven's matrices problems (Stevenson & Gernsbacher, 2013). Thus, when matched at lower IQ levels, performance of the ASD group on Raven's matrices may be superior to that of the typically developing group.

Limitations of current knowledge about analogical reasoning in ASD

Our literature search and meta-analyses highlight some important gaps in the existing literature on analogical reasoning in autism. The majority of the studies

focused on Raven's matrices. This was the case both for the studies that we initially identified as potentially relevant for our review, and for the studies that we included in our meta-analysis. Although Raven's matrices provide an important measure of fluid intelligence, which correlates highly with success on standardized analogy tests, Raven's matrices by no means provide a comprehensive assessment of analogical ability. Indeed, some studies that were returned by our search did not use Raven's matrices as a measure of analogical reasoning ability, but rather as an intelligence measure. This fact partially explains why we had to exclude so many of these studies (i.e., because they did not include additional measures of IQ).

Just four of the studies included in our meta-analysis used pictorial analogy problems other than (or in addition to) Raven's matrices; one of these reported a performance deficit for ASD participants (Tzuriel & Groman, 2017) and three reported no reliable differences between ASD and typically developing groups (Morsanyi & Holyoak, 2010; Green et al., 2017; Sahyoun, Soulières, Belliveau, Mottron, & Mody, 2009). Notably, the first two studies examined analogies based on pictured scenes, some of which depicted interactions between people. For example, the test used by Green et al. includes an analogy between two photos, each showing a teacher making a presentation to a group of students (with different people in each photo). On the face of it, the lack of an ASD deficit for problems with social content is particularly unexpected given the general picture of ASD as a disorder of social understanding. However, it is unclear to what extent social understanding (as opposed to more basic recognition of visual properties and relations) was actually required to solve the specific analogies used in laboratory experiments. It has not been established whether these pictorial analogies depend for their solutions on recognition of human intentions or emotions.

While studies using pictorial analogies other than Raven's matrices are scarce, studies examining *verbal* analogies are almost non-existent. ASD might be associated with language impairments, which is likely to impact performance on verbal analogy tests. One study (Tzuriel & Groman, 2017) included a verbal analogy test as one of a set of measures to assess proverb understanding. Relative to the typically developing group, performance of ASD participants was substantially impaired on this verbal test of analogical mapping. Nevertheless, as we have seen in the case of Raven's matrices, group differences might be moderated by some aspects of cognitive functioning (in the case of verbal analogies, most likely those related to language skills). It will therefore be important to further investigate verbal analogical reasoning in ASD.

In the related domain of metaphor comprehension, work with typically developing populations has led to vigorous debates concerning the possible role that may be played by analogical reasoning (for a review see Holyoak & Stamenković, 2018). In the literature on ASD, efforts have been made to distinguish possible problems with comprehension of metaphor and other types of figurative language from more general impairments of core language competencies (for reviews see Kalandadze et al., 2018; Vulchanova, Saldaña, Chauboun, & Vulchanov, 2015). Additional research that examines both analogical reasoning and metaphor

comprehension abilities in matched ASD and typically developing groups could help to disentangle the interrelated cognitive processes involved in these tasks.

The range of analogy problems used in ASD studies appears even more limited when considered in light of the four basic stages of analogical reasoning we sketched earlier: retrieval, mapping, inference, and generalization. Regardless of format, all the tests so far used have tapped some aspect of mapping (finding correspondences between source and target) and/or inference (identifying the best completion). Conspicuously lacking are any studies that have examined the ability to retrieve a relevant source when a target is encountered, or the ability to learn a more general schema in the aftermath of drawing an analogy between a source and target. The analogy tests used in ASD research have also been tightly structured. It is an open question whether ASD may lead to impairments in less constrained situations in which spontaneous analogical transfer (or spontaneous generalization) is possible. Indeed, it has been proposed that autism is characterized by reduced generalization (Plaisted, 2001).

In addition to using a broader range of tasks, it would also be advantageous to measure reaction times (e.g., Keane et al., 1994) to further investigate the efficiency of analogical reasoning in autism, given that previous studies often have found no differences between groups in accuracy.

