Understanding how deficits in sub- and higher-order cognitive processes impact problem-solving abilities in childhood

Sinéad Mary O'Brien

This dissertation is submitted for the degree of *Doctor of Philosophy*

MRC Cognition & Brain Sciences Unit

Trinity Hall

University of Cambridge

December 2021

Dedicated to my parents John and Marie

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit set by the School of Clinical Medicine and the Board of Graduate Studies.

Sinéad O'Brien

December 2021

Some of the work presented in *Chapter 4* has been submitted for publication and received a revise and resubmit from the *Quarterly Journal of Experimental Psychology*: **O'Brien, S.**, Mitchell, D., Duncan, J. & Holmes, J, "Cognitive segmentation and fluid reasoning in childhood".

A manuscript closely related to *Chapter 5* has been submitted for publication to the *Quarterly Journal of Experimental Psychology* as: **O'Brien, S.,** Mitchell, D.J., Dalmaijer, E.S., Duncan, J., & Holmes, J. "Contributions of working memory, processing speed and cognitive segmentation to children's fluid intelligence".

My research outlined in *Chapter 6* was undertaken as part of a collaborative research project and has been published as: **O'Brien, S**., Ng-Cordell, E., Astle, D. E., Scerif, G., & Baker, K. (2019). STXBP1-associated neurodevelopmental disorder: a comparative study of behavioural characteristics. *Journal of neurodevelopmental disorders*, *11*(1), 1-11.

Understanding how deficits in sub- and higher-order cognitive processes impact problem-solving abilities in childhood

ABSTRACT

Fluid Intelligence describes the ability to solve complex problems under novel conditions and predicts success in a wide range of areas. The overarching aim of this thesis was to explore the cognitive processes necessary for the completion of complex cognitive tasks in children with and without intellectual disabilities.

Cognitive segmentation, the ability to separate a complex problem into component parts, is vital for the successful completion of fluid intelligence tasks (Duncan, 2013). In the first empirical chapter, Chapter 2, I describe the development of a child-appropriate measure of cognitive segmentation that is used in the experimental work in Chapters 4 and 5. Two versions of the task were developed. In one, participants were presented with a traditional 2x2 matrix reasoning problem and asked to draw the missing matrix item in a response box below. In the second, the problem was broken down into its component features across three separate cells, reducing the need for participants to segment the problem. The task development process took 12 months and involved the creation of multiple stimuli, several rounds of revisions, and two pilot sessions. This study presents the first attempt to develop a child-appropriate version of Duncan et al.'s (2017) cognitive segmentation task.

Chapter 3 outlines the methodology used in experimental Chapters 4 and 5. The second empirical chapter, Chapter 4, builds on Duncan et al.'s (2017) work to test whether cognitive segmentation improves performance on complex fluid intelligence problems in children aged 6-10 years in the same way as it does for some adults. The effects of age were explored to test whether younger children fail to segment more than older children. The influence of additional within-task characteristics (e.g., the number and type of rules in each item) on performance were also investigated. The results revealed that cognitive segmentation is crucial for problem-solving success in children, irrespective of age and ability. The implications of the within-task characteristics analyses for the design of fluid intelligence tasks are discussed.

The aim of the third empirical chapter, Chapter 5, was to explore the relative contribution of cognitive segmentation to fluid reasoning alongside other cognitive processes

commonly implicated in problem-solving, namely working memory and processing speed. I hypothesised that performance would be correlated across all tasks and that each would predict performance on a fluid reasoning task. The main question of interest was whether each cognitive skill would make unique contributions to fluid intelligence, and crucially whether cognitive segmentation contributed anything above and beyond working memory and processing speed. The results revealed that cognitive segmentation does make a significant, unique contribution to problem-solving, above that accounted for by the other cognitive processes.

The final empirical chapter, Chapter 6, presents a phenotyping study of a group of individuals characterised by intellectual disability (ID); individuals with a rare *de novo* mutation on the STXBP1 gene. This was the first study conducted in my PhD, which inspired my interest in understanding the processes involved in "intelligence". It is presented as the last empirical chapter for coherence.

Together, these studies highlight the challenges faced by individuals with ID and demonstrate that breaking problems down (cognitive segmentation) can aide problem-solving in children of all abilities. They underscore the importance of breaking complex problems down in the classroom, and re-structuring multi-step tasks into separate independent steps to support children's learning and classroom success. Using these techniques with individuals with ID may help them complete the everyday challenges they face.

Sinéad Mary O'Brien

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Joni Holmes. I have thoroughly enjoyed our discussions about research and life. Whenever I felt overwhelmed, you provided me with perspective and confidence. I think it is testament to your positive outlook that I left every meeting feeling happier and more motivated. To the past and present members of the Joni Holmes Lab, thank you for your valuable insights and feedback on my research.

I am grateful to my second supervisor, Danny Mitchell, for your guidance and patience. Your questions during supervisions have broadened my understanding of my findings and pushed me to think about their real word application. Thanks also to John Duncan for providing me with the theoretical and methodological platform from which my research ideas grew.

I would also like to thank Duncan Astle, Kate Baker, and Gaia Scerif for giving me the opportunity to work on the *Brain and Behaviour in Intellectual Disability of known Genetic Origin* (BINGO) study. A special word of thanks to Elise for your support on our many adventures visiting families around the UK, and to Diandra for your work on bringing the Farm App data to life. To the families who participated in the BINGO study, thank you for providing a window into the lived experience of rare genetic disorders. The experience fundamentally shaped my research.

I would like to take this opportunity to thank the staff and students who took part in my research. I am very grateful to the teachers for sharing their knowledge and perspectives on education with me. The highlight of my PhD has undoubtedly been my time with the students. Their curiosity, enthusiasm, and willingness to try new things was inspiring. A special word of thanks goes to their parents for seeing the worth in my research and agreeing to let them take part.

I feel very lucky to have had the opportunity to undertake my PhD at the MRC Cognition and Brain Sciences Unit (CBU), University of Cambridge. Being surrounded by such exceptional researchers has been an invaluable experience. In particular, I would like to take this opportunity to thank Edwin Dalmaijer and Susan Howell for ensuring every iPad made it safely to Ireland – much of the research would not have been possible without your support. I have many happy memories of the CBU, thanks in no small part to my PhD cohort: Alex I., Alex Q., Andrea, Carol, Daisy, Giacomo, Guilia, Ivan, Silvana, and Subbu.

To my friends, thank you for being a great source of fun and distraction from PhD life. I would especially like to thank Fánchea Daly, Jen Cosgrove, Sarah Bishop, and Wiebke Lamping for filling me with words of encouragement during my toughest days. To my roommates: Adella, Caroline, Henri (Hon.), Simone, Susie, and Marc, thank you for providing me with a home away from home, filled with great food and endless chats.

Special thanks go to my family. I do not think any of us ever envisioned that I would end up studying at the University of Cambridge. Your phone calls and trips to see me were a lifeline, especially during the final stages of my PhD. To my sister Claire and brother-in-law Stephen McKeogh, I cannot thank you enough for always taking care of me and feeding me up every time I went back to Ireland. To my sister Orla, thank you for your words of encouragement and unwavering positivity. To my sister Eithne, thank you for always being at the end of the phone to chat through the trials and tribulations of academia. To my parents John and Marie, I find it incredibly difficult to sum up what your support means to me. I will forever be indebted to you for the education you have given me - I owe this accomplishment to you.

Finally, to Alexander Smith, thank you for being my constant companion. I would not be writing these words if it were not for your kindness, patience, and support. Thank you for never failing to see the whole of the moon, even when I could not.

THESIS RATIONALE

We are required to follow multiple steps to complete many everyday activities, including cooking, following directions, assembling new furniture, or learning how to use new technology. For example, cooking a new dish requires us to mix and cook multiple ingredients in the right order, at the right time. To help us complete a complex task like this, we follow recipes that break the method into simpler parts. Many classroom activities also require children to break tasks into simpler parts or steps.

Learning is process that relies on children completing individual tasks to acquire knowledge, and each of these individual tasks often requires the decomposition of a whole into its constituent parts. For example, when solving a maths problem or completing a chemistry experiment, children must form a ‘recipe’ in their mind, breaking the task into its component parts to focus on one part at time, in order to complete it.

Children who struggle to break learning activities down into parts may fall behind their peers because they are at increased risk of missing key parts of a learning activity or failing to successfully complete it. Those who fail to complete individual learning tasks are at risk of failing to acquire skills or knowledge that are important for learning more complex material or completing more complex tasks, and over this time this may manifest as slow rates of learning. It is therefore vital to identify the factors (e.g., age and cognitive skills) that constrain children’s abilities to break down problems by themselves. Understanding these factors will enable teachers to better assess the level of support needed by individual children. Scaffolding a child’s learning by breaking a problem down may allow them to focus on ‘following the recipe’ until they acquire the ability to ‘write the recipe book’ themselves.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AC	Aiming and Catching
ACMG	American College of Medical Genetics and Genomics
ADHD	Attention Deficit/ Hyperactivity Disorder
APA	American Psychological Association
AS	Angelman Syndrome
BAL	Balance
DBC	Developmental Behavioural Checklist
DSM	Diagnostic and Statistical Manual of Mental Disorders
ERP	Event-related potential
gc	Crystallised intelligence
gf	Fluid intelligence
GLM	General linear model
ID	Intellectual disability
IQ	Intelligence quotient
Leiter nvIQ	Leiter non-verbal IQ
M-ABC	Movement ABC
MD	Manual dexterity
SDA	Standard dynamic assessment
SRS	Social Responsiveness Scale
STXBP1	Syntaxin binding protein 1
TDA	Tailored dynamic assessment
RT	Reaction time
TBPS	Total Behaviour Problem Score
VABS	Vineland Adaptive Behaviour Scales
VSWM	Visuo-spatial working memory
WASI	Wechsler Abbreviated Scale of Intelligence

CHAPTER 1 LITERATURE REVIEW

1.1 FLUID INTELLIGENCE

Fluid intelligence (gf) describes the ability to solve complex problems under novel conditions where learned skills and knowledge are minimised (Cattell, 1971). Fluid intelligence is different from crystallised intelligence (gc), which represents learned knowledge and skills that have been accumulated through experience e.g., formal education. Fluid reasoning is a key driving force behind developmental changes in crystallised abilities (Ferrer et al., 2007; Ferrer & McArdle, 2004). Early fluid reasoning skills predict later educational achievement, employment prospects, and both mental and physical health outcomes (Batty & Deary, 2004; Chen et al., 2017; Deary, 2012; Postlethwaite, 2011; Sternberg et al., 2001; Zammit et al., 2004). Therefore, it is no surprise that research has sought to understand what supports and constrains fluid intelligence.

Francis Galton (1883) first proposed the theory of 'general cognitive ability', which enables people to learn and acquire new skills. However, it was Spearman (1904) who first demonstrated this 'general cognitive ability' by applying factor analysis to a range of academic and sensory discrimination data from school-age children. The results showed that if a child's performance was high on one task, it was generally high on another task, a psychometric property known as the 'positive manifold'. Spearman (1927) referred to this general ability as the 'g' factor. In addition to 'g', he proposed that performance on different tasks also required specific abilities, known as the 's' factor. His theory was that, in tasks that are correlated, 'g' varies from person to person, but for each person 'g' is the same for each task. On the other hand, the 'specific' or 's' factor also varies from person to person but for each person 's' additionally varies from one task to another (i.e., 's' represents the task-specific variance not accounted for by 'g' (Spearman, 1927, p. 75).

Similar tasks will show high correlations because they require similar 'g' and 's' factors (Duncan, 2010), but some tasks require greater input from 'g' compared to 's' (Spearman, 1927). If participants are assessed on a broad battery of tasks, the 'g-saturation' of each task can be calculated. This is a metric of how well a task measures 'g', as opposed to the specific or ‘s’ factors associated with the task. The ‘g saturation’ of a task is calculated by averaging a task’s correlations with all other tasks in an assessment battery (Duncan, 2010).

Spearman’s ‘g’ has since been decomposed into two factors: crystallised intelligence (gc) and fluid intelligence (gf). As outlined above, crystallised intelligence (gc) represents learned knowledge and skills; whereas fluid intelligence (gf) represents our ability to solve complex novel problems, like those designed by Raven (Raven & Court, 1938; 1962; 1965; Raven, Court, Raven, 1988). Cattell (1971, p. 99) defined gf as “an expression of the level of complexity of relationships which an individual can perceive and act upon when he does not have recourse to such complex issues already sorted in memory”. Spearman’s ‘g’ and Cattell’s ‘gf’ are actually very similar, and the focus of the work presented in this thesis. The two terms are used interchangeably throughout this thesis to refer to fluid intelligence: our ability to solve novel complex problems where we cannot easily draw upon learned knowledge or skills.

In the following sections, I will cover the measurement of fluid intelligence, as it relates to my thesis, and the strategies and processes involved in problem-solving. I will discuss how age- and problem-complexity influence the strategies people use. I will then focus more specifically on problem-solving in children, covering the main stages of analogical reasoning development and the role of relational knowledge in advancing development. I will then cover analogical reasoning in individuals with intellectual disability. Following this, I will consider the characteristics of analogical reasoning tasks that can influence performance in children and adults, focussing on the number of features in a problem (sometimes called complexity), the visual arrangement of the problem (perceptual organisation) and the type of rules governing the relationship between features. Finally, I will consider the cognitive processes thought to support performance on analogical reasoning tasks, focussing on working memory, processing speed and cognitive segmentation.

1.2 MEASUREMENT OF FLUID INTELLIGENCE

Although Spearman did not clearly define ‘g’, he believed it involved “the education of relations and correlates” (Spearman, 1927, pp. 165 – 166). A student of Spearman, Raven, confirmed that the task with the highest ‘g saturation’ (approximately 0.8) was an abstract reasoning task. Since then, fluid intelligence has been measured predominantly using novel, nonverbal analogy tasks. Analogy tasks can be organised in matrices of 3x3 or 2x2 (See Figure 1). In the 2x2 format, objects or shapes occupy the top row (cells A and B) and the left column (cells A and C), with the square in the lower right cell of the matrix (cell D) left empty. The task is solved by identifying the response that goes in the D cell, thus completing the analogy. Typically, multiple-choice paradigms are used, where participants must choose their answer from a range of response options below the matrix (e.g., Carpenter et al., 1990; Chen et al., 2016; Mulholland et al., 1980; Primi, 2001; Siegler & Svetina, 2002; Sternberg, 1977).

In these analogical reasoning tasks (see Figure 1), the rules and relations governing the shapes in a matrix vary systematically across the rows and columns. To solve the analogy, the rules and relations connecting each shape in the top row of the matrix (Figure 1: Cells A and B) must be identified and encoded, and mental representations of these relations created. These rules are then applied to the bottom row (Figure 1: Cell C) to create an appropriate representation of the pattern that fills the blank cell (Figure 1: Cell D). The corresponding response is then selected from the set of response options (Primi, 2001; Sternberg, 1977).

Note that although there are other ways to measure fluid intelligence (e.g., the Block Design task of the Wechsler, 2008; the odd-one-out and series completion tasks of the Cattell Culture Fair test), analogical reasoning tasks in matrix format are the focus of much of the experimental work presented in this thesis. In *Chapter 2*, I develop a modified analogical reasoning task that is based on the matrix reasoning tests of novel complex problem solving. In this thesis 'elements' or 'attributes', that is the shapes appearing in the matrices, are referred to as features. The combination of features presented in each cell is referred to as an 'item'.

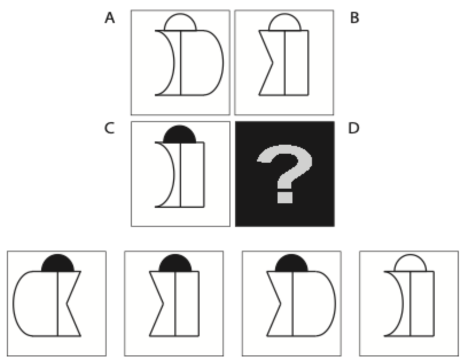


Figure 1: Example of a traditional reasoning task item (modified with permission from Duncan et. al (2017)).

1.3 STAGES OF ANALOGICAL PROBLEM SOLVING

Broadly, theorists have identified three component processes in analogical problem-solving: encoding and creating a mental representation of the problem information; mapping parallels between the features to generalise the rules that relate them; and creating or generating a representation of the solution (Bethell-Fox et al., 1984; Evans, 1968; Gentner & Forbus, 2011; Gentner & Smith, 2012; Glady, French, & Thibaut, 2017; Glady et al., 2016; Goswami, 1991; Hummel & Holyoak, 2005; Mulholland et al., 1980; Primi, 2001; Sternberg, 1977; Stevenson et al., 2016; Stevenson & Hickendorff, 2018; Vakil, Lifshitz, Tzuriel, Weiss, & Arzuoan, 2011). In this section, I provide an overview of the main theories and history of the subprocesses thought to be important for analogical problem-solving in adults and then in children.

Sternberg's early theory of analogical problem-solving involved five sub-processes: (1) encoding to compare relations in an analogy; (2) inference to deduce the rules governing the relations; (3) mapping the A:B pair to the C element and the potential responses (D); (4) applying the A:B inference to the C term; (5) justifying, if needed, to check whether an error was made or whether additional information was needed to find the solution (Sternberg, 1977).

Two earlier alternative theories of the stages of analogical problem-solving had been proposed. The first included all of Sternberg's sub-processes, except for 'mapping' (Johnson, 1962; Spearman, 1923). The second excluded 'application' (Evans, 1968; Winston, 1970), and was based on a simplification of the task requirement steps performed by a computer program. This program, 'ANALOGY' (Evans, 1968), included three core phases: (1) decomposition of the patterns of an analogy into sub patterns to compare the relations in the A:B pair; (2) determining the relations common to A and B; and (3) searching the response alternatives to find a logical C:D pairing. By directly comparing the A:B pair to potential C:D pairs, some of the 'D' response alternatives are eliminated early in the solution process. The transformations present in the 'A:B' pairs are then mapped on to remaining 'C:D' pairings. The 'D' response chosen is the one that contains the transformational rules apparent in the 'A:B' pair (Evans, 1968).

Sternberg compared his theory to these alternatives in four solution strategy models, using latency and error rate data from analogical reasoning measures. 'Mapping' was identified as a key problem-solving process, supporting both Sternberg's (1977) and Evans'

(1968) theories. In these models, both mapping and application were self-terminating processes (i.e., participants stopped mapping and applying features when the target response was identified, that is before all features had been examined). However, the theories differed in terms of 'inference' – the process of inferring the rules governing the relations. This process was exhaustive in Sternberg's theory (i.e., all encoded features are compared to select the rule governing each of the features), but self-terminating in Evans's theory. Data from Sternberg's comparison study fit models of both accounts, but Sternberg (1977, p. 376) concluded that his model 'seemed slightly superior' as the processes described could be generalised to solve a range of measures including verbal and geometric analogies.

In the 1980's, Mulholland, Pellegrino, and Glaser (1980) further explored the stages of problem-solving in geometric analogies. They highlighted the compatibility of Sternberg's (1977) and Evans's (1968) theories, but their findings aligned more with Evans's (1968) three stage simplified model. They referred to their modified model as the 'infer, infer, compare' model (Mulholland, Pellegrino, & Glaser, 1980) and proposed that when solving geometric analogies, participants first inferred the relationship between the A and B items, then inferred the relationship between C and potential D items, and then compared these sets of relationships before choosing a response option for D. This contrasted with Sternberg's (1977) 'infer-apply-test' model where participants first inferred the relations between features in a matrix, then constructed a mental representation of 'D' before searching for a compatible response option.

Using eye-tracking, Bethell-Fox, Lohman, and Snow (1984) confirmed that adults employ both Mulholland et al.'s (1980) and Sternberg's (1977) solution strategies, depending on the complexity of the problem. They proposed that in some situations participants solved analogies using 'constructive matching'. This strategy aligned with Sternberg's (1977) 'infer-apply-test' framework and involved participants first constructing an idealised mental representation of the response before comparing this to the response options. At other times, participants would use a 'response elimination' strategy, which chimed with the 'infer-infer-compare' model (Evans, 1968; Mulholland et al.'s, 1980; Pellegrino & Glaser, 1980). Participants who used this strategy attended to response options early, excluding obviously wrong response alternatives first. Participants appeared to switch strategies in response to item complexity. Complexity could refer here to the number of response alternatives (2 or 4), or the number of features, or the number of rules governing the features. 'Constructive matching' was apparent on simpler items, whereas 'response elimination' was used on more complex items where the rules were less easy to infer. Furthermore, 'response elimination' was used more often by participants with low fluid intelligence, whereas individuals with high fluid intelligence were more likely to use ‘constructive matching’. These findings demonstrated, for the first time, that participants would use different strategies and apply different processes to different problems: the sub-processes underlying problem-solving behaviour were shown to be dynamic, changing in response to the complexity.

The processes and strategies people apply differ with age, as well as complexity. As one example, Sternberg & Rifkin (1979) applied Sternberg’s (1977) framework to compare children and young adults on a picture analogy task (8-, 10-, 12-, & 19-year-olds). Participants of all ages used the ‘encoding’, ‘inference’, ‘application’, and ‘response’ sub-processes to solve the problems, but there were age- and complexity-related differences in the use of ‘mapping’: 8-year-old children did not use it, and older participants only used it when items were more complex. Unlike adults (Sternberg, 1977) and older children, younger children were more likely to self-terminate ‘encoding’ and ‘inference’. This suggests that younger children may not encode and infer all features in a matrix and the rules governing them. The lack of consideration for all features and possible rules, may reflect children’s reliance on lower-order, perceptual similarities of features rather than higher-order, relational similarities. That is younger children solve analogies based on the appearance or similarity of features in a cell, as opposed to finding a shared relational structure between features (Gentner, 1989; Medin et al., 1993). It might also reflect younger children’s more limited capacities to store the outcomes of encoding and inferring all at once (Niebaum & Munakata, 2021; Glady, 2013). In addition, older children’s performance resembled that of adults suggesting that the application of effective solution strategies may improve with age.

In a more recent example of the interaction between strategy use and complexity, Niebaum and Munakata (2021) explored whether children and adults adapt their strategy use to item complexity. As expected, there was an association between better performance and ‘constructive matching’. In addition, increased complexity resulted in strategy shifting, with both children and adults spending more time scanning rows and columns and consulting response options for more complex problems. The authors argue that the use of ‘constructive matching’ does not reflect better task comprehension, but instead is the cause of better performance. Consistent with this, research using training paradigms with (Stevenson et al., 2016; Stevenson & Hickendorff, 2018) and without item-by-item feedback (Glady, French, & Thibaut, 2016; 2017; Rivollier et al., 2020) have demonstrated that alterations to reasoning tasks, that encourage constructive matching, result in improved performance in children and adults.

1.4 ANALOGICAL PROBLEM-SOLVING IN CHILDREN

1.4.1 PHASES OF PROBLEM-SOLVING

We have already seen two studies that compare problem-solving in children and adults, showing both differences (Sternberg & Rifkin, 1979) and similarities (Niebaum & Munakata, 2021). Nevertheless, it is clear that improvements in nonverbal reasoning occur throughout childhood and adolescence (Christie & Gentner, 2014a; Ferrer et al., 2009; Halford et al., 1998a; Mcardle et al., 2002; Richland et al., 2006; Siegler & Svetina, 2002; Starr et al., 2018; Sternberg & Rifkin, 1979; Stevenson & Hickendorff, 2018) with large individual differences apparent in children as young as six years (Whitaker et al., 2018). In Chapter 4, I will explore the effects of age on children's performance on analogical reasoning tasks: this work is cross sectional but would have been longitudinal if it had not been for the COVID-19 pandemic closing down the study.

Research has shown that as children get older, they move through phases of analogical reasoning development. Broadly speaking, children start learning to solve analogies by duplicating an item in a matrix (the duplication phase), and then move into a phase characterised by the inconsistent use of solution strategies (idiosyncratic phase), before finally settling into a more consistent approach to analogical reasoning (called here the consistent phase) (Gentner, 1988; Gentner & Ratterman, 1991; Piaget & Inhelder, 1964; Rattermann & Genter, 1998; Stevenson & Hickendorff, 2018; Tunteler et al., 2008).

Adults typically focus on the A:B pair of an analogy first, then move to C, and the various response options (Gordon & Moser, 2007; Thibaut & French, 2016). In contrast, children aged up to 8 years orient to C and organise their search around this item and the response options (Thibaut & French, 2016). Orienting the search around the C-term instead of focusing on the A:B pair has been associated with poorer performance in both adults and children (Glady et al., 2016; Starr et al., 2018; Thibaut & French, 2011; Vendetti et al., 2017). It is possible that orienting the search around the C-term, and not properly inferring and encoding the relationship between the A:B terms, may result in children copying a term of the matrix as a response to the item – the duplication stage. This phase is considered the least advanced, is typically observed in younger children (Inhelder & Piaget, 1964; Siegler & Svetina, 2002; Stevenson et al., 2016), and is characterised almost exclusively by children solving analogies through duplicating one of the matrix items. Inhelder and Piaget (1964) were one of the first to discover that duplication errors are prevalent prior to the acquisition of more advanced analogical reasoning skills. Siegler and Svetina (2002) replicated this finding in six- to eight-year-olds, finding that the frequency of duplication errors decreased prior to the application of more advanced solution strategies, but that other drawing errors increased.

Stevenson and Hickendorff (2018) recently analysed four- to 10-year-old children's transition from non-analogical processes to full analogical reasoning. They found that the duplication (sometimes called repetition) phase was the least advanced stage of analogical reasoning, and common in younger children. In line with Siegler and Svetina (2002), an additional 'idiosyncratic' phase was identified in which children used variable solution processes: partial responses - choosing a response option with a partially correct solution - were common on easy items, whereas more complex items elicited varied errors. The most common error in more advanced stages of analogical reasoning was almost exclusively arriving at a response option with a partially correct solution. This demonstrates more consistency in the solution of analogical reasoning problems with age, as shown by others too (Raven, Court, and Raven, 1983; Siegler & Svetina, 2002; Vodegel Matzen et al., 1994). Crucially, while Stevenson and Hickendorff (2018) found that children progressed did not always progress sequentially through these three phases – duplication, idiosyncratic, consistent – there was variability. Some children progressed more quickly through some phases than others.

1.4.2 Development of Relational Knowledge

In addition to progressing through different phases, children's reasoning performance improves with the use of relational knowledge. Although relational reasoning has been documented in children as young as two and three (Cattell, 1971, 1987; Halford et al., 1998), young children tend to focus on perceptual matching rather than relational similarities to identify the relations between features (Christie & Gentner, 2014a; Gentner & Medina, 1998; Sternberg & Nigro, 1980). Supporting this, Goswami (1989) tested children between four and seven years of age on pattern- and proportion-based analogies. She found that children as young as four failed because they did not understand the higher order relations used in analogies. Goswami and Brown (1990) argue that young children employ relational rather than perceptual interpretations only when they understand the causal relations governing features.

Gentner (1988) proposed that this increased understanding occurs due to a period known as the relational shift (Gentner, 1988; Gentner & Ratterman, 1991; Rattermann & Gentner, 1998). The ‘relational shift’ theory proposes that changes in performance on analogical reasoning tasks occurs in childhood with the acquisition of knowledge, and not due to general maturation of domain-general cognitive processing. In other words, children’s analogical reasoning skills develop as their knowledge base grows: children must understand how objects are similar before they can use these similarities to draw analogies (e.g., Gentner, 1988; Goswami & Brown, 1990; Kotovosky & Gentner, 1996). In this context, understanding lower-order relations (i.e., perceptual similarity) precedes knowledge of higher-order relations (relational similarity).

Tying together the literature on phases of analogical reasoning development and the role of relational knowledge, Stevenson, Heiser & Resing (2016) conducted a study to show that the format of a problem could alter children’s errors, encouraging them to see relational similarities and shifting them from the duplication to idiosyncratic phase. In their study, children either completed problem-solving tasks with feedback on multiple choice analogies or feedback on constructed response analogies. Constructed response paradigms force children to independently create their responses rather than choose from a set of predefined options. Their answers, therefore, provide a window into why children fail to solve analogical reasoning problems. Although there was no overall difference in performance, children in the ‘constructed response’ group had more partial solutions correct compared to participants in the ‘multiple-choice’ group, who made more duplication errors. The authors suggested that constructing responses resulted in children gaining a better understanding of the solution process. Forcing children to construct their responses may encourage them to move away from solving items based on perceptual features to instead reasoning about the relations between features. In this way, duplications may represent a lack of understanding of the relations and the solution process, or the period prior to development of the ability to reason about relations.

1.5 Analogical Problem-Solving in Intellectual Disability

Intellectual disabilities (ID) typically begin in childhood and are characterized by cognitive deficits and limitations in adaptive functioning (e.g., difficulties that impact on developmental and sociocultural standards for the individual's independence). The cognitive problems in ID are defined as problems in "reasoning, problem solving, planning, abstract thinking…" (APA, 2013, p.33), confirmed by poor performance on age- and population-standardised intelligence tests (typically standard scores <70, or the extreme low end of the normal IQ distribution), as well as clinical evaluation (see *Chapter 6* for a full discussion of ID). While performance on reasoning tasks is characteristically poor for individuals with intellectual disabilities (ID), there has been some work exploring the strategies they use to solve analogical reasoning tasks.

For example, Vakil, Lifshitz, Tzuriel, Weiss, & Arzuoan (2011) found that, like children and adults, individuals with ID showed eye-movements characteristic of 'constructive matching' for easier items, often changing to 'response elimination' when items became more complex on the Ravens Standard Progressive Matrices test (Raven, Court, Raven, 1988). Furthermore, they found that participants with ID made more switches between items in the matrix and response options on complex items and spent less time looking at the correct response. This suggests they observe the features of the analogy superficially and compare items in the matrix to response options without properly inferring the relations governing the matrix.

Lifshitz, Weiss, Tzuriel, and Tzemach (2011) also explored whether adolescents and adults with moderate to mild ID applied different solution strategies and tested whether performance could be improved within a dynamic testing paradigm. Adolescent and adult groups were divided into two sub-groups: standard dynamic assessment (SDA) and tailored dynamic assessment (TDA). Those in the SDA group received explanations of the relation between the A:B, A:C, and B:C analogy pairs, followed by attempts to elicit the correct response. In contrast, the TDA group received training tailored to their problem-solving behaviours. First, the quality of the TDA group's responses were assessed through questions (e.g., 'what do you see?', 'what colour is the box?'). This group then received three teaching strategies: (1) focusing (selecting, comparing, and associating the features in each item of the matrix); (2) leading questions ('what colour is the box?'); (3) reverse questions ('is the dog a human being?'). Finally, the children were asked to verbally analyse their approach to problem-solving. Training was found to significantly increase performance across all age groups and ID levels. In addition, those who received TDA training showed higher gains than those who received SDA. Just as in Vakil et al. (2011), participants with ID showed difficulties with inference. However, mapping and application were revealed as additional sources of difficulty. Lifshitz et al. (2011) suggest that solving multiple rules, together with the need to hold intermediate responses in mind, poses more difficulty for participants with ID.

Denaes (2012) demonstrated that analogy performance improves in children with mild intellectual disability when the need to hold intermediate responses is reduced. The performance of children with and without mild ID was compared on touchscreen-based traditional and constructed response versions of a figural analogies task. In the constructed response condition, all possible features that make up ‘D’ were presented below the 2 x 2 matrix. Children provided a response by touching a feature below the matrix, which would then slide into the ‘D’ cell. There was no significant difference between the two tasks for typically developing children. However, performance was significantly better on the constructed response task for the children with mild ID. The constructed response format enabled the children to focus on one part of the problem at a time, in contrast to the traditional version in which items had to be processed in their entirety. Denaes (2012, p. 277) suggests that the practical implication of the constructed response format is that ‘the decomposition of elements was more beneficial for individuals with mild ID’.

Together these studies show that while individuals with ID struggle on reasoning tasks, they apply different response strategies to items with different levels of complexity, in similar ways to typically developing children and adults. Moreover, these studies show that those with ID require different forms of support depending on the complexity of the items and their own ability levels.

1.6 PROBLEM-SOLVING: WITHIN-TASK CHARACTERISTICS

A growing body of literature demonstrates that different factors within analogical reasoning tasks impact on our ability to infer, encode, and map the features in a matrix to find the correct solution. In this section, several ‘within-task’ characteristics will be presented, and their impact on problem-solving abilities will be considered. I explore the impact of within-task characteristics on children’s task performance in *Chapter 4*.

1.6.1 NUMBER OF FEATURES

Mulholland et al. (1980, p. 256) argue that the processes underlying performance “should be a function of the amount of information that must be processed to solve a given item.” Consistent with this, many studies have shown that the complexity – for example the number of features – of an item relates to performance, with performance generally decreasing as the number of features increases. In one early study, Bethell-Fox, Lohman, & Snow (1984) demonstrated that the number of features was significantly associated with the difficulty of an item, particularly for adults with low ability. Similarly, Carpenter et al. (1990) examined adults' performance on two standardised subsets of the Ravens Advanced Progressive Matrices (I and II; Raven, 1962; 1965), and found that the complexity of items (defined by the number of features), accounted for 45% of the variance in error rates. Vodegel Matzen et al. (1994) replicated Carpenter et al.'s (1990) study in children and found that the number of features in an item accounted for a similar amount of variance (47%) in performance.

More recently, Crone et al. (2009) analysed adults and children's performance on a 3x3 relational reasoning task. The task involved three levels of complexity (0-, 1-, 2-relations). 0-relations simply involved perceptual matching of items within the matrices. The 1-relation condition involved a change of one feature, either on the horizontal or vertical dimension of the matrices. In the 2-relation condition, features changed along the vertical and the horizontal. There was no difference in performance between the 0- and 1-relation conditions. However, more errors were made in the more complex 2-relation condition. In comparison to adult participants, children produced more errors, and this difference was more pronounced in the 2-relation condition compared to the 1-relation condition, demonstrating that age interacts with complexity.

1.6.2 Perceptual Organisation

An additional constraint on the understanding of relations between features in an analogical reasoning problem is the visual arrangement of features in a matrix. To unpack the influence of perceptual organisation on item difficulty, Primi (2001) created 3x3 matrix items that contained various numbers of features with different rule types assigned. The items had two task formats: 'harmonious' or 'non-harmonious'. Harmonious items, or congruent items, had features that could be easily grouped across the cells in a matrix based on colour or form. In contrast, nonharmonic, or incongruent items, were designed with misaligned features that "interrupt natural perceptual continuity of the elements" (Primi, 2001, p. 52). They contained misleading cues about the rule types governing features, thus making it difficult to identify groupings across the cells in matrix. Overall, Primi (2001) found that two main factors influenced item difficulty: the amount of information (number of rules and elements) and the organisation of information (types of rules and the perceptual organisation or 'harmony' of items). However, the perceptual organisation of features in an item had the largest influence on item difficulty, accounting for 53% of the variance.

Meo et al. (2007) argued that Primi's investigation predominantly focused on the overall organisation of items (number of features and types of rules together with perceptual complexity) and did not consider the importance of the 'identifiability' of the features within the items. 'Indentifiability' refers to the ability to encode separate features of an item (Meo et al., 2007). Embedding and overlapping features increases complexity, thus negatively affecting performance as it is harder to distinguish them. In other words, the arrangement of features (overlapping or separate) may be crucial to task performance: the visual complexity of the task likely affects performance.

1.6.3 Rule Type

There is a growing body of literature demonstrating that rule types differentially affect performance on analogy tests in both adults and children (Bethell-Fox et al., 1984; Blum et al., 2016; Green & Kleuver, 1992; Kirchhoff & Holling (n.d.); Novick & Tversky, 1987; Odom et al., 1975; Siegler & Svetina, 2002; Stevenson et al., 2011, 2014, 2016; Stevenson & Hickendorff, 2018; Vodegel Matzen et al., 1994; Whiteley & Schneider, 1981). Rule types refer to the types of rules governing the feature pairs in a matrix. Stevenson et al. (2014) argue that the impact of different rule types on performance should be considered when analysing performance on geometric or figural analogy tests. In this section, I provide a short overview of different studies that support this by demonstrating how rule type affects performance.

First, Mulholland et al. (1980) found that although increasing the number of features increased the overall solution time, it did not affect error rates. However, the complexity of the transformation applied to elements did affect both latency and error rates, with more complex transformations leading to slow completion times and more errors. Whiteley and Schneider (1981) showed that different rule types differentially influenced performance. They examined the effects of spatial rule changes on performance on a geometric analogies task. 'Spatial distortion' referred to rules that involved a change in the size, shade, shape, and number of features across an item, whereas 'spatial transformation' referred to rules that result in the rotation or reflection of the features. The findings revealed that spatial transformations increased item difficulty, relative to spatial distortions.

Many studies with young children have also shown that spatial rules are less salient compared to other types of rules, such as form (e.g., change in shapes), colour, and quantity, as they are for adults (Odom et al., 1975; Siegler & Svetina, 2002; Stevenson et al., 2011). For example, Stevenson et al. (2014) found that children between four and twelve years of age, performed poorest on 'where' (spatial rules) compared to 'what' rules (i.e., rules involving changes in object, size, or quantity). However, the odds of solving a 'where' rule increased more with age, particularly during seven and nine years. The authors suggest that this may be related to a 'relational shift' in ability, characterised by development in the dorsal visual pathway responsible for processing spatial information that occurs around this age (unlike ventral pathway which matures earlier and is responsible for processing 'what' information; Parrish et al., 2005; Spencer et al., 2000; Taylor et al., 2009)

Carpenter et al. (1990) also showed how different rule types differentially impact on performance. They reviewed items from a standardised task, Ravens Advanced Progressive Matrices (Raven, 1983). Each feature was assigned a rule type from their taxonomy of rules. There were five rule types: constant in a row (the same element occurred across a row but differed down a column); quantitative pairwise progressions (involving the increment or decrement of a feature e.g., halving a feature); figure addition and subtraction (involving adding or subtracting a feature); distribution of three values (the same three features appear in the three rows or columns of the matrix e.g., a square, a diamond, and a triangle); and finally, distribution of two values (the same feature appears in two of the three cells in a row or column of the matrix). Based on previous research, the rules were ranked from easiest to hardest. Constant in a row was thought to be the easiest rule because it only required perceptual matching: the same element occurred across a row but differed down a column. Carpenter et al. (1990) found that the difficulty of an item was reflected in the number of rules, *and* the types of rules governing the features. The ranked rule taxonomy was confirmed by the data and more complex rules made 'correspondence finding' between features more difficult to achieve. This has since been replicated in children. Vodegel Matzen, van der Molen, and Dudink (1994) classified the features from the Standard Progressive Matrices (SPM; Raven, Court, Raven, 1988) using the taxonomy outlined by Carpenter et al. (1990), replicating the order of difficulty with children aged eight to twelve years: the easiest rule was constant in a row, followed by quantitative pairwise progression, addition or subtraction, distribution of three values, and distribution of two values (the hardest).

Different rule types also affect the sub-processes employed in problem-solving. Bethell-Fox, Lohman, and Snow (1984) created a new figural analogy task in which each element of a matrix was governed by figural or spatial rules. Similar to Whiteley and Schneider's (1981) rule types, 'figural' rules involved transformations such as halving, doubling, and size change, whereas 'spatial' rules involved rigid-body transformations such as rotation and reflections of features. Bethell-Fox, Lohman, and Snow (1984) discovered that two additional problem-solving processes (in addition to the five outlined by Sternberg (1977)) were required for spatial items only: 'spatial inference' involving the inference of spatial transformations between the A:B pair, and 'spatial application' referring to the application of these transformations to the C-term of the matrix. They concluded that the types of rules governing the features in a matrix, in particular spatial rules, differentially influenced the problem-solving processes applied to each item. Over the last four decades, a number of other studies have replicated this finding (Blum et al., 2016; Green & Kluever, 1992; Kirchhoff & Holling (n.d.); Novick & Tversky, 1987). Collectively, these studies demonstrate that different rule types impact on performance in both children and adults. It is important to note that many of these studies used a multiple-choice task design. A participant's true understanding of the rules governing features in a matrix may be masked by the necessity to select an option from a finite number of response alternatives. Ability testing using 'constructed' response formats – as explained above, problems in which participants construct their own responses independently – provides advantages for determining which aspects of a problem result in an incorrect response (Birenbaum & Tatsuoka, 1987; Stevenson et al., 2016; Stevenson & Hickendorff, 2018). They might therefore facilitate a clearer examination of whether the systematic errors children or adults make are in response to specific rule types (Stevenson et al., 2016; Stevenson & Hickendorff, 2018).

1.7 Analogical Problem-Solving: Cognitive Processes

Research with adults and children suggests that certain cognitive processes constrain performance on tests of fluid intelligence (Duncan et al., 2017; Engel de Abreu et al., 2010; Fry & Hale, 2000a; Gonthier & Thomassin, 2015; Hornung et al., 2011a; Kail, 2007; R. G. Morrison et al., 2006; Niebaum & Munakata, 2021; Richland et al., 2006; H. L. Swanson, 2008; J. P. Thibaut et al., 2010; J. P. Thibaut & French, 2016). The development of these processes likely acts alongside the development of relational understanding in contributing to analogical reasoning skills in childhood. In this section, I will discuss how short-term and working memory, processing speed, and cognitive segmentation are related to fluid reasoning performance. The contribution of these abilities to children's fluid intelligence is examined in *Chapter 5*.

1.7.1 Short-term and Working Memory

Working memory has substantial power to predict performance on a number of complex cognitive ability and learning measures (Unsworth et al., 2014a) such as following instructions (Gathercole et al., 2008; Jaroslawska et al., 2016), reading comprehension (Daneman & Carpenter, 1980), and maths (Holmes & Adams, 2006). A consistent finding is the significant association between working memory and performance on tests of fluid intelligence (Chuderski, 2013; Colom et al., 2004, 2008a; Colom, Flores-Mendoza, et al., 2005; Engle et al., 1999; Frischkorn et al., 2019a; Kane et al., 2005; Kyllonen & Christal, 1990; Oberauer et al., 2005, 2008; Shipstead et al., 2014; Süß et al., 2002; Unsworth et al., 2014a).

Working memory is a dynamic system that simultaneously stores and processes task-relevant information (Baddeley & Hitch, 1974, Miyake & Shah, 1999; Oberauer, 2019). There are a variety of conceptualizations of the nature, structure, and function of working memory (see Conway et al., 2007, for review). One important distinction between models is whether working memory is conceived as a distinct multi-store workspace that includes an attentional component (e.g., Baddeley & Hitch, 1974; Baddeley, 2000) or is embedded within a broader limited-capacity system of controlled attention (e.g., Kane et al., 2001). In general, though, both frameworks provide adequate accounts of large bodies of empirical evidence (e.g., Cowan, 1995) and there is broad consensus that there are short-term memory (simple storage) and attentional control functions within working memory (e.g., Cowan, 2008; Miyake & Shah, 1999). Storage refers to the passive maintenance of information over brief periods. Processing, in simple short-term memory tasks, refers to mechanisms that maintain relevant information (e.g., rehearsal), whereas processing in complex working memory tasks involves the manipulation and transformation of information. I do not align with any single model of working memory within this thesis: I view working memory as the ability to store and process task-relevant information and I discuss the relationship between working memory and fluid reasoning with respect to different theories of working memory and their application.

There have been several mechanistic accounts of the association between working memory and fluid intelligence. One assumes that both constructs rely on the ability to control attention (executive control) in order to ignore distracting information (Engle, 2018; Kane et al., 2007; Shipstead et al., 2016a). An alternative proposal suggests that short-term memory provides the link, storing information necessary for the completion of both working memory and fluid reasoning tasks (Colom et al., 2006, 2008a). A third hypothesis suggests that the working memory processes of building, maintaining, and manipulating arbitrary bindings between items or pieces of information underpins performance on matrix reasoning-type tasks (Oberauer et al., 2008). These three accounts are described below.

1.7.1.1 Controlled Attention

Kane and Engle (Engle, 2002; Engle & Kane, 2004) proposed the *controlled attention model of working memory*. Similar to Cowan's *embedded processes model*, working memory is thought to have an upper capacity limit and is viewed as an intentionally activated subset of long-term memory (Cowan, 1999; Cowan et al., 2005). This framework is distinguished from others by its emphasis on protecting the contents of working memory from interference using inhibitory processes. According to this framework, working memory limitations arise through difficulties in maintaining and retrieving task relevant information in the face of interference from competing information, task demands, and representations (Engle, 2002, 2018; Kane & Engle, 2002; Heitz et al., 2005). Therefore, the capacity of working memory reflects the ability to control attention, by inhibiting irrelevant information, to maintain relevant information in an active, quickly retrievable state (Engle, 2002; Kane & Engle, 2002). To achieve this, two working memory subsystems have been proposed: primary and secondary memory (Unsworth et al., 2010; Unsworth & Engle, 2007). Primary memory is responsible for the active maintenance of information over a short period of time. Items only remain in primary memory if they are actively attended to, if not, they must be retrieved from secondary memory. Thus, secondary memory, or long-term memory refers to the second subsystem which stores encoded information (Unsworth et al., 2010).

Kane and Engle (2003) substantiated this model by experimentally manipulating task-irrelevant information: they demonstrated that requiring participants to complete an additional task before the recall of a list of words in a complex span task impaired performance. Participants with lower working memory capacities showed more proactive interference as a result of the additional task being presented after the primary task, compared to those with higher spans. However, when participants completed the additional task at the same time as the primary task, both groups performed similarly. Performance of high working memory capacity participants decreased and low working memory capacity performance remained relatively unchanged. Kane and Engle (2003) argue that under normal conditions, individuals with high working memory capacity use executive control to combat proactive interference. On the other hand, individuals with poorer working memory capacity do not allocate attention to resist interference. These findings were replicated with a range of other tasks that required participants to exert attentional control to avoid conflicting responses (antisaccade: Conway et al., 2001; dichotic listening paradigms: Conway et al., 2002; and stroop: Kane & Engle, 2003). This has led Kane, Engle, and colleagues to propose that two important working memory functions of executive attention underlie the relationship between working memory and fluid intelligence: the maintenance of goal-relevant information and the ability to deal with proactive interference through inhibiting prepotent responses (Heitz et al., 2005).

According to Kane et al.'s theory, the association between working memory and fluid intelligence occurs because they both rely on the ability to control attention in order to ignore distracting information (Engle, 2018; Kane et al., 2007; Shipstead et al., 2016a). Supporting this, Engle, Tuholski, Laughlin, and Conway (1999) found that when the variances common to working memory and short-term memory (i.e., memory processes) were statistically removed, working memory still correlated with fluid intelligence, whereas the residual short-term memory variance did not. They argued this showed it was the executive elements of working memory and fluid intelligence that were related.

1.7.1.2 Short-term Memory

An alternative proposal suggests that short-term memory provides the link between working memory and fluid intelligence i.e., storing information is necessary for the completion of both working memory and fluid reasoning tasks (Colom, Rubio, Shih, & Santacreu, 2006; Colom et al., 2008). In a series of experiments, Colom, Abad, et al. (2005) and Colom, Flores-Mendoza, et al. (2005) explored the relationships between short-term memory, working memory, and fluid intelligence. The association between working memory and short-term memory was strong, and working memory was found to predict more variance in fluid intelligence scores than short-term memory. Unlike Engle, Tuholski, Laughlin, and Conway (1999), when the shared variance between the constructs of short-term memory and working memory was removed, the predictive power of working memory was reduced, leading the authors to conclude that the short-term storage component of working memory was responsible for the relationship between working memory and fluid intelligence. In an additional experiment (Colom, Abad, et al., 2005), the contributions of working memory and short-term memory to fluid intelligence were considered alongside other key constructs from alternative models, such as mental speed, updating, and the control of attention. Interestingly, working memory and fluid intelligence were found to be almost isomorphic. The storage component accounted for the largest amount of variance in working memory. When this was partialed out of the working memory factor, the relationship between working memory and fluid intelligence disappeared. This led the authors to conclude that it is the short-term storage aspect of working memory that underlies the relationship between working memory and fluid intelligence.

1.7.1.3 Short-term Memory and Executive Control in Children

Theories of the relationship between the short-term storage and executive (controlled attention) aspects of working memory and fluid intelligence have been primarily based on adult data so the extent to which they generalize to children is unclear (Giofrè et al., 2013; Hornung et al., 2011a). Some researchers argue that the processes supporting performance on both short-term storage and more complex executive tasks are distinct in children as they are in adults (e.g., Alloway et al., 2004, 2006; Gathercole et al., 2004; Kail & Hall, 2001; Swanson, 2008). In contrast, other studies have shown that the executive, cognitive control aspects of working memory are recruited even for simple storage tasks in young children. This suggests that the more executive aspects of working memory might be captured in children using tasks which are typically known to assess short-term storage capacity in adults (Hornung et al., 2011a).

Importantly, research has shown that both short-term memory and the cognitive control aspects of working memory predict performance on tests of fluid intelligence in children. Some studies have found that both cognitive control and short-term memory make unique contributions to fluid intelligence (e.g., Bayliss et al., 2003; Giofrè et al., 2013; Tillman et al., 2008). In contrast Swanson (2008) found a distinction between these constructs in children aged six to nine years, supporting the theory that cognitive control primarily underlies age-related changes in fluid intelligence. This finding was replicated by Engle, de Abreau, Conway, and Gathercole (2010) in a sample of children followed from preschool to approximately eight years of age. Others, however, have demonstrated that the short-term memory component of working memory primarily predicts fluid intelligence in children. For example, Hornung et al. (2011a) found that models emphasising storage-only most accurately represented working memory in children, reflecting the focus of attention in order to hold information, and driving the working memory-fluid intelligence relationship.

1.7.1.4 Binding Hypothesis

Oberauer et al. (2000, 2003) explored the presence of three working memory components: simultaneous storage and processing, supervision, and coordination. In his model (2003, p. 169), processing referred to 'the transformation of information or the derivation of new information'. Storage represented the ability to retain new information for a short period of time. Supervision, reflecting executive attention, is described as the activation and prioritization of goal-relevant representations, and the inhibition of irrelevant information. Co-ordination – referred to as 'relational integration' and defined as the ability to build new relations between elements to create mental representations - was proposed as a key aspect of working memory capacity by Oberauer (2003). This was later confirmed by other studies (Buehner et al., 2005; Oberauer, 2019; Oberauer et al., 2008).

Buehner et al. (2005) found that the 'storage and processing' and 'coordination' components of Oberauer's model predicted 95% of variance in fluid reasoning abilities. They found that the high correlation between working memory and fluid intelligence was more pronounced when measures of 'storage in the context of processing' and 'coordination' were included (Buehner et al., 2005, p. 269). Oberauer et al. (2008), found that 'relational integration' and 'storage and processing' were strongly correlated, but that only 'relational integration' significantly predicted reasoning abilities.

This led to the binding hypothesis, which attributes the relationship between working memory and fluid intelligence to the cognitive control aspects of working memory responsible for co-ordination. Unlike the attentional / executive control account, the limiting factor is not the selective maintenance of task relevant information and inhibition of irrelevant information. Rather, it is the working memory processes of building, maintaining, and manipulating arbitrary bindings between items or pieces of information that explains the links between performance on working memory and fluid intelligence tests (Oberauer et al., 2008). In other words, the strong relationship between working memory capacity and fluid intelligence derives from both tasks requiring temporary information bindings to be quickly created and updated (Buehner et al., 2005; Chuderski, 2013; Oberauer, 2019; Oberauer et al., 2008).

1.7.1.5 Reconciling the Executive and Binding Hypotheses

Shipstead et al. (2016a) have proposed a framework that reconciles the executive attention and binding hypothesis models of working memory. Like other theories, the fundamental component of complex cognition is cognitive control. However, according to Shipstead's account, cognitive control is applied via two broadly defined mechanisms: (1) intentional maintenance and (2) intentional disengagement. According to this framework, working memory capacity can represent not just the maintenance of goals but also the intentional disengagement from irrelevant information. Furthermore, Shipstead et al. (2016a) align with Oberauer's model of working memory capacity (Oberauer, 2002; Oberauer et al., 2008), suggesting that working memory creates and maintains representations (bound elements) in which hypotheses can be tested. By this account, individual differences in working memory capacity reflect differences in the binding process.

The explanation given for the strong, but less than perfect correlation between working memory capacity and fluid intelligence by Shipstead's account, is that both tasks require maintenance and disengagement. Working memory is thought to place more emphasis on the maintenance of information whereas performance on tests of fluid intelligence relies more on the ability to disengage from irrelevant information (Shipstead et al., 2016a). In contrast to Oberauer's theory (2002; Oberauer et al., 2008), Shipstead et al., (2016a) attribute the hallmark of fluid intelligence to the unbinding process, as opposed to binding process. They argue that fluid intelligence reflects a person's ability to intentionally unbind from incorrect or irrelevant associations between pieces of information.

Supporting evidence for this idea is provided by a study using an n-back paradigm alongside a fluid intelligence task (Shipstead et al., 2016). In an n-back task, participants are required to respond to a stimulus on screen that matches a stimulus presented n-positions back in the sequence. Lures can be introduced into the sequences (e.g., a stimulus the same as the target can be presented close to the n-position of the target). In this study, the lures were manipulated to be close to the target stimulus (2-, 4-, 5-positions back from the target) or distant to the target (positioned at 7-, 8-, 9- positions away from the target). Performance on distant trials was better than performance on close trials. In addition, as the distance between the 'to be recalled' number and the lure increased, the relationship between working memory capacity and fluid intelligence became stronger. Shipstead, Harrison, and Engle (2016) argued that as the lure position increased, individuals with higher fluid intelligence were able to disengage from the irrelevant information (lures), while lower fluid intelligence participants continued to produce false alarms due to their inability to disengage from this outdated information. This led them to conclude that fluid intelligence performance largely relies on the disengagement from irrelevant information, a process related to working memory.

1.7.2 Processing Speed

Processing speed (PS) or mental speed reflects the ability to quickly process information. Typically, it is measured by identification, discrimination, computing speed, coding speed, judgement tasks, or basic reaction times (Coyle et al., 2011; Dang et al., 2015; Jensen, 2006). Previous research has shown that people with higher fluid intelligence demonstrate higher processing speed (Ackerman et al., 2002; Coyle et al., 2011; Demetriou et al., 2013; Detterman, 2002; Fry & Hale, 2000b; Jensen, 2006; Kail, 2000; Salthouse, 1996). Various theories have been put forward to explain the relationship between processing speed and fluid intelligence, which I outline in the following sections.

1.7.2.1 Information Processing

According to information processing theories, faster processing speeds enable individuals to represent, interpret, and integrate information more quickly in working memory before the relevant information decays (e.g., Fry & Hale, 2000; Kail, Robert & Salthouse, 1994; Kail, 1991; 2007; Salthouse, 1996; Van Der Maas et al., 2006). Both processing speed and fluid intelligence have been shown to increase with age throughout childhood and adolescence, suggesting that processing speed may be a core contributor to the development of fluid intelligence (Coyle et al., 2011; Deary et al., 2010; Jensen, 2006; Kail, 1991, 2000). Likewise, Salthouse (1996) proposed that age-related decreases in processing speed predict reduced cognitive functioning, and declines in fluid intelligence, because the necessary cognitive functions are not able to be executed within a time limit (time mechanism). The products of early processing may also decay before later processing is complete, meaning task-relevant information cannot be synchronised (simultaneity mechanism). By this account, age-related declines in cognitive functioning can be attributed to slower processing speed, which constrains the efficiency of working memory, in turn impacting on decision-making (Kail, 1991; Salthouse, 1996).

1.7.2.2 Developmental Cascade

Coyle et al. (2011) found that the total effect of age on fluid intelligence was almost fully mediated by the indirect effect of processing speed in adolescents aged 13 to 17 years. Cognitive developmental theorists, such as Fry and Hale (1996), explain this in terms of a ‘developmental cascade’, whereby age-related changes in processing speed precede changes in working memory, that in turn lead to improved fluid intelligence (Demetriou et al., 2012; Halford et al., 1998b; Kail, Robert & Salthouse, 1994; Kail, 1992). They found, in children aged 7 – 19 years, that age-related improvements in processing speed resulted in improved working memory. This age-related change was also linked to higher fluid intelligence. Kail (2007) confirmed these findings and demonstrated that they held longitudinally over the course of one year.

Nettleback & Burns (2010) extended this work using cross-sectional, latent variable analyses of a large sample of participants aged between 8 and 87 years, grouped into four age categories. Processing speed, working memory, and reasoning abilities improved from 8 to 45 years and declined thereafter. Processing speed moderated the relationship between age and working memory in children and adolescents and accounted for a large proportion of the variance in working memory, lending support to the developmental cascade theory. The adult data, however, revealed that other unidentified variables exerted a direct influence on working memory other than processing speed, and did not show the expected pattern of a mirror image of the child data. The authors suggested that age-related declines in reasoning performance may instead be influenced by a decline in short-term memory or the attentional aspects of working memory instead of processing speed. However, Dang et al. (2015) examined the role of short-term memory and processing speed in a large sample of adults and found that the processing speed component of working memory capacity accounted for the largest proportion of variance in fluid intelligence, supporting the theory that processing speed underlies the correlation between working memory capacity and intelligence by enabling faster rehearsal of information in working memory.

1.7.3. Additional Accounts of the Relationship Between Working Memory, Processing Speed and Intelligence

1.7.3.1 Schubert et al: Evidence Accumulation

An alternative account of the relationship between processing speed, working memory and fluid intelligence is provided by Schubert and colleagues (Schubert et al., 2017; Schubert & Frischkorn, 2020). They suggest that greater speed of processing increases the efficiency of information transfer from attentional processes to storage, increasing the availability of information in working memory, which in turn improves performance on fluid intelligence tests. According to Schubert & Frischkorn (2020), reaction times do not represent a single cognitive process, rather they reflect the time taken to complete several processes such as the encoding of information, decision-making, and motor execution. Therefore, one of the main theoretical questions driving their work is: do people with higher intelligence have greater speed of processing in all or some of these processes? To examine this, the neurocognitive-psychometrics account of mental speed combines mathematical models of cognition with neurophysiological data from EEG-recorded event-related potentials. These complementary approaches enable information processing to be segmented, to examine speed differences across individuals with a range of IQ, at different processing stages. Mathematical diffusion models decompose reaction time data into different parameters. The one of interest to the current debate is drift rate - the speed at which evidence is accumulated through sensory input and memory. It is thought to be a stable measure as it does not include variance from confounding variables typically included in reaction time data, such as speed/accuracy trade-offs, speed of encoding, or motor response speed (Frischkorn et al., 2019b; Schubert & Frischkorn, 2020). Drift rates are consistently strongly associated with IQ: Frischkorn & Shubert (2020) argue that individuals with higher intelligence benefit more from the speed of evidence accumulation.

Just as with diffusion models, neurophysiological data can be used to decompose information processing. Event-related potentials (ERP) reflect neural activity related to stimulus processing and can be associated with specific cognitive processes. Schubert, Hageman & Frischkorn (2017) found that certain ERP components explained about 80% of the variance in intelligence scores. The strongest association was found for the P300 ERP, which has been associated, amongst other things, with the inhibition of irrelevant information needed to facilitate the processing of information in working memory (Polich, 2007). In line with attentional control theories of intelligence, Schubert, Hageman & Frischkorn (2017) used this as evidence to propose that greater processing speed increases working memory capacity, and in turn intelligence, by increasing the efficiency of selective attention and memory updating (Unsworth et al., 2014a). They also proposed that increased processing speed may increase capacity or improve secondary memory by facilitating the creation and dissolution of temporary bindings. This is in line with other theories that propose the executive attention aspect of working memory is important for maintaining robust representations of the bindings between stimuli and responses in speeded tasks, and that these bindings in working memory are also necessary for building relational representations in reasoning tasks (Kane & Engle, 2003; Wilhelm & Oberauer, 2006)

1.7.3.2 Chuderski: Learning and Relation Processing

Chuderski (2013) offers an additional explanation for the association between working memory, processing speed and intelligence in terms of the role of relational representations and temporary bindings in fluid intelligence test performance. He analysed data from timed and untimed intelligence tests and found that when no time limit was applied, working memory capacity only explained approximately a third of the variance in fluid intelligence performance. Further analysis demonstrated that having unlimited time allowed low-capacity participants time to learn from previous test items and transfer this knowledge to subsequent item performance. Chuderski (2013) proposed that relational learning accounted for the additional proportion of variance in intelligence, unexplained by processing speed and working memory capacity. In other words, individuals who can learn and process relational representations better are more equipped to identify, transform, and apply the rules governing the relations in subsequent test items.

As previously discussed, fluid intelligence tasks require individuals to complete many processes, such as to encode stimuli, identify rules, integrate relations, form mental representations, and subsequently apply relations to response options. However, further work by Chuderski (2019) demonstrated that the ability to process a single relation or binding may be a core component of fluid intelligence. Three measures assessing the ability to process single relations were used to measure relation processing at the latent level. Relation processing explained approximately 90% of the variance in fluid intelligence and mediated the relationship between working memory capacity and fluid intelligence. Furthermore, relation processing and fluid intelligence shared approximately 66% of the variance unexplained by working memory. Therefore, Chuderski (2019, pg. 7) argues that fluid intelligence is equivalent to the “validation of correlates”, which is the ability to determine whether a given option does or does not satisfy a target relation. However, as this study was correlational, the findings cannot explain the exact cognitive mechanisms underlying the association between relation processing and fluid intelligence. Moreover, although working memory capacity was controlled for, other cognitive factors that have been found to explain variance in fluid intelligence were not included in the study, such as attention control, short- and long-term memory, and processing speed. Therefore, the possible contribution of these factors to relation processing and fluid intelligence could not be examined.

1.7.4 COGNITIVE SEGMENTATION

An additional cognitive constraint on intelligence test performance, long described in the literature but rarely isolated and examined, is cognitive segmentation. This refers to the ability to separate a complex problem into component parts that can be systematically attended to and processed, based on knowledge of the problem and its features. Both Egeth (1966) and Nickerson (1967) found "that individuals decompose geometric patterns by separately isolating and comparing the individual elements or dimensions" [In Mulholland et al., (1980), p. 257]. Similarly, Carpenter et al.'s (1990) eye-movement data showed that participants decompose the problem into smaller sub-problems and incrementally organise these sub-problems to find a solution. Carpenter et al. (1990, p. 429) concluded that decomposing problems and iterating through each sub-problem is the 'common ability' required for performance on all tests of intelligence. Furthermore, Luria & Tsvetkova's (1964) work with brain injury patients further emphasised the important contribution of cognitive segmentation to fluid intelligence. Using a geometric pattern-construction task, they found that correct performance was impossible if participants did not engage in the 'breaking down of the elements…into constructional units, and synthesis[e]…the necessary forms from the constructional elements' (Luria & Tsvetkova, 1964, p. 96).

While these ideas have been around for a long time, it is only recently that the role of cognitive segmentation in fluid intelligence was tested directly. This came about in part through Duncan and colleagues work (Duncan, 2013; Duncan et al., 2017, 2020), which called for a shift in our understanding of the link between intelligence and cognitive control functions. They argue that while aspects of control such as maintenance in working memory, attentional biasing, and inhibition are important, they do not provide sufficient accounts of even simple instances of organised cognition and behaviour. Guided by problem-solving programs in artificial intelligence, Duncan, Assem & Shashidhara (2020) argue that our understanding of 'control' must be broadened. Complex problems cannot be solved in one step. Instead, a goal-subgoal hierarchy is created in which the completion of subgoals moves the problem solver closer to completing the overall goal.

One of the key brain regions associated with intelligence is the multiple demand (MD) network, which consists of a broad set of frontoparietal brain regions (Assem et al., 2021; Duncan, 2005, 2013; Duncan et al., 1995; Tschentscher et al., 2017a; Woolgar et al., 2010, 2018). Activation in this network is suggested to coordinate activity in other brain regions so that novel, complex problems can be solved in adults (Assem et al., 2020; Duncan, 2013; Tschentscher, Mitchell, & Duncan, 2017) and children (Crone et al., 2009; Ferrer, O'Hare & Bunge, 2009; Wright et al., 2008). Patterns of activity differ in response to items of varying complexity in children and adults (Crone et al., 2009; Ferrer, O'Hare & Bunge, 2009; Wright et al., 2008), and it has been argued that activation in these regions becomes more precise across development (Ferrer, O'Hare, & Bunge, 2009).

Duncan and colleagues (Duncan, 2005; Duncan et al., 2017, 2020) suggest that one of the principal functions of the MD network is cognitive segmentation. They argue that the creation of attentional episodes that enable each component of a complex problem to be systematically encoded places a core limit on nonverbal reasoning task performance. They suggest that the segmentation of a complex problem into constituent parts or subgoals, and the integration of the correct combinations of information in each subgoal, gives rise to performance on fluid intelligence tests, and to all organised cognition and behaviour (Duncan, 2013; Duncan et al., 2017, 2020).

To test this theory, Duncan et al. (2017) modified a traditional geometric analogy task, making segmentation easy or hard to achieve, while also reducing demands on working memory and processing speed. In traditional matrix reasoning tasks (Figure 1), participants must consider the features that vary across the columns and rows to select the correct answer from a set of options. This requires multiple processes: identifying the individual features that connect the cells; establishing how they vary; holding in working memory different parts of the solution while working on others; and integrating different parts of the solution to choose between the possible answers, which may rely on processing speed to ensure all aspects of the problem are available simultaneously (Salthouse, 1996). Duncan et al. (2017) minimised the demands on speed, integration, and working memory by modifying three-feature matrix reasoning problems such that the only significant requirement was to break the three-feature problem into one-feature parts: to focus on one soluble part at a time.

Adults completed two different versions of this modified task. In the Combined format (Figure 2 a) participants were presented with a 2x2 matrix and asked to draw the missing matrix item into a response box below. Drawing the solution (rather than selecting an answer from a set of options) allowed participants to focus on one feature at a time, reducing demands on integration and working memory. In one experiment, the task was administered without time constraints to eliminate processing speed demands. In the Separated format (Figure 2 b) the problem was broken down into its component features across three separate cells, removing the need for participants to segment the problem themselves. Despite the reduced demands on working memory and processing speed, performance on the Combined version of the task remained poor for adults with low fluid intelligence. These errors largely vanished in the Separated condition when it became trivial to segment the problem into its component parts / features. These data indicate that when working memory, integration and processing speed demands are reduced in nonverbal reasoning tasks, the ability to cognitively segment remains critical to success.

Recent research demonstrating that performance can be improved when problems are broken down also supports the notion of cognitive segmentation. For example, Glady, French, & Thibaut (2016, 2017) found that alterations to the task presentation forced children to focus on the A:B relation prior to displaying C, which resulted in improved performance. Similarly, Rivollier et al. (2020) found that manipulating the amount of information made available to adult participants at any one time (i.e., presenting the matrix while withholding response options) encouraged the use of more efficient strategies. These findings support the hypothesis that the ability to cognitively segment complex problems into component parts benefits performance on tests of fluid intelligence.

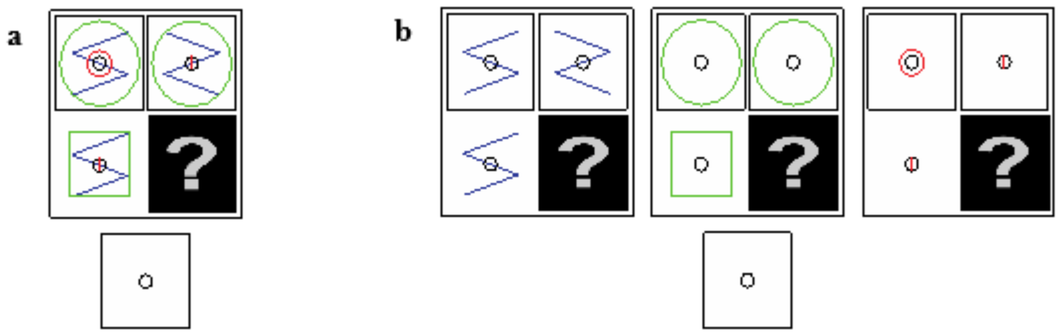


Figure 2: (a) cognitive segmentation Combined item. (b) cognitive segmentation Separated item. Note: Participants see Cognitive Segmentation task items in black and white. Features are shown in colour here for illustration purposes.

1.8 THESIS AIMS

The broad aim of this thesis was to develop a measure of cognitive segmentation for use with children (*Chapter 2*), and to use it to explore whether cognitive segmentation is important for children's analogical reasoning (*Chapters 4* and *5*). The first aim of the study presented in *Chapter 4* was to examine whether segmenting problems into their constituent parts aids fluid reasoning test performance in children aged 6-10 years in the same way as it does for adults with low IQ. A second aim was to explore the contribution of additional within-task characteristics (number of elements or features, and feature type), and problem-solving performance (error types and response times) to better understand why children struggle to solve complex problems. The aim of *Chapter 5* was to explore the relative contribution of cognitive segmentation, alongside working memory and processing speed, to fluid intelligence. The final empirical chapter (*Chapter 6*) presents a departure from cognitive segmentation, but continues to focus on intelligence, with a study in which children with a rare genetic mutation causing intellectual disability (ID) are phenotyped. The purpose of this study was to examine whether mutations on the STXBP1 gene gave rise to specific medical, cognitive, and behavioural phenotypes when compared to an ID-matched comparison sample. This study was the first piece of empirical work conducted during my PhD, which inspired my interest in understanding the processes that underlie our ability to engage with, understand, and solve complex problems. It is presented last to aide coherence. *Chapter 7* provides a summary of the thesis. The main findings and conclusions are discussed, in addition to the theoretical, methodological, and practical implications of this work, the limitations and potential areas for future research.

Chapter 2 Cognitive Segmentation Task Development

As discussed in Chapter 1 Section 1.7.4, recent research has demonstrated that when working memory and processing speed demands are minimised in tests of fluid intelligence, cognitive segmentation places a core limit on adults' performance. In this chapter, I develop a child-appropriate measure of cognitive segmentation. Task development took place over a period of 12 months. It was an iterative process involving the creation of multiple stimuli, several rounds of revisions, followed by two piloting sessions. The final version of the Cognitive Segmentation task used in subsequent experimental chapters, the task administration information, and scoring procedures are detailed at the end of the chapter.

2.1 INTRODUCTION

Performance on tests of fluid intelligence is influenced by a number of factors: the number of features, number of rules, type of rules, and the perceptual organisation of the problem. In a typical matrix reasoning problem, the number of features refers to the number of geometric shapes in a single cell of a matrix. The number of rules refers to the number of rules governing how each feature changes across the cells. As discussed in *Chapter 1, Section 1.6.1*, increasing the number of features and rules affects performance accuracy and increases the processing time required to solve the problem (Crone et al., 2009; Mulholland et al., 1980; Primi, 2001). In addition, as discussed in *Chapter 1 Section 1.6.3*, research has shown that the types of rules governing the feature pairs in a matrix may have differential effects on performance (Carpenter et al., 1990; Embretson, 1998; Green & Kluever, 1992; Hornke & Habon, 1986; Jacobs & Vandeventer, 1972; Mulholland et al., 1980; Primi, 2001; Whitely & Schneider, 1981). The elements presented in each cell are referred to as an item, and as discussed in *Chapter 1 Section 1.6.2*, their perceptual organisation can be "harmonic" or "nonharmonic" (Primi, 2001). Harmonic items, sometimes referred to as congruent, display a congruency in the arrangement of their features. The elements that make up harmonic items can be easily grouped across the cells in a matrix based on colour or form. Nonharmonic, or incongruent items, are designed with misaligned elements that " interrupt natural perceptual continuity of the elements" (Primi, 2001, p. 52) making it difficult to identify groupings across the cells in matrix.

While research has consistently attributed the management of these task factors to specific cognitive processes, such as working memory (Carpenter et al., 1990; Colom et al., 2008b; Kyllonen & Christal, 1990; Mulholland et al., 1980; Primi, 2001; Whitely & Schneider, 1981) and processing speed (Salthouse, 1996), Duncan et al. (2017) demonstrated that cognitive segmentation also places a core limit on our fluid reasoning abilities. As outlined in *Chapter 1 Section 1.7.4*, they used a modified nonverbal reasoning task that reduced working memory, integration and processing speed demands while manipulating the requirement for cognitive segmentation and found that the ability to cognitively segment remained critical to success.

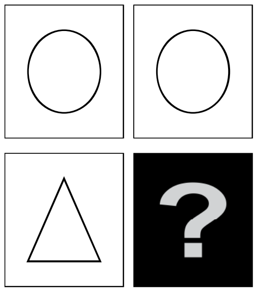
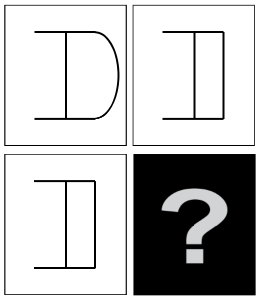
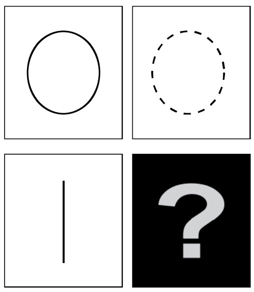
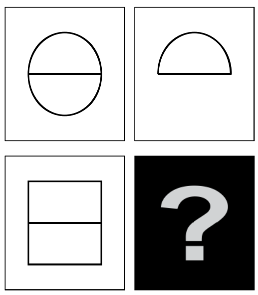
A central aim of this thesis is to use Duncan et al.'s (2017) paradigm with children to: i) test whether segmenting problems into their constituent parts aids performance in the same way as it does for adults with low IQ; and ii) explore the effects of additional task characteristics, such as the number of features and types of rules, on children's performance. This thesis presents the first attempt to examine the role of cognitive segmentation and complexity factors in relation to children's fluid reasoning abilities using Duncan et al.'s (2017) experimental approach. To use this approach, it was necessary to adapt the modified matrix reasoning tasks used by Duncan et al. (2017), to ensure they were appropriate for use with children. The aim of this *Chapter* was to develop this task.

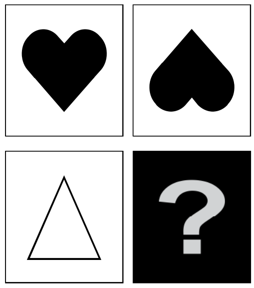
Following Duncan et al.'s design (2017), two formats of the task were developed to isolate the effects of cognitive segmentation: Combined (Figure 2 a, p.27 of this thesis) and Separated (Figure 2 b). The Combined format is similar to a traditional A:B::C:D analogical reasoning task, with each multi-feature item presented in a 2x2 matrix. In the Separated format each feature of an item is presented in its own matrix. Duncan et al.'s task contained 20 items, with a Combined and Separated version of each. Each item had 3-features.

There were several key factors to consider in re-designing this task for use with children. First, the geometric shapes used in the items had to be recognisable to children aged six to ten years (e.g., triangle, square, circle, line etc.; Aslan & Arnas, 2007). Second, it was important to test whether children could complete 3-feature items, and to develop new 1-, and 2-feature items to systematically examine the effect of the number of features on performance. Although 1-feature items do not provide information about children's abilities to segment, they do provide both a means of testing children's ability to reason relationally and an indication of their understanding of the types of rules governing the items. Third, to assess the role of rule type on children's performance, it was important to develop an *a priori* taxonomy of rules (see Table 1). These were developed using information from previous studies that have described and categorised different rule types for analogical reasoning tasks (e.g., Carpenter et al., 1990; Jacobs & Vandeventer, 1972; Primi, 2001). Similar to Duncan et al. (2017), but unlike many other previous studies using geometric analogy tasks (Crone et al., 2009; Mulholland et al., 1980; Primi, 2001), each feature of an item, in the newly developed task, was governed by a single rule type. This allowed the role of segmentation to be isolated from other factors that could influence performance: segmentation had to be the only factor differing across the Combined and Separated conditions, such that once a segmentation had been found, each feature could be processed independently. That is, the first feature could be drawn immediately, and the processing of one feature did not influence the processing of another. If additional rule types were collapsed on to a single feature, the Combined and Separated conditions would have differed in terms of re-evaluation as well as segmentation, and performance would have been more variable in terms of both error rates and reaction times (RTs). To eliminate this variability, and noise in the data, the new task was developed with one rule governing each feature. During task development and piloting, any items identified as having features governed by multiple rules, or which could be interpreted as such, were redesigned, or eliminated.

The development of the task was an iterative process, which involved the creation of four batches of stimuli prior to piloting, and then two rounds of revisions following three pilot sessions. This chapter describes the task development process, which took place over a period of 12 months.

Table 1: Taxonomy of rules governing the changes features across the matrix.

Rule Type	Specific Rule	Rule Description	Example
Constant	Constant in a row	The same feature occurs throughout a row (Carpenter et al., 1990).	
Design	Shape Change	Shape changes across a row	
	Design Change	Design of the shape changes across a row.	
Figural	Quantitative Pairwise Progression	Feature changes in size or quantity across a row e.g. halving or doubling (Bethell-Fox et al., 1984; Carpenter et al., 1990)	

Spatial	Rotation	An element rotates across a row (e.g.45º, 90º, 180 º (Bethell-Fox et al., 1984)	

2.2 DEVELOPMENT OF TASK ITEMS

2.2.1 BATCH 1.0

This first set of stimuli (Batch 1.0) consisted of 14 items: 2 x 1-feature, 12 x 2-feature. I designed the first batch of items in collaboration with architect Susie Newman. The items were then independently reviewed by two researchers (DM and JH). These items were not piloted with any children. The team concluded that most 2-feature items contained features that corresponded to different rule types and therefore could be incorrectly interpreted. For example, the Separated condition in Figure 3 demonstrates how each feature changes in size horizontally (big to small triangle; small to big square) and also in shape vertically (triangle to circle; square to triangle). Participants could identify either the features changing size horizontally (A:B), or the features changing shape vertically (A:C). I decided to redesign the items, increase the size of the shapes so they were more child-friendly, and remove shading from larger shapes to avoid the children spending a disproportionate amount of time colouring in the features. The number of items developed in the subsequent set was increased.

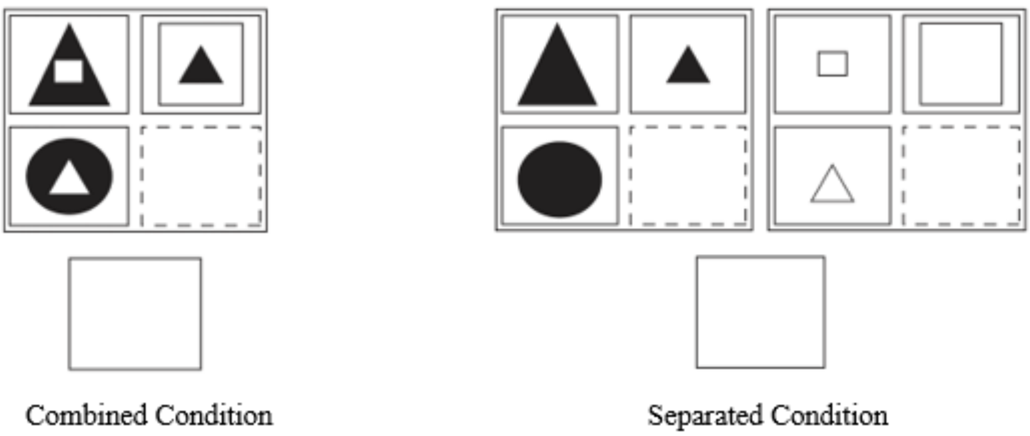


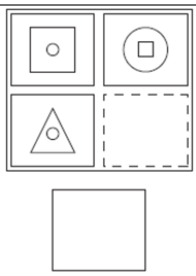
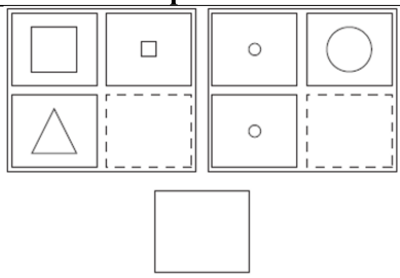
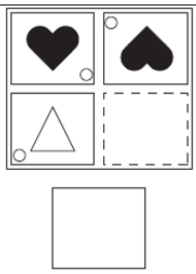
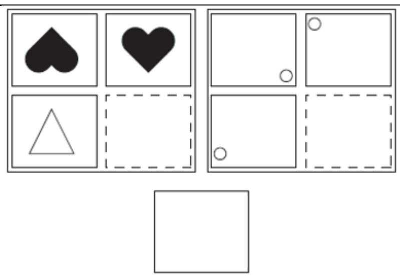
Figure 3: Batch 1.0, Item 5

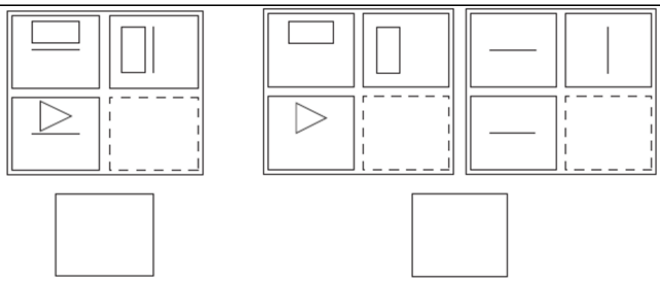
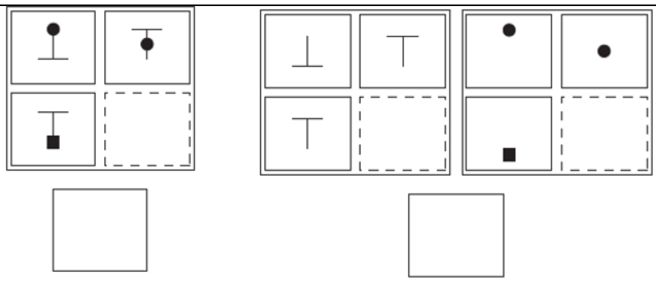
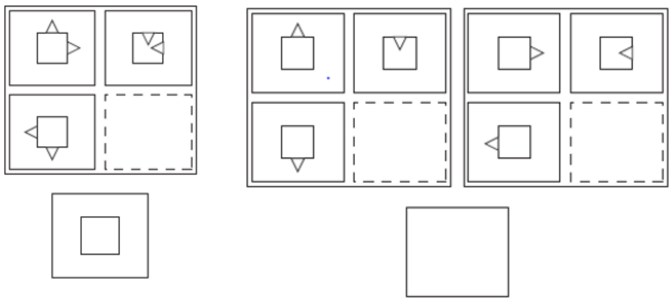
2.2.2 BATCH 2.0

In Batch 2.0, 46 items were created: 14 x 1-feature, 29 x 2-feature, 3 x 3-feature. The existing 1- and 2-feature items were updated in response to feedback from the independent reviewers (DM and JH), and a set of 3-feature items were created. The items were not piloted with any children.

The complete new set of stimuli were independently reviewed by four researchers (DM, JH, DA, FM). DM identified 3 items that were repeats of existing items and one item with misaligned features. The two main issues identified with Batch 2.0 in addition to the specific comments from DM (Table 2) were: i) Rule type ambiguity (see Table 2: items 19, 20, 28, and 41), and ii) the conflation of multiple features into a single feature due to the same transformation being applied to all features in an item (see Table 2, item 25).

Table 2: Reviewer comments Batch 2.0

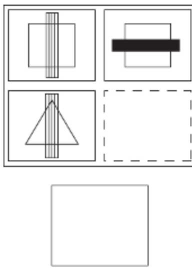
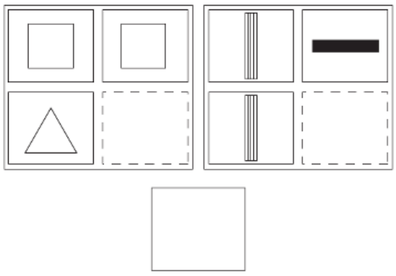
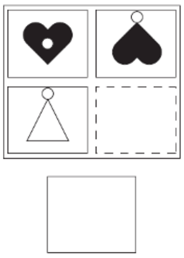
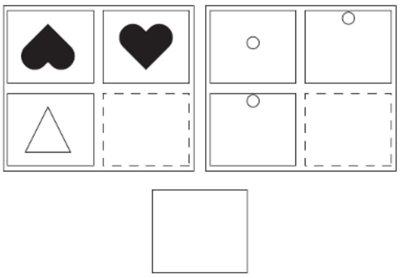
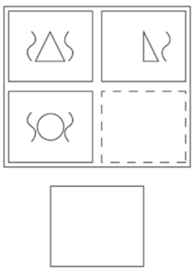
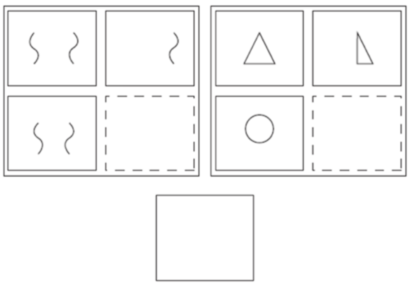
Item	Combined	Separated	Reviewer Comment
19			Combined: if participants focus on the small shapes first, they may put a small square in the bottom right. When they look at the larger shape, they would get stuck. This could be part of the challenge but could also be unfair because the little square *could* be a valid part answer until the rest is attended.
20			The 'best' answer is top right (180-degree rotation). However, the bottom left is also valid (reflection about y=x)?

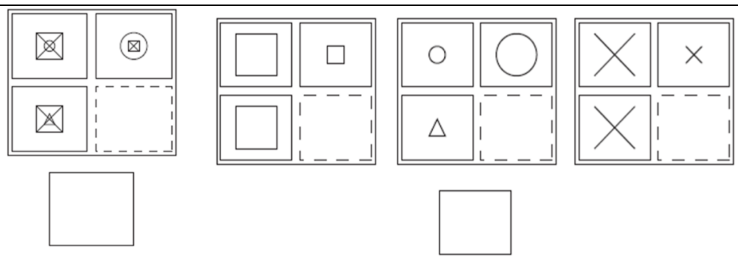
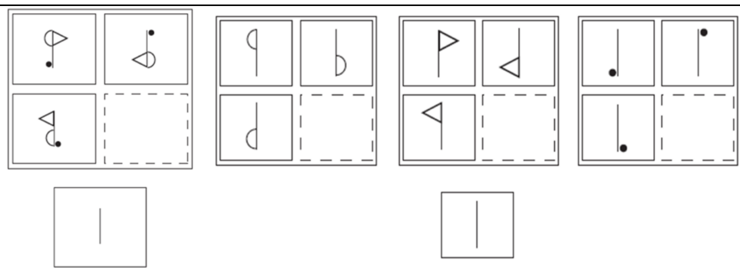
25		These are not necessarily 2 features: In the combined case the whole thing can just be rotated (or doubled) together. This also applies to items: 26, 29, 34, 36, 42, 44, 45, 46
28		It is obvious what the intended answer is but could also be valid to shift the square down a bit.
41		It is easy to think of this as a single rule: Fold all triangles across their attached edge or, if people look at the vertical relationship, rotate the item by 180 degrees.

2.2.3 BATCH 3.0

Batch 3.0 excluded eight items that were identified as problematic in Batch 2.0, and a further eight were updated (see Table 3). Batch 3.0 stimuli were independently reviewed by one researcher (DM). Three items in this batch were identified as collapsing multiple features into one (see Table 3, items 41 and 43). Some concerns were raised about Item 19, Batch 2.0. For example, if children decided to focus on the smaller features first, a valid answer would be a small square, participants would not realise that this is incorrect until the larger items were considered. While the complexity of the item is valid, this implicit increased complexity may have resulted in disproportionately higher reaction times or overall drawing times. If children realised that they were incorrect (in drawing a small square), they would have to redraw their answer, thus increasing overall drawing time. This item was eventually deleted from the final stimuli set.

Table 3: Reviewer comments Batch 3.0

Item	Combined	Separated	Reviewer Comment
14			It might be simpler if participants don’t need to add shading/texture (E.g., the left items could be filled, and right items hollow). Just because people differ a lot in how thoroughly and precisely, they think they need to do the shading.
20			The “opposite rule” is too difficult and unnecessary.
39			Features don’t match across conditions (orientation of upper left squiggles). Also, the compound rule for the first feature is unnecessarily difficult. Go with halving of the central shape, but only reflection of the line(s).

41		There are only 2 features/rules here: A crossed box that gets smaller, and an inner shape that gets bigger.
43		This is just a single rule: Rotate the whole thing by 180 degrees.

2.2.4 BATCH 4.0

Batch 4.0 included all but four items from Batch 3.0. Seven of the retained items were updated. These stimuli were independently reviewed by three researchers (DM, JH, and JD). Some minor amendments were suggested, including positioning all features in a consistent position in each matrix across both conditions (Combined and Separated) to eliminate the possibility that this could influence performance (DM), and developing an additional ‘pure’ 1-feature item set for piloting to ensure young children could understand the task instructions, layout, and the types of rules governing each feature (JD). The team agreed that following the implementation of these changes, the stimuli were ready for piloting in the next phase.

2.3 PILOTING

2.3.1 PILOT 1

The purpose of Pilot 1 was to test whether very young children could understand the task instructions, layout, and rules in the Combined condition. It also provided an opportunity to identify any potential design problems missed during the development phase, and to assess young children’s drawing abilities. Pilot 1 included six participants aged between 4 and 9 years, who each completed 50 items: 26 x 1-feature items, 21 x 2-feature items, and 3 x 3-feature items. Items were presented in the Combined format only.

Participant 01, a four-year-old, answered 22 out of the 26 1-feature items correctly, indicating they understood the task rules and were able to draw the responses. Four incorrect responses were due to two drawing errors, one copying a matrix term, and one instance of partially completing an item. The participant struggled with the 2-feature items, scoring 12 out of 21. However, one feature was correctly drawn in six of the eight items that the child got incorrect. This suggests an ability to isolate and attend to separate features of the problem, but that the increased complexity (i.e., 2-features) made it difficult to correctly solve the items. This was further supported by performance on 3-feature items. Although all items were incorrect, the participant did correctly draw one feature correct in each of the items. This young child could not finish all items in a single session, so completed them over several sittings.

For the following five participants, one 1-feature item was excluded due to a design flaw. Similar to participant 01, participants 02 and 03, aged four years, struggled to complete the task. Participant 02 refused to continue beyond the first two items. Participant 03 completed 21 1-feature items and answered 18 correctly. Three participants (Participant 04, 05 and 06), who were aged 6, 8, and 9 years respectively, completed the three complexity conditions (1-, 2-, 3-features). For these participants (see Table 4), almost all the 1-feature items were correct. In the 2-feature condition, all participants answered between 13 and 14 items correctly. None of the participants answered all 3-feature items correctly.

Table 4: Pilot participants number of items correct

Participant	Age	N correct		
		1-feature (/25)	2-feature (/21)	3-feature (/3)
01	4	22*	12	0
02	4	2	-	-
03	4	18	-	-
04	6	24	13	2
05	8	25	14	0
06	9	25	14	0

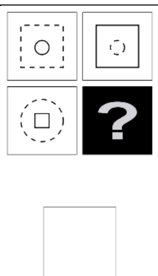
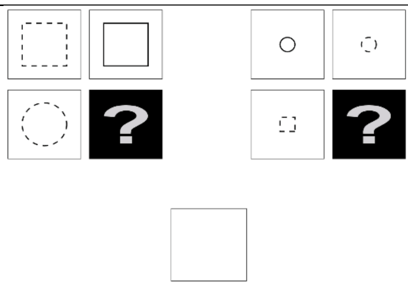
*Note: Participant received 26 items.

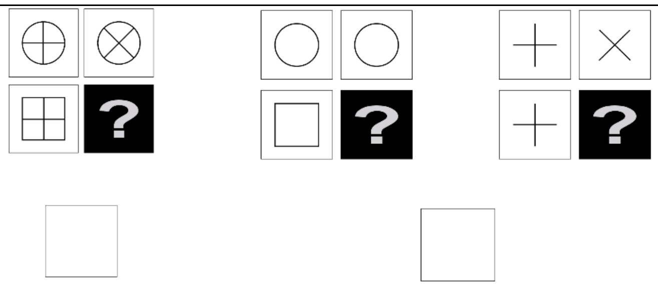
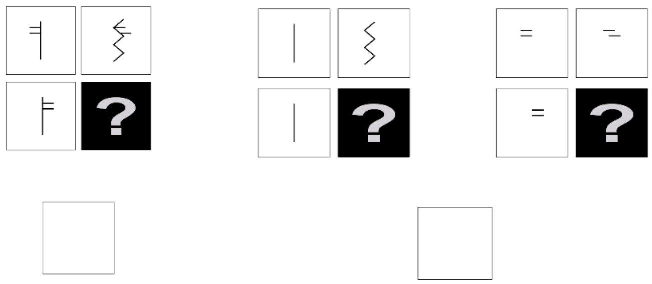
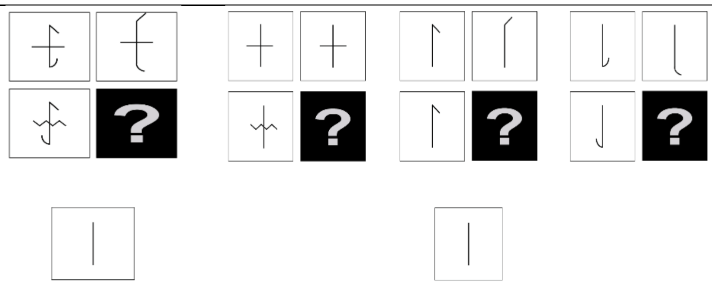
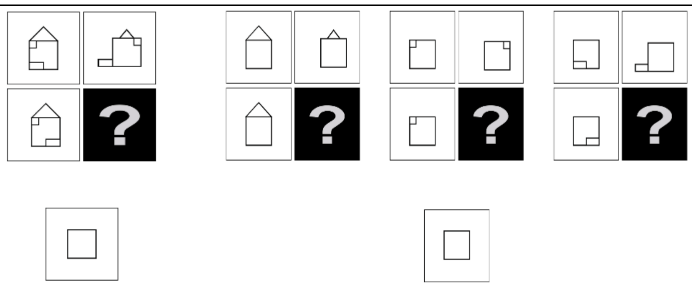
Overall, the task took the older participants (04, 05 and 06) approximately 40 minutes to complete. Two researchers (JH and DM) independently reviewed the task items and pilot performance. Item-related feedback is outlined in Table 5. To summarise, the research team identified three 2-feature items as problematic due to: collapsing features into one (2 items), and difficulty defining the number of features (1 item). One pilot participant (04) suggested that the zigzag shape used in a 1-feature item should be more pronounced. All of these items were updated in line with feedback ahead of the next Pilot.

Despite instructing children to draw the response in the answer box provided, all participants drew responses in the ‘D-term’ cell of the matrix. To ensure answers were placed in the answer box below the matrix, the task was modified prior to the next Pilot phase; a black square with a white question mark was designed and placed in the ‘D-term’ cell (see Table 5). Participants were often approximate with absolute sizes and the position of features in their responses (see Table 5 for positioning errors). Prior to the next Pilot phase, and in an attempt to resolve this issue, an ‘anchor shape’ was added that was common to all four entries in the matrix and already drawn in the answer box (see Table 3 item 41 with no anchor and item 43 for item with an anchor). In addition, it was decided that developing a detailed instructions script for the researcher would help alleviate confusion or misinterpretation of the rules. It would also ensure that each participant received the same information and task instructions.

Overall, piloting demonstrated that children as young as four could reason by analogy and understand both the task instructions and layout. However, performance indicated that children 4 – 9 years of age struggled to isolate and identify the correct features as the complexity of the task items increased. In addition, children aged 4 struggled to remain on task for the duration of the testing session. Performance was not at ceiling on 2-feature items and participant 04 showed that participants as young as 6 have the capacity to manage complex problems involving 3-feature items. Therefore, the decision was made to recruit children between 5 – 10 years of age in the experimental studies that follow in this thesis.

Table 5: Reviewer comments on Batch 4.0 items following piloting

Item	Combined	Separated	Reviewer Comment
30			Item 30 and 31 do not match – the bottom left square is solid when combined and dotted when separated. It is, perhaps best to avoid them needing to draw a small, dotted square, as this could be tricky.

36		This is a 1-rule item (rotate everything by 45 degrees). If subjects do this, then they will get a different answer to the separated version.
63		This might be tricky especially in the separated case because the offset of the horizontal lines does not really make sense without the context of the vertical zigzag. Perhaps the bottom part of the vertical line could get a kink while the top bit that the horizontal lines attach to could stay straight.
74		I think the rule for the bottom feature might be interpreted in a couple of ways. When I did the combined item, I rotated it 90 degrees clockwise, so it looked a bit like the tail of a 'q'. When I did the separated item, my instinct was to imagine it dropping under gravity, so it looked instead like the tail of a 'j'.
92		Elements that look like they should be in the same place across cells move around a bit. The big square moves slightly to the right, in both. It also moves slightly down in the separated version.

2.3.2 PILOT 2

The purpose of Pilot 2 was to: (1) pilot the task instructions script developed by SOB and DM; (2) time the task; (3) check understanding of Combined and Separated task layout, with the inclusion of the question mark feature in the 'D-term' cell; (4) examine the use of the anchor shape.

One 7-year-old child (participant 07) completed 60 items: 6 practice items (2 x 1-feature; 2 x 2-feature combined; 2 x 2-feature separated), and 45 test items (8 x 1-feature items; 12 x 2-feature combined items; 12 x 2-feature separated items; 11 x 3-feature combined items; 11 x 3-feature separated items). Participant 07 completed the same items in both the Combined and Separated format to explore whether the task instructions and items were understood. The participant answered all 1-feature items correctly. Performance on the Separated condition was higher than performance in the Combined condition for 2- (combined: 3/12, separated 8/12) and 3-feature items (combined: 1/11, separated 7/11).

The child enjoyed the task. However, they remarked that the task should have fewer items because 'it is too long and tiring'. The participant understood both task layout and instructions. The question mark feature ensured that responses were drawn in the answer box provided. The participant also understood the purpose of the anchor shape and used it.

2.4 PILOTING OUTCOMES

As outlined above, piloting performance led to the decision to restrict recruitment to participants aged 5 to 10 years, mainly because 4-year-olds struggled to stay on task. The number of task items in each task set was reduced (to 41 items, se *Section 2.5.2*). Piloting highlighted the need to monitor the performance of 5- and 6-year-olds closely as they struggled with some of the 2-feature and most of the 3-feature items. The decision was made to discontinue the task administration after the set of 2-feature items, and not administer 3-feature items, for 5-year-olds, and some 6-year-olds who struggled with 2-feature items. In addition, if they performed well on 2-feature items but expressed fatigue, they would be asked to complete 5 instead of 10 of the 3-feature items in both format conditions (Combined and Separated). Moreover, it became clear that including an anchor item in all items would help ensure consistency in drawing and scoring. Therefore, prior to testing, 2-feature items were updated to include an anchor feature. Participant 07, from Pilot 2, confirmed that the updated task instructions and task layout were understood. Overall, piloting revealed that children had the ability to understand and complete the task and enjoyed doing so.

The details of the final task are described in the following section. This task was used in the experimental studies in this thesis that investigate cognitive segmentation (*Chapters 4* and *5*) to capture children's performance on Combined and Separated versions of analogical reasoning tasks.

2.5 COGNITIVE SEGMENTATION FINAL TASK

2.5.1 ITEMS AND ITEM LISTS

Forty-six items, with a Combined and Separated layout, were developed in total (6 x practice items, 10 x 1-feature, 10 x 2-feature, 20 x 3-feature items). The forty test items (i.e., 1-, 2-, 3-feature items) were split into two lists of 20 test items (list I and II), with a Combined and Separated versions of each (e.g., Combined list I, Combined list II, Separated list I and Separated list II). These four lists were used to create two task sets with non-overlapping items in the Combined and Separated conditions: set A consisted of Combined items from list I and Separated items from list II; set B consisted of Combined items from list II and Separated item from list I. The presentation order of the items was counterbalanced: for half of set A and half of set B the Combined items were presented first (A1 and B1, see Appendix A and C), and for the other half the Separated items were presented first (A2 and B2, see Appendix B and D). Each task set (A1, B1, A2, and B2) contained the same practice items, and the same 2-feature practice items were used in the Combined and Separated format conditions. The purpose of the practice trials was not to test participants, but to ensure they understood the rules for each condition (Combined and Separated). Children did not receive practice items for conditions with 3-features. They were instructed that: 'The puzzles will get a little harder – you might even see three shapes!' Before trying 3-feature items, participants were told 'Now you will see puzzles with 3-shapes. Answer the puzzle in the same way as you did before.'

2.5.2 ADMINISTRATION

As outlined above, participants completed one of four task sets (A1, A2, B1 or B2; see Appendices A - D) containing 41 items: 6 practice items (2 x 1-feature; 2 x 2-feature combined; 2 x 2-feature separated), and 35 test items (5 x 1-feature items; 5 x 2-feature combined items;

5 x 2-feature separated items; 10 x 3-feature combined items; 10 x 3-feature separated items). Each task item was presented on an A4 page in landscape orientation. Participants were given unlimited time to solve each item. The reason for this is two-fold: 1) Duncan et al., (2017) found no significant difference in performance between participants who were timed and those that were untimed; and 2) one purpose of the task was to reduce the load on working memory, integration, and processing speed. Chuderski (2013) has shown that introducing any time constraints to fluid intelligence tests increases the relationship between working memory and reasoning (see *Chapter 1*, *Section 1.6.2.3.2*). The full task instructions are outlined in Appendices E and F.

2.5.3 Scoring

2.5.3.1 Item and Feature

For both the Separated and Combined tasks, the proportion of correct items and features was scored for each complexity condition (1-, 2-, 3-features) and each format condition (Combined and Separated). The proportion of correct items in each complexity condition was scored by summing the correct items (i.e., all features in an item are correct) and dividing by the overall number of items in the complexity condition (i.e., for 3-feature items the number of correct items was summed and divided by 10). The proportion of correct features in each complexity condition (1-, 2-, 3-features) was calculated by scoring the number of features correct per item (i.e., out of 3 in the 3-feature condition), and then summing the features correct across items to provide a total number of features correct (out of a total of 30; 10 items with a possible 3 features per item). This was converted to a proportion correct (total correct / 30).

2.5.3.2 Rule Type

To assess whether the rule types have a differential effect on performance the number of errors for each rule type was recorded. Each feature was assigned a rule category (see Table 2). The total number of errors within each rule category (constant, design, figural, and spatial) was calculated for each participant.

2.5.3.3 Error Type

For each feature, errors were classified into one of five categories: Wrong Alternative; Omission of Part; Other Drawing Error; Item Not Attempted; or Copying the C-Term. ‘Wrong Alternative’ errors occur when the participant draws the alternative feature that is presented in the matrix but is not the correct response feature. Participants who either forget, or fail to draw, all three features commit an 'Omission of part' error. An 'Other Drawing Error' refers to instances where a participant draws a feature that does not appear in the matrix, or when participants apply an extra transformation such as reflection. 'Item Not Attempted' errors occur when participants refuse to solve the item at all. 'Copying the C-Term' errors are instances where participants copy all features of the C-Term matrix item as their response. The error rate within each of these five categories was calculated for both the Combined and Separated format conditions.

2.5.3.4 Drawing Time

Participants' start time and end time was recorded using a stopwatch. As per Duncan et al. (2017) task protocol, both the start time or time to first stroke and the end time was noted. Time to first stroke gives an indication of the amount of time children take to analyse the analogy before responding. The overall drawing time is calculated by subtracting the start time from the end time. This measure gives an indication of the amount of time children take to draw each item.

2.6 CHAPTER SUMMARY

In this chapter, I outlined the processes involved in the creation of a child-appropriate measure of cognitive segmentation. The task development took place over a period of 12 months, requiring modifications to the adult task used by Duncan et al. (2017). This was achieved through careful stimuli design and two rounds of piloting. This task was created for use in subsequent chapters to assess both the role of cognitive segmentation in children's fluid reasoning skills (*Chapter 4*), and its contribution to fluid intelligence alongside other cognitive processes such as working memory and processing speed (*Chapter 5*).

CHAPTER 3 METHODS

In this chapter I present elements of the methodology that were the same for the empirical work in Chapters 4 and 5. Details specific to each experiment are provided within each Chapter, but here I provide a detailed description of the sample, data collection procedures, and tasks, along with information about the methods used to identify and correct for missing values. As a broad overview, a large sample of children (N=171) were invited to participate in the study. Each child completed three testing sessions. In Session 1, they completed an age-standardised measure of fluid intelligence, the Leiter International Performance Scale, 3rd edition. In Session 2, the Cognitive Segmentation task outlined in Chapter 2 was administered. Finally, in Session 3, children completed multiple tasks assessing working memory and processing speed.