An additional limitation of the studies that we reviewed is that, given the requirement of a matched, typically developing control sample, we mostly focused on studies using participants with IQs in the normal or above average ranges. About 44% of individuals with ASD have normal or above average IQs, with an additional 25% in the borderline range of IQs between 71–85 (Baio et al., 2018). Our review therefore cannot give a full picture of analogical reasoning in autism. Although we might expect that individuals with very severe intellectual disabilities will not be capable of analogical reasoning, a potential future direction could be to recruit ASD participants with Raven's performance in the normal range, and match them to a typically developing sample on this basis. These samples could then be compared on various other aspects of intellectual functioning. On the basis of the findings such as those of Dawson et al. (2007) and Soulières et al. (2011), we might expect that individuals with ASD who perform in the normal range on Raven's matrices may nevertheless exhibit some deficits in other aspects of intellectual functioning.

Implications for theories of analogical reasoning and of ASD

The fact that basic analogy processes (at least for mapping and inference with pictorial problems) are spared in ASD has implications both for understanding the functions of analogical reasoning in the typically developing population and for understanding the nature of cognitive functioning in individuals with ASD. With respect to the former, some theorists (e.g., Gentner, 2010) have argued that analogy is a fundamental cognitive process that underlies cognitive and linguistic development from infancy on. From this perspective it would seem surprising that a large group of people – those with ASD – can seemingly exhibit intact analogy ability, yet

show striking impairments in acquiring social understanding and in language comprehension. Without denying that analogy plays important roles in normal development, it seems clear that analogy by itself (at least the type of formal and explicit analogical reasoning preserved in ASD) is far from sufficient to ensure successful development of social understanding and language. Other forms of learning from experience, such as gradual accumulation of predictive regularities, may be more important than explicit analogical reasoning in guiding typical cognitive and linguistic development.

The spared analogical abilities observed in ASD also suggest refinements in current models of autistic functioning. On the face of it, solving problems of the type exemplified by Raven's matrices requires attention to visuospatial relations between geometrical forms (rather than solely perceptual details of individual forms), and also requires integration of constraints provided by multiple relations. Thus if autism is characterized in terms of "weak central coherence" (e.g., Frith & Happé, 1994), analogical reasoning would appear to provide an island of spared coherence (cf. Morsanyi & Holyoak, 2010).

More recent developments in general theories of autistic functioning have focused on the possibility that ASD is primarily a disorder of a system for *predictive coding* (e.g., Pellicano & Burr, 2012; Sinha et al., 2014; van Boxtel & Lu, 2013). The phenomenology of ASD is sometimes characterized as living in a state of perpetual surprise (e.g., Sinha et al., 2014). Unless an event is virtually an exact repetition of one previously experienced, it is likely to be interpreted in isolation, without drawing connections to past events. Interpreted within a Bayesian framework, ASD is characterized by weak priors (i.e., previous beliefs and experience fail to guide current processing) and strong likelihoods (i.e., focusing on the immediate perceptual experience). The result can be a kind of "hyper-realism" because the data at hand is interpreted directly, without being biased by prior expectations. However, normal perception is usually guided by priors that help to resolve ambiguities and fill in perceptual gaps. Lacking the ability to anticipate perceptual inputs, a person with ASD is prone to experience anxiety (which may be alleviated by tactics such as production of repetitive behaviors).

The predictive coding account of ASD focuses on the impact of the neurological condition on perception. The salient symptoms involving deficits in social understanding are attributed to the particular dependence of this ability on use of subtle priors to deal with dynamic events involving hidden causes (such as the intentions of others). It has been argued that the islands of proficiency often reported in people with ASD (e.g., spared or enhanced abilities in visual search, mathematics, calendar calculation, music, and drawing) tend to involve domains that are strongly rule-based (Sinha et al., 2014), minimizing uncertainty of outcomes. Explicit analogical reasoning, within a context that establishes clear criteria for problem solution, has this rule-based quality. For example, a Raven's matrix problem has a definite solution derivable by consistent application of rules for generating the missing cell in the matrix. The problem is static, and does not require drawing any connection with prior semantic knowledge.

From the perspective of the predictive coding hypothesis, analogical reasoning would be expected to become more challenging for people with ASD if the solution is less clearly rule-based, if prior semantic knowledge is relevant, or if the context does not establish a task set to reason by analogy. The evidence that analogical reasoning by ASD participants is not impaired even for meaningful pictorial analogies based in part on social relations is encouraging, implying that analogical abilities in ASD are not confined to purely formal problems. Nonetheless, as we noted above, there is almost no evidence yet available concerning the capabilities of people with ASD when faced with situations that require spontaneous retrieval of a source analog in the absence of a context that elicits explicit search for one. In this respect, it is worth noting that studies of high-level reasoning processes (other than analogical reasoning) have revealed a reduced tendency in ASD individuals to spontaneously retrieve knowledge from memory and use this information during reasoning and judgment, as compared to typically developing controls (e.g., McKenzie, Evans, & Handley, 2010; Morsanyi, Handley, & Evans, 2010; Pijnacker, Geurts, van Lambalgen, Buitelaar, & Hagoort, 2010; Pijnacker et al., 2009). ASD individuals have also been shown to struggle with reasoning about imaginary content (e.g., Leevers & Harris, 2000; Morsanyi & Handley, 2012; Scott, Baron-Cohen, & Leslie, 1999).

But even if spontaneous analogical reasoning proves to be difficult for people with ASD, the fact that they are capable of explicit use of analogy when suitably cued offers promise for the development of therapeutic interventions and compensation strategies (Green et al., 2017). For example, it has been proposed that conceiving of the mind as a camera inside one's head (i.e., using a picture-in-the-head analogy) could help children with autism to better reason about other people's mental states (McGregor et al., 1998; Swettenham, 1996). More generally, even if much of cognitive development normally proceeds by implicit learning of probabilistic cues, coupled with near-automatic use of priors to interpret experience, it may still be possible to teach people with ASD to use explicit analogical reasoning as an alternative mode of thinking and learning. Analogy may be a particularly valuable island of preserved cognitive competence, from which people with ASD (like those who are typically developing) can build bridges to explore the relative unknown.

Acknowledgments

Preparation of this chapter was supported by a project grant from the Ministry of Education, Science, and Technological Development of the Republic of Serbia (179013) to Dušan Stamenković, and by NSF Grant BCS-1827374 to Keith Holyoak.

Notes

- 1 Although there are similarities between analogies and metaphors, there is currently a debate about the extent to which the production and understanding of metaphors and analogies are based on similar processes (see Holyoak & Stamenković, 2018, for a review).

- 2 The Wechsler Intelligence Scale has a child and an adult version, both including several verbal and nonverbal subtests. The most relevant subtests for our discussion are block design, object assembly, and verbal comprehension. In the case of the block design subtest (which is part of both the child and adult scales), participants are presented with blocks with various color patterns on different sides. Their task is to arrange the blocks in a way that corresponds to a set pattern as quickly as possible. The block design test measures spatial visualization and motor skill. The object assembly subtest is only included in the child version of the scale; it requires children to complete pieces of a puzzle to form a recognizable object (such as a ball or a calf) as quickly as possible. This subtest measures similar skills to the block design test, but uses everyday shapes rather than abstract patterns. The verbal comprehension subtest is included in both the adult and child version of the scale. The items require participants to answer questions about what should be done in certain circumstances, the meaning of proverbs, why certain societal practices are followed, etc. This subtest assesses practical judgment, common sense, and the ability to understand and adapt to social customs.
- 3 Although the sample as a whole did not represent participants with intellectual disabilities (which would be defined as the 2nd percentile or lower), the average verbal comprehension score of these children was between one and two standard deviations below the population mean, indicating significant difficulties.
- 4 Morsanyi and Holyoak (2010) and Sahyoun et al. (2009) collected data for both pictorial analogy tests and Raven's matrices. In order to have a clean separation by test type, only the data for pictorial tests is included in Table 4.1. To ensure that each study contributed a single independent set of analogy results, scores for the pictorial analogy tests in both studies were averaged to create a composite score.
- 5 Bodner et al. (2014) tested both child and adult participants, but we only included the results from the child sample because the adult ASD and control samples were not matched on IQ. The child samples were matched on verbal IQ only, but there was a significant difference between the groups in full-scale Wechsler IQ, with lower scores in the ASD group.
- 6 One might argue that it is difficult to draw firm conclusions regarding pictorial analogies on the basis of these analyses, as only four studies included pictorial analogies. Nevertheless, of these four studies, only one reported a significant group difference, which is in line with our finding of no group differences for pictorial analogies. The overall sample size from these four studies (138 autistic and 143 typically developing participants) is also reasonably large.
- 7 For the Raven's matrices meta-analysis, we included the Raven's matrices data reported by Morsanyi and Holyoak (2010) and Sahyoun et al. (2009), rather than their data for pictorial analogies used in the analyses shown in Figure 4.4.
- 8 See further discussion of the type of analogical reasoning that the RPM measures in the next section.
- 9 In either a fixed-effect or a random-effect analysis, each study is weighted by the inverse of its variance (i.e., precision). The difference is that in random-effect models the variance includes both the original (within-studies) variance and the estimate of the between-studies variance. When the heterogeneity of effect sizes between studies is large, the weighting of within-study variance decreases, and the result will be more similar to the unweighted average effect size across the studies (Borenstein et al., 2009).