3.1 DATA COLLECTION

3.1.1 PARTICIPANTS

A total of 171 children aged between 5 and 10 years were recruited from primary schools located in the South-West Region of the Republic of Ireland. Of these, 168 took part in the study. However, only 119 completed the full testing protocol (see *Section 3.1.3* for protocol; see *Section 3.2* for missing data information). Three children moved school prior to testing starting so were unable to take part in the study. There were some incomplete or missing data for those who completed the full protocol; details are provided in *Section 3.2*. The purpose of the experiments in this thesis was to examine 'novel' problem-solving abilities. Therefore, care was taken to ensure that the children recruited to the study had no prior experience of completing analogical reasoning tasks similar to those included in this study. In addition, to obtain a sample of children reflecting the socio-economic (SES) spread of Ireland, the schools selected for recruitment included children from a range of SES backgrounds (see Table 6). Deprivation indices were retrieved from the Pobal HP Deprivation Index (**https://maps.pobal.ie/**: Haase & Pratschke, 2012). Exclusion criteria for the study were pre-existing genetic or neurological conditions and having experience of nonverbal analogical reasoning tasks. Parents/carers provided informed written consent and children provided assent. Children were not paid for their participation. Ethical approval was granted by the Cambridge University Psychology Research Ethics Committee (CPREC reference: PRE.2018.051, see Appendix G for approval letter).

Table 6: School and main sample demographics

		School 1	School 2	School 3	School 4	Total
N		10	50	54	54	168
Gender	N females (%)	4 (40)	25 (50)	25 (46.3)	27 (50)	81 (48.2)
Age (mths)	Mean	113.6	98.14	92.30	94.94	96.15
	SD	4.65	15.67	17.30	19.89	17.84
	Range	106 - 120	69 - 131	63 - 123	62 - 130	62 - 131
						N (%)
SES	Very affluent	-	1	-	-	1 (0.67)
	Affluent	5	1	1	5	12 (8.11)
	Marginally above average	2	18	24	27	71 (47.97)
	Marginally below average	2	23	19	10	54 (36.49)
	Disadvantaged	-	1	2	4	7 (4.73)
	Very Disadvantaged	-	-	-	3	3 (2.03)
	Missing SES data	1	6	8	5	20

3.1.2 PROCEDURE

Each child participated in three testing sessions. Sessions 1 and 2 were completed on a one-to-one basis in a quiet area of the child's school. In Session 1, participants completed the Cognitive Segmentation task that was developed in *Chapter 2*. In Session 2, they completed an age-standardised test of fluid intelligence. Each of these sessions lasted approximately 40 minutes. Session 3 was conducted in groups of between six and 16 participants, depending on the number of children recruited from each school and the available testing space. Each child received an iPad that was pre-loaded with an app containing all the assessments. The children wore headsets so that they could listen to sounds and the spoken instructions. The headsets and privacy screen covers ensured that the children could not see or hear each other's iPads. I invigilated all testing sessions and answered questions where necessary, with the app developers on hand remotely to troubleshoot any technical difficulties. Each app task was preceded by child-friendly spoken instructions and many tasks had interactive practice trials. The app battery included five tasks, presented in a fixed order as follows: Cancellation; Digit Span; Go/No-Go; Dot Matrix; and Cattell. A game was included at the end of the battery that was intended as padding to prevent the participants who had finished early from distracting other children in the group. The app completion time varied between 40 and 60 minutes.

The session order was fixed: participants completed Session 1 first, Session 2 three to five days later, and Session 3 approximately one to two weeks later. Condition order was counterbalanced for the Cognitive Segmentation task: 53% completed the Combined condition first, and 47% the Separated condition first. The subscales of the standardised test of fluid intelligence administered in Session 2, and the measures of cognition and learning administered in Session 3 were administered in a fixed order.

3.1.3 MEASURES

3.1.3.1 Cognitive Segmentation

Both conditions of the Cognitive Segmentation task developed in *Chapter 2* were administered: Combined and Separated. To summarise the task here, it is a modified matrix reasoning task. The Combined format is like the original matrix reasoning task, with each item presented in a 2x2 matrix, while in the Separated format each feature of the item is presented in its own matrix. Unlike a traditional matrix reasoning task where participants select answers from multiple options, the children were asked to draw the answers in a box provided below each matrix for both conditions.

Children completed one of four task sets (A1, A2, B1 or B2, see Appendix A - D) containing 41 Items: two x 1-feature practice items; five x 1-feature items; two x Combined 2-feature practice items; five x Combined 2-feature items; two x Separated 2-feature practice items; five x Separated 2-feature items; 10 x Combined 3-feature items; 10 x Separated 3-feature items. Children who received Set A1 or Set B1 completed 2- and 3-feature Combined items first, followed by 2- and 3-feature Separated items. If children completed task set A2 or B2 the condition order was reversed i.e., Separated items first, followed by Combined.

Each task item was presented on an A4 page in landscape orientation. Participants were given unlimited time to solve each item. Time to first stroke (i.e., start time) and overall drawing time were recorded using a stopwatch. For both the Separated and Combined tasks, the proportion of correct features was scored. This was calculated by scoring the number of features correct per item (i.e., out of 2 or out of 3), and then summing the features correct across items to provide a total number of features correct (out of a total of 5, 10, or 30; i.e., 5 items with a possible 1 or 2 features per item, and 10 items with a possible 3 features per item). This was converted to a proportion correct (total correct: /5; /10; /30). The rule type and error types were also recorded for each feature.

3.1.3.2 Fluid Intelligence

3.1.3.2.1 Leiter Performance Scales, 3rd Edition

The Leiter Performance Scales, 3rd Edition (Leiter-3; Roid, Miller, Pomplun, & Koch, 2013) is a standardised test battery that assesses nonverbal reasoning abilities, attention, and memory in individuals aged 3 to 75+ years. Participants completed four standardised subtests from which the Leiter non-verbal IQ score, from hereon in the thesis referred to as Leiter nvIQ, was derived. The four subtests included: Figure Ground, Form Completion, Classification Analogies, and Sequential Order. Figure Ground is a visual inference task that requires participants to identify particular stimuli that are embedded in increasingly complex scenes. Children were given cards, each with a single image on it, and asked to match the cards to the correct image on a scene in the stimulus book. The scenes in the stimulus book become increasingly complex. Form Completion requires the synthesis of fragmented pieces of a stimulus in order to correctly identify the whole stimulus from a series of images. Participants were presented with a set of foam shapes arranged in a disorganised way. They had to manipulate the shapes to form a pattern that was the mirror image of a pattern presented in the stimulus book. As the task complexity increased, participants were given blocks instead of foam shapes. The images on the blocks were broken into pieces and arranged in a disorganised way. The children had to mentally arrange the fragmented image on the blocks and match each block to an image on the stimulus book. The block was ‘matched’ to the image by placing it in a plastic holder placed on the table in front of the stimulus book. The block was correctly ‘matched’ if it was placed directly under the correct image. Classification Analogies measures abstraction and rule generation. Participants had to sort and classify visual items according to different principles. For early items, participants classified items based on perceptual features such as colour and shape. For example, in the practice item participants were given yellow, blue, and red circles and the researcher indicated (non-verbally) that the foam shapes had to be sorted by colour. As the task became increasingly more difficult, items had to be classified based on semantic rather visual features i.e., a picture of a shoe and a foot, followed by a hand, where participants would select a glove. Sequential Order measures the ability to perceive sequential patterns and understand the rules and relationships between stimuli to find the missing element. Participants had to select the correct image (shown on a foam shape or block) to complete the sequence and place it in the correct space in the plastic holder in front of the stimulus book. Raw scores (sum of the correct responses) were converted to age-scaled scores for each subtest (range from 0 – 20, $M = 10$, $SD = 3$). A non-verbal fluid intelligence (nvIQ) score was calculated by summing the four subtest-scaled scores and converting them to age-scaled nvIQ scores using a table of norms.

3.1.3.2.2 Cattell Culture Fair Test

As part of Session 3, children completed two subtests adapted from the Cattell Culture Fair test, administered as part of the Resilience in Education and Development App (RED App Ireland (version 1.3.4): Bignardi et al., 2020; Dalmaijer et al., 2021). These were based on the Series and Classification subtests of the Cattell (Scale 2, Form A: Cattell, 1940; IPAT, 1973a, 1973b). In the Series subtest, children were presented with 12 items, each containing a series of three abstract figures and one empty box. For each item, children were required to choose which of five abstract figures completed the series. In the Classification subtest, children were presented with 14 items and asked, for each item, to identify which of five abstract figures was different from the others. Performance on both subtests was measured as the proportion of correct responses out of the total number of items (Series: 12 items, Classification: 14 items). The average proportion of correct responses from the two subscales was used as a proxy measure of fluid reasoning.

3.1.3.3 Processing Speed

Metrics of processing speed were derived from two reaction time-based tasks, Cat Cancellation and Go/No-go, presented on an iPad within the RED App (RED App Ireland (version 1.3.4): Bignardi et al., 2020; Dalmaijer et al., 2021). These tasks are used to measure higher-level cognitive processes, selective attention and inhibition respectively, but the indices derived from them for the current experiments measure speed of responses and were shown to reliably measure speed of information processing in the development of the RED App Ireland (version 1.3.4): Bignardi et al., 2020; Dalmaijer et al., 2021).

3.1.3.3.1 Cancellation

The Cat Cancellation task is a multi-target visual search task in which children are required to tap on all the targets on screen (cats with a smiling face) as quickly as possible. There are a total of 40 targets amongst 40 distractors. All stimuli were presented on a single screen. The task ended either when all targets had been found, or when two minutes had passed. Two versions of the task were administered: cancellation marked, followed by cancellation unmarked. In the marked version, tapping a target produced auditory feedback ("meow!") over headphones, and visibly marked it with a red cross. Only the auditory feedback was provided in the unmarked version. For both versions, the inverse median time taken between successful cancellations was used an index of processing speed (i.e., correct cancellations per second).

3.1.3.3.2 Go/No-Go

In the Go/No-Go task, stimuli appeared sequentially in the centre of the screen, and children were asked to tap each target stimulus (dog) and avoid pressing the distractors (poop). Participants heard a barking noise when a target stimulus was tapped and received 50 points. They heard a fart noise and were penalised by 150 points when a distractor stimulus was tapped. Stimuli remained on-screen until tapped, or until a timeout of 1.5 seconds occurred. The time between the offset of one stimulus and the onset of the next was 300-700ms (randomly drawn from a uniform distribution). Stimulus type was chosen randomly, with a probability of 0.8 for targets and 0.2 for distractors. The feedback score (added or subtracted points) was shown upon the tapping of a stimulus, and the total score was presented in the top right-hand corner of the screen. The inverse median reaction time for target stimuli was used as an index of processing speed.

3.1.3.4 Working Memory

Children completed two tests of working memory from the RED App Ireland (version 1.3.4; Bignardi et al., 2020), a Digit Recall and a Dot Matrix task.

3.1.3.4.1 Digit Span

The digit recall task required participants to recall lists of digits in serial order. The digits were presented one at a time, visually in the centre of the screen with simultaneous audio (e.g., "one" was heard as the digit 1 was presented). Each digit was presented on screen for 1 second, with a 1 second inter-stimulus interval between each digit. Participants entered responses into a number pad (including a delete button) which was presented on screen 1 second after the offset of the final digit. There was no response timeout. Digit sequences were randomised during task programming, ensuring that each participant was presented with the same sequences. The sequences started at span length three (three digits to recall). There were six trials at each sequence length. Span length increased by 1 digit if participants correctly recalled four or more of the six sequences presented. The task stopped when participants failed to recall three or more sequences correctly, or when they reached the maximum span length of nine. The number of trials correct (whole sequences recalled in serial order) was scored.

3.1.3.4.2 Dot Matrix

The dot matrix task required participants to recall spatial locations in serial order (cells that lit up one at a time in a 4x4 grid on screen). Each location lit up for 1 second, with a 1 second interstimulus interval between the presentation of each location. Participants were required to recall the sequence by tapping the memorised locations in serial order on the same grid at the end of each trial. Each location flashed for 100ms to indicate a response was recorded. Sequences were randomised during the programming of the task, so were the same for every participant. The task started at a span length of three (three cells lit up one at a time). There were six trials at each span length, and the difficulty increased by a span length of one if at least four out of six trials were correct. The task stopped if three or more trials were incorrect at a sequence length, or when the maximum span length of eight was reached. The number of trials correct (whole sequences recalled in serial order) was scored.

3.2 DATA MANAGEMENT

Overall, the percent of participants with missing data, who completed the full testing protocol was relatively low (N = 119; missing data: 0 – 8.4%; see Table 7). To ensure unbiased results, careful management of the missing values was required. Methods for managing missing data have moved beyond pairwise or listwise deletion or replacing values with measures of central tendency. Multiple imputation methods have demonstrated accuracy and robustness, even when the missing data rate is above 10% (Marshall et al., 2010; Newman, 2014; Johnson, 2018). missForest is a non-parametric imputation method that is implemented using a random forest algorithm (Stekhoven & Bühlmann, 2012). I chose this algorithm because it accommodates mixed data types and does not require tuning or specification of a parametric model (Liao et al., 2014; Stekhoven & Bühlmann, 2012; Tang & Ishwaran, 2017; Waljee et al., 2013).

Table 7: Percent of missing data for variables included in analyses

Variable	n_miss	pct_miss	Reason for missing data
Cognitive Segmentation	0	0	
Leiter_nvIQ	0	0	
Cattell series	7	5.88	Data not recorded for 3 participants and 4 missed the iPad session.

Cattell classification	10	8.40	Data not recorded for 6 participants and 4 missed the iPad session.
Go/No-Go RT Hit	5	4.20	Data not recorded for 1 participant and 4 missed the iPad session.
Cancellation marked intertime	6	5.04	Data not recorded for 2 participants and 4 missed the iPad session.
Cancellation unmarked intertime	5	4.20	Data not recorded for 1 participant and 4 missed the iPad session.
Digit Span n correct	6	5.04	Data not recorded for 2 participants and 4 missed the iPad session.
Dot Matrix n correct	6	5.04	Data not recorded for 2 participants and 4 missed the iPad session.

The missForest algorithm uses an interactive imputation process of training a random forest on observed values. First, the missing values for variables are filled in using either the median or mode imputation. Variables with missing values are treated as a response variable which borrow information from other variables to accurately predict it's value. The rows of data are marked as either training (data collected) or predicted values (median/mode imputed missing values). The data is then fed into a random forest model trained to predict missing values using resampling-based classification and regression trees which grows a random forest for the final prediction. This process continues iteratively, each time improving the prediction of the missing data until the imputed values reach convergence (Liao et al., 2014; Stekhoven & Bühlmann, 2012; Ye, 2020). This method was applied to the data using the missForest package in R (Stekhoven, 2013). To avoid underestimating the strength of the relationships between variables in further statistical analyses (Moons et al., 2006; Newman, 2014), all variables in the dataset were included in the imputation.

Data from 49 of the total 168 participants was excluded from the imputation due to large amounts of missing data. 45 of these were unable to complete the full testing schedule, predominantly due to age; the other four participants had incomplete data due to school absence. The resulting dataset was imputed six times. It is important to note that analysis in the following chapters will be conducted on the imputed dataset (*N* = 119).

3.3 Chapter Summary

In this chapter, I outlined the methods and tasks that were used to collect the data for the experiments presented in *Chapters 4* and *5*. I was due to repeat these assessments a year later to collect longitudinal data, but the schools closed due to the COVID-19 pandemic, meaning I was unable to collect these data.

CHAPTER 4 COGNITIVE SEGMENTATION AND FLUID REASONING IN CHILDHOOD

Some of the work presented in this chapter has been submitted for publication and received a revise and resubmit from the Quarterly Journal of Experimental Psychology: O'Brien, S., Mitchell, D., Duncan, J. & Holmes, J, "Cognitive segmentation and fluid reasoning in childhood". This chapter contains extensive additional analyses not presented in the submitted manuscript. These explore the effects of additional within-task characteristics on children's performance and investigate additional measures of children's problem-solving behaviour.

Previous research has shown that the ability to cognitively segment complex problems into smaller parts constrains nonverbal reasoning in adults. This chapter examines whether the same is true for children, by testing whether cognitively segmenting problems improves their nonverbal reasoning task performance. Using the age-appropriate Cognitive Segmentation task developed in Chapter 2, I collected data from a large community sample of children from the Republic of Ireland. Children completed both versions of the task, Combined and Separated (see Chapter 2 for details).

I hypothesised that performance in the Combined and Separated conditions of the Cognitive Segmentation task would be associated with fluid intelligence. Just as with adults, I hypothesised that children would perform better on the Separated than the Combined versions of the task. As a first step towards thinking about when cognitive segmentation develops, I explored the effects of age on performance within complexity conditions (1-, 2-, 3-features). I also examined the effects of task order to test whether completing the Separated items first influenced performance on subsequent Combined items. Finally, to understand how within-task complexity influences performance, data from the number of rules, types of rules, types of errors and overall drawing times were analysed.

4.1 INTRODUCTION

Understanding whether children can spontaneously segment complex problems, and how segmentation ability develops, has important implications for classroom practice. The ability to break down complex problems is crucial for classroom learning and educational attainment: children need to be able to decompose multi-step instructions, and break individual learning tasks into their component parts for success (e.g. Jaroslawska, Gathercole, Allen & Holmes, 2016). Knowing whether and when children can break down problems by themselves, and the beneficial effects of help in segmenting complex cognitive tasks, can provide a useful guide for teachers in terms of gauging the best way to present material, and the level of support they need to provide. The broad aim of this study was to investigate children's fluid intelligence abilities and their segmentation skills, testing: i) whether making segmentation easy to achieve aides performance on nonverbal analogical reasoning tasks; ii) whether age or IQ interact with the ability to segment; iii) how different within-task characteristics impact performance on nonverbal reasoning problems; and iv) how different problem-solving behaviours may inform us about why children fail to solve complex problems.

4.1.1 ANALOGICAL REASONING

As discussed in *Chapters 1 and 2*, fluid intelligence is often measured using novel, nonverbal analogy tasks, with an A:B::C:D format, organised in matrices of 3x3 or 2x2. To re-cap here, in an A:B::C:D analogy, the shapes and their relations vary systematically across the rows and columns, with the bottom right cell of the matrix left blank. In traditional reasoning tasks, participants must choose the answer from response options below the matrix (see Figure 4; a replication of Figure 1 in *Chapter 1*).

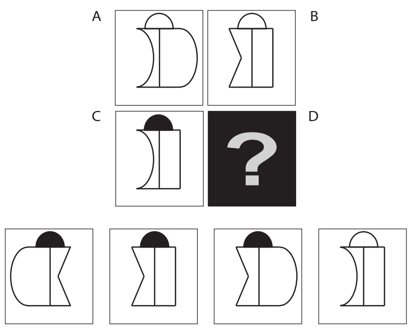


Figure 4: Example of a traditional matrix reasoning task item depicting the A:B::C:D format, with four possible response options below the matrix.

Previous research has shown that within-task characteristics, known as complexity factors (Primi, 2001), differentially affect how people manage the features in a matrix and their relations (see *Chapter 2*, *Section 1.6* for a full discussion). To summarise the main effects here: i) items with perceptually harmonious or congruent elements, or easily separated elements, are easier than items with elements that are hard to group or require dissociation from a whole to identify separate elements (e.g, Hornke & Habon, 1986; Primi, 2001); ii) increasing the number of features (elements and rules) negatively affects performance (Bethell-Fox, Lohman, & Snow, 1984; Carpenter, Just, & Shell, 1990; Crone et al., 2009; Freund, Hofer, & Holling, 2008; Mulholland et al., 1980; Primi, 2001) and; iii) the types of rules governing the relationships between different features of a matrix have differential effects on performance (Bethell-Fox et al., 1984; Blum et al., 2016; Green & Kluever, 1992; Kirchhoff & Holling, n.d; Whitely & Schneider, 1981).

4.1.2 DEVELOPMENT OF ANALOGICAL REASONING

As discussed in *Chapter 1 Section 1.4*, improvements in the ability to understand and manage relations and complex problems occur throughout childhood and adolescence (Chen et al., 2016; Christie & Gentner, 2014b; Ferrer et al., 2009; Mcardle et al., 2002; Richland et al., 2006; Starr et al., 2018; Sternberg & Rifkin, 1979), with large individual differences apparent in children as young as six years (Steveson et al., 2016; Whitaker et al., 2018). Young children tend to focus on perceptual matching rather than relational similarities (Christie & Gentner, 2014; Gentner & Medina, 1998; Sternberg & Nigro, 1980). It is not until five years plus, that children reliably use relational rather than perceptual or attributional interpretations. This is a period known as the relational shift (Gentner, 1988; Rattermann & Gentner, 1998), which is related to the theory that children's analogical reasoning skills develop as their knowledge base grows: children must know about conceptual similarities in order to draw analogies (e.g, Gentner, 1988; Goswami & Brown, 1990a). An alternative account explains the development of analogical reasoning through the development of executive functions, such as working memory and inhibition (e.g., Morrison et al., 2006; Richland et al., 2006; Thibaut et al., 2010; Thibaut & French, 2016). It is worth noting that there is no inherent contradiction between these accounts; cognitive maturation theories acknowledge a strong role for knowledge acquisition in the development of analogical reasoning skills. In the current study, I test the idea that cognitive segmentation places an additional constraint on the development of children's analogical reasoning performance.

4.1.3 Cognitive Segmentation

Cognitive segmentation has rarely been isolated and examined in relation to performance on nonverbal reasoning tasks. The seminal study in this area, conducted by Duncan and colleagues and described in full in *Chapter 1 Section 1.7.4* and in *Chapter 2*, found that modifying the layout of a traditional fluid intelligence task to aid segmentation, improved task performance in adults, particularly for those with low IQ (Duncan et al., 2017). This study demonstrated that when working memory, integration, and processing speed demands were reduced in nonverbal reasoning tasks, so that the only significant requirement is to break the multi-feature problem into one-feature parts, the ability to cognitively segment remained critical to success.

4.1.4 Cognitive Segmentation in Children

We propose that cognitive segmentation abilities will constrain children's nonverbal reasoning, and that this ability develops with age. This hypothesis is based on a body of research demonstrating that children do not appear to take a systematic approach to breaking down A:B::C:D problems. Adults typically focus first on the A:B pair, then move to C and the various response options (Gordon & Moser, 2007; Thibaut & French, 2016). In contrast, children aged up to 8 years orient to C and organise their search around this item and the response options (Thibaut & French, 2016). Thibault et al. propose that children orient to C because they cannot inhibit the primary task goal to "find what goes with C" (p. 24). Adults, however, are able to set aside the main goal, and break down the problem to focus first on the A–B pair. In other words, adults seem to segment the whole problem into sub goals, while children do not. Alternatively, children might be able to segment a problem into sub-goals but may struggle to inhibit or delay the second goal while acting on the first.

Orienting the search around the C-term instead of focusing on the A:B pair is also associated with poorer performance in both children and adults (Glady et al., 2016; Starr et al., 2018; Thibaut & French, 2011; Vendetti et al., 2017). Crucially, breaking down the problems to encourage participants to focus on one part of the problem at a time improves performance. Glady, French, & Thibaut (2016, 2017) found that children's performance could be improved by forcing them to focus on the A:B relation prior to displaying C. Similarly, Rivollier et al. (2020) found that manipulating when information was made available to adult participants (i.e., presenting the matrix while withholding response options) encouraged the use of more efficient strategies. These studies suggest that breaking down complex analogies into their component parts benefits performance. The benefits of segmentation appear to be particularly high for adults with low IQ (Duncan et al., 2017), and for children whose predominant response pattern is to focus on the final goal of complex problems.

4.1.5 Chapter Aims

The primary aim of this study is to extend Duncan et al.'s (2017) work to test whether segmentation constrains children's performance on nonverbal analogical reasoning tasks, as it does for adults. To do this, the child-appropriate version of Duncan et al.'s modified matrix reasoning paradigm developed in *Chapter 2* was used. Duncan et al. (2017) used data from 3-feature problems only. The child-appropriate version developed in *Chapter 2* contains items with 1-, 2 and 3-features, with 3-feature items skipped in the youngest and lower-performing individuals.

The main analyses, exploring whether segmentation benefits children's performance, and whether it interacts with age and fluid intelligence, were first conducted on the 3-feature data only, to replicate Duncan et al.'s (2017) approach. Data with 3-feature items were available for 115 participants, and the dataset used is referred to as the '1-2-3 feature' dataset from hereon. This dataset was used to explore the effects of age (do younger children fail to segment more than older children?) and to test for task order effects (does completing the Separated version of the modified matrix reasoning task first improve performance in the Combined condition)? I then replicated the analyses on 2-feature data to test whether the same patterns emerged: 2-feature data was available for 152 children (this dataset is referred to as the '1-2 feature" dataset from hereon). I hypothesised that children, like adults, would perform better on the Separated rather than the Combined conditions. This is because eye-tracking studies show children's performance can be improved by forcing them to focus on the A:B relation prior to displaying C (Glady, French, & Thibaut, 2017), and because children's performance on complex tasks can be improved when "useful scaffolding" is used, that is, when others divide complex tasks into simpler more manageable parts (e.g., Neitzel & Stright, 2003; Duncan et al., 2021).

An additional aim of the current study was to examine the effects on performance of other within-task characteristics, such as complexity and rule types. Based on previous research (see *Chapter 1, Section 1.6*), I hypothesised that: i) increasing the complexity of items (i.e., the number of features and their associated rules) would negatively affect children's performance on the cognitive segmentation task and ii) the types of rules governing the relationships between features in an item will have differential effects on performance, namely that constant rules would be easiest and spatial rules hardest to solve.

Finally, the types of errors children make, and their response times to complex problems are explored. These analyses were intended to be exploratory, therefore there are no formal hypotheses. Duncan et al. (2017) demonstrated that the majority of errors committed by adults were 'Wrong Alternative' errors i.e., when participants drew the alternative feature that was presented in the matrix but was not the correct response feature. In addition, Duncan et al. (2017) found that adults' time to first stroke (i.e., start time) was significantly longer in the Combined condition compared to the Separated condition; whereas total time spent drawing an answer was similar in both conditions. I investigated whether these findings would be replicated in children.

4.2 METHOD

Full methodological details are provided in *Chapter 3.* Details specific to this study are described in the following sections.

4.2.1 PARTICIPANTS

The analyses predominantly focus on 115 participants from the '1-2-3 feature' dataset (four of the 119 participants were identified as outliers), and, where appropriate, replication analyses are conducted with data from 152 participants from the '1-2-feature' dataset. The composition of these datasets are described below. Note that the 115 children who form the '1-2-3 feature' dataset are a subset of the larger '1-2 feature' dataset who were able to complete 3-feature items.

4.2.1.1 '1-2-3 feature' Dataset

A total of 115 participants aged between six and 10 years (age in months, *M* = 104.79, *SD* = 13.36, range = 81 - 131; n females = 58) were included in the analysis of 1-, 2-, and 3-feature items: 10 participants were recruited from school 1 (age in months, *M* = 113.60, *SD* = 4.65, range = 106 - 120; n females = 4); 39 participants from school 2 (age in months, *M* = 101.85, *SD* = 14.38, range = 82 - 131; n females = 22); 31 participants from school 3 (age in months, *M* = 107.08, *SD* = 11.65, range = 84 - 130; n females = 13); and 35 participants were recruited from school 4 (age in months, *M* = 103.29, *SD* = 14.16, range = 81 - 129; n females = 19). The participants recruited from these schools were from a range of socioeconomic backgrounds: affluent (10.4%); marginally above average (48.7%); marginally below average (35.7%);

disadvantaged (5.2%). Socioeconomic data (SES) was collected from the Pobal HP Deprivation Index system (https://maps.pobal.ie/).

4.2.1.2 '1-2 feature' Dataset

A total of 152 participants, aged between five and 10 years (M=98.56; *SD*=16.62; range=63-131) were included in the analyses of 1- and 2-feature items: 10 participants were included from school 1 (age in months, *M* = 113.60, *SD* = 4.65, range = 106-120; n female = 4); 47 participants from school 2 (age in months: *M* = 98.77, *SD* = 15.34, range = 71-131; n female = 25); 49 participants from school 3 (age in months: *M* = 96.53, *SD* = 17.70, range = 63-130; n female = 22); and 46 participants were included from school 4 (age in months: M = 97.24, *SD* = 17.05, range = 66-129; n female: 24). SES data: affluent (8.6%); marginally above average (45.4%); marginally below average (32.9%); disadvantaged (5.9%), data were missing for 11 participants.

4.2.2 PROCEDURE

To re-cap the important details from *Chapter 3*, participants completed three testing sessions. In Session 1 they completed all four subscales of the Leiter Performance Scales-3 non-verbal IQ assessment (Leiter Performance Scales (Leiter-3; Roid, Miller, Pomplun, & Koch, 2013), lasing approximately 40 minutes. Children completed both the Combined and Separated versions of the Cognitive Segmentation task in Session 2, on a one-to-one basis in a quiet area of their school. This session lasted approximately 40 minutes. In Session 3, the participants completed two subscales of the Cattell (series and classification) as part of the Resilience in Education and Development App protocol (RED App Ireland (version 1.3.4): Bignardi et al., 2020; Dalmaijer et al., 2021). Session order was fixed: all participants completed Session 1 first, Session 2 three to five days later, followed by Session 3 approximately one to two weeks later.

4.2.3 DESIGN

This was a within-participants design: all participants completed all tasks. Task order was counterbalanced for the Cognitive Segmentation task: 51.3% completed the Combined condition first and 48.7% completed the Separated condition first. Task order was fixed for the Leiter (Figure Ground, Form Completion, Classification Analogies, and Sequential Order) and the Cattell (Series then Classification).

4.2.4 MEASURES

4.2.4.1 Fluid Intelligence: Cattell Culture Fair Test

Participants completed two subscales of the widely used Cattell Culture Fair tests: series (12 items) and classification (14 items) (Scale 2, Form A: Cattell, 1940; IPAT, 1973a, 1973b). The average proportion of correct responses from the two subscales was used as a proxy measure of fluid intelligence. This measure was not age-standardised.

4.2.4.2 Fluid Intelligence: Leiter International Performance Scale-Third Edition (Leiter-3)

The Leiter-3 (Roid, Miller, Pomplun, & Koch, 2013) was administered to provide an age-standardised measure of fluid intelligence. Participants completed four subscales: Figure Ground, Form Completion, Classification Analogies, and Sequential Order. For each subscale, raw scores were converted to age-standardised scaled scores ranging from 0 – 20 (M = 10, *SD* = 3). The Leiter non-verbal fluid intelligence (nvIQ) score was obtained by combining the four subtest-scaled scores and converted using a table of norms.

4.2.4.3 Cognitive Segmentation Task

A detailed description of the Cognitive Segmentation task is outlined in *Chapter 2.* Participants completed one of four task sets (A1, A2, B1 or B2; see Appendices A - D). Each task item was presented on an A4 page in landscape orientation. Participants were given unlimited time to solve each item. Time to first stroke and overall drawing time were recorded using a stopwatch. To re-cap, the number of items per feature-condition are presented below.

4.2.4.3.1 3-feature Items

Participants completed 10 items in the Combined and Separated 3-feature conditions. Performance was measured as the proportion of correct features across all 10 items in each condition. This was calculated by scoring the number of features correct per item (out of 3), and then summing the features correct across items to provide a total number of features correct (out of a total of 30; 10 items with a possible 3 features per item). This was converted to a proportion correct (total correct / 30).

4.2.4.3.2 2-feature Items

Participants completed 2 practice items followed by 5 test items in both the Combined and Separated 2-feature condition. Performance was measured as the proportion correct i.e., summing the number of correct features across all items (out of a total of 10; 5 items with a possible 2 features per item) and dividing by the total number of items (total correct / 10).

4.2.4.3.3 1-feature Items

Participants completed 2 practice items followed by 5 test items in the 1-feature condition. Again, performance was measured as the proportion correct. This was calculated by summing the number of correct items (out of a total of 5) and dividing by the total number of items (total correct / 5).

4.2.4.3.4 Within-Task Characteristics Measures

Each feature was assigned a rule type: Constant, Design, Figural, and Spatial. Performance on each rule type was measured by calculating the proportion incorrect for each person, for each rule type.

Each feature that was incorrectly drawn by a participant was classified into one of five categories: Wrong Alternative; Omission of Part; Other Drawing Error; Item Not Attempted; or Copying the C-term. ‘Wrong Alternative’ errors occurred when the participant drew an alternative feature that was presented in the matrix but was not the correct response feature. Participants who either forgot, or could not draw, all three features committed an ‘Omission of Part’ error. An ‘Other Drawing Error’ referred to instances when a participant drew a feature that did not appear in the matrix, or drew the feature incorrectly (e.g., applying an extra transformation such as reflection). ‘Item Not Attempted’ differs from ‘omission of part’ as it refers to participants’ refusal to attempt any part of the item. ‘Copying the C-term’ refers to errors that arise due to the participants copying the item in the C-cell of the matrix.

Response times were recorded in seconds using a standard stopwatch. Time to the first drawing stroke and overall drawing time were analysed using the mean time in seconds.

4.2.5 ANALYSIS PLAN

Three sets of analyses were conducted. The first replicated Duncan et al.’s (2017) study with adults, investigating whether cognitive segmentation constrained children’s accuracy on nonverbal reasoning tasks. The second explored the impact of additional within-task characteristics (complexity and rule type) on performance. The third explored additional measures of performance (error types and response times). The analysis plan and results are split into these three sections accordingly.

Analysis 1: Cognitive Segmentation - Replication of Duncan et al. (2017)

The purpose of these analyses was to replicate Duncan et al.'s (2017) study. Fluid intelligence was initially indexed using the proportion correct score from the Cattell subtests, for consistency with Duncan et al., who had also used the Cattell Culture Fair test. However, because this measure was not age-standardised, all analyses were re-run using the Leiter-3 (Roid, Miller, Pomplun, & Koch, 2013), an age-standardised child measure of Fluid Intelligence. The primary analyses were conducted on 3-feature data from the '1-2-3 feature' dataset to be consistent with Duncan et al. (2017). They were also replicated on 2-feature data from the '1-2 feature' dataset' (N=152) to explore whether the same patterns emerged. The '1-2 feature' dataset included all participants who completed the 3-feature items, plus a group of predominantly younger children who were unable to complete the 3-feature items. Many of these additional children were too young to complete the Cattell fluid intelligence task, so the analyses on the 2-feature data were only run on the Leiter-3 non-verbal IQ (nvIQ).

Prior to data collection, I pre-registered the analysis plan on AsPredicted (#13338; see Appendix H). This stated that the analyses relevant to the central research questions posed in this study (does segmentation facilitate children's performance on a modified matrix reasoning task, and is there a relationship with age and fluid intelligence ability) would include: i) regression-based general linear models (GLM) to predict proportion of correct responses by condition (Combined and Separated) and fluid intelligence ability, and ii) correlation and regression analyses to test associations between cognitive segmentation and age. The primary analyses followed these steps.

To replicate the analyses of Duncan et al.'s (2017) study with adults, I conducted additional analyses not stated in the pre-registration. Order effects were tested to explore whether completing the Separated condition first enhanced performance in the Combined condition. Simple order effects were tested using a GLM with Condition as the within-subjects variable and task order as the between-subjects variable. To test whether any beneficial effect of receiving the Separated condition first was greater for younger children or those with low fluid intelligence, a GLM was run with Condition as a within-subjects factor, task order as a between-subjects factor, and mean-centred fluid intelligence and mean-centred age as between-subjects covariates.

Analysis 2: Additional Within-task Characteristics

The purpose of these analyses was to explore how two additional within-task characteristics, Complexity and Rule Type, influence children's performance on the modified matrix reasoning task.

Complexity refers to the number of rules children must solve. The design of the Cognitive Segmentation task meant that each feature was assigned one rule. Therefore, the analysis of the effects of Complexity examines the influence of additional features and rules on performance. A repeated measures ANOVA explored the impact of increasing the complexity of items in both conditions. Following this, I explored whether performance on different complexity levels varied with Condition (Combined and Separated), Fluid Intelligence, and Age.

Rule Type refers to the type of rule governing the feature. Each feature was assigned to a Rule Type: Constant, Design, Figural, and Spatial (see *Chapter 2* for details). A repeated measures ANCOVA determined whether different types of rules differentially influenced performance and whether this varied with Condition, Fluid Intelligence, and Age.

Analysis 3: Additional Dependent Measures

In this section of the analysis, I explored two additional dependent measures: Error Type and Response Times.

Every incorrect response made by a participant was categorised as one of five Error Types: Wrong Alternative, Omission of Part, Other Drawing Error, Item Not Attempted, and Copying the C-Term. I was interested in whether the types of errors children make may provide clues to why they fail to solve complex problems. This was intended to be exploratory with a view to igniting a discussion about what different types of errors may reflect in terms of children's problem-solving difficulties i.e., goal neglect, attention difficulties, or working memory difficulties. A repeated measures ANCOVA examined whether the proportion of each error type varied with Condition, Fluid Intelligence, and Age.

The final set of analyses explored whether children's response times varied across conditions. Children's Start Time (or time to begin drawing their answer) and overall Drawing Time were recorded. As before, a repeated measures ANCOVA examined whether children's response times to task items varied with Condition, Fluid Intelligence, and Age.

4.3 RESULTS

4.3.1 ANALYSIS 1: COGNITIVE SEGMENTATION IN CHILDREN

The first part of this analysis was conducted with 3-feature items from the '1-2-3 feature' dataset, as per Duncan et al., (2017). The Cattell fluid intelligence task was used as the primary measure of fluid intelligence, but the analyses are repeated using the Leiter non-verbal IQ, as an age-standardised measure of fluid intelligence. The analyses include general linear models (GLMs) predicting proportion of correct responses by condition (Combined and Separated) with fluid intelligence ability. Improvements in the ability to reason about relations develop throughout childhood (Christie & Gentner, 2014; Ferrer et al., 2009; Mcardle et al., 2002; Richland et al., 2006; Starr et al., 2018; Sternberg & Rifkin, 1979). Therefore, as a first step towards thinking about when cognitive segmentation develops, the effects of Age will also be examined in this section. As stated previously, the original plan was to collect longitudinal data for these purposes, but this was not possible due to school closures arising during the COVID-19 pandemic.

4.3.1.1 Cognitive Segmentation and Fluid Intelligence: 3-feature items

Performance was higher in the Separated (M = 0.89, SD = 0.16) than the Combined condition (M = 0.56, SD = 0.19). Mean performance was close to ceiling in the Separated condition. Average proportion correct on the fluid intelligence task was M = 0.52, SD = 0.14.
There was a significant association between performance in the Combined condition and Fluid Intelligence, r = 0.38, p <.01, and between the Separated condition and Fluid Intelligence, r = 0.41, p <.01. To examine if there was a significant difference in the strength of the correlations between these associations the correlation coefficients were transformed to z-scores. The z-scores were compared using an online calculator for comparing correlations (https://www.psychometrica.de/korrelation.html). The comparison of correlations from the dependent samples test revealed no significant differences in the strength of these correlations (z = -0.34, p = 0.37).

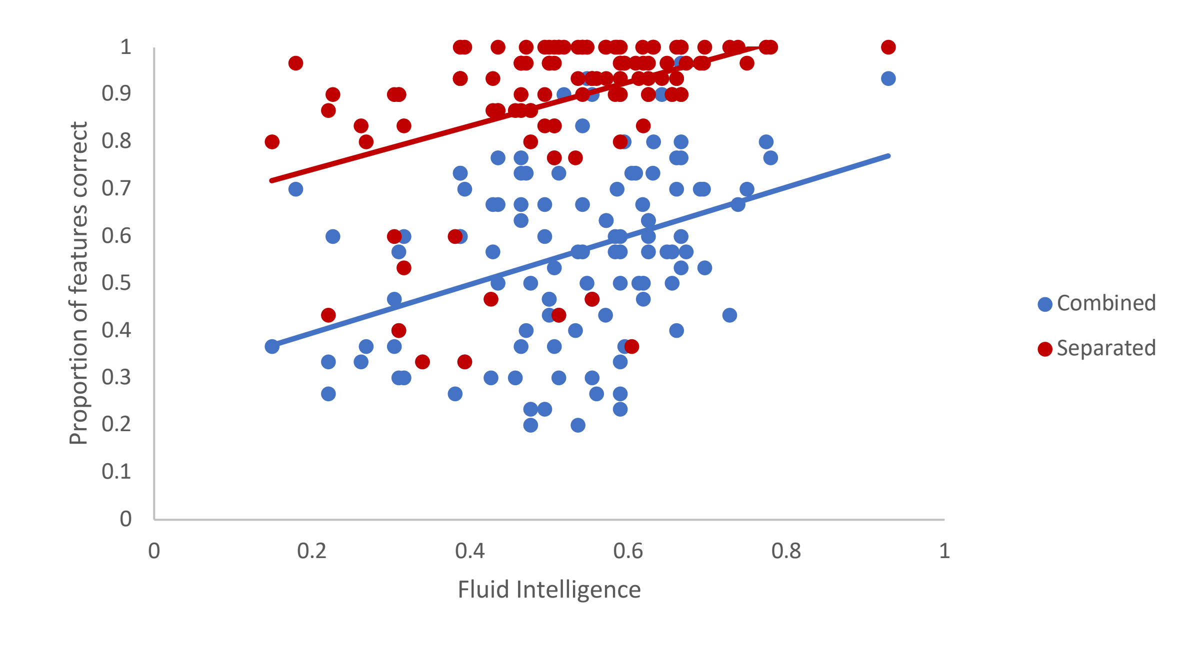


Figure 5: Association between performance in the 3-feature Combined and Separated conditions and Fluid intelligence, measured by the Cattell subtests.

A GLM predicting proportion of correct features from Condition (Combined and Separated) and mean-centred Fluid Intelligence revealed a significant main effect of Condition (F(1, 113) = 379.02, p < 0.001, ηp^2 = 0.77; see Figure 5) and Fluid Intelligence (F(1, 113) = 29.18, p <.001, ηp^2 = 0.21). Children performed better in the Separated condition, and those with higher Fluid Intelligence scores achieved better scores in both the Separated and Combined conditions relative to those with lower Fluid Intelligence scores. There was no interaction between Condition and Fluid Intelligence (F(1, 113) = 0.14, p = 0.71, ηp^2 = 0.001), hence there was no evidence that the Separated condition was more helpful for children with lower fluid intelligence scores.

The above analyses were re-rerun using the age-standardised measure of Fluid Intelligence, The Leiter Performance Scales, 3rd Edition (Leiter-3; Roid, Miller, Pomplun, & Koch, 2013). The same pattern of results emerged using this scale. To summarise, there was a significant association between performance in the Combined condition and Fluid Intelligence (Leiter non-verbal IQ: Leiter nvIQ), r = 0.43, p <.001, and between the Separated condition and Fluid Intelligence, r = 0.50, p <.001. Again, there was no significant difference in the strength of these correlations (z = -0.81, p = 0.21). There was a significant main effect of Condition (F(1, 113) = 379.15, p < 0.001, ηp^2 = 0.77) and Fluid Intelligence (F(1, 113) = 44.65, p <.001, ηp^2 = 0.28), but there was no interaction between Condition and Fluid Intelligence (F(1, 113) = 0.20, p = 0.66, ηp^2 = 0.002; see Figure 6).

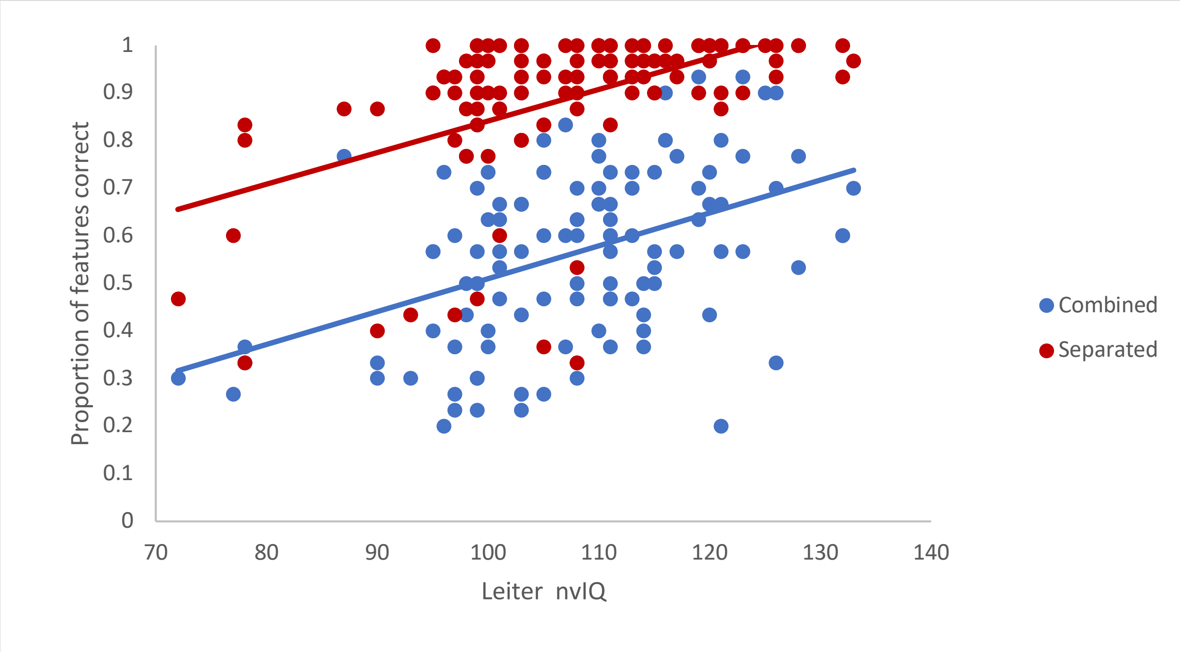


Figure 6: Association between performance in the 3-feature Combined and Separated conditions and Leiter nvIQ.

4.3.1.2 Cognitive Segmentation and Age: 3-feature items

There was a significant but weak association between performance in the Combined condition and Age in months, $r = 0.17$, p <.05, and between the Separated condition and Age, $r = 0.38$, p <.01. A comparison of the correlations demonstrated that the association between Age and the Separated condition was significantly stronger than the association between Age and the Combined condition ($z = -2.21$, $p < 0.05$).

A GLM predicting proportion of correct features from Condition (Combined and Separated) and mean-centred Age revealed a significant main effect of Age ($F(1, 113) = 11.73$, $p < .001$, $\eta p^2 = 0.09$). Children performed better in the Separated condition, but there was no interaction between Condition and Age ($F(1, 113) = 2.63$, $p = 0.11$, $\eta p^2 = 0.02$; see Figure 7) hence no evidence that the Separated condition was more helpful for younger or older children.

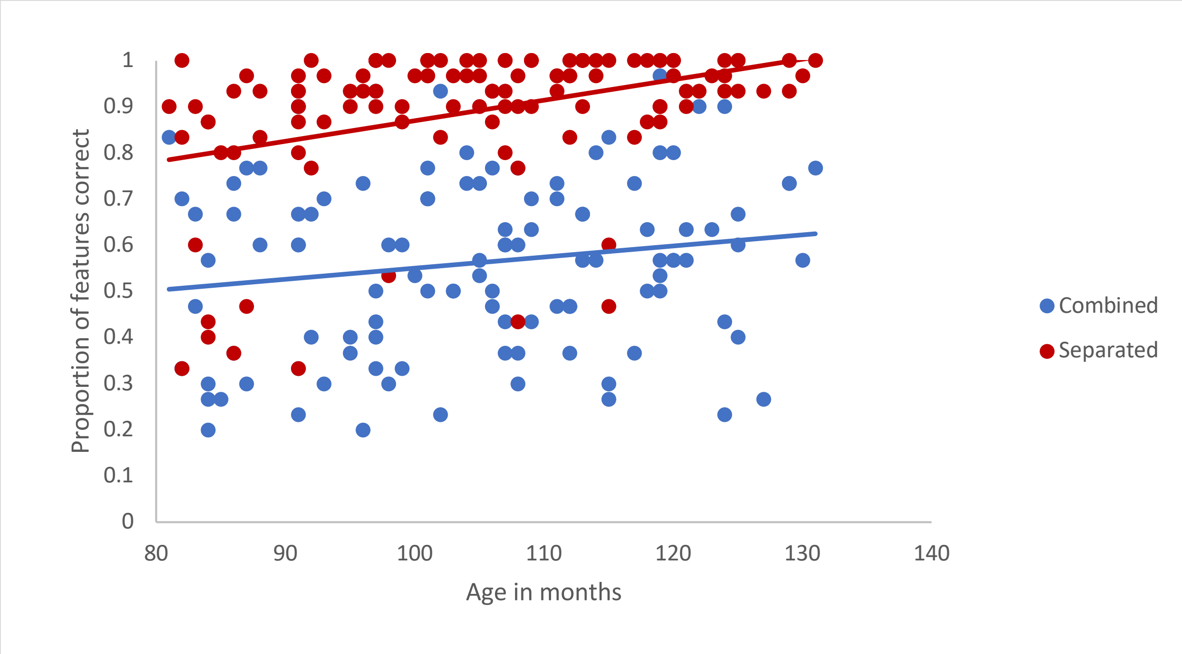


Figure 7: Association between performance in the 3-feature Combined and Separated conditions and age. Note: R2 values come from simple correlations per condition.

4.3.1.3 Cognitive Segmentation Order Effects: 3-feature items

To explore whether there were practice effects associated with completing the Separated condition first, order effects were investigated. Performance was independent of condition order (F(1, 113) = 0.01, p = 0.94, ηp^2 = 0.000). There was also no Condition x Order interaction (F(1, 113) = 0.55, p = 0.46, ηp^2 = 0.005). These conclusions held when fluid intelligence (indexed by either the Cattell or Leiter nvIQ measures) and Age were added as covariates. For all higher-order interactions of Condition x Order with these covariates, F < 0.97, p >0.76, ηp^2 <0.009. Therefore, there was no evidence for task order effects, or that these might depend on fluid intelligence or age.

4.3.1.4 Cognitive Segmentation: 2-feature items only

To examine whether the same pattern of results emerged using less complex problems, the analyses reported in sections 4.3.1.1., 4.3.1.2, and 4.3.1.3 were replicated using data from the 2-feature items (the '1-2 dataset', n=152). As a reminder, the measure of fluid intelligence used in these analyses was the Leiter nvIQ. Overall, the pattern of effects revealed using 2-feature items was the same as that for 3-feature items. A summary is provided below.

4.3.1.4.1 Cognitive Segmentation and Fluid Intelligence: 2-feature items

There was a significant association between performance in the 2-feature Combined condition and Fluid Intelligence (Leiter nvIQ), $r = 0.50$, $p < .01$, and between the 2-feature Separated condition and Fluid Intelligence $r = 0.55$, $p < .01$. A comparison of the correlations revealed no significant difference in the strength of these correlations ($z = -0.95$, $p = 0.17$). There was a significant main effect of Condition ($F(1, 150) = 61.52$, $p < 0.001$, $\eta p^2 = 0.29$) and mean-centred Fluid Intelligence ($F(1, 150) = 71.64$, $p < .001$, $\eta p^2 = 0.32$) but no significant interaction between Condition and Fluid Intelligence ($F(1, 150) = 0.72$, $p = 0.39$, $\eta p^2 = 0.005$; see Figure 8).

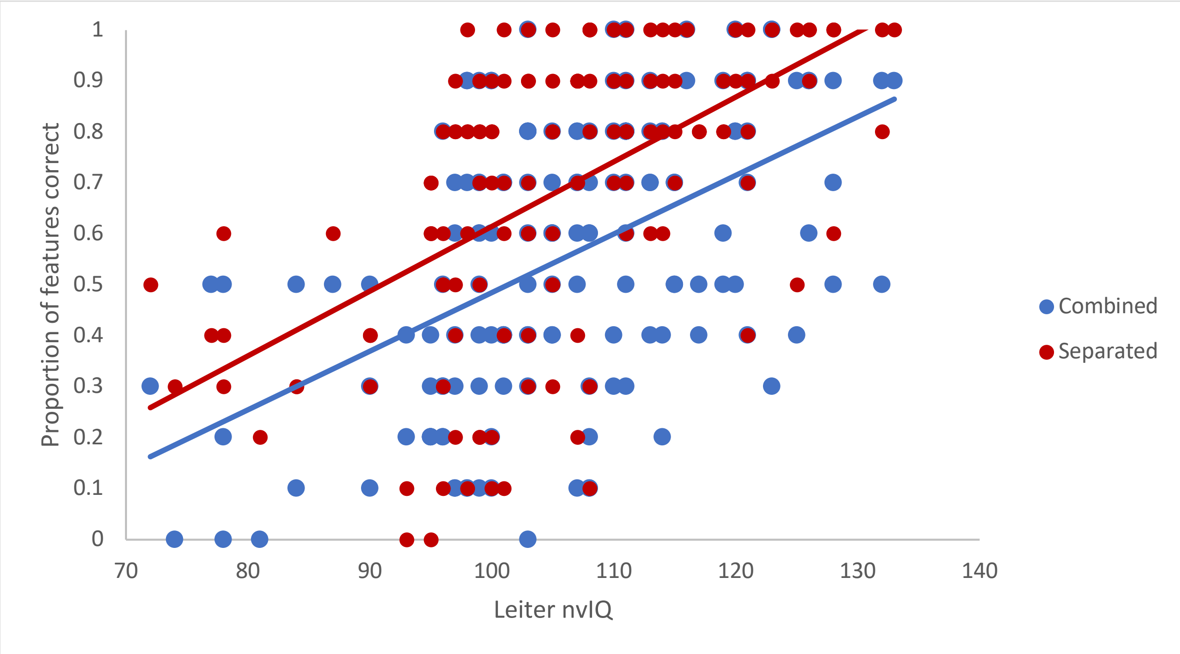


Figure 8: Association between performance in the 2-feature Combined and Separated conditions and Leiter nvIQ.

4.3.1.4.2 Cognitive Segmentation and Age: 2-feature items only

There was a significant association between performance in the Combined condition and Age in months ($r = 0.50$, $p < .01$), and between the Separated condition and Age ($r = 0.54$, $p < .01$). These correlations were not significantly different ($z = -0.75$, $p = 0.23$). As with 3-feature items, there was a significant main effect of mean-centred Age ($F(1, 150) = 61.92$, $p < .001$, $\eta p^2 = 0.29$) but there was no interaction between Condition (2-feature Combined and Separated) and Age ($F(1, 150) = 1.71$, $p = 0.19$, $\eta p^2 = 0.01$; see Figure 9).

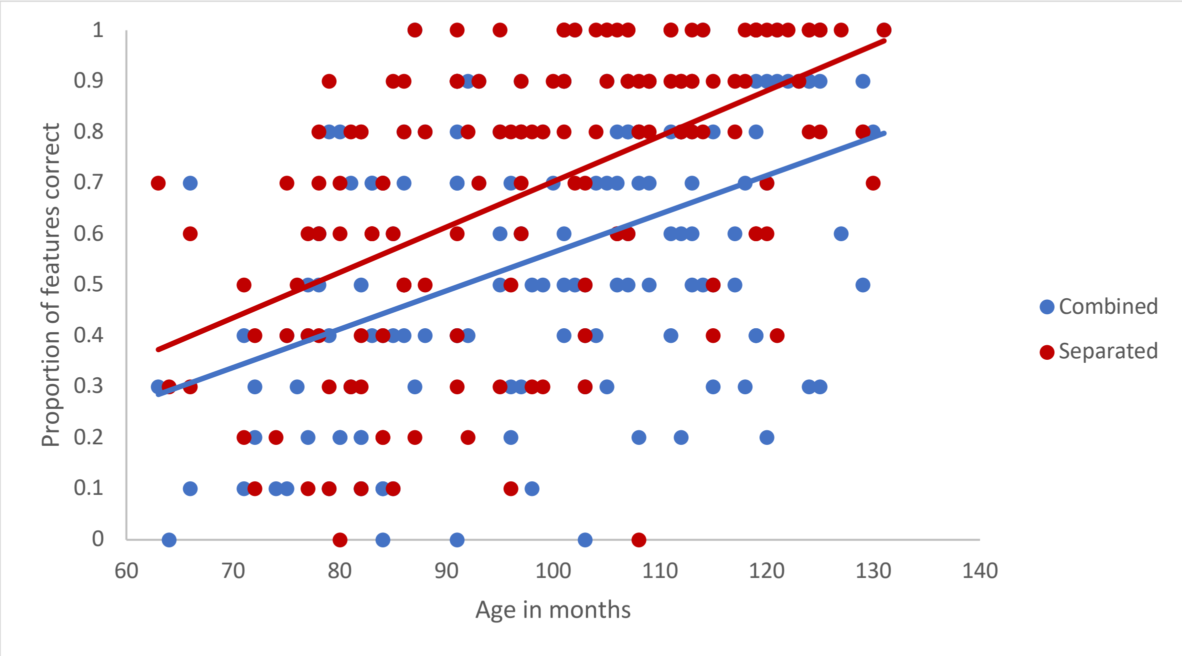


Figure 9: Association between performance in the 2-feature Combined and Separated conditions and age in months.

4.3.1.4.3 Cognitive Segmentation Order Effects: 2-feature items

Again, performance was independent of condition order ($F(1, 150) = 0.87$, $p = 0.35$, $\eta p^2 = 0.006$). There was no significant Condition x Order interaction ($F(1, 150) = 3.85$, $p = 0.052$, $\eta p^2 = 0.025$), and these conclusions held when mean-centred Leiter nvIQ and mean-centred Age were added as covariates. There were no significant higher-order interactions of Condition by Order with these covariates, $F< 2.99$, $p >0.86$, $\eta p^2 <0.020$. Therefore, as for the 3-feature items, there was no evidence for effects of task order.

4.3.2 Analysis 2: Additional Within-Task Characteristics

The analyses in this section explore how two additional within-task characteristics, Complexity and Rule Type, influence children's performance on the two conditions of the Cognitive Segmentation task.

Complexity refers to the number of rules children must isolate and solve. As previously outlined, the task was designed such that each feature was assigned a single rule (see *Chapter 2* for details and rationale). Thus, the number of rules is equal to the number of features in each item (e.g., there are 2 rules for 2-feature items, and 3 for 3-feature items). The effects of complexity reported here reflect the effects of increasing both the number of rules and number of features. I use the term number of features for parsimony in writing. First, I explore whether increasing the complexity of items, by introducing an additional feature, impacts performance. Following this, I explore whether there are significant differences between the Complexity conditions, and if this varies by Condition (Combined and Separated), Fluid Intelligence, or Age.

Each feature was also assigned to a Rule Type: Constant, Design, Figural, and Spatial (see *Chapter 2* for details). I examine whether rule types differentially affect performance, and whether this varies with Condition (Combined and Separated), Fluid Intelligence, and Age.

4.3.2.1 Complexity

Performance was at ceiling on 1-feature items, which also cannot be defined as either "Combined" or "Separated"; therefore, analysis of Complexity was restricted to performance on 2- and 3-feature items from the Combined and Separated conditions.

4.3.2.1.1 Combined

There was a significant main effect of Complexity on performance in the Combined condition ($F(1, 114) = 9.76$, $p < 0.01$, $\eta p^2 = 0.079$). As expected, participants performed significantly better on 2-feature ($M = 0.62$, $SD = 0.25$) compared to 3-feature items ($M = 0.56$, $SD = 0.19$; see Figure 10)

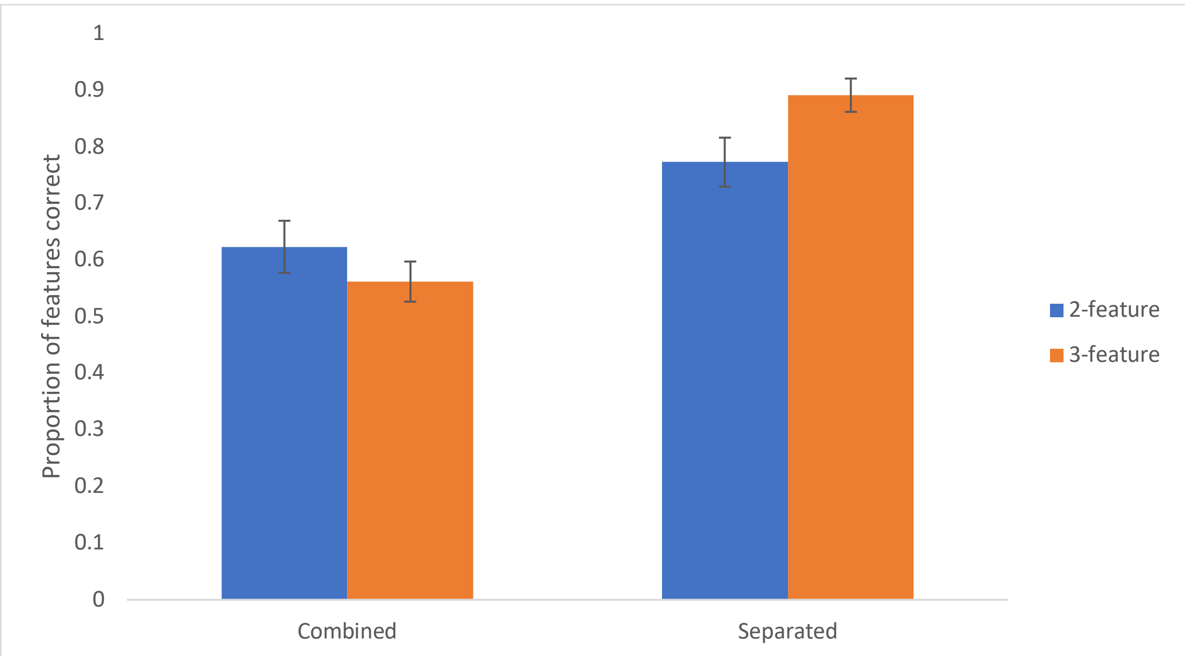


Figure 10: Proportion of features correct on 2- and 3-feature items in the Combined and Separated conditions. Note: error bars are 95% confidence intervals.

4.3.2.1.2 Separated

There was also a significant main effect of Complexity on performance in the Separated condition ($F(1, 114) = 66.31$, $p < 0.001$, $\eta p^2 = 0.368$). However, the results were reversed; performance was significantly better on 3-feature items ($M = 0.89$, $SD = 0.16$) compared to 2-feature items ($M = 0.77$, $SD = 0.23$; see Figure 10). This reversed pattern of results may be due to practice (i.e., participants complete 2-feature Separated items before 3-feature Separated items).

4.3.2.1.3 Complexity by Condition with Fluid Intelligence and Age

Examining the effects of Complexity there was a significant main effect ($F(1, 111) = 5.17$, $p < 0.05$, $\eta p^2 = 0.045$), and a significant Complexity x Condition interaction ($F(1, 111) = 59.18$, $p < 0.001$, $\eta p^2 = 0.348$). As outlined above, performance in the Combined condition was better on 2-feature compared to 3-feature items, whereas the opposite was true of the Separated condition. As expected, the main effect of Condition shows that, on average, the Separated condition is easier than the Combined. Regardless of the reason for better performance on 3 versus 2-feature items when Separated, it is clear that extra features in the Combined condition present particular difficulties.

There was also a significant Complexity x Fluid Intelligence interaction ($F(1, 111) = 7.43$, $p < 0.01$, $\eta p^2 = 0.063$). Figure 11 demonstrates that the effect of Fluid Intelligence is stronger in the 2-feature set.

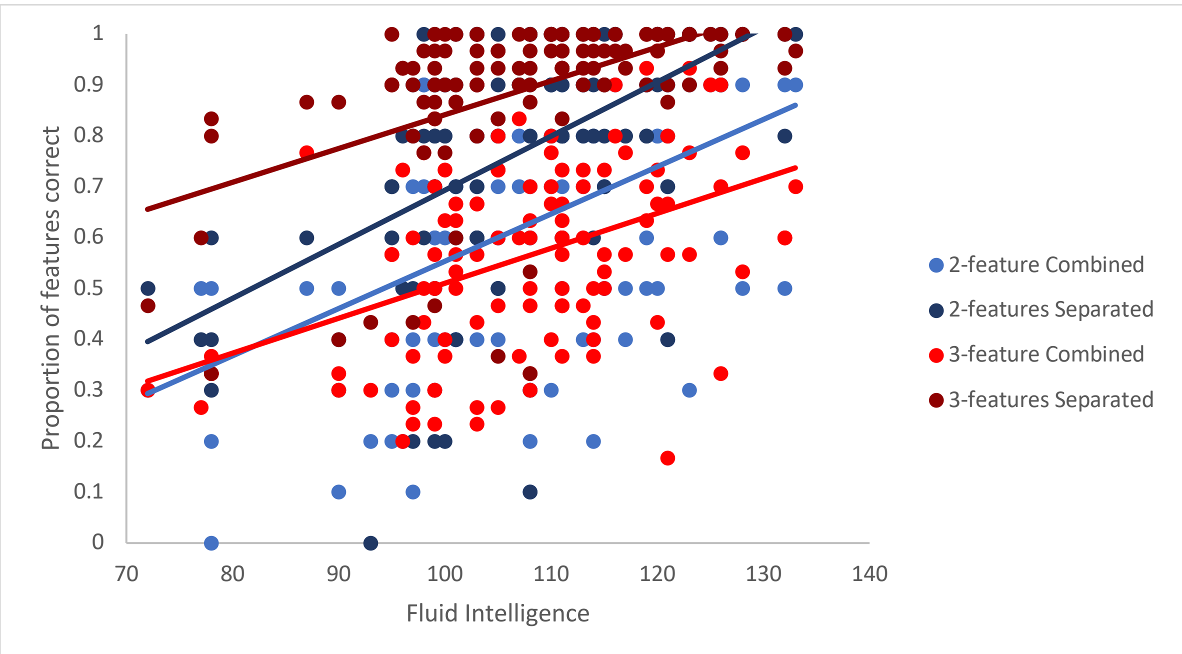


Figure 11: Association between performance in different Complexity conditions and Fluid Intelligence.

There was also a significant Age x Complexity interaction ($F(1, 111) = 5.83$, $p < 0.05$, $\eta p^2 = 0.050$). Figure 12 shows that the effect of age is stronger in the 2-feature set. All higher order interactions were not significant ($F < 2.90$, $p > 0.09$, $\eta p^2 < 0.025$).

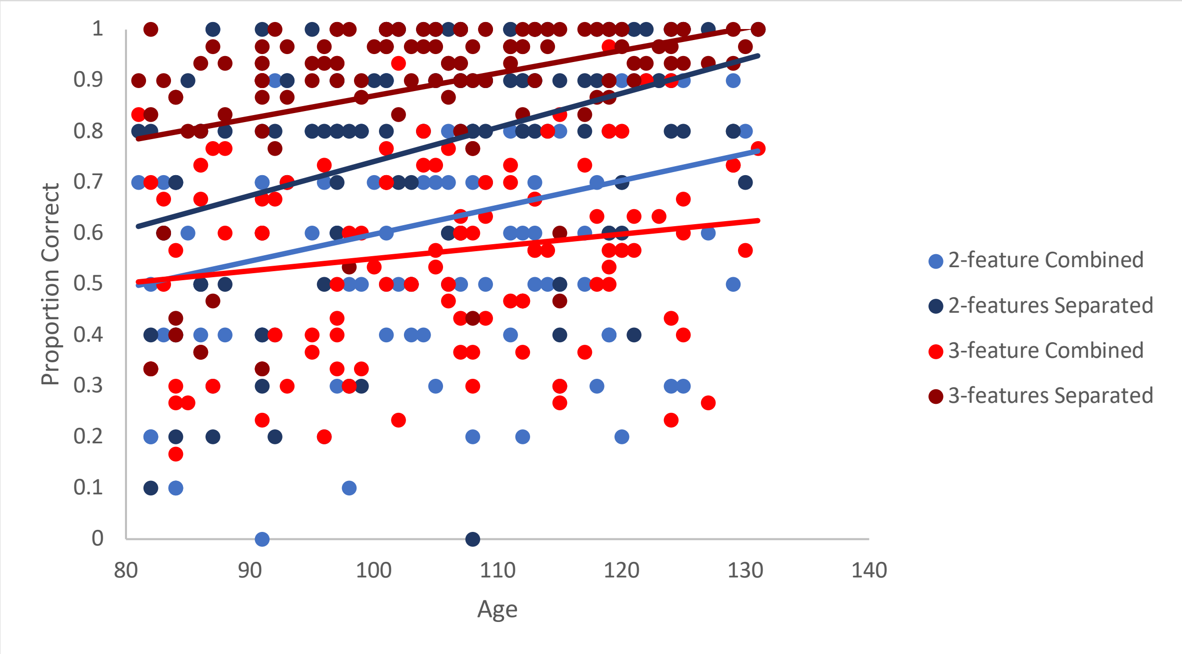


Figure 12: Association between performance in different Complexity conditions and Age.

Overall, the 2-feature items may be better calibrated for measuring individual differences in this age range, while the 3-feature items are better calibrated for measuring the effect of segmentation.

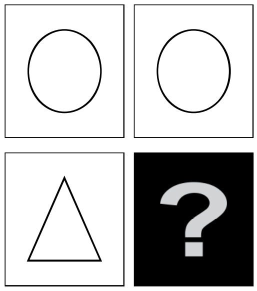
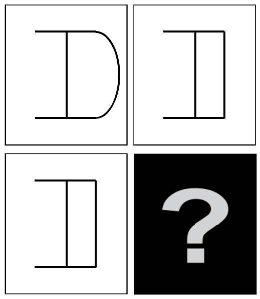
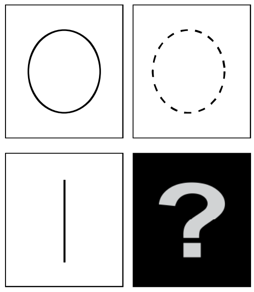
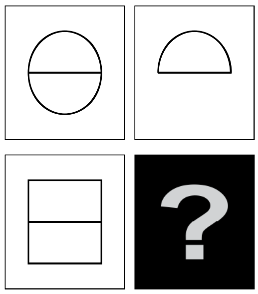
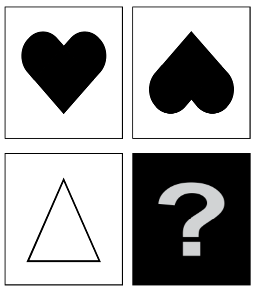
4.3.2.2 Rule Type

A taxonomy of rules (relationships between matrix positions) was created, with four categories: Constant; Design; Figural; Spatial (Table 2, reproduced here as Table 8). For a detailed description of these rules, see *Chapter 2*. Each feature was assigned to one 'Rule Type' category.

Since the confound with practice limits further interpretation of the main effect of Complexity, the Complexity factor is not considered further. Therefore, data from 2- and 3-features were aggregated per condition. This increased the amount of data, thus ensuring that each Rule Type was adequately represented. Rule Type performance data was missing for two participants.

Proportion incorrect was calculated for each rule type by summing the number of features incorrect for each rule type and then dividing by the number of features allocated to the rule type. For example, if participants incorrectly answered six of ten Figural features, from the 2- and 3-feature Combined conditions, their Combined Figural features proportion incorrect would be 0.6.

Table 8: Taxonomy of Rule Type governing the changes in a feature across the matrix

Rule Type	Specific Rule	Rule Description	Example
Constant	Constant in a row	The same feature occurs throughout a row (Carpenter et al., 1990).	
Design	Shape Change	Shape changes across a row	
	Design Change	Design of the shape changes across a row.	
Figural	Quantitative Pairwise Progression	Feature changes in size or quantity across a row e.g. halving or doubling (Bethell-Fox et al., 1984; Carpenter et al., 1990)	
Spatial	Rotation	An element rotates across a row (e.g.45º, 90º, 180 º (Bethell-Fox et al., 1984)	

4.3.2.2.1 Combined: Rule Type performance

A one-way repeated measures ANOVA was used to examine the influence of Rule Type (Constant, Design, Figural, and Spatial) on Combined condition performance. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 15.03$, $p < .05$, therefore the degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .92$; Field, 2013). The results revealed a significant main effect of Rule Type on performance ($F(2.82, 316.34) = 11.76$, $p < 0.001$, $\eta p^2 = 0.095$). Post-hoc analysis with a Bonferroni adjustment showed that performance was significantly better on Constant rules ($M = 0.35$, $SD = 0.22$) compared to Design ($M = 0.49$, $SD = 0.26$) and Figural ($M = 0.45$, $SD = 0.31$) rules. Errors were significantly higher on Design compared to Spatial ($M = 0.40$, $SD = 0.22$) rules. No other pairwise differences were significant (see Figure 13).

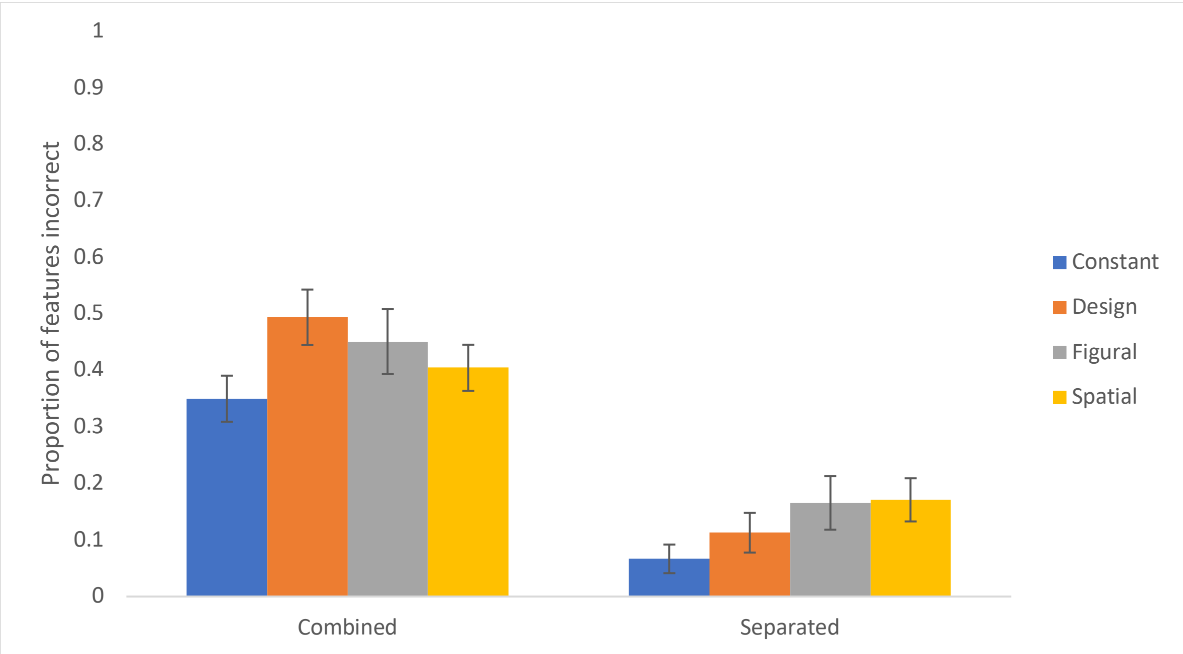


Figure 13: Proportion incorrect for each Rule Type within Combined and Separated conditions.

4.3.2.2.2 Separated: Rule Type performance

The above analyses were repeated for Separated condition performance. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 22.55$, $p < .001$, therefore the degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .87$; Field, 2013). The results revealed a significant main effect of Rule Type ($F(2.69, 301.36) = 16.40$, $p < 0.001$, $\eta p^2 = 0.128$). Post-hoc analysis with a Bonferroni adjustment showed that unlike the Combined condition, there was no significant difference between performance on Constant rules ($M = 0.06$, $SD = 0.14$) and Design rules ($M = 0.11$, $SD = 0.19$). Performance on both of these rule types was significantly better than Figural (M = 0.17, SD = 0.25) and Spatial (M = 0.17, SD = 0.21) rules. No other pairwise differences were significant (see Figure 13).

4.3.2.2.3 Rule Type by Condition with Fluid Intelligence and Age

Examining the effects of Rule Type, there was a significant main effect ($F(1.55, 317.83)$ = 16.12, p < .001, ηp^2 = 0.129), and a significant Rule Type x Condition interaction ($F(2.64, 288.17)$ = 11.21, p < .001, ηp^2 = 0.093). While performance was poorer on all Rule Types in the Combined condition, it seems that the Separated condition was particularly beneficial to performance on Design rules. This suggests that something within the layout of the Separated condition supported children more when solving features with Design Rules.

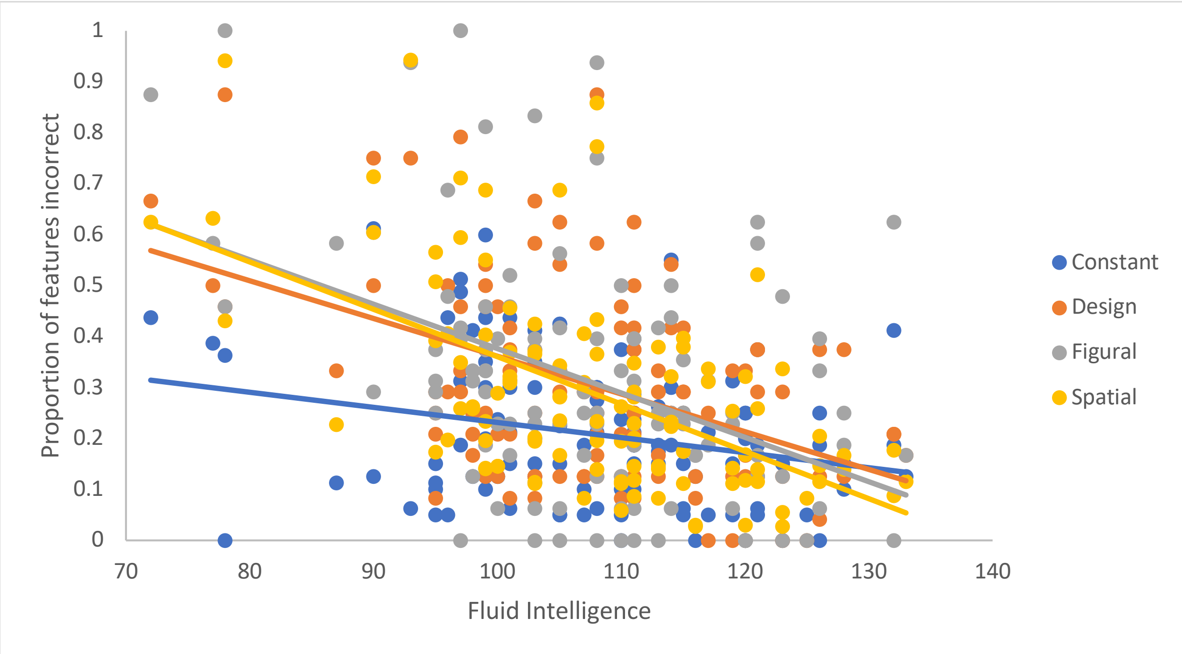


Figure 14: Rule Type by Fluid Intelligence interaction. Note: Data are collapsed across other conditions.

In addition, there was a significant Rule Type x Fluid Intelligence interaction ($F(2.92, 317.83)$ = 6.50, p < .001, ηp^2 = 0.056; see Figure 14), and a significant Rule Type x Age interaction ($F(2.92, 317.83)$ = 2.79, p < .05, ηp^2 = 0.025; see Figure 15). The form of the interactions suggests that the rule types that are harder on average are especially hard for children of younger age or lower Fluid Intelligence. There were no other significant higher order interactions with Rule Type (F < 1.88, p > 0.14, ηp^2 < 0.017).

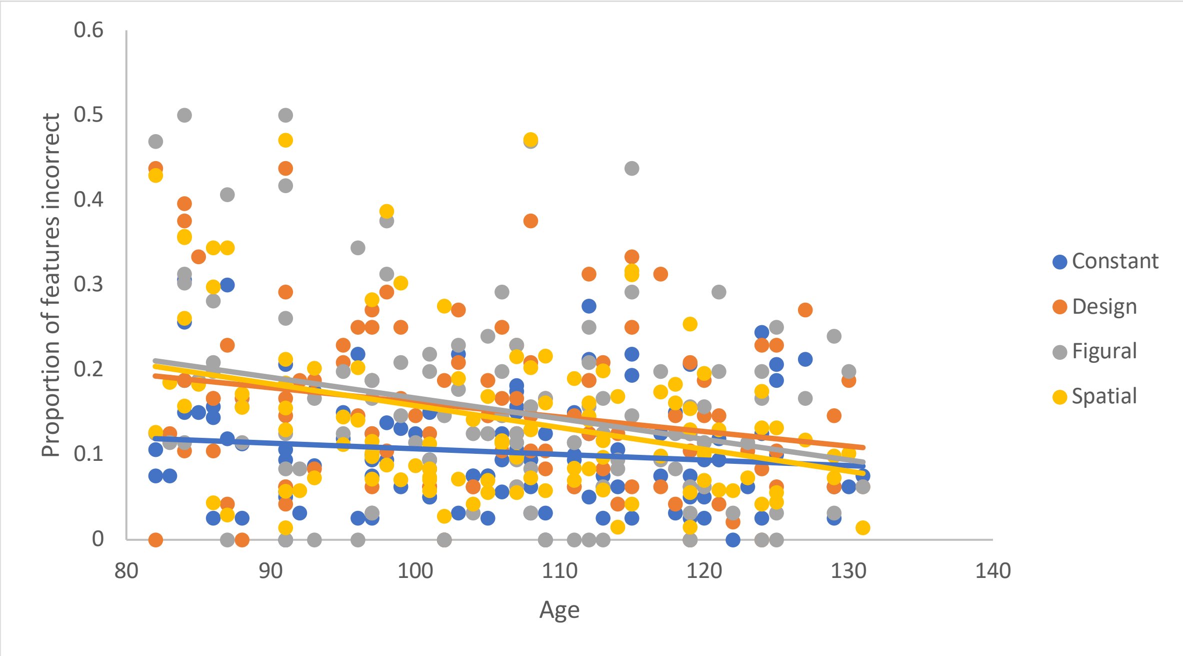


Figure 15: Rule Type by Age interaction. Note: Data are collapsed across other conditions.

4.3.2.2.3 Rule Type by Task Set

The composition of Rule Types differed between the A1/A2 and B1/B2 task sets (see *Chapter 2* for detailed description and Appendices A - D for task sets). Therefore, a one-way repeated measures ANOVA was run with Combined Rule Type, and Task Set (A1/A2 and B1/B2) as a between subjects factor. The results showed no significant Task Set effect ($F(1, 111) = 1.04$, $p = 0.31$, $\eta p^2 = 0.009$) and no Rule Type x Task Set interaction ($F(2.86, 317.08) = 1.18$, $p = 0.32$, $\eta p^2 = 0.010$). These findings were replicated for Separated Rule Type performance (Task Set: ($F(1, 111) = 0.00$, $p = 0.99$, $\eta p^2 = 0.000$; Rule Type x Task Set interaction ($F(2.72, 301.90) = 0.38$, $p = 0.75$, $\eta p^2 = 0.003$). Therefore, Rule Type performance in the Combined and Separated condition did not differ between task sets.

4.3.3 Analysis 3: Additional Dependent Measures

Here, I explore two additional dependent measures: Types of Errors and Response Times. Each incorrect response made by participants was classified into one of five Error Types, as outlined in *Chapter 2*. 'Wrong Alternative' errors occur when the participant draws the alternative feature that is presented in the matrix but is not the correct response feature. Participants who either forget, or fail to draw, a feature commit an 'Omission of Part' error. An 'Other Drawing Error' refers to instances where a participant draws a feature that does not appear in the matrix, or draws a feature incorrectly (e.g., reflected). 'Item Not Attempted' errors occur when participants refuse to solve the item at all. 'Copying the C-Term' errors are instances where participants copy all features of the C-Term matrix item as their response. The purpose of this exploratory analysis was to examine whether there are significant differences between the types of errors children make, and whether the type of errors made vary with Fluid Intelligence, Age, or Condition (Combined and Separated).

Participant's Response Times to Combined and Separated items were recorded. Time to first stroke, which will be referred to as Start Time, is a measure of when children began drawing the first feature of their answer. Drawing Time refers to children's time taken to draw their answer, from first to last stroke. I examine whether there is a significant difference in Start Time and Drawing Time between the Combined and Separated conditions. Response Times are aggregated across 2- and 3-feature items. Finally, I analyse whether Response Times vary with Condition, Fluid Intelligence, or Age.

4.3.3.1 Error Types

Here, I explore whether there is a significant difference in types of errors for Combined and Separated items. Data from 1-feature items were not included due to the low number of errors. Error data from 2- and 3-feature items were aggregated to increase available data.

The proportion of each type of error was calculated as follows: each incorrect answer was categorised as one of five types of errors: Wrong Alternative (see Figure 16); Omission of Part (see Figure 17); Other Drawing Error (see Figure 18); Item Not Attempted; or Copying the C-Term (see Figure 19). The number of each type was then divided by the total number of errors for each participant. Participants who made no errors on the condition being analysed (i.e., Combined or Separated) were excluded from the analysis. Error data were missing for four participants.

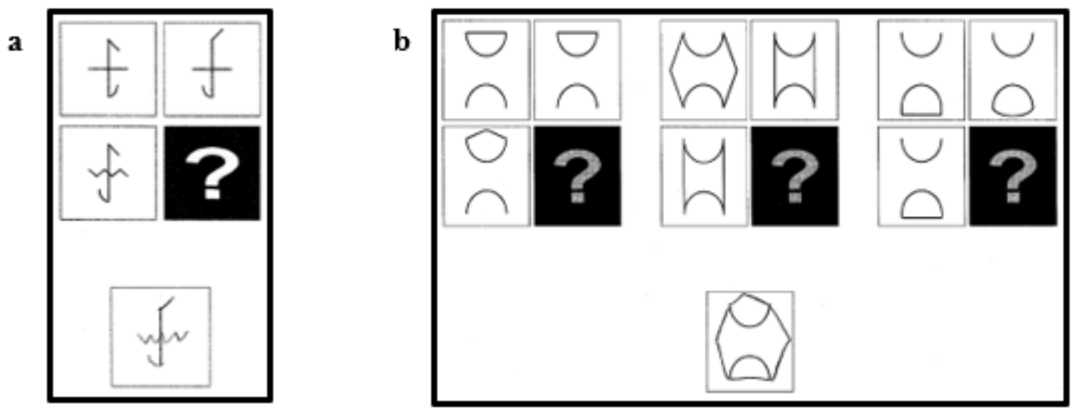


Figure 16: Examples of wrong alternative errors (a) 3-feature Combined (b) 3-feature Separated.

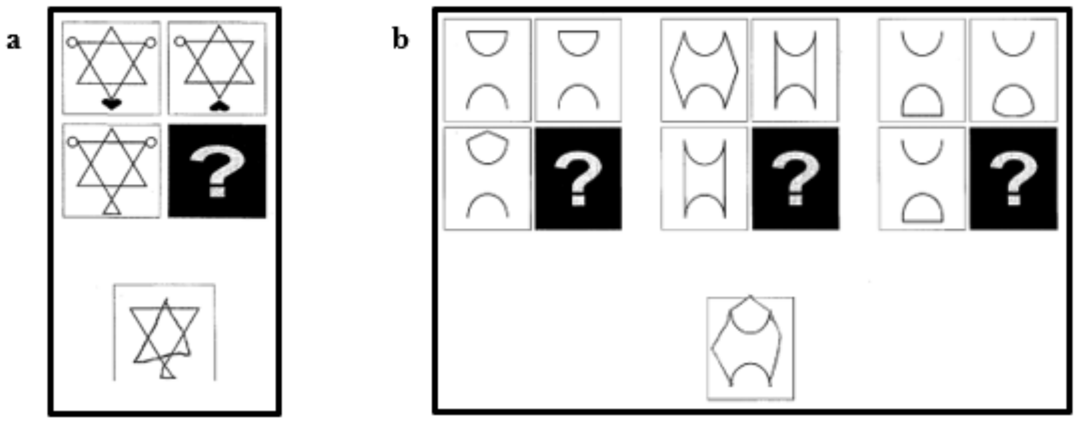


Figure 17: Examples of omission of part errors (a) 3-feature Combined (note: item also includes a wrong alternative error) (b) 3-feature Separated.

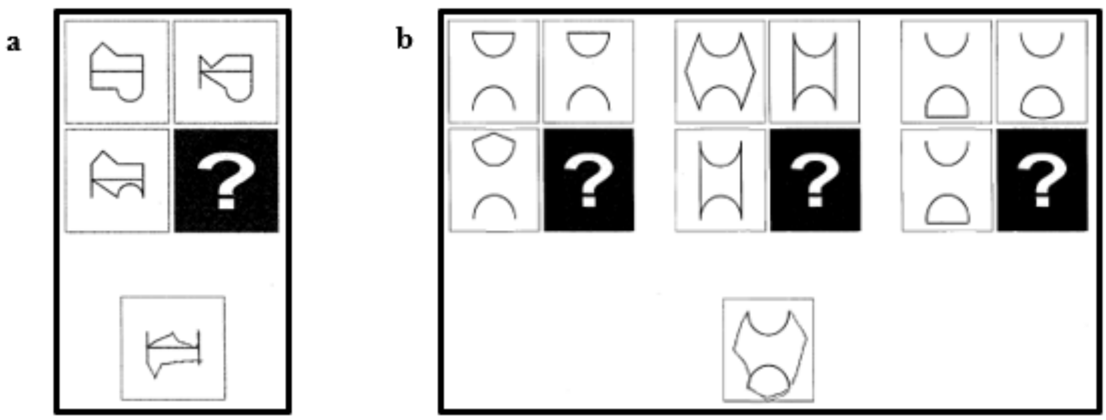


Figure 18: Examples of other drawing errors (a) 3-feature Combined (b) 3-feature Separated (note: item also includes an omission of part error).

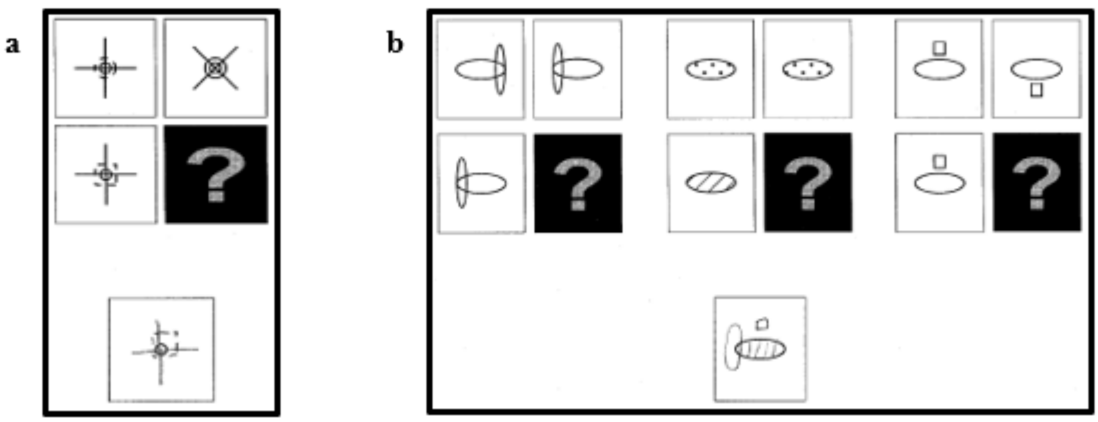


Figure 19: Examples of copying the c-term errors (a) 2-feature Combined (b) 3-feature Separated.

4.3.3.1.1 Combined

111 participants were included in the analysis of Error Types for the Combined condition. A one-way repeated measures ANOVA was used to examine the influence of Error Type on Combined condition performance. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 99.55$, $p < .001$, therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .68$). The results revealed a significant main effect of Error Type ($F(2.72, 298.82) = 236.86$, $p < 0.001$, $\eta p^2 = 0.683$). Post-hoc analysis with a Bonferroni adjustment demonstrated that the proportion of Wrong Alternative errors ($M = 0.66$, $SD = 0.23$) was significantly higher than all other Error types: Omission of Part ($M = 0.07$, $SD = 0.10$); Other Drawing Error ($M = 0.16$, $SD = 0.17$); Item Not Attempted ($M = 0.04$, $SD = 0.12$); and Copying the C-Term ($M = 0.06$, $SD = 0.17$). Other Drawing Errors were significantly higher than Omission of Part, Item Not Attempted and Copying the C-term. No other pairwise differences were significant (see Figure 20).

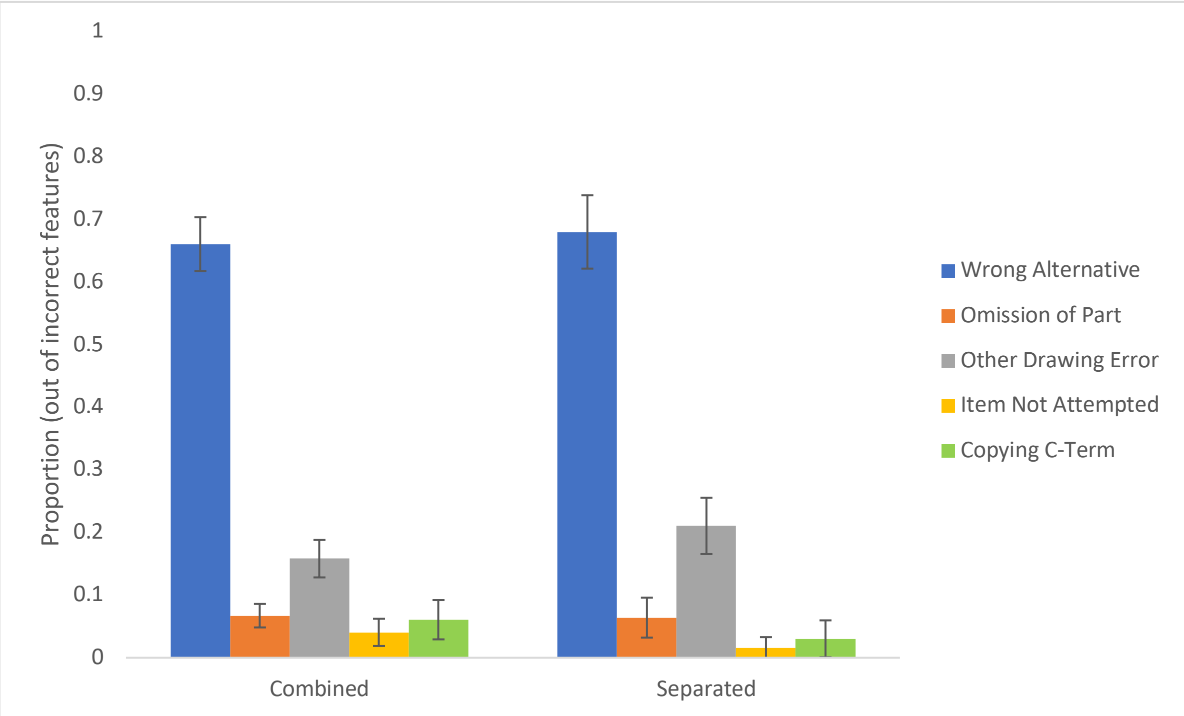


Figure 20: Proportion of each Error Type within the Combined and Separated condition.

4.3.3.1.2 Separated

17 participants correctly answered all Separated features, across 2- and 3-features conditions. Therefore, 98 participants were included in the analysis of Error Types for the Separated condition.

Just as with the Combined condition, Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 156.27$, $p < .001$, therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .55$). As with Combined features, there was a significant main effect of Drawing Error for Separated features ($F(2.22, 215.70) = 138.99$, $p < 0.001$, $\eta p^2 = 0.589$). The same pattern of results emerged for post-hoc analysis (with a Bonferroni adjustment) of Separated Drawing Errors. The proportion of Wrong Alternative errors ($M = 0.68$, $SD = 0.31$) was significantly higher than all other drawing error types: Omission of Part ($M = 0.06$, $SD = 0.17$); Other Drawing Error ($M = 0.21$, $SD = 0.24$); Item Not Attempted ($M = 0.02$, $SD = 0.09$); and Copying the C-Term (0.03, $SD = 0.16$). Again, Other Drawing Error was significantly higher than Omission of Part, Item Not Attempted, and Copy the C-term. No other pairwise differences were significant (see Figure 20).

4.3.3.1.3 Error Type by Condition with Fluid Intelligence and Age

For all analyses in this section, the between-subject main effect is meaningless because error rates per type were scored as a proportion of all errors for each person, meaning they summed to unity for all participants. To explain, only participants with errors were included, and for each participant the proportion of each error type was calculated as a proportion of their total number of errors made. For this reason, the focus of these analyses is on the interaction terms, and not the between-person main effects.

A repeated measures ANCOVA was run to explore if the distribution of error types (Wrong Alternative, Omission of Part, Other Drawing Error, Item Not Attempted, and Copying the C-Term), depended on Condition (Combined and Separated), Fluid Intelligence or Age. There was a significant main effect ($F(2.60, 244.65) = 291.29$, $p < .001$, $\eta p^2 = 0.756$) but there was no Error Type x Condition interaction ($F(2.11, 198.45) = 1.65$, $p = 0.19$, $\eta p^2 = 0.017$). There was a significant Error Type x Fluid Intelligence interaction ($F(2.60, 244.65) = 4.57$, $p < .01$, $\eta p^2 = 0.046$; see Figure 21 below). As outlined in previous sections, Wrong Alternative are the most prevalent Error Type committed across the fluid intelligence range. The form of the interaction between Error Type and Fluid Intelligence suggests that low ability participants are relatively more likely to commit Copying the C-Term errors compared to high ability participants (with a correspondingly lower proportion of Wrong Alternative errors).

There was no Error Type x Age interaction ($F(2.60, 244.65) = 1.77$, $p = 0.16$ $\eta p^2 = 0.018$). All higher order interactions were not significant ($F < 0.81$, $p > 0.054$, $\eta p^2 < 0.030$).

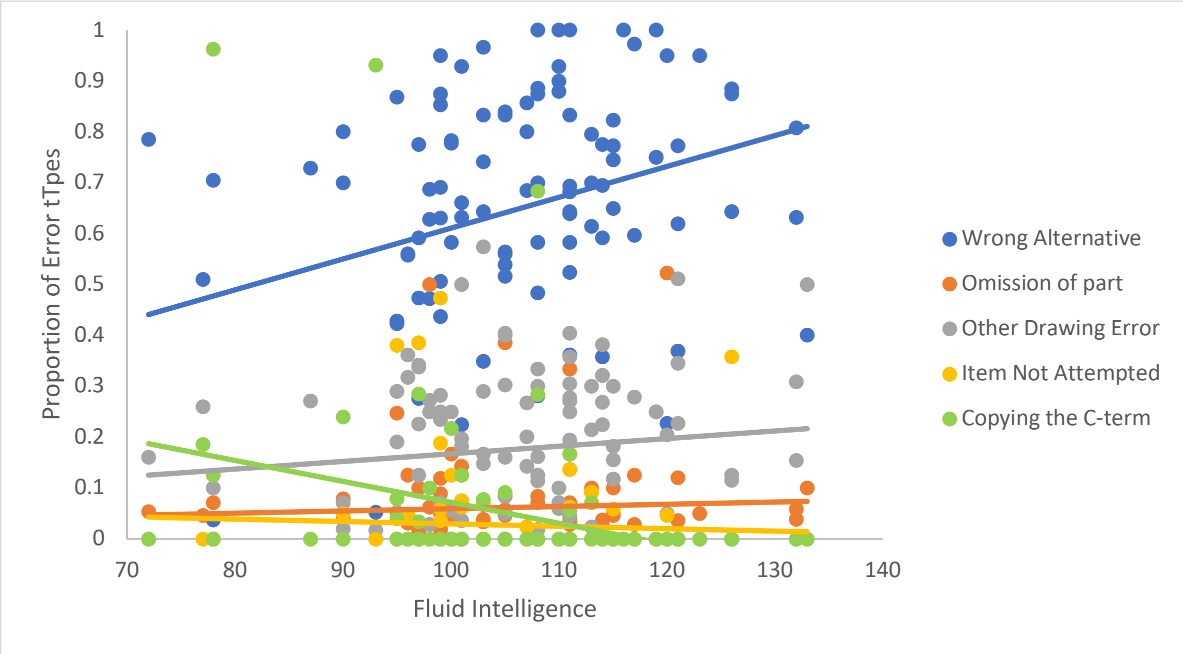


Figure 21: Proportion of Error Types by Fluid Intelligence. Note: Da*ta are collapsed across other conditions.*

4.3.3.2 Response Times

In these analyses, data for each Response Time measure (Start Time and Drawing Time) were aggregated across 2- and 3-features. First, I examine whether there is a significant difference in time to first stroke (Start Time) and time spent drawing (Drawing Time) between the Combined and Separated conditions. I then explore whether there is a Response time by Condition effect, and if this varies with Fluid Intelligence and Age. I could not analyse whether Drawing Times differ across different Rule Types or error types, because timings were measured per item rather than per feature.

4.3.3.2.1 Start Time

Start time data in the Combined condition was available for 95 participants, and for 109 in the Separated Condition. No data points were identified as outliers using the Outlier Labelling Rule (Hoaglin, Iglewicz, & Tukey, 1986; Hoaglin, & Iglewicz, 1987: Upper Range: Q3+2.2*(Q3-Q1), Lower Range: Q3-2.2*(Q3-Q1)).

A Mann Whitney U test revealed that the time from problem presentation to first stroke was significantly longer in the Combined (M = 11.04, SD = 5.08) than the Separated condition (M = 6.93, SD = 2.46), U = -7.86, p <.001 (see Figure 22).

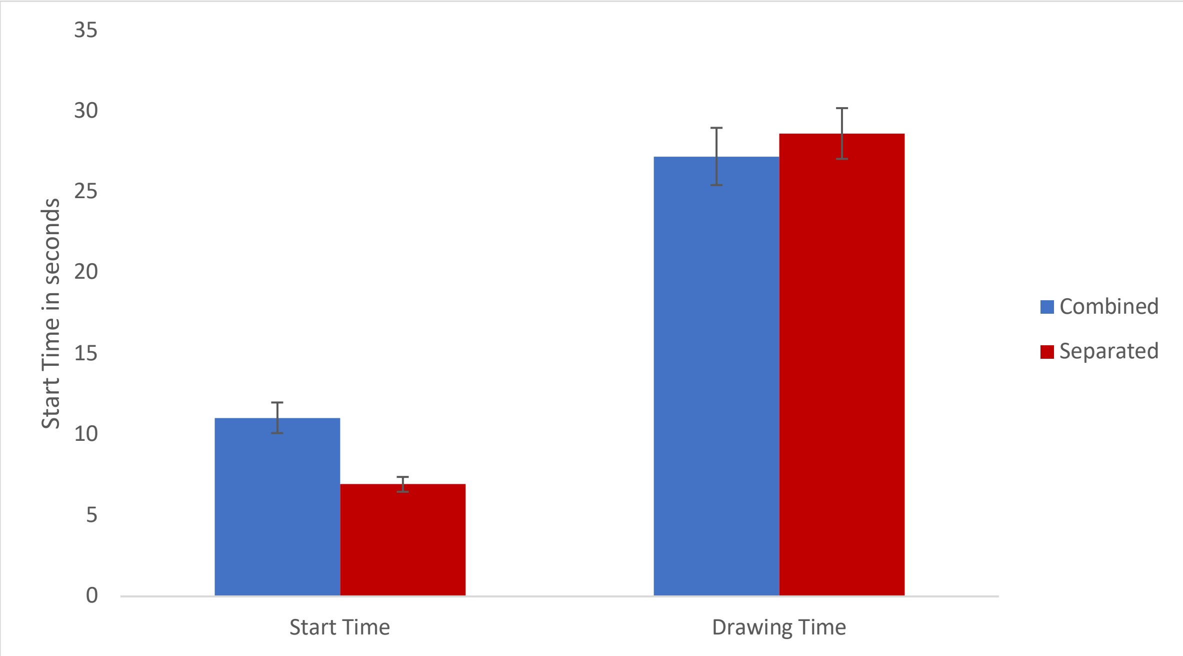


Figure 22: Response Times per condition. Note: Error bars represent 95% confidence intervals.

4.3.3.2.2 Drawing Time

Drawing time data in the Combined condition was available for 111 participants, and for 114 in the Separated Condition. One data point, in the Combined condition, was identified as an outlier, using the Outlier Labelling Rule outlined above (Hoaglin, Iglewicz, & Tukey, 1986; Hoaglin, & Iglewicz, 1987). However, the datapoint was just 1.7 seconds beyond the Upper Limit (60.69s) so the decision was made to retain it in further analyses.

A Mann Whitney U test revealed that there was no significant difference in Drawing Time between the Combined ($M = 27.51$, $SD = 10.02$) and Separated condition ($M = 28.61$, $SD = 8.41$), $U = -1.77$, $p = 0.08$ (see Figure 22).

4.3.3.2.3 Response Times by Condition with Fluid Intelligence and Age

There was a significant main effect of Response Time ($F(1, 86) = 1262.21$ $p < 0.001$, $\eta p^2 = 0.936$), Condition ($F(1, 86) = 10.17$, $p < 0.01$, $\eta p^2 = 0.106$), and a significant Response Time x Condition interaction ($F(1, 86) = 82.73$, $p < 0.001$, $\eta p^2 = 0.490$). The main effect of Condition suggests that the Separated condition has a net response time benefit. This is in addition to the accuracy benefit outlined in Analysis 1. The interaction effect suggests that when problems are broken down into their component parts, the time to first stroke is affected more than the overall time taken to draw the answer.