References

- Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., ... & Durkin, M. S. (2018). Prevalence of autism spectrum disorder among children aged

- 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR Surveillance Summaries*, 67, 1–23.
- Barbeau, E. B., Soulières, I., Dawson, M., Zeffiro, T. A., & Mottron, L. (2013). The level and nature of autistic intelligence III: Inspection time. *Journal of Abnormal Psychology*, 122, 295–301.
- Bartak, L., Rutter, M., & Cox, A. (1975). A comparative study of infantile autism and specific development receptive language disorder. I. The children. *The British Journal of Psychiatry*, 126, 127–145.
- Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. *Brain Research*, 1380, 42–77.
- Bodner, K. E., Williams, D. L., Engelhardt, C. R., & Minshew, N. J. (2014). A comparison of measures for assessing the level and nature of intelligence in verbal children and adults with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 8, 1434–1442.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley.
- Brown, A. L., Kane, M. J., & Echols, C. H. (1986). Young children's mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development*, 1, 103–121.
- Brynskov, C., Krojgaard, P., & Eigsti, I. M. (2016). Language and communication in children with autism: Do research and clinical practice converge in Denmark? *Nordic Psychology*, 68, 100–113.
- Bunge, S. A., Helskog, E. H., & Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *NeuroImage*, 46, 338–342.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97, 404–431.
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1147–1156.
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton-Mifflin.
- Chan, J., & Schunn, C. D. (2015). The impact of analogies on creative concept generation: Lessons from an *in vivo* study in engineering design. *Cognitive Science*, 39, 126–155.
- Chen, F., Planche, P., & Lemonnier, E. (2010). Superior nonverbal intelligence in children with high-functioning autism or Asperger's syndrome. *Research in Autism Spectrum Disorders*, 4, 457–460.
- Cho, S., Moody, T. D., Ferdinandino, L., Mumford, J. A., Poldrack, R. A., Cannon, T. D., Knowlton, B. J., & Holyoak, K. J. (2010). Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. *Cerebral Cortex*, 20, 524–533.
- Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, 14, 1136–1149.
- Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, 20, 775–803.
- Dawson, M., Soulières, I., Gernsbacher, M. A., & Mottron, L. (2007). The level and nature of autistic intelligence. *Psychological Science*, 18, 657–662.
- Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson (Eds.), *Mechanisms of insight* (pp. 365–395). Cambridge, MA: MIT Press.

- Ehlers, S., Nydén, A., Gillberg, C., Sandberg, A. D., Dahlgren, S. O., Hjelmqvist, E., & Odén, A. (1997). Asperger syndrome, autism and attention disorders: A comparative study of the cognitive profiles of 120 children. *Journal of Child Psychology and Psychiatry*, *38*, 207–217.
- Eigsti, I. M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, *5*, 681–691.
- Frith, U. (1989). Autism and “theory of mind”. In C. Gillberg (Ed.), *Diagnosis and treatment of autism* (pp. 33–52). New York, NY: Plenum Press.
- Frith, U., & Happé, F. (1994). Autism: Beyond “theory of mind”. *Cognition*, *50*, 115–132.
- Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*, 752–775.
- Gentner, D., & Rattermann, M. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). London, England: Cambridge University Press.
- Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*, 277–300.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306–355.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38.
- Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cognitive Development*, *4*, 251–268.
- Green, A. E., Kenworthy, L., Gallagher, N. M., Antezana, L., Mosner, M. G., Krieg, S., Dudley, K., Ratto, A., & Yerys, B. E. (2017). Social analogical reasoning in school-aged children with autism spectrum disorder and typically developing peers. *Autism*, *21*, 403–411.
- Green, A. E., Kraemer, D. J., Fugelsang, A. J., Gray, J. R., & Dunbar, K. N. (2010). Connecting long distance: Semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cerebral Cortex*, *20*, 70–76.
- Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*, *14*, 497–505.
- Happé, F. (1999). Autism: Cognitive deficit or cognitive style? *Trends in Cognitive Sciences*, *3*, 216–222.
- Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*, 5–25.
- Hayashi, M., Kato, M., Igarashi, K., & Kashima, H. (2008). Superior fluid intelligence in children with Asperger’s disorder. *Brain and Cognition*, *66*, 306–310.
- Hedges, L. V. (1981). Distribution theory for Glass’s estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107–128.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hermann, I., Haser, V., van Elst, L. T., Ebert, D., Müller-Feldmeth, D., Riedel, A., & Konieczny, L. (2013). Automatic metaphor processing in adults with Asperger syndrome: A metaphor interference effect task. *European Archives of Psychiatry and Clinical Neuroscience*, *263*, 177–187.
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560.

- Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, 8, 26–32.
- Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 234–259). New York, NY: Oxford University Press.
- Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, 15, 323–340.
- Holyoak, K. J., Junn, E. N., & Billman, D. O. (1984). Development of analogical problem-solving skill. *Child Development*, 55, 2042–2055.
- Holyoak, K. J., & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin*, 144, 641–671.
- Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.
- Holyoak, K. J., Junn, E. N., & Billman, D. O. (1984). Development of analogical problem-solving skill. *Child Development*, 55, 2042–2055.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104, 427–466.
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220–263.
- Jarrold, C., & Brock, J. (2004). To match or not to match? Methodological issues in autism-related research. *Journal of Autism and Developmental Disorders*, 34, 81–86.
- Jones, E. J. H., Gliga, T., Bedford, R., Charman, T., & Johnson, M. H. (2014). Developmental pathways to autism: A review of prospective studies of infants at risk. *Neuroscience and Biobehavioral Reviews*, 39C, 1–33.
- Kalandadze, T., Norbury, C., Nærland, T., & Næss, K.-A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism*, 22, 99–117.
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217–250.
- Keane, M. (1985). On drawing analogies when solving problems: A theory and test of solution generation in an analogical problem-solving task. *British Journal of Psychology*, 76(4), 449–458.
- Keane, M. (1987). On retrieving analogues when solving problems. *Quarterly Journal of Experimental Psychology Section A*, 39(1), 29–41.
- Keane, M. T., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science*, 18(3), 387–438.
- Knowlton, B. J., Morrison, R. G., Hummel, J. E., & Holyoak, K. J. (2012). A neurocomputational system for relational reasoning. *Trends in Cognitive Sciences*, 16, 373–381.
- Krawczyk, D. C. (2012). The cognition and neuroscience of relational reasoning. *Brain Research*, 1428, 13–23.
- Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., Miller, B. L., & Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia*, 46, 2020–2032.
- Kubricht, J. R., Lu, H., & Holyoak, K. J. (2017). Individual differences in spontaneous analogical transfer. *Memory & Cognition*, 45, 576–588.
- Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *BMJ*, 333, 597–600.
- Leevers, H. J., & Harris, P. L. (2000). Counterfactual syllogistic reasoning in normal 4-year-olds, children with learning disabilities, and children with autism. *Journal of Experimental Child Psychology*, 76, 64–87.