There was no main effect of Fluid Intelligence ($F(1, 86) = 0.006$, $p = 0.941$, $\eta p^2 = 0.000$), and no Response Time x Fluid Intelligence interaction ($F(1, 86) = 0.428$, $p = 0.514$, $\eta p^2 = 0.005$). There was a main effect of Age ($F(1, 86) = 16.82$, $p < 0.001$, $\eta p^2 = 0.164$) and a significant Response Time x Age interaction ($F(1, 86) = 27.65$, $p < .001$, $\eta p^2 = 0.243$). The interaction plotted in Figure 23 below suggests that while younger children are slower on average, they tend to spend longer drawing but not much longer in thinking before starting to draw. All higher order interactions were not significant ($F < 2.50$, $p > 0.111$, $\eta p^2 < 0.029$).

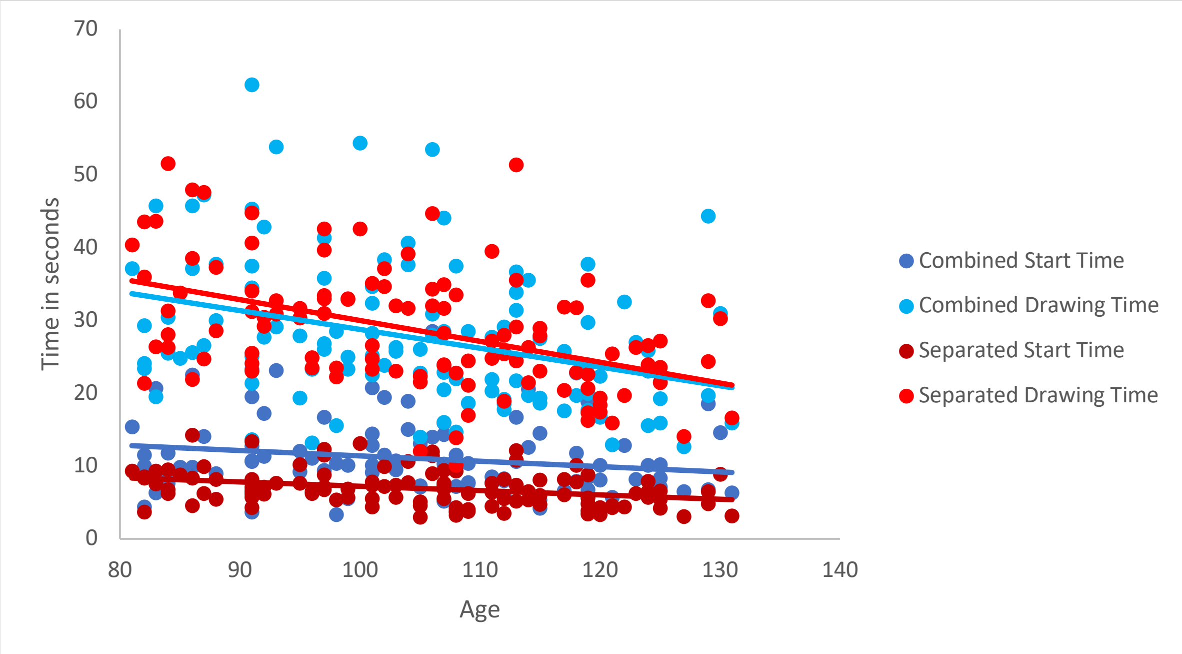


Figure 23: Response Time measures (in seconds) per condition by Age.

4.4 Discussion

The primary aim of this study was to extend Duncan et al.'s (2017) work to test whether segmentation constrains children's performance on nonverbal analogical reasoning tasks, as it does for adults. This study is the first replication of Duncan et al.'s (2017) study, using the child-friendly version of the Cognitive Segmentation task outlined in detail in *Chapter 2*. As with adults, performance on both tasks was related to fluid intelligence, and better in the condition in which the problems were separated into component parts. Unlike adults, children with lower fluid intelligence did not benefit more in the separated condition than children with higher fluid intelligence. Older children performed better than younger children in both conditions, but there was no evidence that breaking problems down was more or less beneficial for younger than older children. A secondary aim was to examine the effects of additional within-task characteristics and explore problem solving behaviours to better understand why children fail to solve some complex problems. These results are discussed in turn below.

4.4.1 Cognitive Segmentation and Fluid Intelligence

Duncan et al. (2017) linked adult problem-solving skills to the principle of compositionality and the attentional control functions of the frontal and parietal cortex (Duncan, 2013; Duncan et al., 2020; Tschentscher et al., 2017a) when they found that performance on a modified matrix reasoning task could be improved by breaking problems down into smaller parts. Our data reveal the same behavioural pattern is true for children. When the working memory and processing speed demands of traditional matrix reasoning problems were minimised by removing time constraints and allowing children to draw the answers, children's performance in the Combined condition was worse than in the Separated condition in which the problem was broken down into its component features across three separate cells. The same pattern of results emerged with 2-feature data, and when using different measures of fluid intelligence. This demonstrates that cognitive segmentation is a critical component of solving analogical reasoning tasks in children, as it is for adults.

While the segmented task was designed to be less complex than the combined task there was no difference in the associations between the two tasks and fluid intelligence. This suggests that the segmented task was still cognitively complex. The demands on working memory and processing speed were reduced in the segmented task by removing the time constraints and allowing children to draw the responses, but it is still possible that it required cognitive effort and that it drew on other cognitive skills not assessed in the current study. For example, the children still had to keep track of their place in the task, which will have created a memory load, they will have needed to inhibit drawing the wrong feature, and they will have needed to draw on their planning skills to complete the task. Future work is needed to expand our understanding of the range of cognitive skills that contribute to fluid intelligence and to quantify the extent to which the segmented and combined tasks used in this study draw on different higher-level cognitive skills. Inhibition will be a cognitive skill of particular interest because it has been shown to play a key role in children's abilities to complete traditional analogical reasoning tasks (e.g., Thibaut et al., 2010; Thibaut & French, 2016).

Performance in both conditions was linked to fluid intelligence, as expected both because nonverbal reasoning is linked to performance on most cognitive tasks (Deary, 2012; Kovacs & Conway, 2016; Jensen, 1998; Spearman, 1927), and because performance on the fluid intelligence tasks requires focussed attention on the right things at the right time (i.e., cognitive segmentation, Duncan et al., 2017). Duncan et al. (2017) found that adults with low IQ benefitted more from the Separated condition than adults with high IQ. Unlike adults, there was no interaction between performance in the two conditions and children's fluid intelligence, suggesting that children with lower fluid intelligence did not benefit more or less from the problems being broken down for them. One possible explanation is that all six to 10-year-old children perform like adults with lower fluid intelligence and struggle to spontaneously segment problems in the Combined condition. In contrast to children and adults with lower fluid intelligence, adults with higher IQ might spontaneously segment, leading to much improved performance in the Combined condition, and accordingly less benefit in the Separated condition.

4.4.2 Cognitive Segmentation and Age

There was a main effect of age: older children performed better than younger children in both the Separated and Combined conditions. Age-related improvements in the Combined condition in our sample of children aged six plus are consistent with widely reported developmental improvements in reasoning (Christie & Gentner, 2014; Ferrer et al., 2009; McArdle et al., 2002; Richland et al., 2006; Starr et al., 2018; Sternberg & Rifkin, 1979; Stevenson et al., 2016) that are more pronounced after the relational shift (Gentner, 1988; Rattermann & Gentner, 1998). These developments are likely related to both increases in children's knowledge about conceptual similarities (e.g., Gentner, 1988; Goswami & Brown, 1990), and developmental increases in other cognitive abilities vital for performance on analogical reasoning tasks (e.g., Morrison et al., 2011; Richland et al., 2006; Thibaut et al., 2010). There was no interaction between age and condition, suggesting that age-related improvements are unlikely to reflect an increase in the ability to break problems down. They may instead be driven by other factors such as an increased ability to understand the task instructions, relational reasoning ability, or age-related changes in attentional focus or motivation.

As children get older, they adopt a more adult-like approach to solving matrix problems, focussing on the A:B relation prior to moving to C (Starr et al., 2018; Thibaut & French, 2011). It is therefore perhaps surprising that guiding children's search strategies by separating the problems was not more useful for younger children. There are several reasons why this might be the case. First, the data presented are cross-sectional. Longitudinal data might be more sensitive to subtle age effects, and it was my intention to collect these data had it not been for the COVID-19 pandemic. Second, the age range included (six to 10 years) might be too narrow to detect cross-sectional differences. Third, previous studies demonstrating that segmentation might help children (e.g., Starr et al., 2018; Thibaut & French, 2011) segmented the A:B::C:D cells *within* a single feature, whereas I separated features that otherwise co-occur per cell. Finally, within the age range studied, the ability to spontaneously segment may not have started to develop or may be masked by concurrent development of analogical reasoning ability. A shift to basing responses on relational similarities rather than perceptual matches occurs around age five (Gentner & Ratterman, 1991; Goswami, 1992). If the younger children are yet to fully understand matrix relationships, they would fail whether the features were segmented or not, and so the potential for a segmentation benefit would increase with age as analogical reasoning develops. This could then counteract a decreasing segmentation benefit with age due to the overlapping development of spontaneous segmentation ability.

4.4.3 Cognitive Segmentation and Order Effects

I speculated that completing the Separated condition first might benefit performance in the Combined condition by providing a useful strategy to decompose the problems, and practice in applying it. However, just as with adults, children's performance in the Combined condition did not benefit more from completing the Separated condition first. As suggested by Duncan et al. (2017), the instructions and practice trials for both conditions may have minimised order effects because they emphasised applying the same procedure to both conditions – focussing on one part after another. In other words, although instructed and encouraged to break a combined problem into parts, both children and low IQ adults fail to do so effectively and are not helped by having seen and briefly practised such decomposition.

4.4.4 Additional Within-task Characteristics

The results supported the hypothesis that increasing the complexity of items negatively affects children's performance on the Cognitive Segmentation task (Bethell-Fox, Lohman, & Snow, 1984; Carpenter, Just, & Shell, 1990; Crone et al., 2009; Freund, Hofer, & Holling, 2008; Mulholland et al., 1980; Primi, 2001). However, this was only true in the Combined condition. Increasing the complexity of the items by adding an additional feature increased the number of errors children made. In the Separated condition, the reverse pattern of results emerged. This is most likely due to practice effects because children completed the 2-feature Separated items before the 3-feature Separated items. If there were similar practice effects in the Combined condition, they were not strong enough to overcome the added difficulty of increasing the number of combined features.

The hypothesis that the types of rules governing the relationships between features in an item will have differential effects on performance was supported. The results were consistent with previous research demonstrating that problems with Constant rules are the easiest to solve (Carpenter et al., 1990). Previous research with adults has consistently shown that spatial rules present the greatest difficulty (Bethell-Fox et al., 1984; Blum et al., 2016; Green & Kluever, 1992; Kirchhoff & Holling, n.d.; Whitely & Schneider, 1981). In my data, different patterns of results emerged in the Combined and Separated conditions. There was no significant difference between performance on Constant and Spatial rules in the Combined condition, whereas spatial rules had the highest proportion of errors in the Separated condition. Moreover, Design rules (i.e., transformations involving a change of shape or design) gave rise to the largest proportion of errors in the Combined condition but there was no significant difference between Design and Constant rules in the Separated condition. A potential reason for these findings is that there is something within the layout of the Separated condition that supports children more with solving features, especially those governed by Design Rules. Furthermore, the interactions of Rule Type with Fluid Intelligence and Age imply that the rules that are harder to decipher on average are especially hard for younger children and those with lower Fluid Intelligence. Thus, there did not appear to be a qualitative shift in the types of rules that the younger or lower IQ children struggled with.

4.4.5 Additional Dependent Measures

The final set of analyses were exploratory in nature. The purpose of examining the types of errors children made was to understand more about problem solving behaviours. Previous studies have shown that when problems are presented to children in a segmented form performance improves (Glady et al., 2017, 2016). Performance also improved in the Separated condition relative to the Combined condition in this study, but errors remained in the Separated condition. This suggests it may be necessary to provide additional help to assist children in solving complex problems, even in their segmented form. Siegler and Svetina (2002) found that 6-year-olds performance on matrix completion tasks improved when they were presented with the answer after each item and asked to explain why the answer was correct. Similarly, Chen et al. (2016) argue that errors occur when children fail to engage the problem-solving strategies that facilitate performance in adults. They found that poor performance in 5- and 6-year-olds was eliminated if adequate performance feedback was given. They argue that discussing the rules underlying the relations following the completion of an item helps facilitate the encoding and integration of relations on subsequent items. It would be interesting in future studies to investigate whether segmentation and feedback have independent or interacting effects on performance, and whether feedback is especially beneficial for certain types of rules or guarding against certain types of errors.

Just as in Duncan et al.'s study (2017), the majority of errors in both the Combined and Separated conditions were Wrong Alternatives i.e., drawing the alternative feature that was presented in the matrix, but was not the correct response feature. This might suggest that the children were easily confused when selecting which features to attend to and select from cells A, B or C, and is consistent with the broader literature on children's problem-solving skills, suggesting children do not systematically orient and organise their searches across the cells in matrix reasoning problems (Glady, French, & Thibaut, 2017; Glady et al., 2016; Starr et al., 2018; Thibaut & French, 2011; Vendetti et al., 2017).

Lower ability children had a higher proportion of copying the C-term errors. This is consistent with being in the earliest and least advanced phase of analogical reasoning development (see *Chapter 1 Section 1.4.1*), in which children commit more duplications of other items in a cell (e.g., Stevenson et al., 2016). This might be linked to maturational limitations that constrain children's problem-solving abilities: such errors might occur due to an inability to activate and inhibit information in working memory necessary for mapping mental representations to the response (Richland et al., 2006; Thibaut & French, 2016).

Response time results replicated those of Duncan et al.'s study (2017). Like adults, children took longer to begin drawing their answer in the Combined condition compared to the Separated condition, which may suggest they spend longer thinking about the problem and mentally decomposing it before they start drawing. There was no significant difference in the overall time taken to draw Combined and Separated items. Examining the effects of Age revealed that younger children spend more time drawing their responses but did not spend more time thinking before starting their responses. It is not possible to disentangle the reasons for this, but speculatively, it could reflect that younger children are physically slower at drawing or that they require longer to engage in problem-solving processes such as encoding and inferring relations.

4.5 CHAPTER SUMMARY

In this chapter I demonstrate that cognitive segmentation is a critical component of complex problem-solving in children, as it is for adults, and that performance can be influenced by different within-task characteristics. By forcing children to focus their attention on separate parts of a complex visual problem, their performance can be dramatically improved. This is akin to scaffolding children's behaviour by dividing complex tasks into simpler steps, and guiding attention to each in turn, which has long been regarded as an effective way to aid children's learning and development (Duncan et al., 2021; Neitzel & Stright, 2003; Wood et al., 1976; Vygotsky, 1978). Interestingly, cognitive segmentation appears to be equally beneficial to all children aged six to 10 years, with no greater or lesser effects for children with lower or higher fluid intelligence or of different ages within this range. These data underscore the importance of breaking complex problems down in the classroom, for children of all abilities, and re-structuring multi-step tasks into separate independent steps to support children's learning and classroom success.

CHAPTER 5 WORKING MEMORY, PROCESSING SPEED, COGNITIVE SEGMENTATION, AND FLUID INTELLIGENCE

A manuscript closely related to this chapter has been submitted for publication to the *Quarterly Journal of Experimental Psychology* as: **O'Brien, S.**, Mitchell, D.J., Dalmaijer, E.S., Duncan, J., & Holmes, J. "Contributions of working memory, processing speed and cognitive segmentation to children's fluid intelligence".

In this study, I explore the relative contributions of cognitive segmentation to fluid reasoning alongside other cognitive processes commonly implicated in problem-solving, namely working memory and processing speed. The analysis was conducted on the sample of 115 children described in Chapter 3 and included in the analysis in Chapter 4. Participants completed two tests of working memory and two of processing speed. I hypothesised that performance would be correlated across all tasks and that each would predict performance on a fluid reasoning task. The main question of interest was whether each cognitive skill would make unique contributions to fluid intelligence, and crucially whether cognitive segmentation contributed anything above and beyond working memory and processing speed. The results revealed that cognitive segmentation does make a significant, unique contribution to problem-solving, above that accounted for by the other cognitive processes. These data add to accumulating evidence suggesting that the ability to break complex problems into their constituent parts is vital for nonverbal reasoning.

5.1 INTRODUCTION

Individual differences in working memory (Colom et al., 2008b; Engle, 2010; Kyllonen & Christal, 1990) and processing speed (Jungeblut et al., 2020; Salthouse, 1996) are strongly related to individual differences in fluid intelligence. As demonstrated in *Chapter 4*, cognitive segmentation, the ability to separate problems into their component parts, is also critical for solving nonverbal reasoning problems. The aim of this study was to examine the relative contributions of working memory, processing speed and cognitive segmentation to children's fluid intelligence. I begin this Chapter by reviewing and summarising the evidence for the roles of working memory and processing speed in fluid reasoning. A fuller account of this literature is provided in *Chapter 1, Section 1.7.*

5.1.1 WORKING MEMORY

Working memory, the ability to simultaneously store and process information, is closely linked to fluid intelligence (e.g. Colom, Flores-Mendoza, Quiroga, & Privado, 2005; Colom et al., 2008; Engle, Laughlin, Tuholski, & Conway, 1999; Frischkorn, Schubert, & Hagemann, 2019; Kane, Hambrick, & Conway, 2005; Kyllonen & Christal, 1990; Oberauer et al., 2005; Oberauer, Süß, Wilhelm, & Wittmann, 2008; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002; Unsworth, Fukuda, Awh, & Vogel, 2014). As described in *Chapter 1 Section 1.7.1*, there have been several mechanistic accounts of this association. One assumes that both working memory and fluid intelligence rely on the ability to control attention in order to ignore distracting information (Engle, 2018; Kane et al., 2007; Shipstead et al., 2016b). An alternative proposal suggests that the working memory processes of building, maintaining, and manipulating arbitrary bindings between items or pieces of information underpins performance on matrix reasoning-type tasks (Oberauer et al., 2008). A third hypothesis suggests that short-term memory provides the link, storing information necessary for the completion of both working memory and fluid reasoning tasks (Colom, Rubio, Shih, & Santacreu, 2006; Colom et al., 2008).

5.1.2 PROCESSING SPEED

Processing speed is also strongly related to fluid intelligence (Coyle, Pillow, Snyder, &, Kochunov, 2011; Detterman, 2002; Fry & Hale, 2000; Jensen, 2006; Salthouse, 1996). There are several theoretical accounts of the relationship between speed of information processing and fluid intelligence, as reviewed in full in *Chapter 1, Section 1.7.2*. Some suggest that faster processing speed allows information to be processed more quickly in working memory, thereby preventing decay, and allowing items on fluid intelligence tests to be processed before the necessary information is lost (e.g. Kail, Robert & Salthouse, 1994; Salthouse, 1996; Van Der Maas et al., 2006). Others argue that working memory is important for maintaining robust representations of the bindings between stimuli and responses in speeded tasks, and that these bindings in working memory are also necessary for building relational representations in reasoning tasks (Kane & Engle, 2003; Wilhelm & Oberauer, 2006). A more recent proposal suggests that greater processing speed increases the efficiency of information transfer from attentional processes to storage, increasing the availability of information in working memory and thereby improving performance on fluid intelligence tests (Dang et al., 2015; Schubert et al., 2017; Schubert & Frischkorn, 2020).

5.1.3 COGNITIVE SEGMENTATION

Duncan et al. (2017) demonstrated that when demands on working memory and processing speed were reduced, the ability to decompose each problem into parts remained critical to success on nonverbal reasoning tasks in adults. In *Chapter* 4, I replicated this finding with children aged six to ten years. Participants performed better in the condition in which the problems were segmented, indicating that when demands on working memory and processing speed are reduced, cognitive segmentation ability constrains performance on nonverbal reasoning tasks for children, just as it does for adults (Duncan et al., 2017).

5.1.4 CHAPTER AIMS

While individual differences in working memory and processing speed are undoubtedly related to individual differences in performance on tests of fluid intelligence, the relationship between these abilities and cognitive segmentation, and their relative contributions to fluid intelligence alongside cognitive segmentation, are not yet known. In this study, multiple measures of working memory and processing speed were administered, and principal component analysis was used to derive composite measures of each construct.

Two working memory tasks, digit recall and dot matrix, were administered. Digit recall is used widely to measure verbal aspects of working memory. Some argue that the forward recall of digits taps into short-term memory, and not the cognitive control aspects of working memory, and that the backward recall of digits better captures the more attentionally or executively demanding aspects of working memory (e.g., Alloway et al., 2006). However, many studies have shown that forward and backward recall tap into similar resources. For example, Rosen and Engle (1997) found that forward and backward recall did not differ in terms of predicting performance on standardised academic tasks, and latent variable studies have revealed that forward and backward recall load onto the same factor during factor analysis (e.g., Colom, Abad, et al., 2005; Engle et al., 1999). Thus, forward digit recall tasks might be attentionally and executively demanding, especially in younger children (< 10 years) who are not yet able to group and rehearse stimuli efficiently (Cowan et al., 2005). We were agnostic as to which aspect of working memory this task measured. The dot matrix task, derived from the Corsi block task (Corsi, 1972; Milner, 1971), assesses visuo-spatial working memory, and has also been shown to draw on executive aspects of working memory, particularly in children (e.g., Alloway et al., 2006b; Cowan et al., 2005).

Processing speed or mental speed reflects the ability to quickly process information. It can be measured in many ways, for example using speed of identification, discrimination, computing speed, coding speed, judgement tasks, or basic reaction times (Coyle et al., 2011; Dang et al., 2015; Jensen, 2006). In the current study, metrics of processing speed were derived from two reaction time-based tasks, a Cancellation and a Go/No-go task, each from the RED App (RED App Ireland (version 1.3.4): Bignardi et al., 2020; Dalmaijer et al., 2021). These tasks were primarily designed to measure higher-level cognitive processes, selective attention, and inhibition respectively, but the indices derived from them for the current experiments measure simple response speed and were shown to reliably measure speed of information processing in the development of the RED App Ireland ((version 1.3.4): Bignardi et al., 2020; Dalmaijer et al., 2021).

The Cognitive Segmentation task used in the current study has been described in detail in *Chapters 2, 3, and 4,* and the Fluid Intelligence task and the age-standardised Leiter-3 measure described in *Chapters 3 and 4*.

The key question of interest was whether each cognitive skill would make unique contributions to fluid intelligence, and in particular whether cognitive segmentation added anything beyond working memory and processing speed.

5.2 METHOD

5.2.1 PARTICIPANTS

Data from the same 115 participants, outlined in detail in *Chapter 4*, were included in this study.

5.2.2 PROCEDURE AND DESIGN

The procedure and design are outlined in *Chapter 4*. The tasks used in the current study are detailed in the following section.

5.2.3 MEASURES

The Leiter-3 was used to measure Fluid Intelligence. Cognitive segmentation ability was measured using accuracy (proportion of features correct) from the 3-feature Combined and Separated conditions. To obtain a measure of each child's relative ability to segment problems when factors contributing to variance in performance on the Separated condition are considered, performance in the Separated condition (proportion correct) was regressed from performance in the Combined condition (proportion correct). Larger residuals, or 'segmentation ability' score reflects a child being more able to independently segment complex problems. Both the Leiter- 3 and Cognitive Segmentation task are described in detail in *Chapters 3 and 4*.

5.2.3.1 Working Memory

Children completed the two tests of working memory from the RED App Ireland (version 1.3.4; Bignardi et al., 2020) described in *Chapter 3*. To summarise here, one was a digit recall task that required participants to recall sequences of numbers in serial order. The other was a dot matrix task that required participants to recall spatial locations in serial order (cells that lit up one at a time in a 4x4 grid on screen). For both digit span and dot matrix, the number of trials correct (whole sequences recalled in serial order) was scored and used in the analysis.

5.2.3.2 Processing Speed

The measures of speed were taken from the Cancellation and Go/ No-go tasks from the RED App (RED App Ireland (version 1.3.4): Bignardi et al., 2020; Dalmaijer et al., 2021). The indices derived are described in detail in *Chapter 3*. To re-cap here, two versions of the Cancellation task were administered: cancellation marked, and cancellation unmarked. In the marked version, tapping a target produced auditory feedback ("meow!") over headphones, and visibly marked it with a red cross. Only the auditory feedback was provided in the unmarked version. For both versions, the inverse median time taken between successful cancellations was used an index of processing speed. In the Go-No-Go task, children were asked to tap each target stimulus (dog) and avoid pressing the distractors (poop). The inverse median reaction time for target stimuli was used as an index of processing speed.

5.2.4 Analysis plan

To control for age, raw scores were regressed on age and the standardized residuals were used for the measures of working memory, processing speed and cognitive segmentation. Higher scores were associated with better performance. Age-scaled scores from the Leiter-3 were used to index fluid intelligence.

First, to derive robust composite indices of working memory and processing speed, the five standardized scores from the tasks measuring these abilities were submitted to a principal components analysis. Next, to explore the associations between these constructs and both cognitive segmentation and fluid intelligence, simple correlations were run. Simultaneous multiple regression analyses were then conducted to test whether working memory, processing speed, and cognitive segmentation made unique contributions to fluid intelligence.

5.3 RESULTS

Descriptive statistics for each measure are presented in Table 9.

Table 9: Descriptive statistics for raw and age-adjusted scores

	Statistic	Mean	St. Dev.	Min	Max
Raw scores					
	*Leiter Total Raw	47.0	7.9	26.0	65.0
	Combined	0.56	0.19	0.20	0.97
	Separated	0.89	0.16	0.33	1.00
	Segmentation ability	0.00	0.17	-0.39	0.47
	GoNoGo	1.51	0.20	1.07	2.15
	Cancellation marked	1.75	0.45	1.00	3.00
	Cancellation unmarked	1.79	0.51	0.92	3.75
	Digit Span	10.71	3.55	0.00	19.0
	Dot Matrix	6.40	4.24	0.00	17.0
Age-adjusted scores					
	Leiter nvIQ	107.5	11.9	72.0	133.0
	r.Segmentation ability	0.00	0.17	-0.39	0.47
	r.GoNoGo	0.00	0.19	-0.40	0.60
	r.Cancellation marked	0.00	0.41	-0.92	1.49
	r.Cancellation unmarked	0.00	0.49	-0.92	1.91
	r.Digit Span	0.00	3.43	-11.53	7.48
	r.Dot Matrix	0.00	3.82	-7.25	8.90

Note: *Leiter Total Raw is a combination of the scaled scores from the four subtests (Figure Ground, Form Completion, Classification Analogies, and Sequential Order). Raw score units of measurement: *Combined* and *Separated* = proportion of features correct (n. correct/30 features). *Segmentation ability* = separated score regressed from combined score. *GoNoGo* = inverse median RT to target stimuli. *Cancellation marked and unmarked* = inverse median time between successful cancellations. *Digit Span* and *Dot Matrix* = number of correctly recalled sequences. Age-adjusted scores: *Leiter nvIQ*: Standardised Leiter score. *r.* = age residualised raw scores.

5.3.1 INDICES OF WORKING MEMORY AND PROCESSING SPEED

A principal component analysis with varimax rotation was applied to the age-adjusted and z-standardised residual scores from the measures of working memory and processing speed (digit span, dot matrix, go-no-go, cancellation marked, and cancellation unmarked) to derive robust indices of performance for each construct. Assessment of Mahalanobis distance, with a threshold value of $p<0.001$ (Tabachnick, Fidell, & Ullmann, 2007), indicated two multivariate outliers, which were removed. The principal component analysis was conducted in *R* using the psych package (Revelle, 2016). A two-factor PCA solution was specified and explained 62% variance. Component loadings are shown in Table 10. The tasks with a speeded component all loaded on component 1, and those with memory and executive demands on component 2. Note that the Go/No-Go task loaded on both components, presumably because it tapped into both processing speed and higher-order cognitive processes. Based on the loadings, component 1 likely corresponds to processing speed, and component 2 to working memory. Factor scores were saved for each child for each component and used in the subsequent analyses.

Table 10: Principal components reflecting processing speed and working memory

	Component 1	**Component 2**
GoNoGo	0.44	0.62
Cancellation Marked	0.86	
Cancellation Unmarked	0.89	
Digit Span		0.76
Dot Matrix		0.62

Note: loadings <.3 are not shown

5.3.2 Associations between Working Memory, Processing Speed, Cognitive Segmentation, and Fluid Intelligence

Correlation analyses were run to explore the associations between the key cognitive constructs. Leiter nvIQ was significantly correlated with working memory, and with cognitive segmentation, but not with processing speed (see Table 11). Cognitive segmentation was not significantly correlated with either working memory or processing speed (working memory and processing speed were uncorrelated as a consequence of their construction).

Table 11: Pearson Correlations with 95% confidence intervals

Variable	1	2	3
1. Leiter nvIQ			
2. Processing Speed	.08		
	[-.11, .26]		
3. Working Memory	.32**	-.00	
	[.15, .48]	[-.18, .18]	
4. Cognitive Segmentation	.23*	-.00	.09
	[.05, .40]	[-.19, .18]	[-.09, .27]

Note: Values in square brackets indicate the 95% confidence interval for each correlation. *indicates $p < .05$. ** indicates $p < .01$.

5.3.3 PREDICTORS OF FLUID INTELLIGENCE

A check for outliers revealed that no data points exceeded a Cook's distance greater than 1. However, a more conservative threshold was applied using $4/(N - k - 1)$, where N is the sample size, and k is the number of independent variables. This indicated four potentially influential points (Hair, Anderson, Tatham, & Black, 1998). To examine the effect of each outlier on regression estimates, a series of regression analyses were run (predicting Leiter nvIQ from working memory, cognitive segmentation, and processing speed), excluding each outlier in turn. The changes in regression estimates were minimal, therefore, the decision was made to retain all 113 data points for further analyses (Note: $N = 113$ because two datapoints were identified as problematic multivariate outliers prior to the construction of components. See *Section 5.3.1* for further details).

A multiple linear regression with all predictors (working memory, processing speed and cognitive segmentation) was significant $F(1, 109) = 6.49$, $p < 0.001$. Working memory and cognitive segmentation, but not processing speed, made significant unique contributions to the prediction of fluid intelligence, accounting for ~13% of the variance (see Table 12).

Table 12: Multiple regression predicting Leiter nvIQ scores from cognitive measures

	B	*SE*	*B*	*p*	R^2
Processing Speed	0.97	1.06	0.08	0.36	
Working Memory	3.67	1.06	0.31	<0.001***	
Cognitive Segmentation	2.41	1.06	0.20	0.03*	
					0.13

Note: B = unstandardised beta; SE = standard error; *B* = standardised beta; *N* = 113; $^{*}p < 0.05$, $^{***}p < 0.001$

A model including both working memory and cognitive segmentation (model B) accounted for significantly more variance than a model with working memory alone, $F(1, 110) = 5.14$, $p < 0.05$, $\Delta R^2 = 0.03$ (model A) (see Table 13), indicating that cognitive segmentation contributes to fluid intelligence after the variance accounted for by working memory is taken into account.

Table 13: Model comparisons for cognitive measures identified as significant predictors of Leiter nvIQ scores in simultaneous linear regressions

Model	Cognitive measure	B	*p*	*F*	Adjusted R^2	Difference
A	Working Memory	3.89	<0.001***	13.04	0.10	
B	Working Memory	3.67	<0.001***			
	Cognitive Segmentation	2.40	0.03*	9.33	0.13	$\Delta R^2 = 0.03$
						F = 5.14*

Note: B = unstandardised beta; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; *F* = F-statistic; Adjusted R^2 = R^2 that is adjusted for the number of predictors in a model; ΔR^2 = difference in adjusted R^2 between the two models.

5.4 DISCUSSION

The relationship between individual differences in both working memory and processing speed and nonverbal reasoning skills is well-established (e.g., Engel de Abreu et al., 2010; Fry & Hale, 2000b; Hornung et al., 2011b; Kail, 2000; Swanson, 2008). The aim of this study was to test their relative contributions to children's fluid intelligence alongside cognitive segmentation - the ability to break complex problems into their constituent parts. Longstanding associations between working memory and fluid intelligence were replicated (Engel de Abreu et al., 2010), but no direct links were found between processing speed and fluid intelligence. Notably, cognitive segmentation was not significantly associated with working memory or processing speed yet predicted unique variance in fluid intelligence. This suggests that cognitive segmentation may be a distinct skill in children, which supports nonverbal reasoning.

As discussed throughout this thesis, the role of cognitive segmentation in fluid reasoning has been previously tested by modifying traditional matrix reasoning problems to make segmentation easy or difficult to achieve and, at the same time, minimising any major role for other factors such as working memory or processing speed (e.g., Duncan et al., 2017). Breaking the problems down and allowing answers for each part to be drawn in turn removed the requirement to store and manipulate intermediate results, reducing demands on working memory, and using untimed problems reduced demands on processing speed. Across these studies, children and adults performed better when segmentation was easy to achieve, indicating that when demands on working memory and processing speed were reduced, cognitive segmentation constrained performance on nonverbal reasoning tasks.

The current study extends these results by directly testing the relative contribution of cognitive segmentation to children's fluid intelligence *alongside* measures of working memory and processing speed. The data reveal that children's abilities to break down complex problems predicted their fluid reasoning scores, even when individual differences in working memory and processing speed were controlled for. There was no evidence that segmentation skills were related to working memory or processing speed, which is likely explained by the deliberate attempt to reduce memory and speed demands from the Cognitive Segmentation task paradigms. Together, these findings indicate that the ability to isolate and attend to separate parts of a complex whole may be distinct from processing speed and working memory, and that it makes an independent contribution to the successful completion of nonverbal reasoning problems.

Working memory was highly predictive of children's fluid intelligence, explaining a substantial portion of variance in performance. This is consistent with previous reports of strong associations between working memory and nonverbal reasoning in children (e.g., Alloway et al., 2004; Bayliss et al., 2003, 2005; Engel de Abreu et al., 2010; Fry & Hale, 2000b; Giofrè et al., 2013; Hornung et al., 2011b; Swanson, 2008). The two measures used to derive a latent working memory score, digit span and dot matrix, are often classified respectively as measures of the verbal and visuo-spatial short-term storage aspects of working memory in children (e.g., Alloway et al., 2008; Bayliss et al., 2003; Gathercole et al., 2004; Gathercole & Pickering, 2000). By this account, the data provide support for theories that explain the association between working memory and fluid intelligence in terms of the role of short-term memory in maintaining information necessary for the completion of both types of tasks (e.g., Colom et al., 2006, 2008; Hornung et al., 2011). However, both tasks might be attentionally and executively demanding in younger children (e.g., Alloway et al., 2006; Cowan et al., 2005). As such, it is not possible to rule out the possibility that the cognitive control aspects of working memory support performance on fluid intelligence tasks by facilitating the inhibition of distracting or irrelevant information (Engle, 2018; Kane et al., 2007; Shipstead et al., 2016), or facilitating arbitrary bindings between items or pieces of information (Oberauer et al., 2008). Replicating the current study with additional measures of working memory that explicitly tap into the executive aspects of working memory (e.g., listening or counting span) would be a valuable avenue for future research to disentangle the contributions of different aspects of working memory to fluid intelligence, particularly in relation to the contributions of cognitive segmentation.

I found no evidence that the measure of processing speed was related to fluid intelligence or working memory. This is surprising given the substantial literature reporting strong associations between processing speed and both working memory and intelligence (Fry & Hale, 2000a; Jensen, 2006; Jungeblut et al., 2020; Salthouse, 1996), and evidence for developmental cascade theories that assume increases in processing speed drive improvements in fluid reasoning both directly and indirectly through improving the efficiency of working memory(Coyle, Pillow, Snyder, & Kochunov, 2011; Fry & Hale, 1996; Kail, 2007; Nettelbeck & Burns, 2010; Tourva & Spanoudis, 2020). There are at least two potential explanations for the absence of these associations in this study. First, neither the working memory nor the fluid reasoning tasks were timed. Studies have shown that introducing any time constraints to fluid intelligence tests increases the relationship between working memory and reasoning (Chuderski, 2013). Thus, the anticipated links between processing speed and other cognitive tasks may have emerged if time constraints were imposed.

Second, processing speed is typically measured by simple or choice reaction time tasks that minimise decision-making. The indices of speed used in the current study were based on reaction times (RTs) derived from lower-level conditions of more complex cognitive tasks (cancellation and go/no-go), and as such may have been influenced by additional cognitive processes. Indeed, Schubert et al. (2020) demonstrated that behaviourally measured RTs encapsulate several processes, including the encoding of information, decision-making, and motor execution, making it difficult to identify which components of the stream of processing accurately account for the relationship between processing speed and fluid intelligence (Frischkorn et al., 2019b). Additional methods for assessing processing speed would thus be useful in future studies to further explore its relative contribution to fluid intelligence alongside cognitive segmentation and working memory. For example, using parameters from ExGaussian distributions of reaction times or time course data from EEG recordings would enable behavioural and neural processing speed data to be decomposed into distinct stages of information processing (e.g., Schubert & Frischkorn, 2020).

5.5 CHAPTER SUMMARY

In summary, the present study demonstrates that individual differences in children's working memory and cognitive segmentation ability make independent contributions to children's fluid intelligence. These findings add to growing evidence that cognitive segmentation is a critical component of complex problem-solving in children (Glady, Yannick, French, & Thibaut, 2017; Glady et al., 2016), and support earlier studies showing that executive functions including working memory support children's performance on fluid intelligence tasks (Richland et al., 2006; Thibaut & French, 2016).

CHAPTER 6 STXBP1-ASSOCIATED MEDICAL, COGNITIVE, AND BEHAVIOURAL PHENOTYPE

A manuscript based on the work presented in this chapter has been published: **O'Brien, S.**, Ng-Cordell, E., Astle, D. E., Scerif, G., & Baker, K. (2019). STXBP1-associated neurodevelopmental disorder: a comparative study of behavioural characteristics. *Journal of neurodevelopmental disorders*, *11*(1), 1-11. DOI: https://doi.org/10.1186/s11689-019-9278-9.

The work presented in this Chapter contains additional analyses not presented in the paper, which focusses on the cognitive profile of the children.
See Appendix I for published manuscript

The work presented so far in this thesis focuses on the cognitive processes supporting performance on fluid intelligence tasks. This work was inspired by a study I conducted in the first year of my PhD, which attempted to characterise the medical, cognitive, behavioural, social, and emotional phenotype of a group of individuals with a genetic mutation causing intellectual disability. I present this study as a stand-alone piece of work in this Chapter but discuss the associations between this work and the other empirical studies presented in this thesis in the General Discussion in Chapter 7.

The aim of this study was to characterise the behavioural phenotype of a group of children and adolescents diagnosed with STXBP1 a rare genetic disorder that has recently been associated with intellectual disability. I compare their profiles to a group of children with intellectual disabilities (ID) caused by other known genetic mutations.

Medical, cognitive, behavioural, emotional, and social data were collected from 14 children and adolescents with STXBP1. The characteristics associated with STXBP1 variants are first summarised. This is followed by a comparison of a subsample of the STXBP1 group (N = 8) to two groups of participants with intellectual disability associated with mutations on other genes. The 'ID Comparison' group was comprised of 33 individuals. Eight individuals *from this group were selected as ID-matched controls, referred to as the 'ID Subsample'. The purpose of selecting an ID-matched subsample was to explore the specificity of STXBP1 associated behavioural characteristics. The comparison analysis was restricted to the following parent questionnaire measures: Medical Questionnaire (Epilepsy details only), VABS-II (Sparrow, Cicchetti, & Balla, 2005), Movement ABC Parent (Henderson, Sugden, & Barnett, 2007), Conners (Conners, 2008), Developmental Behaviour Checklist (Einfeld & Tonge, 2002), and Social Responsiveness Scale (Constantino & Gruber, 2005).*

The results show that the 'STXBP1' group demonstrated impairments across all assessed domains. When compared to the ID Comparison group, the STXBP1 group had more severe global adaptive impairments, fine motor difficulties, and hyperactivity. The 'STXBP1 Subsample', in comparison to the 'ID Subsample' had more severe receptive language and social impairments. A striking unique feature of the 'STXBP1 Subsample' group was preservation of social motivation.

6.1 INTRODUCTION

6.1.1 INTELLECTUAL DISABILITIES

Intellectual disabilities (ID) are defined by the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5; American Psychiatric Association, 2013) as neurodevelopmental disorders that begin in childhood and are characterized by cognitive deficits and limitations in adaptive functioning (e.g., difficulties that impact on developmental and sociocultural standards for the individual's independence). Crucially, and relevant to the current thesis, an individual's cognitive problems are defined as problems in "reasoning, problem solving, planning, abstract thinking…" (APA, 2013, p.33) as confirmed by poor performance on age- and population-standardised intelligence tests (typically standard scores <70, or the extreme low end of the normal IQ distribution), as well as clinical evaluation. It is estimated that between 1% and 3% of the population worldwide have a diagnosis of ID, although it is difficult to ascertain accurate estimates due to the lack of epidemiological information in some developing countries (Durkin, 2002; Fujiura et al., 2005; Maulik et al., 2011; WHO, 2007). ID is caused by both environmental (e.g., pre- and post-natal toxins, nutrition, brain-injury) and genetic factors. ID is highly heritable (Plomin et al., 2012; Reichenberg et al., 2016), but here I focus on non-inherited (*de novo*) mutations as a cause of ID.

6.1.2 GENETIC DIAGNOSES OF INTELLECTUAL DISABILITY

Our genes encode numerous proteins that must be functionally active in the right place at the right time for the systems within our cells to function correctly. It is therefore unsurprising that any changes affecting our genes can result in a cascade of severe disruptions to brain development and cognitive functioning (Vissers et al., 2016). Genetic technologies, such as next generation sequencing, have uncovered many variations in our genetic materials that cause ID, and new ID-associated genes are being identified in rapid succession (Chiurazzi & Pirozzi, 2016; Harripaul et al., 2017; Ilyas et al., 2019). As of September 2021, the Genomics England database has recorded over 2300 genes associated with ID (https://panelapp.genomicsengland.co.uk/panels/285/), and with potential diagnostic yields now exceeding 50%, it is estimated that approximately 350,000 to 750,000 individuals in the

UK have an ID-associated genetic disorder (Gilissen et al., 2014; Vissers et al., 2016; Waite et al., 2014).

De novo mutations, or those that are not inherited, are increasingly associated with neurodevelopmental disorders (Guo et al., 2018; Ku et al., 2013; Turner & Eichler, 2019; Wright et al., 2015). Unlike inherited mutations, *de novo* mutations occur by chance, as the result of a variant or mutation in either a germ cell (egg or sperm) or in the fertilized egg early during embryonic development. It is difficult to predict the functional effect of *de novo* mutations because they do not follow typical genetic inheritance rules, (Cardoso et al., 2019; Veltman & Brunner, 2012). Therefore, for every new candidate ID-gene identified due to *de novo* mutations, there is an imperative to characterise the associated medical, cognitive, and behavioural features to provide patients with evidence-based prognostic advice, targeted family support, and ultimately personalised interventions. A number of *de novo* ID-genes have been identified (e.g., ARID1B, EHMT1, CASK, SHANK1, SHANK3, etc). Here I focus on *STXBP1*.

6.1.3 STXBP1

6.1.3.1 Location, Function, and Prevalence

Syntaxin-Binding Protein 1 (*STXBP1*, formally known as MUNC18-1) is a protein encoded by the *STXBP1* gene, located on chromosome 9q34.11. (Hamdan et al., 2009; Stamberger et al., 2016; Swanson et al., 1998). *STXBP1* is expressed ubiquitously throughout the brain during development (see Figure 24: Accessed on: 10/09/2021; https://hbatlas.org/pages/hbtd). *STXBP1* is part of the SEC1/MUNC18 (SM) family of proteins. Although the exact mechanism of SM proteins is not fully understood (Shanks et al., 2012), it is clear that they are involved in membrane-trafficking, interacting with SNARE (SNARE stands for SNAP receptor: SNAP stands for soluble NSF attachment protein, and NSF stands for N-ethylmaleimide sensitive factor) proteins to facilitate the docking and fusion of synaptic vesicles to presynaptic active zones, thus, enabling the synaptic release of neurotransmitters (Rizo, 2018; Yugi et al., 2018). Synaptic function disturbances contribute to many features of neurological, neurodevelopmental, and psychiatric disorders (Brose et al., 2010; John et al., 2021), highlighting the potential role of genes and proteins encoding their function in cognitive and behavioural difficulties.

Pathogenic *de novo STXBP1* variants were first identified in five patients diagnosed with Ohtahara Syndrome, also known as early infantile epileptic encephalopathy (Saitsu et al., 2008). The epilepsy spectrum associated with *STXBP1* variants has since expanded to include, Dravet Syndrome, Lennox-Gastaut Syndrome, West Syndrome, infantile spasms, and other types of early-onset epilepsy (Allen et al., 2013; Carvill et al., 2014; Deprez et al., 2010; Di Meglio et al., 2015; Stamberger et al., 2016). More recently, *STXBP1* variants have been discovered in individuals with ID-related neurodevelopmental disorders, both with and without a comorbid diagnosis of epilepsy (Barcia et al., 2014; Gburek-Augustat et al., 2016; Hamdan et al., 2009; Hamdan, Gauthier, Dobrzeniecka, Lortie, Mottron, Vanasse, D'anjou, et al., 2011; Rauch et al., 2012; Stamberger et al., 2016; Suri et al., 2017; Yuge et al., 2018).

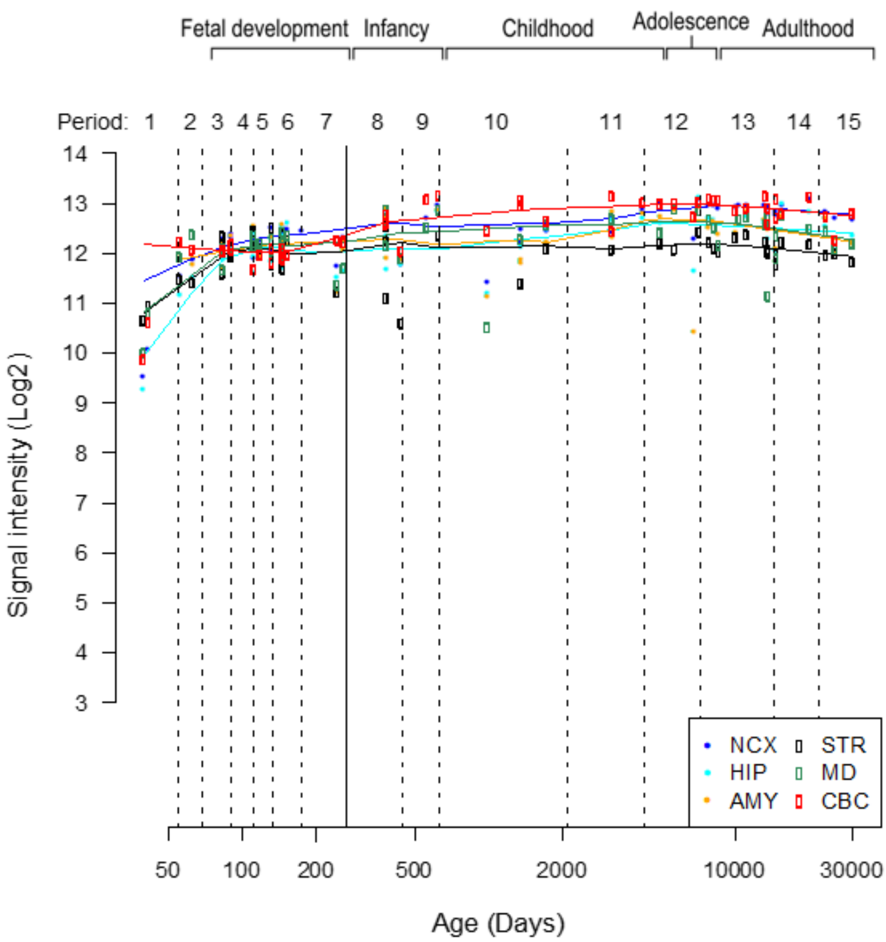


Figure 24: STXBP1 gene expression along entire development and adulthood in the cerebellar cortex (CBC), mediodorsal nucleus of the thalamus (MD), striatum (STR), amygdala (AMY), hippocampus (HIP) and 11 areas of neocortex (NCX). Age (Days): Postconceptional days

The prevalence of pathogenic *STXBP1* variants is estimated at 3% for those with severe childhood epilepsy and between 0.25 – 0.5% for individuals with unspecified developmental disorders (Deciphering Developmental Disorders Study, 2017; Uddin et al., 2017; Wright et al., 2015). Given the increasing prevalence of ID, and due to associated advances in genetic testing, it is thought that the number of individuals diagnosed with an *STXBP1*-associated ID-related neurodevelopmental disorder will rise substantially over the coming years.

6.1.3.2 Existing Clinical Phenotype

Up to 95% of patients with an *STXBP1* mutation have epilepsy, and many have severe to profound ID, but the age of seizure onset does not predict the severity of ID (Stamberger et al., 2016). There has been no systematic phenotyping, but collation of existing clinical information reveals considerable variability in the types and severity of neurodevelopmental problems associated with *STXBP1* variants. Reported problems include movement impairments such as ataxia, tremor, dystonia, and hypotonia, as well as language delay and behavioural symptoms including hyperactivity, stereotypies, and autistic features (Hamdan et al., 2011; Mastrangelo et al., 2013; Milh et al., 2011; Stamberger et al., 2016; Suri et al., 2017; Uddin et al., 2017; Yamamoto et al., 2016).

Basic science and animal models have contributed greatly to our understanding of the neurodevelopmental characteristics associated with *STXBP1* variants. Cultured neurons from human embryonic stem cells demonstrated that homozygous mutations resulted in neurodegeneration, neuronal apoptosis, and early fatality by impeding synaptic transmission (Grone et al., 2016; Patzke et al., 2015; Toonen & Verhage, 2007; Verhage, 2000). Heterozygous mutations have reportedly milder impacts. A study by Kovačević et al. (2018) found that mice with these mutations demonstrated twitches, jerks, and jumps that often coincided with EEG spike wave discharges reminiscent of a human seizure phenotype. They also showed impaired cognitive performance, hyperactivity, and anxiety-like traits, but preserved social behaviours.

6.1.4 AIMS

Despite the increasing number of patients diagnosed with *STXBP1* variants, no study has attempted to systematically characterise the developmental and behavioural aspects of the *STXBP1*-associated phenotype in human patients post genetic diagnosis. The primary aim of the current study was to describe in detail the neurodevelopmental phenotype of children and adolescents with *STXBP1* variants, ascertained as broadly as possible. A secondary aim was to compare the profile of individuals with *STXBP1* variants to an ID-matched group with variants in other ID-associated genes to determine which, if any, characteristics are specific to *STXBP1*. This comparison was conducted to add specificity to the *STXBP1* prognostic information available to families and clinicians, and to set hypotheses for future investigation of molecular, neural, and cognitive mechanisms.

6.2 METHODS

6.2.1 PARTICIPANTS

6.2.1.1 STXBP1

Participants were identified through clinical genetics services, paediatric neurology services, and clinical testing laboratories from across the UK. A recruitment pack was sent to the family via their clinical genetics or neurology service. The pack contained a consent form and stamped addressed envelope which the family returned to the research team. This method of recruitment resulted in a total of 14 individuals with a pathogenic *de novo* variant in *STXBP1,* and their families, taking part (age in years, *M* = 9.93, *SD* = 5.84, range = 0.15 -17; 10 females). Confirmation of the diagnosis was received through the Deciphering Developmental Disorders research study (Wright et al., 2015) for eight participants, and six participants had been diagnosed via hospital-based clinical testing. For each participant, the pathogenicity of variants had been evaluated and confirmed by the regional genetics service according to American College of Medical Genetics and Genomics (ACMG) guidelines and local confirmation procedures. For additional participant information see Table 21 at the end of this chapter.

6.2.1.2 STXBP1 Subsample

Three participants were too young for the questionnaires, and data was excluded from a further three participants with partially completed questionnaire packs. Therefore, eight participants from the 'STXBP1' group, with complete questionnaire data, were included in the 'STXBP1 Subsample' group (age in years, *M* = 13.63, *SD* = 2.33, range = 10 - 17; 5 females).

6.2.1.3 ID Comparison

A comparison sample of 33 individuals (age in years, *M* = 13.91, *SD* = 5.55, range = 5 - 25; 17 females) with variants in other ID-associated genes (see Table 22 at the end of this chapter) was recruited in the same way as the STXBP1 group, mainly through clinical genetic services. Their diagnoses and pathogenicity of variants were received and checked in the same way as the STXBP1 group. There were no significant differences in age or gender between the 'STXBP1' and 'ID Comparison' groups. However, there was a significant difference in global adaptive function, as measured by the Vineland Adaptive Behaviour Scales-II (Sparrow, 2005; see Table 14).

6.2.1.4 ID Subsample

To explore the specificity of behavioural characteristics associated with *STXBP1* mutation. A subset of eight participants from the 'ID Comparison' group, referred to as the 'ID Subsample' (age in years, M = 14.88, SD = 3.27, range = 8 - 18; 5 females) was selected and compared to the 'STXBP1 Subsample'. Participants were first matched on the Vineland Adaptive Behaviour Composite (V-ABC) Scores (VABS-II; Sparrow, 2005; see Table 14). This 'ID subsample' was checked to make sure it contained the same gender distribution and a similar age range. Comparison analysis confirmed that there were no significant differences in age and gender between the 'STXBP1 Subsample' and 'ID Subsample' groups (see Table 14).

Table 14: Demographic information

		STXBP1	ID Comparison		STXBP1 Subsample	ID Subsample	
Age	Mean	9.93	13.91	$U = 158.00$, $p = 0.09$	13.63	14.88	$U = 20.50$, $p = 0.23$
	SD	5.84	5.55		2.33	3.24	
	Range(yrs)	1 – 17	5 – 25		10 – 17	8 – 18	
Gender	Males %(n)	28.60 (4)	48.50 (16)	$\chi(1) = 1.60$, $p = 0.21$	37.50 (3)	37.50 (3)	$\chi(1) = 0.00$, $p = 1.00$
Epilepsy (ever)	% (n)	92.90 (13)	39.40 (13)	$\chi(1) = 11.37$, $p < .001$	87.50 (7)	50.00 (4)	$\chi(1) = 2.62$, $p = 0.12$
Epilepsy (current)	% (n)	92.30 (12)	61.50 (8)	$\chi(2) = 4.80$, $p = 0.09$	85.70 (6)	75.00 (3)	$\chi(2) = 2.36$, $p = 0.31$
Age of onset(n)	Mean	49.67 (12)*	70.73 (11)*	$U = 51.50$, $p = 0.38$	84.14 (7)	12.67 (3)*	$U = 4.00$, $p = 0.18$
	SD	67.06	76.35		70.12	9.87	
	Range	1 – 156	1 – 204		1 – 155	6 – 24	
V-ABC	Mean	45.86	58.30	$U = 130.00$, $p < 0.05$	41.00	48.00	$U = 14.00$, $p = 0.07$
	SD	14.99	15.66		12.13	9.59	
	Range	26 – 69	20 – 96		31 – 67	36 – 68	

Note: *Age of epilepsy onset data was missing for some participants. **SES calculated using "English Indices of Deprivation" (2015). V-ABC: Vineland Adaptive Behaviour Composite Score.

6.2.2 PROCEDURE

Participants were assessed in a quiet space in their homes for between four and six hours. Prior to the home visit, the parent or carer were telephoned to obtain general information about the participant's cognitive, verbal, and motor abilities. The home visit protocol included two standardised measures of intelligence and a test of manual dexterity, balance, and movement. Parents completed interviews assessing their child's medical history and adaptive behaviours, as well as a set of questionnaires measuring different aspects of their child's development and behaviour. Additional cognitive data was collected via an iPad app, the FarmApp (Brkić et al., 2022), over the course of two weeks following the initial home visit. During the home visit, participants practised these tasks with the researcher. Participants were encouraged to use the FarmApp for at least 30 minutes, twice a day, over the following two-week period. Their parents returned the iPads to the research team using a stamped-addressed box provided, at the end of the two-week period. The task order was not fixed: flexibility was required to engage the participants, and make them feel comfortable, to enable as much data as possible to be collected.

6.2.3 MATERIALS

6.2.3.1 Child Assessments

6.2.3.1.1 Intelligence Tests

Wechsler Abbreviated Scale of Intelligence (WASI)

The Wechsler Abbreviated Scale of Intelligence Second Edition (WASI-II; Wechsler, 2011) is comprised of four subscales: Block Design, Vocabulary, Matrix Reasoning, and Similarities. Raw scores were converted to age-normed T-scores for each subscale, and these were then combined and converted to form the age-standardised Full-Scale IQ-4 score.

In the Block Design subtest, children viewed a geometric design presented in a stimulus book, which they recreated using one or two-colour blocks within a specified time limit. The blocks had to be arranged so that the pattern on the top of the blocks (facing up) matched the pattern in the stimulus book. There were 13 items in total, with different start points for children of different ages (older children started with more complex items and were credited with the highest scores available for earlier items). Participants were awarded 2 points if they answered items 1-4 correctly on their first attempt, and 1 point if they were correct on their second attempt. For items 5 – 13, participants received between four and seven points based upon their accuracy and speed. The task was discontinued following two consecutive incorrect responses.

In the Vocabulary subtest children were asked to verbally name the word for a picture in the stimulus book, and verbally define words or concepts that were read aloud (and presented visually in a stimulus book for children aged 9 and over). There were 31 items. All participants began on item 4 (word item) but went back to the preceding item if they could not complete this item or item 5, to ensure three correct responses in a row before continuing. A score of zero, one or two was awarded for each item, determined by the quality of the child's description of each word as measured against the test guidelines. The task was discontinued following three consecutive incorrect responses.

The Matrix Reasoning subtest consisted of 30 items. Participants were shown an incomplete matrix and asked to select one item from five response options to complete the matrix. There were different start points for children of different ages, with older children starting at higher items that were more complex (they were credited for earlier items). Each correct item was given a score of one. The task was discontinued following three consecutive incorrect responses.

There were 24 items in the Similarities subtest. Participants were asked to describe the relationship between two concepts or objects that were read aloud to them. Younger children (aged 6- to 8 -years) started the task with three picture-based items before moving on to the orally presented items. Older children complete the oral items but were credited for the picture items. Scores of zero, one, or two were awarded for each item based on the participant's ability to describe the relationship between the stimuli (more points for better descriptions, as per the test guidelines). The task was discontinued following three consecutive incorrect responses.

Leiter International Performance Scale-Revised (Leiter-R).

Four subtests of the Leiter- R Visualisation and Reasoning battery (Roid & Koch, 1997) were administered to provide an age-standardised measure of fluid intelligence (Leiter Brief IQ). Three of the four subscales have been outlined in detail in *Chapter 3* (Figure Ground, Form Completion and Sequential Order). The fourth subscale, Repeated Patterns, required participants to identify the missing piece of a pattern displayed in the stimulus book. Participants provided their responses by selecting foam pieces that completed the pattern. As items became more complex, participants selected the correct picture card that completed the pattern. Participants indicated their answer by placing the foam pieces or picture cards into slots in a plastic tray displayed below the stimulus book. The total raw score for each subscale was converted to an age-scaled score (M = 10, SD = 3). These four subtest-scaled scores were then combined and converted to a total age-standardised fluid intelligence score (Brief IQ) score (M = 100, SD = 15).

6.2.3.1.2 Movement Assessment

The Movement Assessment Battery for Children (M-ABC-2; Henderson, Sugden, & Barnett, 2007) assesses physical movement in individuals between 3 and 16 years of age. It is comprised of three components: Manual Dexterity (MD), Aiming and Catching (AC), and Balance (BAL). There were three measures in the MD component, which assessed speed and accuracy on fine motor tasks (e.g., threading beads and posting coins through a small toy letterbox). The AC component assessed participants' abilities to catch and throw an object with one or two hands, depending on their age. The BAL component examined participants' static (e.g., standing on one leg) and dynamic balance (e.g., hopping).

The test guidelines suggest participants should complete a total of eight tasks appropriate to their age band (3 – 6 years; 7 – 10 years; 11 – 16 years). However, due to movement difficulties, the majority of the STXBP1 sample could not complete the age-appropriate tasks, so a range of stage, rather than age, appropriate tasks were selected. These data were not analysed but will be described in Section 6.2.3.2.3. Details about the tasks selected for each participant are provided in Table 24 at the end of the chapter. Performance on the M-ABC-2 was video-taped to independently code movement subtleties not assessed as part of the M-ABC-2.

6.2.3.1.3 Cognitive Assessments

Age-standardised cognitive assessments likely underestimate the capabilities of individuals with ID. Standardisation samples do not typically include individuals with ID, which limits the sensitivity of measures to detect differences in performance at the lower end. Performance is often recorded as 'at floor' or 'incomplete' for individuals with ID due to difficulties participants with ID have in understanding task goals, attending to task instructions, or processing complex task stimuli. In an attempt to overcome these testing constraints, a relatively newly designed, intuitive, flexible, iPad-based cognitive assessment, FarmApp, was used in this study to measure participants' speed of response and inhibition (a go/no-go type tasks called Sheep Game), visuo-spatial working memory (a Corsi-block-like task called

Chicken Game) and long-term recognition (Memory Game). FarmApp has a game-like structure, minimal verbal instructions, and developmental age- and stage-appropriate materials designed to increase the accessibility of cognitive assessments for individuals with ID and thus maximise opportunities for the accurate measurement of their abilities (Brkić et al., 2022).

The FarmApp platform has three operational modes: Research Assistant (RA), Adult, and Child. RA mode allows tasks to be introduced to participants at a guided pace. It also allows for the remote monitoring of participant engagement and performance. This mode was designed to resemble existing experimenter driven assessments, albeit with more flexibility. Participants were first introduced to an initial 'Training' phase, in RA mode, in which I explained the rules and requirements of the game. This phase was repeated until I was satisfied that the participant understood the game. During the 'Warm-up' phase, the participant attempted several trials with reduced supervision and prompting. The task then progressed to 'Baseline' in which the participant completed their first full-length block of each game. Subsequent phases are referred to as 'Adaptive' because the task difficulty increased or decreased according to each participant's performance.

Once the participant completed the games in RA mode, FarmApp was switched to Child mode. This allowed participants to play the games freely and at their own pace. In Child mode, there are no 'Training' and 'Warm-up' phases, only 'Baseline' and 'Adaptive' phases. Each time the participant opened one of the games, a new "run" for that game was initiated. A run was deemed 'complete' when the participant had progressed through three blocks of a particular phase of the game. Parents and guardians were asked to ensure that the children played the app twice a day for approximately 30 minutes each.

Parents were instructed to enter Adult mode (see Figure 25 (Right)) if they felt the child did not understand the games. Adult mode was accessed by pressing invisible buttons (marked with blue circles in Figure 25 (Left)), for seven seconds, located in the in the top left and right of the screen. Similar to RA mode, parents could select a game to practice with their child. Adult mode also provided additional task instructions and information on the number of runs that each child had played. This was included so that if there was a problem with the iPad or FarmApp, the researcher could monitor the amount of data recorded.

Figure 25: (Left) Hidden buttons to access Adult Mode are marked in blue and outlined in red on the home screen. (Right) Adult Mode Home screen.

Go/No-Go

The Sheep Game, a go/no-go task, required participants to ‘round up’ sheep (go) by touching the stimuli when they appeared in the left-hand corner of the screen. If touched successfully, the sheep appeared in a pen in the top right-hand corner of the screen and participants heard audio feedback of a “baa” sound. Participants had to inhibit their responses and not touch the pigs (no-go) that appeared in the same position as the sheep. A successful inhibition of a response was rewarded with an “oink” sound. During ‘Training’, ‘Warm-up’, and ‘Baseline’ phases, the stimuli were presented for a set duration (“stimulus duration”) of 3000ms. During the ‘Adaptive’ phases, stimulus duration was adjusted according to participants’ accuracy on no-go trials: following a correctly inhibited response, stimulus duration decreased by 20%; after an incorrectly uninhibited response, stimulus duration increased by 20%. See Table 15 for an overview of the structure of the task. To keep the games engaging, minimum, and maximum stimulus durations were capped at 1000ms and 8000ms, respectively. Multiple indices could be derived from this task. However, due to the low performance of the STXBP1 group and their inability to complete the adaptive blocks, only RTs for Go trials are reported in the Results section.

Table 15: Sheep game information

Phase Type	Type of block	No. of blocks	Trials per block
Training	Mixed	1	5: 4 (Go), 1 (NoGo)
Warmup	Mixed	1	5: 4 (Go), 1 (NoGo)
Baseline	Go only and Mixed	2	Block 1: Go only: 20
			Block 2: Mixed: 15 (Go), 5 (NoGo)
Adaptive	Mixed	3	20: 15 (Go), 5 (NoGo)

Visuo-Spatial Working Memory

The Chicken Game is a Corsi block task (see *Chapter 3* for a fuller account of Corsi-block type tasks). In this task, a series of chickens appeared out of eight fixed hutches (locations) on screen. The participants were then required to tap on the hutches in the same sequence in which the chickens had appeared. When the correct chicken hut was touched the chicken reappeared and participants heard a ‘cluck’ sound. If the sequence was correctly recalled, the last chicken layed an egg (reward). The rewards accumulated over multiple runs, appearing in a haystack. During the Adaptive phase, the number of chickens to be recalled adjusted based on participant’s accuracy on the previous block. A threshold was set for all participants such that if 66% of trials were correct, they would move to the next block (i.e., from three to four chickens). See Table 16 for an overview of the task structure. Again, multiple indices could be derived from this task but due to low performance only the average span reached is reported.

Table 16: Chicken game information

Phase Type	Type of block	No. of blocks	Trials per block
Training	2 Chickens	1	3
Warmup	2 Chickens	1	6
Baseline	3 Chickens	1	6
Adaptive	3-8 Chickens	3	6

Recognition

The Memory Game is a measure of long-term recognition. The stimuli (chicken and sheep) in the Sheep and Chicken games are individualised with different hats, hair, clothes etc (see Figure 26) and each participant was randomly assigned a unique set of sheep and chicken stimuli. The

Memory game was 'unlocked' or made available to participants following three successful runs of both the Sheep and Chicken Games.

Figure 26: Example FarmApp stimuli: chickens (top row) and sheep (bottom row).

The Memory game is a forced-choice paradigm in which participants had to select the stimulus they had seen before from a pair of stimuli. The 'Training' phase required participants to select a farm item that they would have seen in the Sheep and Chicken games. The known farm item was presented with a novel farm item. In the 'Baseline' phase participants were presented with pairs of chickens or sheep and had to tap the stimulus in each pair that they had seen before in the Chicken or Sheep games (the stimuli from the Chicken and Sheep games were paired with novel distractor stimuli). Participants completed 10 trials (pairs) per block, and there were three blocks. See Table 17 for an overview of the task. Audio feedback was played to reward any response i.e., it was not based on accuracy. Performance was measured by accuracy (total number of correct responses). This task did not have Warmup or Adaptive phases.

Table 17: Memory Game information

Phase Type	Type of block	No. of blocks	Trials per block
Training	Farm items	1	5
Baseline	Chickens and Sheep	3	10

6.2.3.2 Parent Questionnaires

6.2.3.2.1 Medical History

A structured medical history interview was conducted with each participant's parents. This asked about the child's general health (e.g., mother's prenatal and postnatal health, pregnancy complications, obstetric and neonatal history, past, current, and family history of medical and psychiatric problems (including medication history), and the child's health. It also asked about neurological problems and developmental milestones (A copy of this questionnaire can be found in Appendix J).

6.2.3.2.2 Behaviour

Vineland Adaptive Behaviour Scales-II Survey Form

The Vineland Adaptive Behaviour Scales-II Survey Form (VABS-II; Sparrow, Cicchetti, & Balla, 2005) is a semi-structured interview, validated for individuals from birth to 90 years of age, completed with both parents/primary caregivers. It asks about a child's adaptive behaviour across four domains: communication (receptive, expressive, and written subdomains), daily living skills (personal, domestic, and community subdomains), socialization (interpersonal relationships, play and leisure time, and coping skills), and motor skills (fine and gross motor skills for children up to 6 years of age). It also provides an optional maladaptive behaviour index score (internalising and externalising behaviours). Age equivalent and standard scores were obtained for each domain (M = 15, SD = 3). These scores were then combined and converted using a table of norms to create a total Vineland Adaptive Behaviour Composite (V-ABC) standard score (M = 100, SD = 15), which provides a measure of global impairment. The maladaptive behaviour index scores were not analysed as part of this study.

Developmental Behavioural Checklist (DBC)

The Developmental Behaviour Checklist Primary Carer Version (DBC-P: Einfeld & Tonge, 2002) is a 96-item questionnaire designed to assess children's behavioural and emotional difficulties. The DBC-P has five sub-scales: Disruptive/Antisocial, Self-absorbed, Communication Disturbance, Anxiety, and Social Relating. Parents / caregivers rate each statement on a 3-point scale (0 = 'not true as far as you know', 1 = 'somewhat or sometimes true' and 2 = 'very true or often true'). The Total Behaviour Problem Score (TBPS) is the sum of all item scores (range = 0 to 192). A score greater than 45 indicates clinically significant behavioural and emotional difficulties.

Conners 3rd Edition–Parent Short form

The Conners-3 Parent Short Form (Conners, 2008) is a 43-item questionnaire used to assess symptoms linked to Attention Deficit/Hyperactivity Disorder (ADHD). The questionnaire is composed of six subscales: Inattention, Hyperactivity/Impulsivity, Learning Problems, Executive Functioning, Defiance/Aggression, and Peer Relations. Parents or carers were asked to rate statements based on their child's behaviour in the past month on a 4-point scale: 0 = Not true at all (Never, Seldom), 1 = Just a little true (Occasionally), 2 = Pretty much true (Often, quite a bit), 3 = Very much true (Very often, very frequently). Total raw scores for each subscale were converted to a T-score using standardised norms from the Conners-3 manual.

Social Responsiveness Scale (SRS-2)

The Social Responsiveness Scale-2 (SRS-2: Constantino & Gruber 2005) is a 65-item parent/carer rated questionnaire that assesses social deficits in the child related to autism spectrum disorder (ASD). There are five subscales: Social Awareness, Social Cognition, Social Communication, Social Motivation, and Restricted Interests and Repetitive Behaviour. Respondents were asked to rate statements within in each subscale on a 4-point Likert-type scale (1 = not true, 2 = sometimes true, 3 = often true, 4 = almost always true). T-scores were calculated for each subscale, and a total T-score (SRS-2 T-scores) indexing the severity of social deficits was generated by combining the subscales. T-scores above 60 are suggestive of an autism spectrum disorder diagnosis (normal range $\leq$ 59; clinically significant social impairment in the mild-moderate range = 60–75; severe social impairment $\geq$ 76).

6.2.3.2.3 Movement Assessment

The Movement ABC-2 Parent Checklist (Parent M-ABC-2; Henderson, Sugden, & Barnett, 2007) was completed by parents or carers to measure how their child manages movement related activities. The Parent M-ABC-2 contains three sections: A, B and C. Statements in Section A and B cover child interactions with static or dynamic environments. For the purposes of this study, only data from Section C of the Parent M-ABC-2 was analysed. In Section C, parents were asked about non-motor factors that interfered with their child's ability to complete movement activities (e.g., being overactive, impulsive etc.). Parents and carers responded to each statement by selecting 'yes' = 1 or 'no' = 0. A total score was obtained by summing the responses. High scores on the Parent M-ABC-2 indicate that non-motor factors negatively impact movement.

6.2.3.3 Data Completion Rates

6.2.3.3.1 STXBP1

It is extremely difficult and cognitively tiring for individuals with ID to complete multiple assessments. Of the *STXBP1* participants, two completed the full cognitive and behavioural testing protocol, and another completed all tasks except the WASI (Wechsler, 2011). The protocol was not complete for the remaining 11 children with *STXBP1* for various reasons: for two children, the assessments were terminated due to attention and behavioural problems, for six children the assessments were too difficult, and the remaining three children were too young to complete the tests (they were younger than the standardisation range). Therefore, the descriptive reporting of the *STXBP1* phenotype was largely restricted to measures with available data from all 14 participants (Medical Questionnaire, Vineland, Movement ABC Parent Checklist, Developmental Behaviour Checklist, and Social Responsiveness Scale). Conners (Conners, 2008) data were missing for 6 participants. WASI (Wechsler, 2011) data were available for three participants, and two of these plus an additional one participant completed all Leiter Brief IQ subtests (Roid & Miller, 1997). Four participants completed runs across all three FarmApp (Brkić et al., 2022) tasks: Go/NoGo, Visuo-Spatial Working Memory, and Recognition. Data were available for one additional participant on the Go/NoGo task.

6.2.3.3.2 STXBP1 Subsample

Only parent questionnaire data was analysed in the STXBP1 Subsample. All eight participants had complete data for the following measures: VABS-II (Sparrow, Cicchetti, & Balla, 2005), Movement ABC Parent (Henderson, Sugden, & Barnett, 2007), Conners (Conners, 2008), Developmental Behaviour Checklist (Einfeld & Tonge, 2002), and Social Responsiveness Scale (Constantino & Gruber, 2005).

6.2.3.3.3 ID Comparison and ID Subsample

Only parent questionnaire data from the ID Comparison and ID Subsample groups was included in this analysis. All participants had complete data for the following measures: Medical Questionnaire (Epilepsy details only, see Table 14), VABS-II (Sparrow, Cicchetti, & Balla, 2005), Movement ABC Parent (Henderson, Sugden, & Barnett, 2007), Conners (Conners, 2008), Developmental Behaviour Checklist (Einfeld & Tonge, 2002), and Social Responsiveness Scale (Constantino & Gruber, 2005).

6.3 RESULTS

The first section presents in-depth phenotyping of all *STXBP1* participants and includes all available data for each measure (section 6.3.1). The second section compares the behavioural profile of *STXBP1* participants (Group name: STXBP1) to 33 participants with other genetic mutations causing intellectual disability, referred to as the 'ID Comparison' group (section 6.3.2). Finally, the third section (section 6.3.3) compares data from a subsample of the STXBP1 group (N = 8, referred to as 'STXBP1 Subsample', to an ability-matched subgroup of the 'ID Comparison' group (N = 8), which will be referred to as 'ID subsample' (*see Section 6.2.1.4* for detailed description of how the sub-groups were matched). This was done to explore the specificity of behavioural characteristics associated with *STXBP1* mutation by comparing the group to a subset of the ID comparison group selected to match the range of global adaptive abilities within the *STXBP1* group.

6.3.1 STXBP1 PHENOTYPE

In this section I present the medical, motor, cognitive, and behavioural characteristics of 14 participants with *STXBP1*.

6.3.1.1 Medical History

6.3.1.1.1 General Health

The duration of pregnancies for the STXBP1 group ranged from 32 to 42 weeks. No abnormal ultrasounds or threatened miscarriages were reported. However, five mothers reported persistent vomiting throughout pregnancy. There were concerns about foetal well-being during the birth of six of the participants. Four participants were born by Caesarean section. Only three cases subsequently spent time in the Special Care Baby Unit, with none requiring neonatal intensive care. Three individuals were readmitted to hospital within the first month of life for investigation of suspected seizures.

A large proportion of parents (85%) reported feeding issues during infancy and early childhood, common problems being reflux, choking, and difficulty moving solids. Sleep difficulties were reported in four cases during infancy, characterised by problems falling and staying asleep. In two cases, sleep difficulties persisted into adolescence.

Parents also reported a range of sensory issues with regard to hearing, sight, and touch. Six parents reported sight issues including ptosis, misshapen eye, difficulty focusing, excessive blinking, and long-sightedness. Three parents reported hearing issues: one participant suffered from recurrent ear infections, one had glue ear in both ears, and one participant required hearing aids. A diverse range of sensory issues involving touch and sound were reported in eight cases ($N = 3$ for touch, $N = 3$ for sound, and $N = 2$ for touch and sound). Three participants enjoyed sensory feedback from sensory objects and people (e.g., hugging, and rough play), whereas two did not. One child appeared to enjoy high-pitched sounds such as babies crying, whereas four parents noticed that their children were averse to certain loud sounds such as hoovers, fireworks and weather. One of these children had received a diagnosis of hyperacusis.

6.3.1.1.2 Neurological Problems

Thirteen *STXBP1* participants had a history of seizures, diagnosed as epilepsy by a paediatric neurologist. The age of seizure onset varied from just after birth to 13 years of age. Half ($N = 7$) had a seizure in their first year of life. Four of these participants had a seizure within the first two months of life. For the remaining six individuals with epilepsy, the age of onset of seizures was broadly spread across childhood. Three participants experienced their first seizure during the peripubertal phase. The type of seizure also varied. One participant was diagnosed with infant encephalopathy, two had tonic-clonic seizures, and six experienced focal seizures. Seizure phenotypes were mixed or uncertain for the remaining participants. Epilepsy has been treated with a wide range of anti-epileptic drugs, the most frequently used was Levetiracetam ($N = 4$). Whilst seizure frequency was generally reduced on treatment, no participant was reported to be completely seizure free during the months prior to research assessment. One participant experienced infrequent seizures (approximately three between 5-14 years of age) but was not on medication. Four participants also experienced tremors in their hands, which I observed during assessments.

6.3.1.1.3 Developmental Milestones

Thirteen participants learned to sit independently from 7 to 42 months. Seven out of twelve sat independently before 10 months, which is within the normal to mildly delayed range, whilst the remainder were clearly delayed. The youngest child in the sample (1-year-old) had extremely low muscle tone and was not able to sit independently at the time of assessment. Four children did not learn to crawl but ‘bum-shuffled’ or ‘bunny hopped’. Nine participants were walking independently at the time of assessment. However, eight acquired this skill late: between 25 and 60 months old. One participant, aged 14, could walk a few steps but required constant support. Three participants between 1.9 and 6.8 years of age had not yet learned to walk.

Speech and language acquisition were delayed in all cases. At time of testing, eight participants used some verbal communication, with age of first words being between 18 and 156 months (13 years). The age of participants unable to verbally communicate at time of testing was between 1.1 and 15 years. Of the four participants who could form simple phrases or sentences, three of these acquired this skill late, after 5.5 years of age.

6.3.1.2 Motor Abilities

According to the Motor Skills subscale of the Vineland (Sparrow, Cicchetti, & Balla, 2005), the participant with the highest score, aged ten, could run, jump, hop, skip, and kick, throw and catch a ball. Whereas the participant with the lowest score, aged 6 years, was unable to pull themselves to a standing position. Fine motor abilities also ranged from severely impaired to average for age. Seven participants could turn pages of a book and stack blocks. A further three participants with better fine motor abilities could draw shapes freehand and use scissors. In contrast, three participants with poorer skills were unable to turn pages or stack blocks (see Table 23 at the end of this chapter).

The child assessment of motor abilities, the M-ABC-2 (Henderson, Sugden, & Barnett, 2007), revealed the *STXBP1* participants had a range of gross and fine motor skills. Due to the variability in gross and fine motor skills across the sample, appropriate tasks were selected from the three age bands in the M-ABC-2. The tasks could only be administered with five of the participants. Three children were too young, and an additional three children had not acquired the motor capabilities to attempt the tasks. The measure could not be administered with three participants due to inattention, hyperactivity, and limited time available to complete the full testing schedule. A detailed breakdown of performance is available in Table 24 and the end of this chapter.

Over 85% of parents reported in the Parent MABC-2 (Henderson, Sugden, & Barnett, 2007), that being overactive, distractible, impulsive, disorganised, and hesitant contributed to their child's motor difficulties. In addition, 62.5% of the participants were described as becoming anxious in stressful, movement-related situations.

6.3.1.3 Cognitive Abilities

6.3.1.3.1 Fluid Intelligence

The three participants who completed the WASI-II (Weschler, 2011) had a Full-Scale IQ score of 40, which is at floor, and there was little variability in subscale raw scores (see Table 18). Two of these participants and an additional participant completed all four subscales of the Leiter Brief IQ (Roid & Koch, 1997). Performance was again at floor, but there was greater variability in performance on this measure than the WASI (see Table 18).

All three participants had a Leiter Brief IQ of 36, subscale scores of 1, which is at floor. Leiter subscale raw scores give some indication of the variability in cognitive abilities (see Table 18)

Table 18: Cognitive abilities raw scores for STXBP1 group

Assessment	**Subtest**	**N**	Mean	SD	Range
WASI-II		3			
	Block Design		0.67	1.16	0 – 2
	Vocabulary		1.33	1.53	0 – 3
	Matrix Reasoning		0.00	0.00	-
	Similarities		1.67	1.53	0 – 3
	Full Scale IQ		40.00	0.00	-
Leiter		3			
	Figure Ground		8.00	3.00	5 – 11
	Form Completion		11.67	4.16	7 – 15
	Repeated Patterns		3.67	0.58	3 – 4
	Sequential Order		7.33	0.58	7 – 8
	Leiter Brief IQ		36	0	-
FarmApp					
Go/NoGo		4			
	No. of runs		7.26	6.99	1.13-11.48
	Go RT (seconds)		1.83	0.20	0.80 – 1.27
VSWM		5			
	No. of runs		4.94	3.36	1.00 – 9.83
	Average Span		3.25	2.50	1.00 – 7.00
Recognition		4			
	No. of runs		2.50	0.58	2.00 – 3.00
	Total Response Accuracy		0.66	0.17	0.49 – 0.87
	Sheep Accuracy		0.73	0.21	0.49 – 0.87
	Chicken Accuracy		0.60	0.17	0.42 – 0.83

Note: FarmApp data from Child Mode only. N: number of participants with complete data.

6.3.1.3.2 Inhibition, Visuo-Spatial Working Memory, and Recognition

Go/No-Go

Four participants completed at least one run of the Go/No-Go task, the Sheep Game (*M* = 7.26 runs, *SD* = 6.99, range = 1.13-11.48). The mean reaction time for participants on Go trials in the Sheep game was 1.83 seconds (*SD* = 0.20, range = 0.80-1.27).

Visuo-Spatial Working Memory

Five participants completed at least one run of the visuo-spatial working memory (VSWM) task, the Chicken Game (*M* = 4.94 runs, *SD* = 3.36, range = 1-9.83). The average span reached (number of chickens touched in the correct sequence) was 3.25 (*SD* = 2.5, range = 2-7).

Recognition

Four participants unlocked the recognition task, the Memory Game (*M* = 2.5 runs, *SD* = 0.58, range = 2-3). Children's response accuracy was above chance levels (*M* = 0.66, *SD* = 0.17, range = 0.49 – 0.87). Children were more accurate in recognising Sheep (*M* = 0.73, *SD* = 0.21, range = 0.53 – 0.93) than Chicken stimuli (*M* = 0.60, *SD* = 0.17, range = 0.42 – 0.83).

6.3.1.4 Behaviour

6.3.1.4.1 Vineland Adaptive Behaviour Scales-II Survey Form

The VABS-II (Sparrow, Cicchetti, & Balla, 2005) revealed that participants with *STXBP1* mutations had poor communication skills overall, with scores in the moderately to severely impaired range (see Table 19 below). However, there was huge variability in the receptive and expressive subscales. The participant's receptive communication skills ranged from an inability to listen and attend to speech through to being able to recall and carry out instructions given five minutes previously. Their expressive skills ranged from difficulty with pre-speech expression through to engaging in interactive speech. The Vineland socialisation subscale scores indicated that the STXBP1 group demonstrated social adaptive impairments. The participants expressed a strong interest in social interaction during the testing sessions, despite limited communication skills and social skills as assessed by Vineland Adaptive Behaviour Scales.

6.3.1.4.2 Developmental Behavioural Checklist

Seven out of eight *STXBP1* participants had a DBC Total Problem Behaviour Score (Einfeld & Tonge, 2002) above the clinical cut-off (T score: 46) indicative of the likely presence of psychopathology. Behaviours such as biting, spitting, ripping things, hitting, and running away from caregivers were common, and parents reported that these behaviours intensified when their children were frustrated with something. Two parents remarked that these behaviours had become increasingly difficult to manage following the onset of puberty. High levels of anxiety associated with specific phobias of fireworks, adverse weather, hoovers, lawnmowers and the emotions of others were reported in five cases (4 females, 1 male). Six parents also reported that their child rarely cried and did not appear to feel pain.

6.3.1.4.3 Conners 3rd Edition–Parent Short form

Participants scores were very elevated on the inattention, hyperactivity, learning problems, and peer relations subscales. Two participants had a T-score in the normal range for executive functions, and five participants scored within the normal range on the aggression subscale.

6.3.1.4.4 Social Responsiveness Scale

Three participants had an SRS total score within the mild to moderately impaired range and five had scores indicative of severe social impairment likely to meet criteria for an autism spectrum diagnosis. Participants' scores fell within the moderate to severe range on Social Awareness, Social Cognition, Social Communication and Restricted Interests and Repetitive Behaviours subscales. In contrast, scores on the Social Motivation subscale were notably preserved and within the normal range. Parents typically described a social profile characterised by self-confidence in social settings and an interest in engaging with peers and adults.

Table 19: Vineland Domain and Subdomain Standard scores

		STXBP1	ID Comparison		STXBP1 subsample	ID subsample	#
Communication	Mean	47.64	59.15	$U = 146.5$, $p < 0.05$	43.38	47.50	$U = 20.0$, $p = 0.23$
	SD	14.60	18.65		12.06	7.62	
	Range	30 – 71	21 – 100		33 – 67	38 – 61	
Receptive	Mean	5.93	8.70	$U = 122.0$, $p < 0.05$	4.75	7.13	$U = 11.0$, $p < 0.05$
	SD	5.52	2.99		2.61	1.46	
	Range	2 – 12	1 – 16		2 – 10	5 – 10	
Expressive	Mean	4.64	6.88	$U = 155.5$, $p = 0.08$	3.38	3.00	$U = 32.0$, $p = 0.08$
	SD	3.25	4.02		2.83	1.85	
	Range	1 – 8	1 – 17		1 – 8	1 – 6	
Written	Mean	5.55	7.97	$U = 93.0$, $p < 0.05$	5.38	5.88	$U = 21.5$, $p = 0.28$
	SD	1.75	3.44		1.85	0.99	
	Range	4 – 9	3 – 20		4 – 9	5 – 8	
Daily Living Skills	Mean	45.57	57.36	$U = 150.0$, $p = 0.06$	38.38	43.00	$U = 22.0$, $p = 0.33$
	SD	19.74	18.25		15.53	13.35	
	Range	25 – 77	21 – 98		25 – 73	30 – 71	
Personal	Mean	4.71	6.18	$U = 176.0$, $p = 0.20$	2.88	2.75	$U = 29.0$, $p = 0.80$

	SD	4.48	4.27		3.79	2.55	
	Range	1 – 12	1 – 19		1 – 12	1 – 8	
Domestic	Mean	6.43	8.73	$U = 150.0, p = 0.06$	4.38	5.63	$U = 18.5, p = 0.16$
	SD	4.27	3.74		3.34	3.29	
	Range	2 – 13	3 – 17		2 – 12	3 – 13	
Community	Mean	4.86	6.18	$U = 182.0, p = 0.25$	3.00	3.88	$U = 26.5, p = 0.57$
	SD	3.92	3.63		1.85	2.64	
	Range	1 – 12	1 – 13		1 – 7	1 – 9	
Socialisation	Mean	52.29	63.30	$U = 124.5, p < 0.05$	47.75	56.13	$U = 11.0, p = 0.03$
	SD	12.12	16.52		8.43	9.16	
	Range	34 – 75	20 – 108		40 – 67	46 – 75	
Interpersonal	Mean	6.00	7.94	$U = 146.0, p < 0.05$	4.63	6.63	$U = 15.0, p = 0.08$
	SD	3.33	2.97		2.07	1.85	
	Range	2 – 13	1 – 15		2 – 8	4 – 10	
Play and Leisure	Mean	4.93	7.76	$U = 135.5, p < 0.05$	3.63	5.00	$U = 26.0, p = 0.57$
	SD	3.45	3.65		3.25	3.74	
	Range	1 – 11	1 – 15		1 – 11	1 – 11	
Coping Skills	Mean	6.79	9.30	$U = 116.5, p < 0.05$	6.00	8.00	$U = 9.0, p < 0.05$
	SD	1.85	3.12		1.31	1.31	

	Range	4 – 9	6 – 20		5 – 8	6 – 10	
Motor	Mean	60.43	69.03	$U = 153.5$, $p = 0.07$	63.25	62.63	$U = 29.0$, $p = 0.80$
	SD	16.39	12.48		17.41	5.63	
	Range	37 – 97	56 – 107		40 – 97	56 – 72	
Gross	Mean	8.50	9.58	$U = 151.5$, $p = 0.06$	9.63	9.00	$U = 24.5$, $p = 0.44$
	SD	2.77	1.79		2.97	1.51	
	Range	5 – 15	7 – 16		6 – 15	7 – 11	
Fine	Mean	8.29	10.09	$U = 140.5$, $p < 0.05$	8.00	8.50	$U = 30.0$, $p = 0.87$
	SD	4.01	2.95		3.12	0.93	
	Range	4 – 17	7 – 20		4 – 14	7 – 10	

Comparison to *STXBP1* full dataset group, Mann-Whitney U test, P<0.05

6.3.2 COMPARISON OF STXBP1 GROUP AND ID COMPARISON GROUP

In this section, *STXBP1* participants are compared to the 'ID Comparison' group (N = 33). Complete parent questionnaire data (see Table 20 below) was only available for eight *STXBP1* participants. Therefore, analysis of these data was restricted to the 'STXBP1 Subsample' (N = 8) and the 'ID Comparison' group (N = 33).

6.3.2.1 Motor Function

There was no significant difference in gross motor skills between 'STXBP1' group (N = 14) and 'ID Comparison' group as assessed via Vineland Adaptive Behaviour Scales (see Table 19 above). However, the STBXP1 group had poorer Vineland fine motor subscale scores when compared to the ID Comparison group (U = 140.5, p < 0.05).

The M-ABC Parent data demonstrated that there was also no significant difference between groups in how parents evaluated the contextual factors influencing children's motor performance in the 'STXBP1 Subsample' group compared to the 'ID Comparison' group (see Table 20 below).

6.3.2.2 Emotional and Behavioural Difficulties

The 'STXBP1 Subsample' and ID Comparison groups did not significantly differ on DBC total problem behaviour scores or subscale scores (stratified for ID severity), indicating that emotional and behavioural difficulties in general are not more prevalent or severe in the STXBP1 group than expected for intellectual disability due to other causes. Anxiety and disruptive behaviours were prominent problems for some but not all individuals within the STXBP1 group; however, group mean DBC anxiety subscale scores did not differ from the ID comparison group (U = 99.5, p = 0.29). On the Conners scales, the 'STXBP1 Subsample' were rated as significantly more hyperactive / impulsive compared to the 'ID Comparison' group (U = 52.0, p = 0.03). In summary, the STXBP1 group demonstrated similar total problem behaviour scores to the ID Comparison group, with variation within the group in the severity and types of problems reported for each individual. High hyperactivity / impulsivity scores were more consistent within the STXBP1 group and on average were more severe than reported for individuals with other monogenic causes of ID.

6.3.2.3 Social Functioning

There was a significant difference between the 'STXBP1' (N = 14) and the ID Comparison Group, on Vineland Socialisation scores: the ID group on average displayed stronger skills on all three subscales; Interpersonal Relations ($U = 146.0$, $p < 0.05$), Play and Leisure ($U = 135.5$, $p < 0.05$), and Coping Skills (U = 116.5, p < 0.05).

The 'STXBP1 Subsample' and 'ID Comparison' groups did not differ significantly on SRS Total Score or Social Awareness, Social Cognition, Social Communication and Repeated Interests and Repetitive Behaviours subscales. However, the ID comparison group were somewhat more impaired on Social Motivation, compared to the STXBP1 group who fell within the 'Normal' range for this subscale (U = 74.0, p = 0.06). In summary, the STXBP1 group demonstrated an atypical profile of social functioning in comparison to ID in general, with more severely impaired everyday social behaviour (on Vineland scales), equivalently impaired social cognition (on SRS), but preserved social motivation.

Table 20: Parent Questionnaire data: Movement, behavioural, emotional, and social characteristics

		STXBP1 subsample	ID Comparison		ID subsample	
MABC Parent						
Static environments	Mean	24.50	18.45	$t = -1.57, p = 0.13$	23.00	$t = 0.31, p = 0.67$
(Motor Competence)	SD	12.32	9.16		5.95	
	Range	3 – 42	1 – 37		11 – 31	
Dynamic environments	Mean	26.00	22.12	$t = -1.04, p = 0.30$	25.50	$t = 0.10, p = 0.92$
(Motor Competence)	SD	11.86	8.83		6.76	
	Range	3 – 43	3 – 39		14 – 34	
Total Motor Score	Mean	50.50	40.61	$t = -1.36, p = 0.18$	48.50	$t = 0.22, p = 0.83$
	SD	23.58	17.16		10.94	
	Range	6 – 85	6 – 74		35 – 62	
Conners (T-score)*						
Inattention	Mean	84.75	79.19	$U = 80.5, p = 0.29$	86.00	$U = 30.0, p = 0.88$
	SD	7.48	12.03		5.26	
	Range	72 – 92	55 – 90		79 – 90	
Hyperactivity	Mean	85.25	72.37	$U = 52.0, p = 0.03$	77.25	$U = 22.0, p = 0.33$
	SD	6.52	14.98		13.23	
	Range	73 – 90	40 – 90		56 – 90	

Learning Problems	Mean	86.38	79.44	$U = 70.0$, $p = 0.14$	86.13	$U = 27.5$, $p = 0.65$
	SD	8.31	10.91		4.97	
	Range	66 – 90	57 – 90		78 – 90	
Executive Functions	Mean	76.00	64.56	$U = 61.5$, $p = 0.07$	66.25	$U = 19.0$, $p = 0.20$
	SD	14.80	14.43		13.33	
	Range	54 – 90	43 – 90		48 – 83	
Aggression	Mean	59.88	52.30	$U = 94.0$, $p = 0.60$	47.50	$U = 22.5$, $p = 0.33$
	SD	17.36	10.47		4.75	
	Range	44 – 90	44 – 83		44 – 58	
Peer Relations	Mean	84.13	80.33	$U = 105.5$, $p = 0.92$	81.38	$U = 30.5$, $p = 0.88$
	SD	7.20	14.99		16.11	
	Range	72 – 90	45 – 90		45 – 90	
DBC (Percentile/stratified by ID severity)						
Disruptive	Mean	68.25	50.91	$U = 91.5$, $p = 0.19$	54.00	$U = 22.5$, $p = 0.33$
	SD	24.29	28.82		25.19	
	Range	28 – 90	4 – 98		26 – 90	
Self-Absorbed	Mean	73.75	70.12	$U = 128.5$, $p = 0.91$	77.75	$U = 28.0$, $p = 0.72$
	SD	19.20	25.81		17.93	

	Range	34 – 94	10 – 100		40 – 100	
Communication	Mean	75.5	76.42	$U = 87.0$, $p = 0.15$	78.00	$U = 29.0$, $p = 0.80$
	SD	17.82	23.11		18.79	
	Range	48 – 96	6 – 100		40 – 92	
Anxiety	Mean	42.50	58.30	$U = 99.5$, $p = 0.29$	66.25	$t = -1.73$, $p = 0.11$
	SD	24.49	26.85		30.09	
	Range	12 – 76	10 – 100		10 – 96	
Social Relations	Mean	49.50	58.85	$U = 99.5$, $p = 0.29$	57.00	$t = -0.60$, $p = 0.56$
	SD	22.85	27.96		27.28	
	Range	24 – 96	12 – 100		24 – 98	
Total Problems	Mean	73.50	63.94	$U = 103.5$, $p = 0.36$	67.25	$U = 23.5$, $p = 0.38$
	SD	24.14	27.39		22.27	
	Range	30 – 94	0 – 100		38 – 98	
SRS (T score)						
Social Awareness	Mean	78.75	73.64	$t = -1.24$, $p = 0.22$	76.38	$t = 0.49$, $p = 0.63$
	SD	5.95	11.21		12.29	
	Range	70 – 87	49 – 93		56 – 92	
Social Cognition	Mean	73.88	72.70	$t = -0.27$, $p = 0.79$	74.38	$t = -0.11$, $p = 0.92$
	SD	11.48	11.02		6.63	

	Range	53 – 92	47 – 96		67 – 85	
Social Communication	Mean	76.25	72.70	*t* = -0.80, *p* = 0.43	80.88	*t* = -0.91, *p* = 0.38
	SD	9.17	11.75		11.10	
	Range	66 – 90	50 – 96		67 – 96	
Social Motivation	Mean	58.00	64.52	*U* = 74.00, *p* = 0.06	69.50	*U* = 9.50, *p* = 0.02
	SD	13.55	11.77		13.65	
	Range	45 – 89	41 – 95		58 – 95	
RIRB	Mean	79.00	78.36	*t* = -0.11, *p* = 0.91	87.38	*t* = -1.26, *p* = 0.23
	SD	11.89	15.20		14.64	
	Range	62 – 92	48 – 108		72 – 108	
SRS Total Score	Mean	77.75	75.12	*t* = -0.58, *p* = 0.57	81.75	*t* = -0.79, *p* = 0.44
	SD	9.27	12.01		10.90	
	Range	66 – 94	49 – 98		70 – 98	

Note: *6 participants in the ID Comparison sample had no Conners data.

6.3.3 Comparison of STXBP1 Subsample and ID Subsample Groups

To explore the specificity of behavioural characteristics associated with an *STXBP1* mutation in more detail, secondary comparisons were made to a subset of the 'ID Comparison' group selected to match the range of global adaptive abilities within the STXBP1 group. The profile of participants with complete data in the STXBP1 group, referred to as the 'STXBP1 Subsample' ($N = 8$) was compared to an ability-matched subgroup of the ID comparison group, referred to as the 'ID Subsample' ($N = 8$; see Table 14 for comparison information).

Although the two groups had, by construction, comparable global adaptive function scores (see Table 14 Vineland), communication impairments were more severe in the 'STXBP1 Subsample' group. The 'ID Subsample' group on average demonstrated stronger receptive than expressive abilities whereas the 'STXBP1 Subsample' group showed severe restriction of both receptive and expressive abilities (with significant difference between groups in receptive score: $U = 11.0$, $p < 0.05$). Groups did not significantly differ on Daily Living Skills or Motor Skills.

Behavioural problems assessed via the DBC (total, subscales) did not significantly differ between groups. The 'STXBP1 Subsample' group also did not significantly differ from the 'ID Subsample' group in Conners hyperactivity or impulsivity scores, indicating that although these difficulties are a consistent problem area for the 'STXBP1 Subsample' group, they are not more impaired in these domains than other individuals with equivalent global impairments (see Table 20 *Parent Questionnaire Data*).

Comparison of Vineland scores indicated that everyday social behaviours were more impaired in the 'STXBP1 Subsample' group in comparison to the 'ID Subsample' group (socialisation domain: $U = 11.0$, $p = 0.03$, in particular on Coping Skills: U = 9.0, p = 0.02). Analysis of SRS scores indicated a stable pattern of results – total SRS scores (see Table 20), social cognitive scores, and restricted and repetitive behaviour scores did not significantly differ, but social motivation was preserved in the 'STXBP1 Subsample' and significantly higher than in the 'ID Subsample' group ($U = 9.50$, $p = 0.02$).

6.4 Discussion

In this study, I systematically assessed the behavioural characteristics of children with deficits in fluid intelligence – intellectual disabilities (ID) – caused by a known *de novo* genetic mutation. First, I characterised the phenotype of children and adolescents with variants in *STXBP1*, with a view to improving prognostication and management for individuals with this rare genetic diagnosis. I then compared their performance to a matched-ability group of individuals with ID caused by *de novo* mutations in other genes.

Participants in the STXBP1 group demonstrated complex and persistent difficulties across multiple domains of everyday function and emotional-behavioural development. Consistent problems within the group included severe communication impairments and hyperactivity-impulsivity. Some participants were greatly troubled by anxiety including specific phobia, but these symptoms varied across the group. I observed an additional contrast in some individuals between low social anxiety and high non-social anxiety. In view of moderate to severe ID, it is important to consider whether behavioural characteristics within this group differ from developmental expectations – I found that receptive language ability and socialisation skills (but not hyperactivity) discriminated the STXBP1 and ID comparison groups, suggesting a specific contribution of aetiology to these features.

I was struck by a consistent profile of preserved social behaviour within the STXBP1 group. Participants showed great enjoyment of social interactions with family members and demonstrated reciprocal behaviours such as sharing and turn taking. My experience was that, from the outset of the testing sessions, participants wanted to engage with me (a novel stranger within the participants' home) particularly in social activities such as drawing, dancing, singing, and playing games. Even participants as young as two years held eye contact and were very sociable. On the other hand, parental questionnaire ratings indicated limited understanding of social expectations and difficulties integrating into normal social settings. The observation of significantly stronger social motivation remained true, whether comparing to the whole ID sample or low ability ID sample, indicating that this feature is not simply a correlate of ID severity. Although several individuals scored above cut-off on the SRS for a likely autism spectrum diagnosis, this profile of social behaviours would not be typical for an individual with autism. These findings support other research into the diverse social developmental trajectories contributing to autism-like impairments in individuals with genomic disorders (Mabb et al., 2011; Orock et al., 2018; Richards et al., 2017).

Epilepsy prevalence within the STXBP1 group was high, as previously reported, but the type, severity, age of onset, treatment sensitivity, and remission of seizures varied across participants. A relatively high prevalence of focal seizures was reported, in keeping with previous reports (Stamberger et al., 2016). Further studies are needed to examine the contribution of subclinical seizure activity to developmental progress for individuals with *STXBP1* variants, including interactions between low-level sensory processing, cognitive biases, and the social environment to mediate anxieties or prosocial behaviour. In addition, comparison to other groups with similar epilepsy prevalence and seizure characteristics could tease apart aetiology-specific behavioural features from correlates of, for example, focal seizures.

A further goal of this research was to stimulate future investigation of the mechanisms linking presynaptic dysfunction to cognitive and social development, which may ultimately lead to aetiology-specific therapies. Toward this goal, I compared our observations of the STXBP1 group to published behavioural phenotyping studies of mice with *STXBP1* mutations, to consider consistency across species. The behavioural phenotypes exhibited across different mouse models include elevated anxiety, hyperactivity, impairments in acquiring and maintaining spatial memory and reversal of previously learned strategies (Hager et al., 2014; Kovačević et al., 2018; Orock et al., 2018). Importantly, not every aspect of learning and behaviour is impaired: *STXBP1* haploinsufficient mice demonstrate normal sociability, preference for social novelty, and normal profiles of attentional control without impulsivity (Kovačević et al., 2018). In one study, social behaviours in *STXBP1* mouse models were examined by introducing a novel mouse into the chamber. Regardless of specific *STXBP1* variant, heterozygous mice spent more time around the novel mouse, displayed normal sociability and preference for social novelty. These findings appear, at least on the surface, to be convergent with my observations of young people with *STXBP1* variants, and future research could explore whether similar neurodevelopmental mechanisms underpin social interactions across species.

In considering potential mechanisms underlying this and other aspects of the *STXBP1* phenotype, I have been struck by phenotypic similarity with Angelman Syndrome (AS). AS is characterised by developmental delay, seizure susceptibility including focal seizures, movement disorders, language deficits, impulsivity, short attention span, and specific phobias (Bakke et al., 2018; Horsler & Oliver, 2006; Huang et al., 2013; Mabb et al., 2011; Pelc et al., 2008). The behavioural hallmarks of AS are hyperactivity and hypersociability (Mabb et al., 2011). Just as in *STXBP1*, individuals with AS display increased social motivation, prolonged social interest and also excessive smiling and laughing (Oliver, Woodcock, & Adams, 2010). Socialisation is believed to be underpinned by a network of brain regions including the amygdala, ventral striatum, orbital and ventromedial regions of the prefrontal cortex. However, different regions and networks may have a greater role in specific aspects of sociability (Chevallier et al., 2012). It has been hypothesised that reciprocal changes within striatal circuits give rise to the atypical social novelty profile associated with AS, consistent with experimental evidence for altered striatal dopamine balance in a mouse model (Berrios et al., 2016; Stoppel & Anderson, 2017). Examining the striatum and its dependent functional neural systems, in the context of the wider 'social' brain network, may provide a starting point for understanding the neural mechanisms driving the atypical social profile in *STXBP1*.

To determine whether specific learning deficits characterise the disorder and contribute to behavioural characteristics, it is necessary to accurately measure relevant aspects of cognition in individuals with moderate to severe ID. However, assessment of cognitive function in individuals with *STXBP1* is challenging due to low levels of understanding, and hyperactivity. Only three participants completed the standardised assessments and performance was at floor. While not considered a replacement for existing standardised assessments (Brkić et al., 2022), the FarmApp demonstrated that more children can engage with and complete cognitive tasks if they are intuitive, and developmentally appropriate. However, only 4-5 participants completed a single run of each FarmApp task, so while it's more sensitive than standard assessments it's still too challenging for most children with *STXBP1*. If the cognitive strengths and difficulties of individuals with moderate to severe ID of known genetic origin are to be accurately characterised, future research must develop age and stage appropriate measures that uncover the variability in sub-ordinate cognitive processes. Given the characteristics of participants with *STXBP1* outlined above, engagement with app-based measures may be increased if: (i) task rewards incorporated a social aspect e.g., smile animations; (ii) and if eye-tracking was used as opposed to manual responses, given the prevalence of motor difficulties.