- Lockyer, L., & Rutter, M. (1970). A five- to fifteen-year follow-up study of infantile psychosis. IV. Patterns of cognitive ability. *British Journal of Social and Clinical Psychology*, *9*, 152–163.
- Lovett, A., & Forbus, K. (2017). Modeling visual problem solving as analogical reasoning. *Psychological Review*, *124*, 60–90.
- Mackintosh, N. J. (1998). *IQ and human intelligence*. New York, NY: Oxford University Press.
- McGregor, E., Whiten, A., & Blackburn, P. (1998). Teaching theory of mind by highlighting intention and illustrating thoughts: A comparison of their effectiveness with three-year-olds and autistic subjects. *British Journal of Developmental Psychology*, *16*, 281–300.
- McKenzie, R., Evans, J. St. B. T., & Handley, S. J. (2010). Conditional reasoning in autism: Activation and integration of knowledge and belief. *Developmental Psychology*, *46*, 391–403.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*, e1000097.
- Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in Frontotemporal Lobar Degeneration. *Journal of Cognitive Neuroscience*, *16*, 260–271.
- Morsanyi, K., & Handley, S. J. (2012). Reasoning on the basis of fantasy content: Two studies with high-functioning autistic adolescents. *Journal of Autism and Developmental Disorders*, *42*, 2297–2311.
- Morsanyi, K., & Holyoak, K. J. (2010). Analogical reasoning ability in autistic and typically-developing children. *Developmental Science*, *13*, 578–587.
- Morsanyi, K., Handley, S. J., & Evans, J. St. B. T. (2010). Decontextualised minds: Adolescents with autism are less susceptible to the conjunction fallacy than typically developing adolescents. *Journal of Autism and Developmental Disorders*, *40*, 1378–1388.
- Morsanyi, K., & Holyoak, K. J. (2010). Analogical reasoning ability in autistic and typically-developing children. *Developmental Science*, *13*, 578–587.
- Mottron, L. (2004). Matching strategies in cognitive research with individuals with high-functioning autism: Current practices, instrument biases, and recommendations. *Journal of Autism and Developmental Disorders*, *34*, 19–27.
- Nader, A. M., Courchesne, V., Dawson, M., & Soulières, I. (2016). Does WISC-IV underestimate the intelligence of autistic children? *Journal of Autism and Developmental Disorders*, *46*, 1582–1589.
- Ozonoff, S., South, M., & Miller, J. N. (2000). DSM-IV-defined Asperger syndrome: Cognitive, behavioral and early history differentiation from high-functioning autism. *Autism*, *4*, 29–46.
- Pellicano, E., & Burr, D. (2012). When the world becomes ‘too real’: A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*, 504–510.
- 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*, 77–98.
- Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin’s mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*, 109–178.

- Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, *48*, 2940–2951.
- Pijnacker, J., Geurts, B., van Lambalgen, M., Kan, C. C., Buitelaar, J. K., & Hagoort, P. (2009). Defeasible reasoning in high-functioning adults with autism: Evidence for impaired exception-handling. *Neuropsychologia*, *47*, 644–651.
- Plaisted, K. C. (2001). Reduced generalization in autism: An alternative to weak central coherence. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 149–169). Mahwah, NJ: Erlbaum.
- Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review*, *27*, 224–260.
- Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence, 1938, individual form*. London, England: H. K. Lewis.
- Richland, L. E., & Burchinal, M. (2012). Early executive function predicts reasoning development. *Psychological Science*, *24*, 87–92.
- Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *94*, 249–271.
- Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogy in the mathematics classroom. *Science*, *316*, 1128–1129.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, *85*, 638–641.
- Sahyoun, C. P., Soulières, I., Belliveau, J. W., Mottron, L., & Mody, M. (2009). Cognitive differences in pictorial reasoning between high-functioning autism and Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *39*, 1014–1023.
- Scott, F. J., & Baron-Cohen, S. (1996). Logical, analogical and psychological reasoning in autism: A test of the Cosmides theory. *Development and Psychopathology*, *8*, 235–245.
- Scott, F. J., Baron-Cohen, S., & Leslie, A. (1999). ‘If pigs could fly’: A test of counterfactual reasoning and pretence in children with autism. *British Journal of Developmental Psychology*, *17*, 349–362.
- Simms, N. K., Frausel, R. R., & Richland, L. E. (2018). Working memory predicts children's analogical reasoning. *Journal of Experimental Child Psychology*, *166*, 160–177.
- Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., Diamond, S. P., & Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences, USA*, *111*, 15220–15225.
- Snow, R. E., Kyllonen, C. P., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 47–103). Hillsdale, NJ: Erlbaum.
- So, W. C., Lui, M., Wong, T. K., & Sit, L. T. (2015). The use of hand gestures to communicate about nonpresent objects in mind among children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, *58*, 373–382.
- Soulières, I., Dawson, M., Gernsbacher, M. A., & Mottron, L. (2011). The level and nature of autistic intelligence II: What about asperger syndrome? *PLoS ONE*, *6*, e25372.
- Soulières, I., Dawson, M., Samson, F., Barbeau, E. B., Sahyoun, C. P., Strangman, G. E., ... & Mottron, L. (2009). Enhanced visual processing contributes to matrix reasoning in autism. *Human Brain Mapping*, *30*, 4082–4107.
- Spearman, C. (1923). *The nature of intelligence and the principles of cognition*. London, England: Macmillan.
- Spearman, C. (1927). *The abilities of man*. New York, NY: Macmillan.