In addition to behavioural methods, application of non-invasive physiological approaches such as wireless EEG in semi-naturalistic settings may be feasible and informative. These methods should be developed now and applied in future once a larger number of individuals have been diagnosed via genome-wide testing and can be followed from early childhood through to adult life.

There are several important limitations to this study. The sample size is small and encompasses a wide age range. I aimed to recruit participants as broadly as possible, from multiple regions of the UK and multiple medical specialties. In comparison to previous studies, the STXBP1 participant group had a higher proportion of individuals recruited via clinical genetics rather than paediatric neurology, which may provide a more comprehensive picture of the phenotypic spectrum or may under-represent individuals with early onset epileptic encephalopathy. The data were collected via parent report, and it was not feasible to corroborate or elaborate via review of medical records. It was also not feasible for me to be blind to genetic diagnoses, however my use of standardised parent-report questionnaire measures should mitigate potential bias of diagnosis-informed clinician ratings. Non-parametric statistical analyses are included to enhance my descriptive observations and provide an indication of which measures may discriminate between groups. Given the small sample size and large number of measures, adjusting the p-value threshold for statistical significance to account for multiple comparisons is not practical, and results should be interpreted with due caution. The robustness of these findings will, I hope, be tested in future independent studies with pre-registered hypotheses and predictions building on our exploratory results. In view of the small sample size, within-group analysis of behavioural variation was not justified (for example genotype-phenotype correlations, associations with epilepsy severity or age-of-onset). These analyses would improve the prognostic utility of our results, if a large enough sample could be collected.

6.5 CHAPTER SUMMARY

This research presents a post-diagnostic evaluation of children and adolescents with *STXBP1*-associated neurodevelopmental disorder. Individuals with *STXBP1* variants have a wide range of complex adaptive, social, and behavioural characteristics. Via comparison to individuals with other ID-related genetic diagnoses, I have highlighted a more limited set of characteristics which may discriminate individuals with *STXBP1* from individuals with other monogenic causes of neurodevelopmental disorder. Delineating this spectrum should assist in recognising this disorder and supporting families after diagnosis. Moreover, convergence between *STXBP1* phenotypes across species and across disorders (AS), provides a new starting point for understanding the molecular and neural mechanisms which underlie prosocial development.

Table 21: Additional STXBP1 participant information

	1	2*	3*	4	5	6	7	8	9	10	11	12	13**	14
Age	0.15	1.79	1.79	5.6	6.69	10.52	12.78	13.0	14.29	14.65	15.04	15.12	16.5	17.74
Sex	F	F	F	F	M	M	F	F	F	F	F	M	M	F
Gestation (wks)	40	32	32	39	42	39	39	36	40	41	40	41	33	38
Birth Weight (kgs)	3.09	2.0	1.96	2.92	3.49	2.78	3.37	3.54	3.63	3.22	2.81	3.86	1.91	4.03
ID	Mild	Mild	Mild	Moderate	Severe	Mild	Severe	Moderate	Severe	Moderate	Severe	Severe	Severe	Severe
Floppy	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes
Feeding Issues	Difficulty moving solids	Reflux	Reflux	Reflux	Reflux	No	Reflux	Choking and reflux	Reflux	No	Reflux	No	Choking	Choking
Sleep Issues	No	No	No	Yes	No	Yes	No	Yes	No	Yes	No	No	No	Yes
Sensory Issues	Sight and hearing	Sight	Sight	Sound sensitive	Ear infections	Touch	Sight, Sound sensitive	Sight, Sound sensitive	None reported	Hyperacusis, Touch	Sight and hearing	Sight	Touch	

*Twins
** Participant has a typically developing twin

Table 22: Gene groups and participant numbers recruited to the study

Genetic Mutations	**Participant Numbers**
ARID1B	4
CASK	1
DDX3X	8
DLG3	2
DYRK1A	1
EHMT1	6
PAK3	1
SETD5	6
SHANK1	1
SMARCA2	1
STXBP1	14
TRIO	1
ZDHHC9	1
Total	47

Table 23: Responses to a selection of Vineland items from the Gross and Fine Motor subscales

ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Age at assessment	1.08	1.79	1.79	5.60	6.69	10.52	12.78	13.00	14.29	14.65	15.04	15.12	16.50	17.74
Gross														
Sits without support	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Crawls		yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	yes
Pulls to standing		yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	yes
Takes two steps		no	no	yes	no	yes	no	yes	yes	yes	yes	yes	yes	yes
Stands for 5 mins		no	no	yes	no	yes	no	yes	yes	yes	yes	yes	yes	yes
Throw ball		no	no	yes	no	yes	yes	usually	yes	yes	no	sometimes	yes	yes
catching		no	no	yes	no	yes	no	sometimes	yes	yes	no	sometimes	yes	yes
bicycle		no	no	no	no	yes	no	yes	no	yes	no	no	no	yes
jumps/hops		no	no	yes	no	yes	no	usually	no	yes	no	sometimes	no	yes
stairs		no	no	yes	no	yes	no	yes	no	yes	no	no	up, not down	yes
run		no	no	yes	no	yes	no	yes	no	not smoothly	no	no	not smoothly	not smoothly
Fine														
pick up small objects	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	sometimes	yes	yes
turns pages		yes	yes	yes	no	yes	no	yes	yes	yes	no	yes	yes	yes
stacks blocks		no	no	yes	no	yes	no	yes	no	yes	no	sometimes	yes	yes
Holds pencil		no	no	yes	no	yes	no	yes	no	yes	no	no	no	yes
draws identifiable forms				no	no	yes	no	yes	no	no	no	no	no	yes
uses scissors				no	no	yes	no	sometimes	no	no	no	no	no	yes
cuts out shapes				no	no	yes	no	no	no	no	no	no	no	no

Table 24: Movement ABC data for STXBP1 participants only.

ID		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Age (*yrs*)		1.08	1.79	1.79	5.60	6.69	10.52	12.78	13.00	14.29	14.65	15.04	15.12	16.50	17.74
Band 1: 3-6yrs		NA	NA	NA	NAT	NA		NAT		NA		NA			NAT
Manual Dexterity	Posting coins						yes						With difficulty	no	
	Threading beads						yes						With difficulty	no	
	Drawing						yes						no	no	
Aiming & Catching	Catching beanbag						yes						no	No	
	Throwing beanbag						yes						yes	yes	
Balance	one-leg balance						yes						no	poor balance	
	walking heels raised						no						no		
	jumping on mats						yes						no	yes	
Band 2: 7-10yrs															
Manual Dexterity	Placing pegs														
	Threading lace								With difficulty		With difficulty				
	Drawing										With difficulty				
Aiming & Catching	Catching beanbag (two hands)								no		With difficulty				
	Throwing beanbag (onto a mat)								yes		With difficulty				
Balance	one-leg balance														
	walking heel-to-toe forwards								With difficulty						

	jumping on mats			
	Hopping on mats			No
Band 3: 11-16yrs				
Manual Dexterity	Turning pegs	With difficulty	With difficulty	
	Making a triangle with nuts and bolts			
	Drawing	yes		
Aiming & Catching	Catching beanbag (one hand)			
	Throwing beanbag (at wall target)			
	Balance			
	two board balance	yes (3-5 secs)		
	walking heel-to-toe backwards			
	zig zag hopping	no		

Note: NA – not applicable; NAT – measure not attempted with participant

CHAPTER 7 GENERAL DISCUSSION

The overarching aim of this thesis was to investigate the sub- and higher-order cognitive processes associated with children's problem-solving abilities. The main findings and conclusions of the empirical work presented in this thesis are discussed in Section 7.1. The theoretical, methodological, and practical implications of the findings are outlined in Section 7.2. The limitations and potential future lines of enquiry are discussed in Section 7.3. Finally, a concluding statement is presented in Section 7.4.

7.1 SUMMARY OF RESULTS

7.1.1 COGNITIVE SEGMENTATION IN CHILDREN

In this section I discusses the experimental work presented in *Chapters 2*, *4*, and *5* that investigates the contribution of cognitive segmentation, alongside other cognitive processes, to children's fluid intelligence.

Given the power of fluid intelligence to predict later academic and life outcomes (Batty & Deary, 2004; Chen et al., 2017; Deary, 2012; Duncan et al., 2020; Postlethwaite, 2011; Sternberg et al., 2001; Zammit et al., 2004), it is no surprise that scientists are keen to understand the processes underlying it. Duncan et al.'s (2017) study demonstrated that cognitive segmentation limits fluid intelligence in adults when working memory, processing speed, and integration demands are reduced, and that adults with low IQ benefitted more than adults with high IQ when the need to segment complex reasoning tasks was removed. These findings suggest that adults with high IQ spontaneously segment, while adults with low IQ are less able to apply the principles of cognitive segmentation to complex problem-solving.

The broad aim of the work presented in *Chapters 2-5* was to test whether cognitive segmentation was important for children's problem-solving, and to explore whether it contributed anything unique to reasoning skills above and beyond working memory and processing speed. This involved the development of a child-appropriate measure of cognitive segmentation (*Chapter 2*), which was based on the task for adults, designed by Duncan et al. (2017). There were several factors that had to be considered in the development process: (1) the geometric shapes used as features needed to be recognisable to young children (Aslan & Arnas, 2007); (2) 1-feature items had to be included to check the ability to reason relationally (3) 2-feature items had to be included to examine the effects of increasing complexity; and (4) each feature was assigned a specific rule type so the effects of different transformations on performance could be explored. Piloting was conducted to ensure the stimuli were appropriate and that children understood the task instructions and format, across all levels of complexity. The outcome of this work was the creation of a child-friendly measure of cognitive segmentation.

This task was used in subsequent chapters to assess both the role of cognitive segmentation in children's fluid reasoning skills (*Chapter 4*), and its contribution to fluid intelligence alongside working memory and processing speed (*Chapter 5*). *Chapter 4* presents the first replication of Duncan et al.'s (2017) study with children. Like adults, performance on both formats of the Cognitive Segmentation task, Combined and Separated, was related to fluid intelligence. Performance was better in the condition in which the problems were separated into component parts, which is consistent with previous literature (Duncan et al., 2017; Glady et al., 2017, 2016). Unlike adults, children with lower fluid intelligence did not benefit more in the separated condition than children with higher fluid intelligence. One possible explanation is that all six- to 10-year-old children perform like adults with lower fluid intelligence and struggle to spontaneously segment problems. In contrast, adults with higher IQ might spontaneously segment, resulting in fewer benefits of segmentation.

Older children performed better than younger children in both conditions which is in line with previous findings demonstrating that fluid reasoning abilities increase with age (Christie & Gentner, 2014a; Ferrer et al., 2009; Mcardle et al., 2002; Richland et al., 2006; Starr et al., 2018; Sternberg & Rifkin, 1979; Stevenson et al., 2016). There was no evidence that breaking problems down was more beneficial for younger than older children, which is surprising. As discussed previously the age range included in this study may be too narrow or too old and therefore missed the more pronounced development in analogical reasoning abilities that occurs before six years (Gentner & Ratterman, 1991; Goswami, 1992). It may also be the case that potential age-related increases in the benefits of segmentation (made possible by development of analogical reasoning ability) counteracted concurrent development of spontaneous segmentation abilities (i.e., resulting in a decreased benefit with age). Longitudinal paradigms, and the inclusion of eye-tracking measures, might be more sensitive to uncover subtle age effects and advances in fluid reasoning abilities. Longitudinal data collection was planned for this thesis. However, schools closed across Ireland in response to the COVID-19 pandemic, which resulted in the cancellation of longitudinal data collection.

There is a growing body of literature demonstrating that different factors within analogical reasoning tasks influence problem-solving processes. Therefore, a secondary objective of *Chapter 4* was to explore the effects on performance of other within-task characteristics, such as complexity and rule types. As expected, performance decreased with increasing complexity in the Combined condition. Surprisingly, performance increased between 2- and 3-features in the separated condition. It is possible that this reflected practice effects, which were not detectable in the Combined condition due to the added difficulty of increasing the number of features. This might be investigated in future by randomly intermixing items of different complexity. In addition to complexity, the types of rules governing features had a differential effect in both conditions. Previous research has consistently demonstrated that spatial features present the greatest difficulty to adults and children (Bethell-Fox et al., 1984; Blum et al., 2016; Green & Kluever, 1992; Kirchhoff & Holling, n.d.; Whitely & Schneider, 1981). Although findings from the Separated condition supported this, the Combined condition performance did not. Features transformed by 'Design' (e.g., change of shape or design) gave rise to the greatest number of errors in the Combined condition. I suggest that there may be something within the layout of the Combined condition that masks features governed by Design transformations, whereas the layout of the Separated condition supports children more with solving features with Design Rules.

Constructed response paradigms provide a window into why people fail to solve analogical reasoning problems (Stevenson et al., 2016). Both task formats required children to construct responses, and the errors children made were categorised into one of five error types: Wrong Alternative, Omission of Part, Other Drawing Errors, Copying the C-term, and Item Not Attempted. Most errors in both the Combined and Separated conditions were Wrong Alternatives i.e., drawing the alternative feature that is presented in the matrix but is not the correct response feature. This is consistent with Duncan et al.'s study (2017), and other work with adults showing that the most common error made is almost exclusively partially correct solution (Raven, Court, & Raven, 1983; Siegler & Svetina, 2002; Vodegel Matzen et al., 1994).

Duncan et al. (2017) did not include the latter two error types: Copying the C-Term and Item Not Attempted. This is presumably because they are more characteristic of the errrors made by young children on reasoning tasks (Inhelder & Piaget, 1964; Siegler & Svetina, 2002; Stevenson et al., 2016; Stevenson & Hickendorff, 2018). Interestingly, lower ability children were relatively more likely to commit Copying the C-Term errors. This supports previous research showing that less advanced reasoners typically focus their matrix search on the C-term and do not attend to the A:B section (Glady et al., 2016; Starr et al., 2018; Thibaut & French, 2011; Vendetti et al., 2017), and it may be why young children commit more duplications of items from other cells of the matrix (Inhelder & Piaget, 1964; Siegler & Svetina, 2002; Stevenson et al., 2016; Stevenson & Hickendorff, 2018).

Individual differences in working memory (Colom et al., 2008; Engle, 2010; Kyllonen & Christal, 1990) and processing speed (Jungeblut et al., 2020; Salthouse, 1996) are strongly related to individual differences in fluid intelligence. *Chapter 5* presents the first attempt to examine the combined and independent contributions of cognitive segmentation, working memory, and processing speed to fluid intelligence. Cognitive segmentation was not significantly associated with working memory or processing speed yet predicted unique variance in fluid intelligence. This finding provides further support to Duncan and colleagues' (2005; 2013; 2017; 2020) theory that the ability to isolate and examine component parts of a problem constrains fluid intelligence abilities. Consistent with a large body of literature, the association between working memory and fluid intelligence was replicated (e.g., Alloway et al., 2004; Bayliss et al., 2003, 2005; Engel de Abreu et al., 2010; Fry & Hale, 2000; Giofrè et al., 2013; Hornung et al., 2011; Swanson, 2008). However, no direct links were found between processing speed and fluid intelligence. One interpretation is that neither the working memory nor fluid intelligence tasks were timed. A recent study showed that using timed tasks increases the relationship between working memory and reasoning (Chuderski, 2013). It is possible that the lack of time constraints also diminished the relationship between processing speed and fluid intelligence. An additional interpretation is that the indices of speed used in the current study were based on reaction times (RTs) derived from lower-level conditions of more complex cognitive tasks (cancellation and go/no-go), and as such may have been influenced by additional cognitive processes. Processing speed is typically measured by simple or choice reaction time tasks that minimise decision-making. Therefore, using more complex tasks may have imposed perceptual or executive demands that masked the association between processing speed and both fluid intelligence and working memory. Consistent with this, Schubert et al. (2020) demonstrated that behaviourally measured RTs, even those that are not embedded in more complex tasks, encapsulate several processes, including the encoding of information, decision-making, and motor execution, making it difficult to identify which components of the stream of processing accurately account for the relationship between processing speed and fluid intelligence (Frischkorn et al., 2019b). Additional methods for assessing processing speed would thus be useful in future studies to further explore its relative contribution to fluid intelligence alongside cognitive segmentation and working memory. For example, using (1) simple reaction time tasks, (2) parameters from ExGaussian distributions of reaction times on executively demanding tasks or (3) time course data from EEG recordings, would enable behavioural and neural processing speed data to be decomposed into distinct stages of information processing (e.g., Schubert & Frischkorn, 2020).

7.1.2 Medical, Cognitive and Behavioural Phenotyping of Children with *STXBP1*-related Intellectual Disability

In this section I discuss the final empirical chapter, which was a stand-alone phenotyping study of individuals with STXBP1; a rare genetic mutation causing intellectual disability. This study was conducted in the first year of my PhD. The observations and findings in this study inspired me to try to understand why children struggle to complete complex cognitive tasks, which led into the experimental work presented in *Chapters 2-5*. I had planned to return to individuals with ID at the end of the experimental work to explore their cognitive segmentation skills. However, all testing was cut short due to the COVID-19 pandemic.

The *STXBP1* group was characterised by communication impairments, particularly in receptive language, and elevated levels of hyperactivity-impulsivity. The movement abilities in the group were highly varied, ranging from difficulty sitting unsupported and picking up small items, to running smoothly and drawing recognisable forms. Like previous findings, intellectual disability ranged from moderate to severe and epilepsy prevalence was high, but not pervasive (e.g., Barcia et al., 2014; Gburek-Augustat et al., 2016; Stamberger et al., 2016; Suri et al., 2017). The group were anxious but had preserved social motivation, consistent with STXBP1 mouse models (Kovačević et al., 2018), and other research into the diverse social developmental trajectories contributing to autism-like impairments in individuals with genomic disorders (Mabb et al., 2011; Orock et al., 2018; Richards et al., 2017). To examine whether the behavioural characteristics within the STXBP1 group differed from developmental expectations, I compared them to a matched-ability group with other monogenic causes of ID. Receptive language ability and socialisation skills (but not hyperactivity) discriminated the STXBP1 from the ID comparison groups. Overall, the profile of the STXBP1 group was very similar to the phenotype associated with Angelman Syndrome (Bakke et al., 2018; Horsler & Oliver, 2006; Huang et al., 2013; Mabb et al., 2011; Pelc et al., 2008). A network of brain regions including the amygdala, ventral striatum, orbital and ventromedial regions of the prefrontal cortex are thought to underlie the increased social motivation in individuals with Angelman Syndrome. Although *STXBP1* is expressed ubiquitously in the brain, examining this network of brain regions may provide a starting point for understanding the neural mechanisms driving the atypical social profile in *STXBP1*.

In terms of cognitive abilities, performance was at floor on standardised measures of IQ for the STXBP1 group. This was not entirely surprising given that ID standardisation samples do not typically include individuals with ID, which limits the sensitivity of measures to detect differences in performance at the lower end. A new battery of tasks called the FarmApp (Brkić et al., 2022), was used to capture cognitive abilities: inhibition, visuo-spatial working memory, and recognition. The findings from this battery are only descriptive but there were higher completion rates with the FarmApp than the standardised tasks, which demonstrates that if tasks are intuitive, and developmentally appropriate, children can engage with them and provide valuable cognitive data. The range of scores in each task suggests that there is variability in low level cognitive abilities that can be captured with age and stage appropriate measures.

7.2 IMPLICATIONS

7.2.1 THEORETICAL IMPLICATIONS

The work included in this thesis has enhanced our understanding of the factors that underlie fluid intelligence, most importantly providing compelling evidence in *Chapters 4* and *5* in support of Duncan's (2013) theory that the ability to separate a complex problem into component parts is vital for the successful completion of fluid intelligence tasks. My results show that this principle generalises to children and is observed across different levels of complexity and rule type. Demonstrating that it predicts unique variance in fluid intelligence, over and above working memory and processing speed, is the first suggestion that it is functionally distinct from these other cognitive processes that are already known to constrain fluid intelligence.

The results in *Chapter 5* also add to the large body of literature demonstrating a strong link between working memory and fluid intelligence. The two tasks used to derive the working memory component are traditionally seen as measures of the verbal and visuo-spatial short-term storage aspects of working memory in children (e.g., Alloway et al., 2008; Bayliss et al., 2003; Gathercole et al., 2004; Gathercole & Pickering, 2000), although both tasks are attentionally and executively demanding in younger children (e.g., Alloway et al., 2006; Cowan et al., 2005). It is therefore not possible to delineate between different theories of the association between fluid reasoning and working memory based on the current data. Ultimately, future research should expand on the number of tasks used in this study to extend what we know about the roles of working memory, processing speed, and cognitive segmentation in problem-solving.

In *Chapter 6*, the observation that the symptom profile of STXBP1 disorders has much in common with the better-known Angelman Syndrome suggests that they may share overlapping trans-diagnostic mechanisms. Identifying common cellular and neural mechanisms may then provide a starting point for understanding the progression of STXBP1 disorders and designing potential support strategies and interventions.

7.2.2 Methodological Implications

Within this thesis, an important methodological advancement was made in the form of a new measure of cognitive segmentation that is suitable for young children. This task was developed and used successfully with a large community sample of children to provide novel insights into children's problem-solving skills. The exploration of problem-solving in *Chapter 4* suggests that the 2-feature condition is more sensitive to measuring individual differences in analogical reasoning in children, while the 3-feature condition is more sensitive at capturing the effect of cognitive segmentation. It may be that the 2-feature Separated condition could be further developed and used with even younger children or less-able children (perhaps those with ID) to capture variance in problem-solving skills at the lower end of the spectrum. This would of course involve extensive work in the future.

In *Chapter 6*, I found that all participants with *STXBP1* were at floor on traditional cognitive assessments, although a minority were able to successfully engage with the FarmApp, which was designed to measure cognition more sensitively in low ability groups. This suggests that further work to make such tests even more accessible for this patient group should be possible and could then provide valuable insights. Further development of the tasks might be aided by incorporating my findings on which abilities tend to be spared and impaired in this group. In particular, tasks that minimise motor requirements while maximising social motivation may be beneficial.

7.2.3 Practical Implications

This thesis demonstrates that breaking problems down into small, simple parts can improve complex task performance for children of all ability levels. This has important implications for the classroom where children need to be able to decompose multi-step instructions and break individual learning tasks into their component parts for classroom success (e.g., Jaroslawska, Gathercole, Allen & Holmes, 2016). The evidence from this thesis, which shows children of all abilities aged 6-10 years are likely to struggle to break down problems by themselves, suggests that teachers should be segmenting complex cognitive tasks into smaller parts to aide classroom learning and task success. It also suggests there might be benefits to teaching children how to segment. That said, the data presented in Chapter 4 shows that practice with segmenting problems in the Separated condition did not aide performance in the Combined condition. This suggests short-term instruction or practice with segmentation is unlikely to be sufficient for children to be able to decompose problems themselves without external support.

Previous research has shown that children's performance improves throughout an analogy task when they are provided with item-by-item feedback (Stevenson et al., 2016; Stevenson & Hickendorff, 2018). Moreover, children who received feedback and independently constructed response items, as opposed to selecting a response option, showed greater understanding of the problem-solving process (Stevenson et al., 2016). Taken together, these findings suggest that breaking problems down into parts, allowing children to independently come up with a solution, before providing detailed feedback on the solution process, may be an effective strategy to scaffold children's ability to segment.

Future studies should investigate whether segmentation and feedback have independent or interacting effects on performance. In the classroom, children who struggle with segmentation may (1) ask for the steps of a task to be repeated, (2) may miss steps of a task and fail to complete the task successfully, or (3) may have trouble organising activities because they have difficulty with isolating the subparts of an activity. To aid these children, educators could leave task detailed instructions on the board for longer periods, check-in with children to see if they are keeping track of the steps they have completed and create crib sheets for children to have beside them. As noted in Chapter 4 (pg. 91) there was a group of children who still struggled to complete problems in the separated format. For these children, asking them to begin at step 1 of a task and cover the rest of the steps, only revealing the next step when the previous one is completed, may help them to attend to one step at a time without the need to inhibit potentially distracting information.

Identifying the strengths and difficulties that discriminate STXBP1 from individuals with other monogenic causes of ID in Chapter 6 aides our understanding of the behavioural characteristics associated with specific mutations and can help in several ways. First, it provides a guide about where to focus clinical assessment before and after diagnosis. Second, it provides information to help clinicians manage the genetic diagnosis. Finally, and perhaps most importantly, it provides parents and carers with information to help support individuals with STXBP1 by providing information that is vital to their understanding of what their child is likely to experience or be like. The key information that should be conveyed to parents and carers to help their understanding of the condition is that while there are some consistent characteristics across individuals, the severity of these traits varies widely between individuals. For example, while there is a consistent profile of preserved social behaviour, with individuals enjoying social interactions and activities, some children also have a limited understanding of social expectations and difficulties integrating into normal social settings. Similarly, the prevalence of epilepsy is high, but the type, severity, age of onset, and remission of seizures varies across individuals. Other key pieces of information to convey are that many individuals with STXBP1 have communication impairments, particularly with receptive language, many have elevated levels of hyperactivity-impulsivity, and many are troubled by anxiety. As such, parents and carers may wish to seek interventions targeting these particular characteristics, but they will need to be mindful that each child is different and that some existing interventions may not be suitable for their children due to the complexity of their overlapping areas of need.

7.3 Limitations and Future Directions

A limitation of the research exploring segmentation in children is that the data were cross-sectional and limited to children aged six to 10 years. There are differences in the segmentation abilities of adults with low and high IQ (Duncan et al., 2017), yet my findings suggest that cognitive segmentation ability may still be developing in children of all ability levels. It is likely that cognitive segmentation skills continue to develop throughout childhood, given that working memory, processing speed, and fluid intelligence become more advanced from late childhood to adulthood. Therefore, longitudinal data are needed to accurately pinpoint and track the emergence of cognitive segmentation abilities throughout the lifespan. This type of data would uncover whether cognitive segmentation develops to a static point or whether it fluctuates dynamically across the lifespan. Including additional cognitive measures would also be important to reveal whether changes in cognitive segmentation interact with other cognitive processes across development.

The work presented in this thesis demonstrates the importance of cognitive segmentation for solving matrix reasoning problems in childhood. This replicates and extends Duncan et al.'s (2017) work showing that adults also use cognitive segmentation to solve matrix reasoning problems. While these studies suggest cognitive segmentation is likely important for fluid intelligence, it is to demonstrate its role in other nonverbal reasoning tasks. While fluid reasoning is often measured using tasks that require participants to identify abstract relations between elements and draw analogies to complete a pattern, as in matrix reasoning tasks, other measures such as block design and A:B:C:D proportional reasoning tasks are also widely used. To explore whether segmentation is important for solving all complex problems, or whether there are specific circumstances under which it is required, future research should explore whether it plays a role in performance on other intelligence tests.

Related to this point, the choice of working memory and processing speed tasks limits the theoretical insights that can be drawn. The task choice was constrained by the assessments available in the RED app that was used for task administration. Including additional working memory tasks that specifically tapped the executive aspects of working memory would have helped me delineate between different accounts of the relationship between working memory and fluid reasoning. For example, previous research has shown that measures of attentional control and binding contribute to performance on fluid reasoning tasks, so including these alongside the measures I used would extend our knowledge. The measures of processing speed included in my work did not predict fluid intelligence, which is at odds with a large body of literature. Using additional methods (e.g., time course data from EEG recordings; Schubert & Frischkorn, 2020) or simple reaction time tasks in future work may isolate a 'purer' measure, and yield different results. Therefore, future research should extend the battery of tasks used in this study, to derive more comprehensive constructs reflecting working memory and processing speed.

As part of the planned longitudinal follow-up, had COVID not intervened, I had hoped to also acquire measures of academic performance, including maths and language attainment. Given this work has shown that cognitive segmentation is important for fluid reasoning in childhood, and because fluid intelligence predicts academic performance and later achievement (e.g., Giofre et al., 2013; Marcela Bizama et al., 2019), it is likely that the principles of cognitive segmentation also apply to academic attainment. Two key academic skills, reading comprehension and mathematical problem-solving, involve multiple processes that must be decomposed. For reading comprehension, children must be able to isolate and attend to relevant pieces of text of derive meaning. For mathematical problem-solving children must decompose the steps of a problem and solve each part in turn to reach the solution. It is important for future research to explore experimentally whether cognitive segmentation is important for breaking tasks down in these contexts as it will provide both insight into reasons why children might be struggling at school and guidance for intervention for teachers (see Practical Implications for more on interventions).

Other limitations relate directly to the Cognitive Segmentation task. This is the first time the task has been used on a large scale. Although it was piloted prior to testing, future research would be useful to validate it as a measure of cognitive segmentation. In addition, only behavioural data were collected. While using a constructed response paradigm did provide some level of insight into the problem-solving behaviours of young children, the application of other methods such as eye-tracking could provide further information about the allocation of attention and problem-solving strategies. Indeed, previous research using eye-tracking has demonstrated that children perform poorly because they use variable, less effective strategies when solving analogies (Glady, French, & Thibaut, 2016, 2017; Thibaut & French, 2016).

The phenotyping study in *Chapter 6* is most notably limited by the small sample size that encompasses a wide age range, and the large number of measures. For this reason, the results of the statistical analyses should be interpreted with caution, particularly the non-parametric group comparisons. In addition, the data were collected via parent report, and it was not feasible to verify the data via review of medical records. However, the use of parent questionnaires to derive the STXBP1 phenotype should mitigate against: (1) the fact that I was not blind to the genetic diagnosis and (2) potentially biased diagnosis-informed clinical ratings of symptoms and their severity. Future research into the phenotypes of individuals with rare *de novo* mutations associated with low ability could use Cognitive Segmentation tasks to both measure variation in performance and provide valuable information about how specific genes contribute to the development of the cognitive processes that scaffold learning and more complex cognition.

7.4 CONCLUSION

In everyday life we are presented with an overwhelming number of novel tasks involving multiple parts that must be solved. These are particularly challenging for younger children and for individuals with intellectual disabilities, although disability caused by *de novo* genetic mutations can produce wide variability in performance as demonstrated in the phenotyping of *STXBP1* individuals. In this thesis I have demonstrated that one reason why individuals might struggle with complex tasks is because they lack the ability to separate problems into their component parts. Helping children, adults with low IQ, and individuals with intellectual disabilities to break problems down might provide one practical way to improve their functioning outcomes. Future studies evaluating segmentation-based interventions in clinics, schools and workplaces will be vital to this endeavour.

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual Differences in Working Memory Within a Nomological Network of Cognitive and Perceptual Speed Abilities. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.131.4.567

Allen, A. S., Berkovic, S. F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E. E., Epstein, M. P., Glauser, T., Goldstein, D. B., Han, Y., Heinzen, E. L., Hitomi, Y., Howell, K. B., Johnson, M. R., Kuzniecky, R., Lowenstein, D. H., Lu, Y. F., Madou, M. R. Z., Marson, A. G., … Winawer, M. R. (2013). De novo mutations in epileptic encephalopathies. *Nature*, *501*(7466), 217–221. https://doi.org/10.1038/nature12439

Alloway, T., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the automated working memory assessment. *Educational Psychology*, *28*(7), 725–734. https://doi.org/10.1080/01443410802243828

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006a). Verbal and Visuospatial Short-Term and Working Memory in Children: Are They Separable? *Child Development*, *77*(6), 1698–1716.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006b). Verbal and Visuospatial Short-Term and Working Memory in Children: Are They Separable? *Child Development*, *77*(6), 1698–1716. https://doi.org/10.1111/J.1467-8624.2006.00968.X

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, *87*(2), 85–106. https://doi.org/10.1016/j.jecp.2003.10.002

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

Aslan, D., & Arnas, Y. A. (2007). International Journal of Early Years Education Three-to six-year-old children's recognition of geometric shapes. *International Journal of Early Years Education*, *15*(1), 83–104. https://doi.org/10.1080/09669760601106646

Assem, M., Shashidhara, S., Glasser, M. F., & Duncan, J. (2021). Precise topology of adjacent domain-general and sensory-biased regions in the human brain. *BioRxiv*, 2021.02.21.431622. https://doi.org/10.1101/2021.02.21.431622

Baddeley, AD. (2000). Short-term and working memory. In E. Tulving, & FIM. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 77 - 99). Oxford University Press, New York.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In Psychology of learning andmotivation. In: Bower, G.H. (Ed.). *The Psychology of Learning and Motivation:*

Advances in Research and Theory (pp. 47-89). New York: Academic Press.

Bakke, K. A., Howlin, P., Retterstøl, L., Kanavin, I. J., Heiberg, A., & Nærland, T. (2018). Effect of epilepsy on autism symptoms in Angelman syndrome. *Molecular Autism*, *9*(1), 1–8. https://doi.org/10.1186/s13229-017-0185-1

Barcia, G., Chemaly, N., Gobin, S., Milh, M., Van Bogaert, P., Barnerias, C., Kaminska, A., Dulac, O., Desguerre, I., Cormier, V., Boddaert, N., & Nabbout, R. (2014). Early epileptic encephalopathies associated with STXBP1 mutations: Could we better delineate the phenotype? *European Journal of Medical Genetics*, *57*(1), 15–20. https://doi.org/10.1016/j.ejmg.2013.10.006

Batty, G. D., & Deary, I. J. (2004). Early life intelligence and adult health. *British Medical Journal*, *329*(7466), 585–586. https://doi.org/10.1136/bmj.329.7466.585

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The Complexities of Complex Span: Explaining Individual Differences in Working Memory in Children and Adults. *Journal of Experimental Psychology: General*, *132*(1), 71–92. https://doi.org/10.1037/0096-3445.132.1.71

Bayliss, D. M., Jarrold, C., Gunn, D. M., Baddeley, A. D., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology*, *41*(4), 579–597. https://doi.org/10.1037/0012-1649.41.4.579

Berrios, J., Stamatakis, A. M., Kantak, P. A., Mcelligott, Z. A., Judson, M. C., Aita, M., Rougie, M., Stuber, G. D., & Philpot, B. D. (2016). Loss of UBE3A from TH-expressing neurons suppresses GABA co-release and enhances VTA-NAc optical self-stimulation. *Nature Communications*, *7*, 1–8. https://doi.org/10.1038/ncomms10702

Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence*, *8*(3), 205–238. https://doi.org/10.1016/0160-2896(84)90009-6

Bignardi, G., Dalmaijer, E. S., Anwyl-Irvine, A., & Astle, D. E. (2020). Collecting big data with small screens: Group tests of children's cognition with touchscreen tablets are reliable and valid. *Behavior Research Methods*, 1–15. https://doi.org/10.3758/s13428-020-01503-3

Birenbaum, M., & Tatsuoka, K. K. (1987). Open-ended versus multiple-choise response formats - it does make a difference for diagnostic purposes. *Applied Psychological Measurement*, *11*(4), 385–395.

Blum, D., Holling, H., Galibert, M. S., & Forthmann, B. (2016). Task difficulty prediction of figural analogies. *Intelligence*, *56*, 72–81. https://doi.org/10.1016/j.intell.2016.03.001

Brose, N., Planck, M., O'connor, V., & Skehel, P. (2010). Synaptopathies: Dysfunction of Synaptic Function? *Biochemical Society*, *38*, 443–444.

https://doi.org/10.1042/BST0380443

Brkić, D., Ng-Cordell, E., O'Brien, S., Martin, J., Scerif, G., Astle, D., & Baker, K. (2022). FarmApp: a new assessment of cognitive control and memory for children and young people with neurodevelopmental difficulties. *Child Neuropsychology*, 1-19.

Buehner, M., Krumm, S., & Ziegler, M. (2005). Are Working Memory and Attention Related Constructs? Situational Perception in Situational Judgment Tests: The Role of Context in Situational Judgment Tests View project S^2-Pan [dropout phenomena in higher education with special regard to primary (science) education] View project. *Article in Journal of Individual Differences*. https://doi.org/10.1027/1614-0001.26.3.121

Cardoso, A. R., Lopes-Marques, M., Silva, R. M., Serrano, C., Amorim, A., Prata, M. J., & Azevedo, L. (2019). Essential genetic findings in neurodevelopmental disorders. *Human Genomics*, *13*(31), 1–7. https://doi.org/10.1186/s40246-019-0216-4

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

Carvill, G. L., Weckhuysen, S., McMahon, J. M., Hartmann, C., Møller, R. S., Hjalgrim, H., Cook, J., Geraghty, E., O'Roak, B. J., Petrou, S., Clarke, A., Gill, D., Sadleir, L. G., Muhle, H., Von Spiczak, S., Nikanorova, M., Hodgson, B. L., Gazina, E. V., Suls, A., … Mefford, H. C. (2014). GABRA1 and STXBP1: Novel genetic causes of Dravet syndrome. *Neurology*, *82*(14), 1245–1253. https://doi.org/10.1212/WNL.0000000000000291

Cattell, R. B. (1940). A culture-free intelligence test. I. *Journal of Educational Psychology, 31*(3), 161–179.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton-Mifflin.

Chen, X., Hertzog, C., & Park, D. C. (2017a). Cognitive Predictors of Everyday Problem Solving across the Lifespan. *Gerontology*, *63*, 372–384. https://doi.org/10.1159/000459622

Chen, X., Hertzog, C., & Park, D. C. (2017b). Cognitive Predictors of Everyday Problem Solving across the Lifespan. *Gerontology*, *63*, 372–384. https://doi.org/10.1159/000459622

Chen, Z., Honomichl, R., Kennedy, D., & Tan, E. (2016). Aiming to Complete the Matrix: Eye-Movement Analysis of Processing Strategies in Children's Relational Thinking. *Developmental Psychology*, *52*(6), 867–878. https://doi.org/10.1037/dev0000113

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239.

https://doi.org/10.1016/j.tics.2012.02.007

Chiurazzi, P., & Pirozzi, F. (2016). Advances in understanding - genetic basis of intellectual disability. *F1000Research*, *5*(0), 1–16. https://doi.org/10.12688/f1000research.7134.1

Christie, S., & Gentner, D. (2014a). Language helps children succeed on a classic analogy task. *Cognitive Science*, *38*(2), 383–397. https://doi.org/10.1111/cogs.12099

Christie, S., & Gentner, D. (2014b). Language helps children succeed on a classic analogy task. *Cognitive Science*, *38*(2), 383–397. https://doi.org/10.1111/cogs.12099

Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence*, *41*(4), 244–262. https://doi.org/10.1016/j.intell.2013.04.003

Chuderski, A. (2019). Even a single trivial binding of information is critical for fluid intelligence. *Intelligence*, *77*, 1–9. https://doi.org/10.1016/j.intell.2019.101396

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008a). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*(6), 584–606. https://doi.org/10.1016/j.intell.2008.01.002

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008b). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*(6), 584–606. https://doi.org/10.1016/j.intell.2008.01.002

Colom, R., Abad, F. J., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence*, *33*(6), 623–642. https://doi.org/10.1016/j.intell.2005.05.006

Colom, R., Flores-Mendoza, C., Quiroga, M. Á., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences*, *39*(5), 1005–1014. https://doi.org/10.1016/j.paid.2005.03.020

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence*, *32*, 277–296. https://doi.org/10.1016/j.intell.2003.12.002

Colom, R., Rubio, V. J., Shih, P. C., & Santacreu, J. (2006). Fluid intelligence, working memory and executive functioning. *Psicothema*, *18*(4), 816–821.

Conners, C. K. (2008). *Conners 3*. North Tonawanda, NJ: MHS.

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale second edition (SRS-2): Manual*. Western psychological services (WPS).

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin and Review*, *8*(2), 331–335. https://doi.org/10.3758/BF03196169

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183.

Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (2007). Variation in working memory: An introduction. In A. R. A Conway, C. Jarrold, M. J. Kane (Eds.) & A. Miyake & J. N. Towse (Ed.), *Variation in working memory* (pp. 3–17). Oxford University Press.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Psychology*.

Cowan, N. (1995). *Attention and memory: an integrated framework*. New York: Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. (pp. 62–101). Cambridge, UK: Cambridge University Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9

Cowan, N., Elliott, E. M., Saults, S. J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Coyle, T. R., Pillow, D. R., Snyder, A. C., & Kochunov, P. (2011). Processing Speed Mediates the Development of General Intelligence (g) in Adolescence. *Psychological Science*, *22*(10), 1265–1269. https://doi.org/10.1177/0956797611418243

Crone, E. A., Wendelken, C., Van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, *12*(1), 55–66. https://doi.org/10.1111/j.1467-7687.2008.00743.x

Dalmaijer, E. S., Gibbons, S. G., Bignardi, G., Anwyl-Irvine, A. L., Siugzdaite, R., Smith, T. A., Uh, S., Johnson, A., & Astle, D. E. (2021). Direct and indirect links between children's socio-economic status and education: pathways via mental health, attitude, and cognition. *Current Psychology*, 1–15. https://doi.org/10.1007/s12144-021-02232-2

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, *19*, 450–466. https://doi.org/10.4324/9781315879840-7

Dang, C. P., Braeken, J., Colom, R., Ferrer, E., & Liu, C. (2015). Do processing speed and short-term storage exhaust the relation between working memory capacity and intelligence? *Personality and Individual Differences*, *74*, 241–247. https://doi.org/10.1016/j.paid.2014.10.012

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*, 453–482. https://doi.org/10.1146/annurev-psych-120710-100353

Deary, I. J., Johnson, W., & Starr, J. M. (2010). Are Processing Speed Tasks Biomarkers of Cognitive Aging? *Psychology and Aging*, *25*(1), 219–228. https://doi.org/10.1037/a0017750

Deciphering Developmental Disorders Study. (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature Publishing Group*, *542*(7642), 433. https://doi.org/10.1038/nature21062

Demetriou, A., Spanoudis, G., Shayer, M., Mouyi, A., Kazi, S., & Platsidou, M. (2012). Cycles in speed-working memory-G relations: Towards a developmental-differential theory of the mind ☆. *Intelligence*, *41*, 34–50. https://doi.org/10.1016/j.intell.2012.10.010

Demetriou, A., Spanoudis, G., Shayer, M., Mouyi, A., Kazi, S., & Platsidou, M. (2013). Cycles in speed-working memory-G relations: Towards a developmental-differential theory of the mind. *Intelligence*, *41*, 34–50. https://doi.org/10.1016/j.intell.2012.10.010

Denaes, C. (2012). Analogical Matrices in Young Children and Students with Intellectual Disability: Reasoning by Analogy or Reasoning by Association? *Journal of Applied Research in Intellectual Disabilities*, *25*(3), 271–281. https://doi.org/10.1111/j.1468-3148.2011.00665.x

Deprez, L., Weckhuysen, S., Holmgren, P., Suls, A., Van Dyck, T., Goossens, D., Del-Favero, J., Jansen, A., Verhaert, K., Lagae, L., Jordanova, A., Van Coster, R., Yendle, S., Berkovic, S. F., Scheffer, I., Ceulemans, B., & De Jonghe, P. (2010). Clinical spectrum of early-onset epileptic encephalopathies associated with STXBP1 mutations. *Neurology*, *75*(13), 1159–1165. https://doi.org/10.1212/WNL.0b013e3181f4d7bf

Detterman, D. K. (2002). General intelligence: Cognitive and biological explanations. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 223–243). Lawrence Erlbaum Associates Publishers.

Di Meglio, C., Lesca, G., Villeneuve, N., Lacoste, C., Abidi, A., Cacciagli, P., Altuzarra, C., Roubertie, A., Afenjar, A., Renaldo-Robin, F., Isidor, B., Gautier, A., Husson, M., Cances, C., Metreau, J., Laroche, C., Chouchane, M., Ville, D., Marignier, S., … Milh, M. (2015). Epileptic patients with de novo STXBP1 mutations: Key clinical features based on 24 cases. *Epilepsia*, *56*(12), 1931–1940. https://doi.org/10.1111/epi.13214

Duncan, J. (2005). Frontal lobe function and general intelligence: Why it matters. *Cortex*, *41*(2), 215–217. https://doi.org/10.1016/S0010-9452(08)70896-7

Duncan, J. (2010). *How intelligence happens*. Yale University Press.

Duncan, J. (2013). The Structure of Cognition: Attentional Episodes in Mind and Brain. *Neuron*, *80*(1), 35–50. https://doi.org/10.1016/j.neuron.2013.09.015

Duncan, J., Assem, M., & Shashidhara, S. (2020). Integrated Intelligence from Distributed Brain Activity. *Trends in Cognitive Sciences*, *24*(10), 838–852. https://doi.org/10.1016/j.tics.2020.06.012

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, *33*(3), 261–268. https://doi.org/10.1016/0028-3932(94)00124-8

Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(20), 5295–5299. https://doi.org/10.1073/pnas.1621147114

Duncan, J., Phillips, E., Mitchell, D., J., C., Cooper, P., & Murray, L. (2021). Parental scaffolding during book-sharing predicts child general intelligence. 1–22.

Durkin, M. (2002). The epidemiology of developmental disabilities in low-income countries. *Mental Retardation and Developmental Disabilities Research Reviews*, *8*(3), 206–211. https://doi.org/10.1002/mrdd.10039

Egeth, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, *1*(4), 245–252. https://doi.org/10.3758/BF03207389

Einfeld, S., & Tonge, B. J. (2002). Developmental Behaviour Checklist (DBC). *Victoria*.

Embretson, S. E. (1998). A Cognitive Design System Approach to Generating Valid Tests: Application to Abstract Reasoning. *Psychological Methods*, *3*(3), 380–396. https://doi.org/10.1037/1082-989X.3.3.380

Engel de Abreu, P. M. J., Conway, A. R. A., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence*, *38*(6), 552–561. https://doi.org/10.1016/j.intell.2010.07.003

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W. (2010). Role of Working-Memory Capacity in Cognitive Control. *Current Anthropology*, *51*, 17–26. https://doi.org/10.1086/650572

Engle, R. W. (2018). Working Memory and Executive Attention: A Revisit. *Perspectives on Psychological Science*, *13*(2), 190–193. https://doi.org/10.1177/1745691617720478

Engle, R. W., & Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. *Psychology of Learning and Motivation - Advances in Research and Theory*, *44*, 145–199. https://doi.org/10.1016/S0079-7421(03)44005-X

Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working Memory, Short-Term Memory, and General Fluid Intelligence: A Latent-Variable Approach. In *Journal of Experimental Psychology: General* (Vol. 128, Issue 3).

Evans, T. G. (1968). A program for the solution of geometric analogy intelligence test questions. In M. Minsky (Ed.), *Semantic information processing* (pp. 271 - 353). Cambridge, Mass.: MIT Press.

Ferrer, E., & McArdle, J. J. (2004). An Experimental Analysis of Dynamic Hypotheses about Cognitive Abilities and Achievement from Childhood to Early Adulthood. *Developmental Psychology*, *40*(6), 935–952. https://doi.org/10.1037/0012-1649.40.6.935

Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2007). Longitudinal Models of Developmental Dynamics Between Reading and Cognition From Childhood to Adolescence. *Developmental Psychology*, *43*(6), 1460–1473. https://doi.org/10.1037/0012-1649.43.6.1460

Ferrer, E., O'Hare, E. D., & Bunge, S. A. (2009). Fluid reasoning and the developing brain. In *Frontiers in Neuroscience* (Vol. 3, Issue MAY). https://doi.org/10.3389/neuro.01.003.2009

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications.

Frischkorn, G. T., Schubert, A.-L., & Hagemann, D. (2019a). Processing speed, working memory, and executive functions: Independent or inter-related predictors of general intelligence. *Intelligence*, *75*, 95–110. https://doi.org/10.1016/j.intell.2019.05.003

Frischkorn, G. T., Schubert, A. L., & Hagemann, D. (2019b). Processing speed, working memory, and executive functions: Independent or inter-related predictors of general intelligence. *Intelligence*, *75*(April), 95–110. https://doi.org/10.1016/j.intell.2019.05.003

Fry, A. F., & Hale, S. (1996a). Processing Speed, working memory, and fluid intelligence: Evidence for a developmental cascade. In *Psychological Science* (Vol. 7, Issue 4). https://doi.org/10.1111/j.1467-9639.1993.tb00256.x

Fry, A. F., & Hale, S. (1996b). Processing speed, working memory, and fluid intelligence. *Psychological Science*, *7*(4), 237–241. papers3://publication/uuid/BCE64528-CCCE-416D-84C2-05E69A707BCE

Fry, A. F., & Hale, S. (2000a). Relationships among processing speed, working memory, and fluid intelligence in children. In *Biological Psychology* (Vol. 54). www.elsevier.com/locate/biopsycho

Fry, A. F., & Hale, S. (2000b). Relationships among processing speed, working memory, and fluid intelligence in children. In *Biological Psychology* (Vol. 54). www.elsevier.com/locate/biopsycho

Fujiura, G. T., Park, H. J., & Rutkowski-Kmitta, V. (2005). Disability statistics in the developing world: A reflection on the meanings in our numbers. *Journal of Applied Research in Intellectual Disabilities*, *18*(4), 295–304. https://doi.org/10.1111/j.1468-3148.2005.00268.x

Galton, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology*, *22*(8), 1019–1037. https://doi.org/10.1002/ACP.1407

Gathercole, S. E., & Pickering, S. J. (2000). Assessment of working memory in six- and seven-year-old children. *Journal of Educational Psychology*, *92*(2), 377–390. https://doi.org/10.1037/0022-0663.92.2.377

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gburek-Augustat, J., Beck-Woedl, S., Tzschach, A., Bauer, P., Schoening, M., & Riess, A. (2016). Epilepsy is not a mandatory feature of STXBP1 associated ataxia-tremor-retardation syndrome. *European Journal of Paediatric Neurology*, *20*(4), 661–665. https://doi.org/10.1016/j.ejpn.2016.04.005

Gentner, D. (1988). Metaphor as Structure Mapping : The Relational Shift. *Child Development*, *59*(1), 47–59.

Gentner, Dedre, & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*(2–3), 263–297. https://doi.org/10.1016/S0010-0277(98)00002-X

Gentner, D., & Ratterman, M. J. (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). Cambridge University Press.

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W. M., Willemsen, M. H., Kwint, M., Janssen, I. M., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H. G., de Vries, B. B. A., Kleefstra, T., Brunner, H. G., … Veltman, J. A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature 2014 511:7509*, *511*(7509), 344–347.

https://doi.org/10.1038/nature13394

Giofrè, D., Mammarella, I. C., & Cornoldi, C. (2013). The structure of working memory and how it relates to intelligence in children. *Intelligence*, *41*, 396–406. https://doi.org/10.1016/j.intell.2013.06.006

Glady, Y. (2013). *Analogical reasoning and its development: role of executive functions and the goal of the task* (Doctoral dissertation, Université de Bourgogne).

Glady, Y., French, R. M., & Thibaut, J.-P. (2017). Children's failure in analogical reasoning tasks a problem of focus of attention and information integration. *Frontiers in Psychology*, *8*, 1–13.

Glady, Y., French, R. M., & Thibaut, J.-P. (2016). Explaining children's failure in analogical reasoning tasks : A problem of focus of attention? *Proceedings of the Thirty-Fourth Annual Meeting of the Cognitive Science Society*, 384–389.

Gonthier, C., & Thomassin, N. (2015). Strategy use fully mediates the relationship between working memory capacity and performance on Raven's matrices. *Journal of Experimental Psychology: General*, *144*(5), 916–924. https://doi.org/10.1037/xge0000101

Gordon, P. C., & Moser, S. (2007). Insight into analogies: Evidence from eye movements. *Visual Cognition*, *15*(1), 20–35. https://doi.org/10.1080/13506280600871891

Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cognitive Development*, *4*(3), 251–268. https://doi.org/10.1016/0885-2014(89)90008-7

Goswami, U., & Brown, A. L. (1990). Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition*, *35*(1), 69–95. https://doi.org/10.1016/0010-0277(90)90037-K

Green, K. E., & Kluever, R. C. (1992). Components of Item Difficulty of Raven's Matrices. *The Journal of General Psychology*, *119*(2), 189–199.