- Spearman, C. (1946). Theory of a general factor. *British Journal of Psychology*, *36*, 117–131.
- Stevenson, J. L., & Gernsbacher, M. A. (2013). Abstract spatial reasoning as an autistic strength. *PLoS ONE*, *8*, e59329.
- Swettenham, J. (1996). Can children with autism be taught to understand false belief using computers? *Journal of Child Psychology and Psychiatry*, *37*, 157–165.
- Terzi, A., Marinis, T., & Francis, K. (2016). The interface of syntax with pragmatics and prosody in children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *46*, 2692–2706.
- Tzuriel, D., & Groman, T. (2017). Dynamic assessment of figurative language of children in the autistic spectrum: The relation to some cognitive and language aspects. *Journal of Cognitive Education and Psychology*, *16*, 38–63.
- van Boxtel, J., & Lu, H. (2013). General commentary: A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology*, *4*, article 19.
- Vulchanova, M., Saldaña, D., Chauboun, S., & Vulchanov, V. (2015). Figurative language processing in atypical populations: The ASD perspective. *Frontiers in Human Neuroscience*, *9*, article 24.
- Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., Thomas, C. R., & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, *10*, 119–125.
- Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition*, *28*, 1205–1212.
- Whitaker, K. J., Vendetti, M. S., Wendelken, C., & Bunge, S. A. (2018). Neuroscientific insights into the development of analogical reasoning. *Developmental Science*, *21*, e12531.
- Wing, L., & Wing, J. K. (1976). *Early childhood autism: Clinical, educational and social aspects*. Oxford, England; New York, NY: Pergamon Press.
- Wurst, E. (1976). *Autismus*. Bern, Switzerland: Huber.
- Yamada, T., Ohta, H., Watanabe, H., Kanai, C., Tani, M., Ohno, T., ... & Hashimoto, R. (2012). Functional alterations in neural substrates of geometric reasoning in adults with high-functioning autism. *PLoS ONE*, *7*, e43220.

Appendix: Further details of the methods used in the systematic review and meta-analysis

Search strategy

We searched the following electronic databases: PsycARTICLES[§], PsycINFO[§], ERIC, Linguistics and Language Behavior Abstracts (LLBA)[§], ProQuest Dissertations & Theses Global (all five searched using the ProQuest engine), Web of Science, and Scopus, for studies published up to May 2018. We used the following combinations of search terms: *autis** OR *ASD* OR *Asperger** crossed with *analog** OR *Raven**. The target fields in the searches included titles, abstracts, keywords, topics, subjects, and indexing. We also examined the references of included studies to identify additional papers that were potentially relevant.

Screening process

The search was conducted by two independent raters (D.S. and K.M.), and any disagreements were resolved by discussions among all three authors. When an abstract contained insufficient information, the full-text article was reviewed. The inclusion criteria were applied to

select from the full-text articles the final papers to be included in the meta-analysis. This selection was performed by K.M. and K.H. Any disagreements were resolved by discussion. Further details on the number of papers included in each step, and the reasons for exclusion, are presented in the PRISMA flow diagram (Moher, Liberati, Tetzlaff, & Altman, 2009) in Figure 4.3.

Meta-analytic procedures and analysis

Based on the original data from each study, effect sizes were computed using Hedges g (a variation of Cohen's d that corrects for biases due to small sample sizes; Hedges, 1981; Hedges & Olkin, 1985). The Hedges g value was defined as positive when individuals with ASD had the higher group mean (and vice versa for negative values). The overall effect size was estimated by calculating a weighted average of individual effect sizes,⁹ based on a random effects model that assumes between-study variations in effect sizes not only result from random error, but also from systematic effects of some variables that are likely to vary from study to study (Borenstein et al., 2009). The assumption that effect sizes are heterogeneous allows the possibility that factors beyond an ASD diagnosis impact effect sizes. Heterogeneity of effect sizes was statistically tested using Cochran's Q -statistic. We also report the I^2 statistic, which expresses the percentage of variation in effect sizes across studies that is due to systematic effects of study variables, rather than chance (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003). Meta-regression analyses, using random-effects models, were carried out to test for the effect of possible moderator variables (type of analogy task and the average age and IQ of ASD participants).