Grone, B. P., Marchese, M., Hamling, K. R., Kumar, M. G., Krasniak, C. S., Sicca, F., Santorelli, F. M., Patel, M., & Baraban, S. C. (2016). Epilepsy, behavioral abnormalities, and physiological comorbidities in syntaxin-binding protein 1 (STXBP1) mutant zebrafish. *PLoS ONE*, *11*(3), 1–25. https://doi.org/10.1371/journal.pone.0151148

Guo, H., Wang, T., Wu, H., Long, M., Coe, B. P., Li, H., Xun, G., Ou, J., Chen, B., Duan, G., Bai, T., Zhao, N., Shen, Y., Li, Y., Wang, Y., Zhang, Y., Baker, C., Liu, Y., Pang, N., … Xia, K. (2018). Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. *Molecular Autism*, *9*(64), 1–12. https://doi.org/10.1186/s13229-018-0247-z

Hager, T., Maroteaux, G., Julsing, J., Vliet, R. Van, & Stiedl, O. (2014). Munc18-1 haploinsufficiency results in enhanced anxiety-like behavior as determined by heart rate responses in mice. *Behavioural Brain Research*, *260*, 44–52. https://doi.org/10.1016/j.bbr.2013.11.033

Hair, J., Anderson, R., Tatham, R., & Black, W. (1998). *Multivariate Data Analysis* (5th Ed.). Englewood Cliffs, NJ: Prentice-Hall.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998a). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*(6), 803–831. https://doi.org/10.1017/S0140525X98001769

Halford, G. S., Wilson, W. H., & Phillips, S. (1998b). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*(6), 803–864. https://doi.org/10.1017/S0140525X98001769

Hamdan, F. F., Gauthier, J., Dobrzeniecka, S., Lortie, A., Mottron, L., Vanasse, M., D'anjou, G., Lacaille, J. C., Rouleau, G. A., & Michaud, J. L. (2011). Intellectual disability without epilepsy associated with STXBP1 disruption. *European Journal of Human Genetics*, *19*, 607–609. https://doi.org/10.1038/ejhg.2010.183

Hamdan, F. F., Gauthier, J., Dobrzeniecka, S., Lortie, A., Mottron, L., Vanasse, M., D'Anjou, G., Lacaille, J. C., Rouleau, G. A., & Michaud, J. L. (2011). Intellectual disability without epilepsy associated with STXBP1 disruption. *European Journal of Human Genetics*, *19*(5), 607–609. https://doi.org/10.1038/ejhg.2010.183

Hamdan, F. F., Piton, A., Gauthier, J., Lortie, A., Dubeau, F., Dobrzeniecka, S., Spiegelman, D., Noreau, A., Pellerin, S., Côté, M., Henrion, E., Fombonne, É., Mottron, L., Marineau, C., Drapeau, P., Lafrenière, R. G., Lacaille, J. C., Rouleau, G. A., & Michaud, J. L. (2009). De novo STXBP1 mutations in mental retardation and nonsyndromic epilepsy. *Annals of Neurology*, *65*(6), 748–753. https://doi.org/10.1002/ANA.21625

Harripaul, R., Noor, A., Ayub, M., & Vincent, J. B. (2017). for Research and Diagnostics for Intellectual Disability. *Cold Spring Harbor Perspectives in Medicine*, *7*, 1–13.

Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In L. Nkov, O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 61–77). https://doi.org/10.4135/9781452233529.n5

Henderson, S. E., Sugden, D. A., & Barnett, A. L. (2007*). Movement assessment battery for children* [examiner's manual](2nd Ed.). London: Pearson Assessment.

Hoaglin, D.C., Iglewicz, B., and Tukey, J.W. (1986). Performance of some resistant rules for outlier labeling, *Journal of American Statistical Association*, *81*, 991-999.

Hoaglin, D. C., and Iglewicz, B. (1987), Fine tuning some resistant rules for outlier labeling, *Journal of American Statistical Association*, *82*, 1147-1149.

Holling, H. Figural Analogies: Rule-based Test Construction and Evaluation Using Random Effect Models.

Holmes, J., & Adams, J. W. (2006). Working Memory and Children's Mathematical Skills: Implications for mathematical development and mathematics curricula. *Educational Psychology*, *26*(3), 339–366. https://doi.org/10.1080/01443410500341056

Hornke, L. F., & Habon, M. W. (1986). Rule-Based Item Bank Construction and Evaluation Within the Linear Logistic Framework. *Applied Psychological Measurement*, *10*(4), 369–380. https://doi.org/10.1177/014662168601000405

Hornung, C., Brunner, M., Reuter, R. A., & Martin, R. (2011a). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence*, *39*, 210–221. https://doi.org/10.1016/j.intell.2011.03.002

Hornung, C., Brunner, M., Reuter, R. A. P., & Martin, R. (2011b). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence*, *39*(4), 210–221. https://doi.org/10.1016/j.intell.2011.03.002

Horsler, K., & Oliver, C. (2006). The behavioural phenotype of Angelman syndrome. *Journal of Intellectual Disability Research*, *50*(1), 33–53. https://doi.org/10.1111/j.1365-2788.2005.00730.x

Huang, H. S., Burns, A. J., Nonneman, R. J., Baker, L. K., Riddick, N. V., Nikolova, V. D., Riday, T. T., Yashiro, K., Philpot, B. D., & Moy, S. S. (2013). Behavioral deficits in an Angelman syndrome model: Effects of genetic background and age. *Behavioural Brain Research*, *243*(1), 79–90. https://doi.org/10.1016/j.bbr.2012.12.052

Inhelder, B., and Piaget, J. (1964). *The Early Growth of Logic in the Child: Classification and Seriation*. New York, NY: Routledge.

IPAT (1973a). *Measuring intelligence with the culture fair tests: Manual for scales 2 and 3*. Champaign, IL: Institute for Personality and Ability Testing.

IPAT (1973b). *Technical supplement for the culture fair intelligence tests: Scales 2 and 3*. Champaign, IL: Institute for Personality and Ability Testing.

Ilyas, M., Mir, A., Efthymiou, S., & Houlden, H. (2019). The genetics of intellectual disability: Advancing technology and gene editing. *F1000Research*, *8*, 1–10. https://doi.org/10.12688/f1000research.16315.1

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

Jaroslawska, A. J., Gathercole, S. E., Allen, R. J., & Holmes, J. (2016). Following instructions from working memory: Why does action at encoding and recall help? *Memory and Cognition*, *44*, 1183–1191. https://doi.org/10.3758/s13421-016-0636-5

Jensen, O. (2006). Maintenance of multiple working memory items by temporal segmentation. *Neuroscience*, *139*(1), 237–249. https://doi.org/10.1016/j.neuroscience.2005.06.004

John, A., Ng-Cordell, E., Hanna, N., Brkic, D., & Baker, K. (2021). The neurodevelopmental spectrum of synaptic vesicle cycling disorders. *Journal of Neurochemistry*, *157*(2), 208–228. https://doi.org/10.1111/jnc.15135

Johnson, D. M. (1962). Serial analysis of verbal analogy problems. *Journal of Educational Psychology*, *53*(2), 86–88. https://doi.org/10.1037/h0047441

Johnson, A. (2018). *The environment and child development: A multivariate approach* [Doctoral dissertation, University of Cambridge]. Retrieved from https://www.repository.cam.ac.uk/handle/1810/292061?show=full.

Jungeblut, H. M., Hagemann, D., Löffler, C., & Schubert, A. (2020). An Investigation of the Slope Parameters of Reaction Times and P3 Latencies in the Sternberg Memory Scanning Task – A Fixed-Links Model Approach. In *Journal of Intelligence* (Vol. 8).

Kail, Robert & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*, 199–225. https://doi.org/10.1016/0001-6918(94)90003-5

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin*, *109*(3), 490–501. https://doi.org/10.1037/0033-2909.109.3.490

Kail, R. (1992). Processing Speed, Speech Rate, and Memory. *Developmental Psychology*, *28*(5), 899–904. https://doi.org/10.1037/0012-1649.28.5.899

Kail, R. (2000). Speed of information processing: Developmental change and links to intelligence. *Journal of School Psychology*, *38*(1), 51–61. https://doi.org/10.1016/S0022-4405(99)00036-9

Kail, R., & Hall, L. K. (2001). Distinguishing short-term memory from working memory. *Memory and Cognition*, *29*(1), 1–9. https://doi.org/10.3758/BF03195735

Kail, R. V. (2007). Longitudinal evidence that increases in processing speed and working memory enhance children's reasoning. *Psychological Science*, *18*(4), 312–313. https://doi.org/10.1111/j.1467-9280.2007.01895.x

Kane, M., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169–183.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working Memory, Attention Control, and the N-Back Task: A Question of Construct Validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615–622. https://doi.org/10.1037/0278-7393.33.3.615

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review*, *9*(4), 637–671. https://doi.org/10.3758/BF03196323

Kane, M. J., & Engle, R. W. (2003). Working-Memory Capacity and the Control of Attention: The Contributions of Goal Neglect, Response Competition, and Task Set to Stroop Interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70. https://doi.org/10.1037/0096-3445.132.1.47

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71. https://doi.org/10.1037/0033-2909.131.1.66

Kovačević, J., Maroteaux, G., Schut, D., Loos, M., Dubey, M., Pitsch, J., Remmelink, E., Koopmans, B., Crowley, J., Cornelisse, L. N., Sullivan, P. F., Schoch, S., Toonen, R. F., Stiedl, O., & Verhage, M. (2018). Protein instability, haploinsufficiency, and cortical hyper-excitability underlie STXBP1 encephalopathy. *Brain*, *141*(5), 1350–1374. https://doi.org/10.1093/brain/awy046

Kovacs, K., & Conway, A. R. A. (2016). Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry*, *27*(3), 151–177. https://doi.org/10.1080/1047840X.2016.1153946

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*(7), 2020–2032. https://doi.org/10.1016/j.neuropsychologia.2008.02.001

Ku, C. S., Polychronakos, C., Tan, E. K., Naidoo, N., Pawitan, Y., Roukos, D. H., Mort, M., & Cooper, D. N. (2013). A new paradigm emerges from the study of de novo mutations in the context of neurodevelopmental disease. *Molecular Psychiatry*. https://doi.org/10.1038/mp.2012.58

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

L M Vissers, L. E., Gilissen, C., & Veltman, J. A. (2016). Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics*, *17*(1), 9–18. https://doi.org/10.1038/nrg3999

Liao, S. G., Lin, Y., Kang, D. D., Chandra, D., Bon, J., Kaminski, N., Sciurba, F. C., & Tseng, G. C. (2014). Missing value imputation in high-dimensional phenomic data: imputable or not, and how? *BMC Bioinformatics*, *15*(346), 1–12. https://doi.org/10.1186/s12859-014-0346-6

Lifshitz, H., Weiss, I., Tzuriel, D., & Tzemach, M. (2011). New model of mapping difficulties in solving analogical problems among adolescents and adults with intellectual disability. *Research in Developmental Disabilities*, *32*(1), 326–344. https://doi.org/10.1016/j.ridd.2010.10.010

Luria, A. R., & Tsvetkova, L. S. (1964). The programming of constructive activity in local brain injuries. *Neuropsychologia*, *2*(2), 95–107. https://doi.org/10.1016/0028-3932(64)90015-6

Mabb, A. M., Judson, M. C., Zylka, M. J., & Philpot, B. D. (2011). Angelman syndrome: Insights into genomic imprinting and neurodevelopmental phenotypes. *Trends in Neurosciences*, *34*(6), 293–303. https://doi.org/10.1016/j.tins.2011.04.001

Marcela Bizama, M., Saldaño, D., & Cristina Rodríguez, R. (2019). Fluid intelligence, working memory, reading fluency and comprehension in chilean school children. *Electronic Journal of Research in Educational Psychology*, *17*(48), 295–316. https://doi.org/10.25115/ejrep.v17i48.2251

Marshall, A., Altman, D. G., & Holder, R. L. (2010). Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC Medical Research Methodology*, *10*(112), 1–10. https://doi.org/10.1186/1471-2288-10-112

Mastrangelo, M., Peron, A., Spaccini, L., Novara, F., Scelsa, B., Introvini, P., Raviglione, F., Faiola, S., & Zuffardi, O. (2013). Neonatal suppression-burst without epileptic seizures: Expanding the electroclinical phenotype of STXBP1-related, early-onset encephalopathy. *Epileptic Disorders*, *15*(1), 55–61. https://doi.org/10.1684/epd.2013.0558

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: A meta-analysis of population-based studies. *Research in Developmental Disabilities*, *32*(2), 419–436. https://doi.org/10.1016/j.ridd.2010.12.018

Mcardle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative Longitudinal Structural Analyses of the Growth and Decline of Multiple Intellectual Abilities Over the Life Span. *Developmental Psychology*, *38*(1), 115–142. https://doi.org/10.1037/0012-1649.38.1.115

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for Similarity. *Psychological*

Review, *100*(2), 254–278. https://doi.org/10.1037/0033-295X.100.2.254

Meo, M., Roberts, M. J., & Marucci, F. S. (2007). Element salience as a predictor of item difficulty for Raven's Progressive Matrices. *Intelligence*, *35*(4), 359–368. https://doi.org/10.1016/J.INTELL.2006.10.001

Milh, M., Villeneuve, N., Chouchane, M., Kaminska, A., Laroche, C., Barthez, M. A., Gitiaux, C., Bartoli, C., Borges-Correia, A., Cacciagli, P., Mignon-Ravix, C., Cuberos, H., Chabrol, B., & Villard, L. (2011). Epileptic and nonepileptic features in patients with early onset epileptic encephalopathy and STXBP1 mutations. *Epilepsia*, *52*(10), 1828–1834. https://doi.org/10.1111/j.1528-1167.2011.03181.x

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, *27*(3), 272–277. https://doi.org/10.1093/oxfordjournals.bmb.a070866

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* Cambridge University Press.

Moons, K. G. M., Donders, R. A. R. T., Stijnen, T., & Harrell, F. E. (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology*, *59*(10), 1092–1101. https://doi.org/10.1016/j.jclinepi.2006.01.009

Morrison, R. G., Doumas, A., & Richland, L. E. (2006). The Development of Analogical Reasoning in Children : A Computational Account Multiple Factors in Analogical Development. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 603–608.

Morrison, Robert G., Doumas, L. A. A., & Richland, L. E. (2011). A computational account of children's analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science*, *14*(3), 516–529. https://doi.org/10.1111/j.1467-7687.2010.00999.x

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, *12*(2), 252–284. https://doi.org/10.1016/0010-0285(80)90011-0

Neitzel, C., & Stright, A. D. (2003). Mothers' scaffolding of children's problem solving: Establishing a foundation of academic self-regulatory competence. *Journal of Family Psychology*, *17*(1), 147–159. https://doi.org/10.1037/0893-3200.17.1.147

Nettelbeck, T., & Burns, N. R. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Personality and Individual Differences*, *48*(4), 379–384. https://doi.org/10.1016/j.paid.2009.10.032

Newman, D. A. (2014). Missing Data: Five Practical Guidelines. *Organizational Research Methods*, *17*(4), 372–411. https://doi.org/10.1177/1094428114548590

Nickerson, R. S. (1967). “Same” - “different” response times with multi-attribute stimulus differences. *Perceptual and Motor Skills*, *24*, 543–554.

Niebaum, J., & Munakata, Y. (2021). *The Development of Relational Reasoning: An Eyetracking Analysis of Strategy Use and Adaptation in Children and Adults Performing Matrix Completion*. https://psyarxiv.com/wfz8u/

Novick, L. R., & Tversky, B. (1987). Cognitive Constraints on Ordering Operations: The Case of Geometric Analogies. *Journal of Experimental Psychology: General*, *116*(1), 50–67. https://doi.org/10.1037/0096-3445.116.1.50

Oberauer, K. (2002). Access to Information in Working Memory: Exploring the Focus of Attention. *Journal of Experimental Psychology: Learning Memory and Cognition*, *28*(3), 411–421. https://doi.org/10.1037/0278-7393.28.3.411

Oberauer, K. (2019). Working memory and attention - A conceptual analysis and review. *Journal of Cognition*, *2*(1), 1–23. https://doi.org/10.5334/joc.58

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittman, W. W. (2000). Working memory capacity Ð facets of a cognitive ability construct. *Personality and Individual Differences*, *29*, 1017–1045. www.elsevier.com/locate/paid

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*, 167–193.

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*(6), 641–652. https://doi.org/10.1016/j.intell.2008.01.007

Oberauer, K., Wilhelm, O., Schulze, R., & Süß, H. M. (2005). Working memory and intelligence -Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 61–65. https://doi.org/10.1037/0033-2909.131.1.61

Odom, R. D., Astor, E. C., & Cunningham, J. G. (1975). Effects of Perceptual Salience on the Matrix Task Performance of Four- and Six-Year-Old Children. *Development*, *46*(3), 758–762.

Oliver, C., Woodcock, K., & Adams, D. (2010). The importance of aetiology of intellectual disability The importance of aetiology of intellectual disability . Oliver , C ., Woodcock , K . and Adams , D . Cerebra Centre for Neurodevelopmental Disorders , School of Psychology , University of Birmingham. In *Grant, G., Ramcharan, P., Flynn, M. and Richardson, M. (Eds.) Learning Disability: A life cycle approach to valuing people.* (pp. 135–146). Open University Press/Wiley.

Orock, A., Logan, S., & Deak, F. (2018). Molecular and Cellular Neuroscience Munc18-1 haploinsu ffi ciency impairs learning and memory by reduced synaptic vesicular release in a model of Ohtahara syndrome. *Molecular and Cellular Neuroscience*, *88*(December 2017), 33–42. https://doi.org/10.1016/j.mcn.2017.12.002

Parrish, E. E., Giaschi, D. E., Boden, C., & Dougherty, R. (2005). The maturation of form and motion perception in school age children. *Vision Research*, *45*(7), 827–837. https://doi.org/10.1016/J.VISRES.2004.10.005

Pelc, K., Cheron, G., & Dan, B. (2008). Behavior and neuropsychiatry manifestations in Angelman syndrome. *Neuropsychiatric Disease and Treatment*, *4*(3), 577–584. https://doi.org/10.2147/NDT.S2749

Plomin R., DeFries J. C., Knopik V. S., & Neiderhiser J. M. (2012). *Behavioral Genetics* (6th Ed). New York: Worth Publishers.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Postlethwaite, B. E. (2011). *Fluid ability, crystallized ability, and performance across multiple domains: a meta-analysis*. https://doi.org/10.17077/etd.zopi8wvs

Primi, R. (2001). Complexity of geometric inductive reasoning tasks contribution to the understanding of fluid intelligence. *Intelligence*, *30*(1), 41–70. https://doi.org/10.1016/S0160-2896(01)00067-8

Rattermann, M. J., & Genter, D. (1998). More evidence for a relational shift in the development of analogy: Children’s performance on causal-mapping task. *Cognitive Development*, *13*, 453–478.

Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children’s performance on a causal-mapping task. *Cognitive Development*, *13*(4), 453–478. https://doi.org/10.1016/S0885-2014(98)90003-X

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., Dufke, A., Cremer, K., Hempel, M., Horn, D., Hoyer, J., Joset, P., Röpke, A., Moog, U., Riess, A., … Strom, T. M. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: An exome sequencing study. *The Lancet*, *380*(9854), 1674–1682. https://doi.org/10.1016/S0140-6736(12)61480-9

Raven, J. C. (1962). *Advanced Progressive Matrices, Set II*. London: H. K. Lewis. (Distributed in the United States by The Psychological Corporation, San Antonio, TX)

Raven, J. C. (1965). *Advanced Progressive Matrices, Sets I and II*. London: H. K. Lewis. (Distributed in the United States by The Psychological Corporation, San Antonio, TX)

Raven, J. C., & Court, J. H. (1938). *Raven's progressive matrices*. Los Angeles, CA: Western Psychological Services.

Raven, J. C., Court, J. H., & Raven J. (1983). Manual for Raven’s Progressive Matrices and Vocabulary Scales (Section 4: Advanced Progressive Matrices). London: H. K. Lewis.

Raven, J. C., Court, J. H., & Raven, J. (1988). *Manual for Raven’s Progressive Matrices and Vocabulary Scales. Section 4: Advanced Progressive Matrices, sets I and II*. (1988 ed ed.). Oxford: Oxford Psychologists Press.

Reichenberg, A., Cederlöf, M., McMillan, A., Trzaskowski, M., Kapara, O., Fruchter, E., Ginat, K., Davidson, M., Weiser, M., Larsson, H., Plomin, R., & Lichtenstein, P. (2016). Discontinuity in the genetic and environmental causes of the intellectual disability spectrum. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(4), 1098–1103. https://doi.org/10.1073/PNAS.1508093112/-/DCSUPPLEMENTAL

Richards, C., Powis, L., Moss, J., Stinton, C., Nelson, L., & Oliver, C. (2017). Prospective study of autism phenomenology and the behavioural phenotype of Phelan-McDermid syndrome: Comparison to fragile X syndrome, Down syndrome and idiopathic autism spectrum disorder. *Journal of Neurodevelopmental Disorders*, *9*(1), 1–15. https://doi.org/10.1186/s11689-017-9217-6

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*(3), 249–273. https://doi.org/10.1016/j.jecp.2006.02.002

Rivollier, G., Quinton, J.-C., Gonthier, C., & Smeding, A. (2020). Looking with the (computer) mouse: How to unveil problem-solving strategies in matrix reasoning without eye-tracking. *Behavior Research Methods*, 1–16. https://doi.org/10.3758/s13428-020-01484-3

Rizo, J. (2018). Mechanism of neurotransmitter release coming into focus. *Protein Science*, *27*(8), 1364–1391. https://doi.org/10.1002/pro.3445

Roid, G. H., & Miller, L. J. (1997). *Leiter international performance scale—Revised*. Wood Dale, IL: Stoelting.

Roid, G. H., Miller, L. J., Pomplun, M. & Koch, C. (2013). *Leiter-international performance scale-third edition*. Wood Dale, IL: Stoelting.

Rosen, V. M., & Engle, R. W. (1997). Forward and Backward Serial Recall. *Intelligence*, *25*(1), 37–47. https://doi.org/10.1037/008822

Saitsu, H., Kato, M., Mizuguchi, T., Hamada, K., Osaka, H., Tohyama, J., Uruno, K., Kumada, S., Nishiyama, K., Nishimura, A., Okada, I., Yoshimura, Y., Hirai, S. I., Kumada, T., Hayasaka, K., Fukuda, A., Ogata, K., & Matsumoto, N. (2008). De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. *Nature Genetics*, *40*(6), 782–788. https://doi.org/10.1038/ng.150

Salthouse, T. A. (1996). The Processing-Speed Theory of Adult Age Differences in Cognition. In *Psychological Review* (Vol. 103, Issue 3).

Schubert, A. L., & Frischkorn, G. T. (2020). Neurocognitive Psychometrics of Intelligence: How Measurement Advancements Unveiled the Role of Mental Speed in Intelligence Differences. *Current Directions in Psychological Science*, *29*(2), 140–146. https://doi.org/10.1177/0963721419896365

Schubert, A. L., Hagemann, D., & Frischkorn, G. T. (2017). Is general intelligence little more than the speed of higher-order processing? *Journal of Experimental Psychology: General*, *146*(10), 1498–1512. https://doi.org/10.1037/xge0000325

Shanks, S. G., Carpp, L. N., Struthers, M. S., Mccann, R. K., & Bryant, N. J. (2012). The Sec1/Munc18 Protein Vps45 Regulates Cellular Levels of Its SNARE Binding Partners Tlg2 and Snc2 in Saccharomyces cerevisiae. *PLoS ONE*, *7*(11), 1–7. https://doi.org/10.1371/journal.pone.0049628

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016a). Working Memory Capacity and Fluid Intelligence: Maintenance and Disengagement. *Perspectives on Psychological Science*, *11*(6), 771–799. https://doi.org/10.1177/1745691616650647

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016b). Working Memory Capacity and Fluid Intelligence: Maintenance and Disengagement. *Perspectives on Psychological Science*, *11*(6), 771–799. https://doi.org/10.1177/1745691616650647

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*(1), 116–141. https://doi.org/10.1016/j.jml.2014.01.004

Siegler, R. S., & Svetina, M. (2002). A Microgenetic/Cross-Sectional Study of Matrix Completion: Comparing Short-Term and Long-Term Change. *Child Development*, *73*(3), 793–809. https://doi.org/10.1111/1467-8624.00439

Spearman, C. *The nature of "intelligence" and the principles of cognition.* London: Macmillan, 1923.

Sparrow, S. S., Cicchetti, D., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales, Second Edition (Vineland-II)* [Database record]. APA PsycTests. https://doi.org/10.1037/t15164-000

Spencer, J., O'Brien, J., Riggs, K., Braddick, O., Atkinson, J., & Wattam-Bell, J. (2000). Motion processing in autism: evidence for a dorsal stream defciency. *Cognitive Neuroscience and Neuropsychology*, *11*(12), 2765–2767.

Stamberger, H., Nikanorova, M., Accorsi, P., Angriman, M., Benkel-herrenbrueck, I., Capovilla, G., Erasmus, C. E., Fannemel, M., Giordano, L., Helbig, K. L., Maier, O., & Phalin, J. (2016). STXBP1 encephalopathy: A neurodevelopmental disorder including epilepsy. *Neurology*, *86*(10), 954–962.

Starr, A., Vendetti, M. S., & Bunge, S. A. (2018). Eye movements provide insight into individual differences in children's analogical reasoning strategies. *Acta Psychologica*, *186*(April), 18–26. https://doi.org/10.1016/j.actpsy.2018.04.002

Stekhoven, D. J. (2013). missForest: Nonparametric Missing Value Imputation using Random Forest. R package version 1.4.

Stekhoven, D. J., & Bühlmann, P. (2012). Data and text mining MissForest-non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, *84*(4), 353–378. https://doi.org/10.1037/0033-295X.84.4.353

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly*, *47*(1), 1–41. https://doi.org/10.1353/mpq.2001.0005

Sternberg, R. J., & Nigro, G. (1980). Developmental Patterns in the Solution of Verbal Analogies. *Society for Research in Child Developmentevelopment*, *51*(1), 27–38.

Sternberg, R. J., & Rifkin, B. (1979). The development of analogical reasoning processes. *Journal of Experimental Child Psychology*, *27*, 195–232. https://doi.org/10.1080/19388078409557800

Stevenson, C. E., Alberto, R. A., van den Boom, M. A., & de Boeck, P. A. L. (2014). Visual relations children find easy and difficult to process in figural analogies. *Frontiers in Psychology*, *5*(AUG), 1–9. https://doi.org/10.3389/fpsyg.2014.00827

Stevenson, C. E., Heiser, W. J., & Resing, W. C. M. (2016). Dynamic Testing of Analogical Reasoning in 5- to 6-Year-Olds: Multiple-Choice Versus Constructed-Response Training Items. *Journal of Psychoeducational Assessment*, *34*(6), 550–565. https://doi.org/10.1177/0734282915622912

Stevenson, C. E., & Hickendorff, M. (2018). Learning to solve figural matrix analogies: The paths children take. *Learning and Individual Differences*, *66*(December 2016), 16–28. https://doi.org/10.1016/j.lindif.2018.04.010

Stevenson, C. E., Touw, K. W. J., & Resing, W. C. M. (2011). Computer or paper analogy puzzles: Does assessment mode influence young children's strategy progression. *Educational and Child Psychology*, *28*(2), 67–84.

Stoppel, D. C., & Anderson, M. P. (2017). Hypersociability in the Angelman syndrome mouse model. *Experimental Neurology*, *293*, 137–143. https://doi.org/10.1016/j.expneurol.2017.04.002

Suri, M., Evers, J. M. G., Laskowski, R. A., O'Brien, S., Baker, K., Clayton-Smith, J., Dabir, T., Josifova, D., Joss, S., Kerr, B., Kraus, A., McEntagart, M., Morton, J., Smith, A., Splitt, M., Thornton, J. M., & Wright, C. F. (2017). Protein structure and phenotypic analysis of pathogenic and population missense variants in STXBP1. *Molecular Genetics and Genomic Medicine*, *5*(5), 495–507. https://doi.org/10.1002/mgg3.304

Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability - And a little bit more. *Intelligence*, *30*(3), 261–288. https://doi.org/10.1016/S0160-2896(01)00100-3

Swanson, D. A., Steel, J. M., & Valle, D. (1998). Identification and characterization of the human ortholog of rat STXBP1, a protein implicated in vesicle trafficking and neurotransmitter release. *Genomics*, *48*(3), 373–376. https://doi.org/10.1006/geno.1997.5202

Swanson, H. L. (2008). Working Memory and Intelligence in Children: What Develops? *Journal of Educational Psychology*, *100*(3), 581–602. https://doi.org/10.1037/0022-0663.100.3.581

Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining*, *10*(6), 363–377. https://doi.org/10.1002/sam.11348

Taylor, N. M., Jakobson, L. S., Maurer, D., & Lewis, T. L. (2009). Differential vulnerability of global motion, global form, and biological motion processing in full-term and preterm children. *Neuropsychologia*, *47*(13), 2766–2778. https://doi.org/10.1016/j.neuropsychologia.2009.06.001

Thibaut, J., & French, R. (2011). In the eyes of the beholder: What eye-tracking reveals about analogy-making strategies in children and adults. *Proceedings of the …*, 453–458. http://palm.mindmodeling.org/cogsci2011/papers/0085/paper0085.pdf

Thibaut, J. P., & French, R. M. (2016). Analogical reasoning, control and executive functions: A developmental investigation with eye-tracking. *Cognitive Development*, *38*, 10–26. https://doi.org/10.1016/j.cogdev.2015.12.002

Thibaut, J. P., French, R., & Vezneva, M. (2010). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology*, *106*(1), 1–19. https://doi.org/10.1016/j.jecp.2010.01.001

Thomas R. Coyle, David R. Pillow, Anissa C. Snyder, and, Kochunov, P. (2011). Processing Speed Mediates the Development of General Intelligence (g) in Adolescence. *Psychological Science*, *22*(10), 1265–1269. https://doi.org/10.1177/0956797611418243

Tillman, C. M., Nyberg, L., & Bohlin, G. (2008). Working memory components and intelligence in children. *Intelligence*, *36*(5), 394–402. https://doi.org/10.1016/j.intell.2007.10.001

Toonen, R. F. G., & Verhage, M. (2007). Munc18-1 in secretion: lonely Munc joins SNARE team and takes control. *Trends in Neurosciences*, *30*(11), 564–572. https://doi.org/10.1016/j.tins.2007.08.008

Tschentscher, N., Mitchell, D., & Duncan, J. (2017a). Fluid intelligence predicts novel rule implementation in a distributed frontoparietal control network. *Journal of Neuroscience*, *37*(18), 4841–4847. https://doi.org/10.1523/JNEUROSCI.2478-16.2017

Tschentscher, N., Mitchell, D., & Duncan, J. (2017b). Fluid Intelligence Predicts Novel Rule Implementation in a Distributed Frontoparietal Control Network. *The Journal of Neuroscience*, *37*(18), 4841–4847. https://doi.org/10.1523/JNEUROSCI.2478-16.2017

Tunteler, E., Pronk, C. M. E., & Resing, W. C. M. (2008). Inter-and intra-individual variability in the process of change in the use of analogical strategies to solve geometric tasks in children: A microgenetic analysis. *Learning and Individual Differences*, *18*, 44–60. https://doi.org/10.1016/j.lindif.2007.07.007

Turner, T. N., & Eichler, E. E. (2019). The Role of De Novo Noncoding Regulatory Mutations in Neurodevelopmental Disorders. *Trends in Neurosciences*, *42*(2), 115–127. https://doi.org/10.1016/j.tins.2018.11.002

Uddin, M., Woodbury-Smith, M., Chan, A., Brunga, L., Lamoureux, S., Pellecchia, G., Yuen, R. K. C., Faheem, M., Stavropoulos, D. J., Drake, J., Hahn, C. D., Hawkins, C., Shlien, A., Marshall, C. R., Turner, L. A., Minassian, B. A., Scherer, S. W., & Boelman, C. (2017). Germline and somatic mutations in STXBP1 with diverse neurodevelopmental phenotypes. *Neurol Genet*, *3*, 199. https://doi.org/10.1212/NXG.0000000000000199

Unsworth, N., & Engle, R. W. (2007). On the Division of Short-Term and Working Memory: An Examination of Simple and Complex Span and Their Relation to Higher Order Abilities. *Psychological Bulletin*, *133*(6), 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014a). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014b). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). The Contributions of Primary and Secondary Memory to Working Memory Capacity: An Individual Differences Analysis of Immediate Free Recall. *Journal of Experimental Psychology: Learning Memory and*

Cognition, *36*(1), 240–247. https://doi.org/10.1037/a0017739

Vakil, E., Lifshitz, H., Tzuriel, D., Weiss, I., & Arzuoan, Y. (2011). Analogies solving by individuals with and without intellectual disability: Different cognitive patterns as indicated by eye movements. *Research in Developmental Disabilities*, *32*, 846–856. https://doi.org/10.1016/j.ridd.2010.08.006

Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842

Veltman, J. A., & Brunner, H. G. (2012). De novo mutations in human genetic disease. *Nature Publishing Group*. https://doi.org/10.1038/nrg3241

Vendetti, M. S., Starr, A., Johnson, E. L., Modavi, K., & Bunge, S. A. (2017). Eye movements reveal optimal strategies for analogical reasoning. *Frontiers in Psychology*, *8*(JUN), 2–10. https://doi.org/10.3389/fpsyg.2017.00932

Verhage, M. (2000). Synaptic Assembly of the Brain in the Absence of Neurotransmitter Secretion. *Science (80-.).*, *287*(5454), 864–869. https://doi.org/10.1126/science.287.5454.864

Vodegel Matzen, L. B. L., van der Molen, M. W., & Dudink, A. C. M. (1994). Error analysis of raven test performance. *Personality and Individual Differences*, *16*(3), 433–445. https://doi.org/10.1016/0191-8869(94)90070-1

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* Cambridge, Mass.: Harvard University Press.

Waite, J., Heald, M., Wilde, L., Woodcock, K., Welham, A., Adams, D., & Oliver, C. (2014). The importance of understanding the behavioural phenotypes of genetic syndromes associated with intellectual disability. *Paediatrics and Child Health (United Kingdom)*, *24*(10), 468–472. https://doi.org/10.1016/j.paed.2014.05.002

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Zhu, J., & Higgins, P. D. R. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*, *3*(8). https://doi.org/10.1136/bmjopen-2013-002847

Wechsler, D. (2008). *Wechsler adult intelligence scale-fourth edition*. San Antonio, TX: Pearson.

Wechsler, D. (2011). Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II). San Antonio, TX: NCS Pearson.

Whitaker, K. J., Vendetti, M. S., Wendelken, C., & Bunge, S. A. (2018). Neuroscientific insights into the development of analogical reasoning. *Developmental Science*, *21*(2), 1–11. https://doi.org/10.1111/desc.12531

Whitely, S. E., & Schneider, L. M. (1981). Information Structure for Geometric Analogies: A Test Theory Approach. *Applied Psychological Measurement*, *5*(3), 383–397.

Wilhelm, O., & Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus-response compatibility in choice reaction time tasks. *European Journal of Cognitive Psychology*, *18*(1), 18–50. https://doi.org/10.1080/09541440500215921

Winston, P. H. (1970). *Learning structural descriptions from examples* (Report AI TR-231). Cambridge: Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1970.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. In *J. Child Psychol. Psychiat* (Vol. 17). Pergamon Press.

Woolgar, A., Duncan, J., Manes, F., & Fedorenko, E. (2018). Fluid intelligence is supported by the multiple-demand system not the language system. *Nature Human Behaviour 2018 2:3*, *2*(3), 200–204. https://doi.org/10.1038/s41562-017-0282-3

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., Roca, M., Antoun, N., Manes, F., & Duncan, J. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(33), 14899–14902. https://doi.org/10.1073/pnas.1007928107

World Health Organization. (2007). *Atlas: Global resources for persons with intellectual disabilities*. World Health Organization.

Wright, C. F., Fitzgerald, T. W., Jones, W. D., Clayton, S., McRae, J. F., Van Kogelenberg, M., King, D. A., Ambridge, K., Barrett, D. M., Bayzetinova, T., Bevan, A. P., Bragin, E., Chatzimichali, E. A., Gribble, S., Jones, P., Krishnappa, N., Mason, L. E., Miller, R., Morley, K. I., … Firth, H. V. (2015). Genetic diagnosis of developmental disorders in the DDD study: A scalable analysis of genome-wide research data. *The Lancet*, *385*(9975), 1305–1314. https://doi.org/10.1016/S0140-6736(14)61705-0

Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience*, *1*(MAR), 1–8. https://doi.org/10.3389/neuro.09.008.2007

Yamamoto, T., Shimojima, K., Yano, T., Ueda, Y., Takayama, R., Ikeda, H., & Imai, K. (2016). Loss-of-function mutations of STXBP1 in patients with epileptic encephalopathy. *Brain and Development*, *38*(3), 280–284. https://doi.org/10.1016/j.braindev.2015.09.004

Ye, A. (2020, August 31). *MissForest: The Best Missing Data Imputation Algorithm?* [Blog Post]. Retrieved from: https://towardsdatascience.com/missforest-the-best-missing-data-imputation-algorithm-4d01182aed3

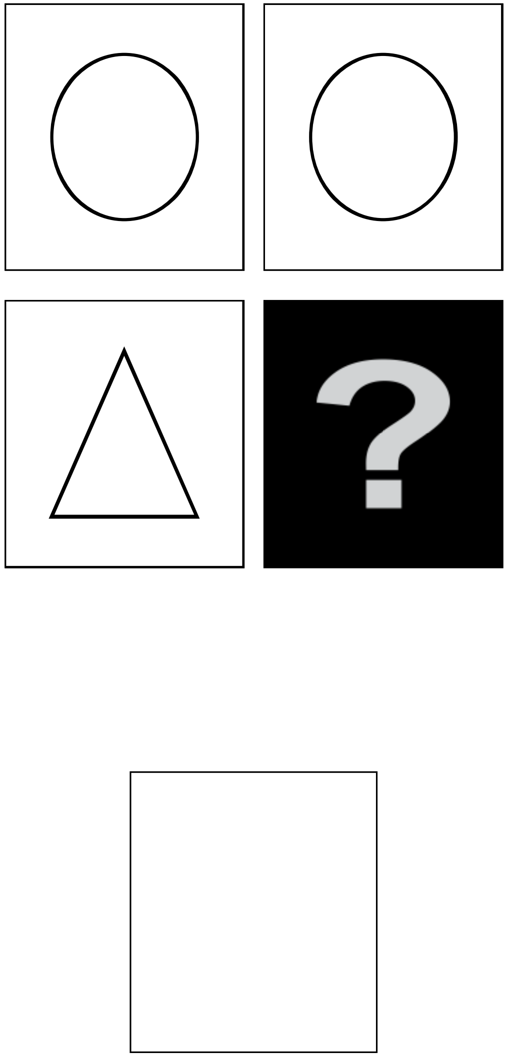
Yuge, K., Iwama, K., Yonee, C., Matsufuji, M., Sano, N., Saikusa, T., Yae, Y., Yamashita, Y., Mizuguchi, T., Matsumoto, N., & Matsuishi, T. (2018). A novel STXBP1 mutation causes typical Rett syndrome in a Japanese girl. *Brain and Development*, *40*(6), 493–497. https://doi.org/10.1016/j.braindev.2018.02.002

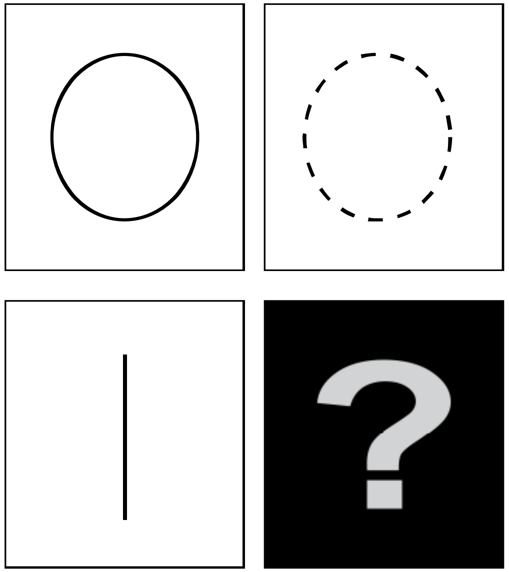
Zammit, S., Allebeck, P., David, A. S., Dalman, C., Hemmingsson, T., Lundberg, I., & Lewis, G. (2004). A Longitudinal Study of Premorbid IQ Score and Risk of Developing Schizophrenia, Bipolar Disorder, Severe Depression, and Other Nonaffective Psychoses. *Archives of General Psychiatry*, *61*(4), 354–360. https://doi.org/10.1001/archpsyc.61.4.354

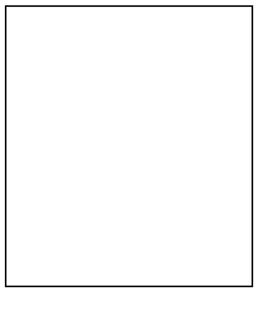
Appendix A Taskset A1

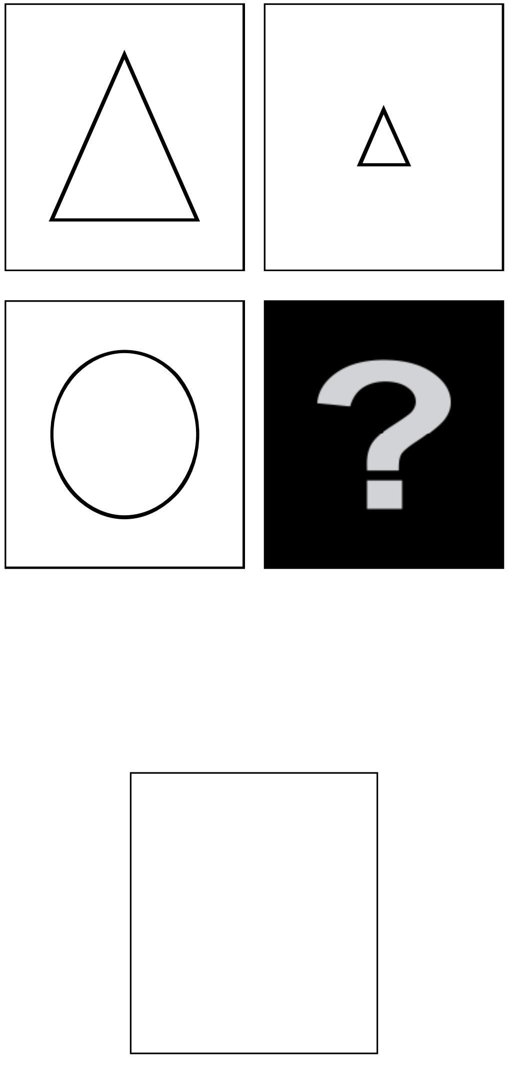
A1

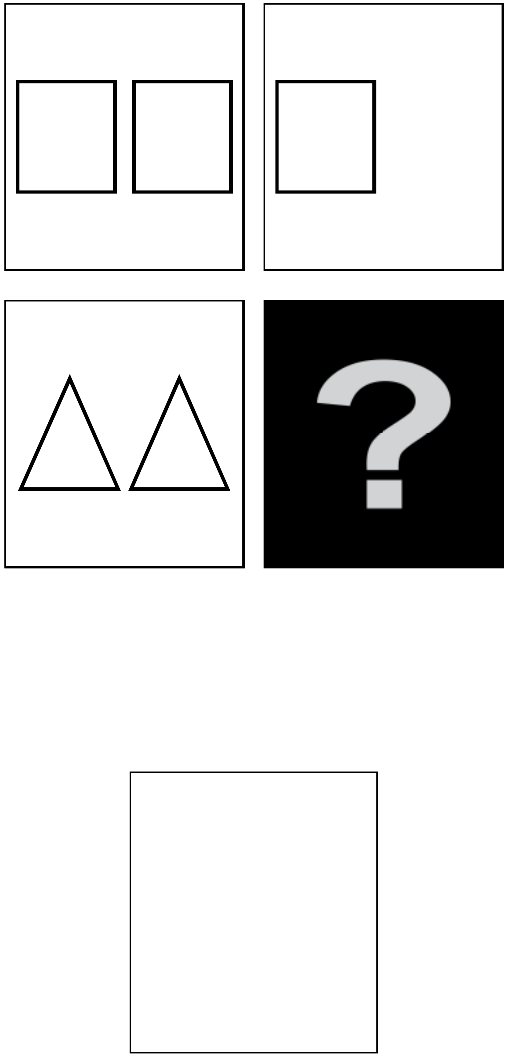
ID: ________________________

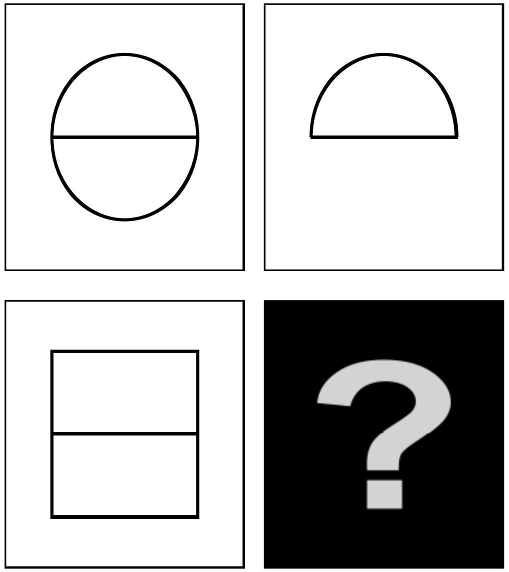


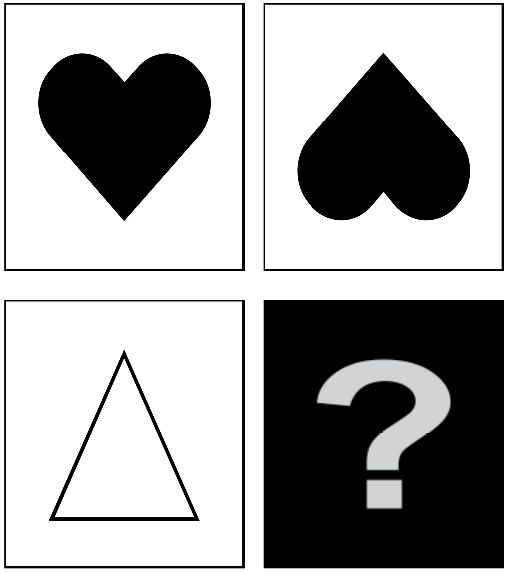


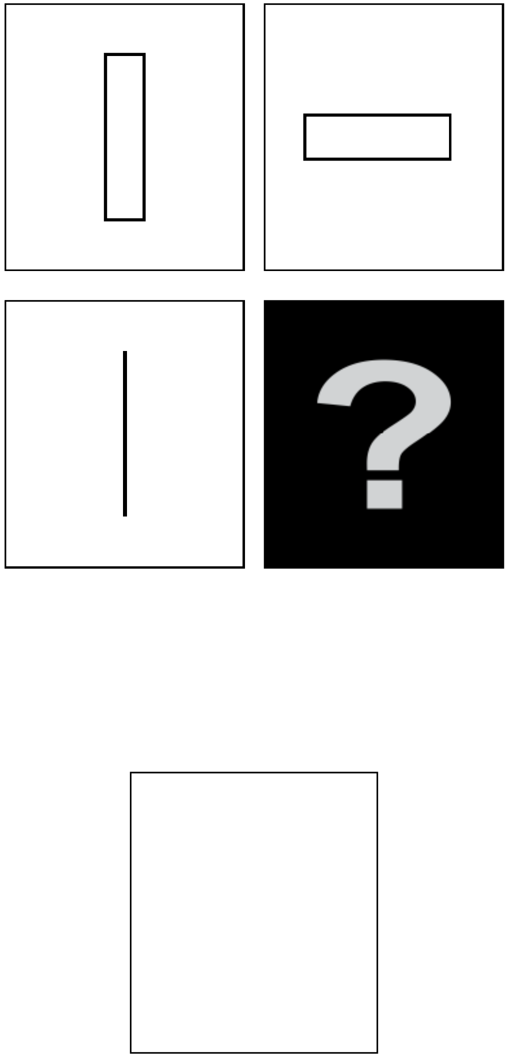




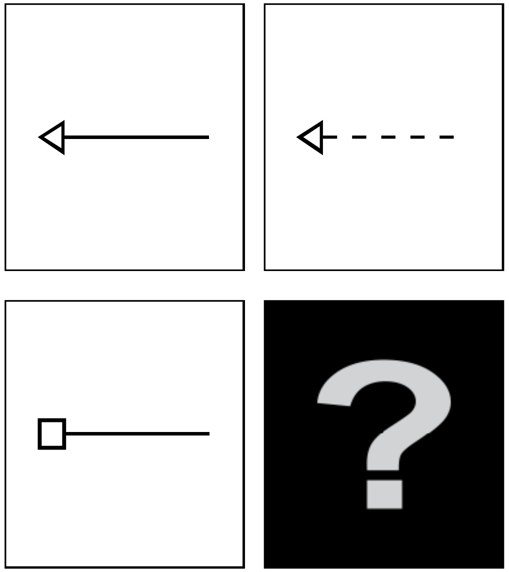


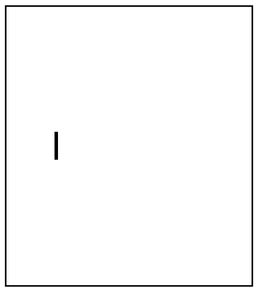


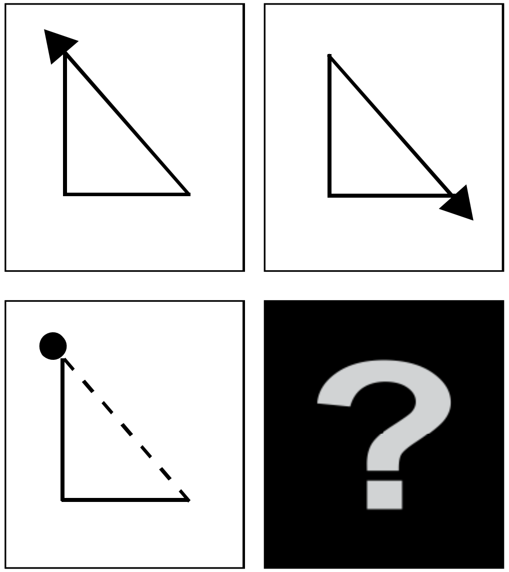


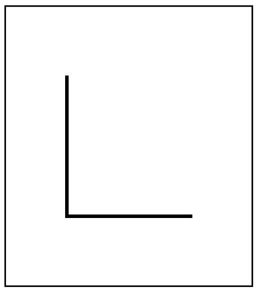


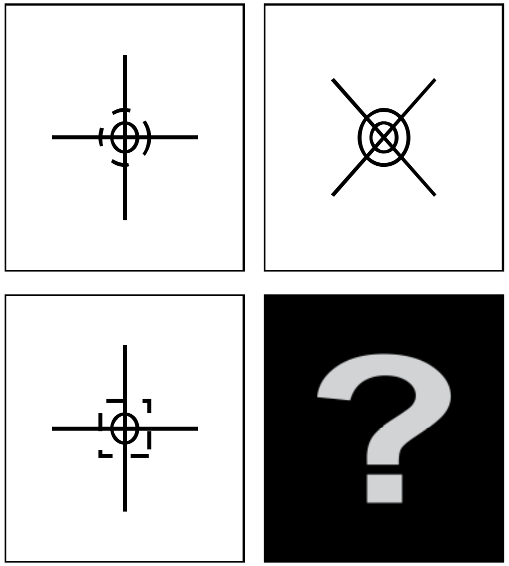
2C

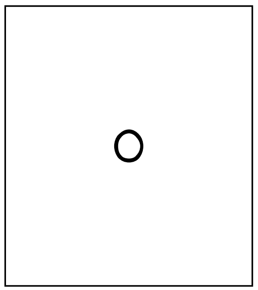


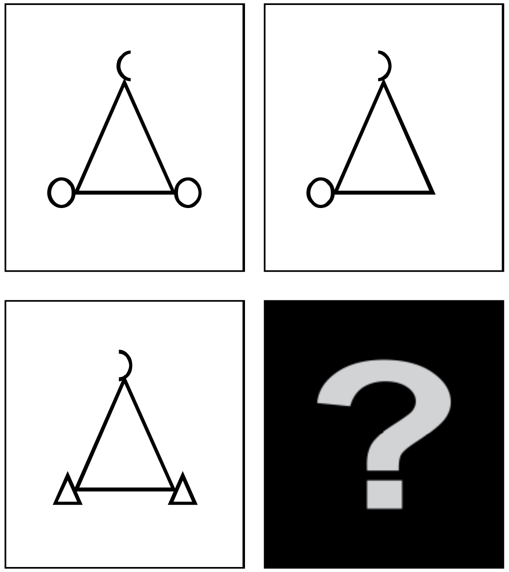


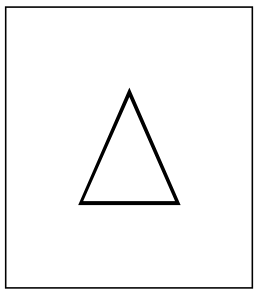


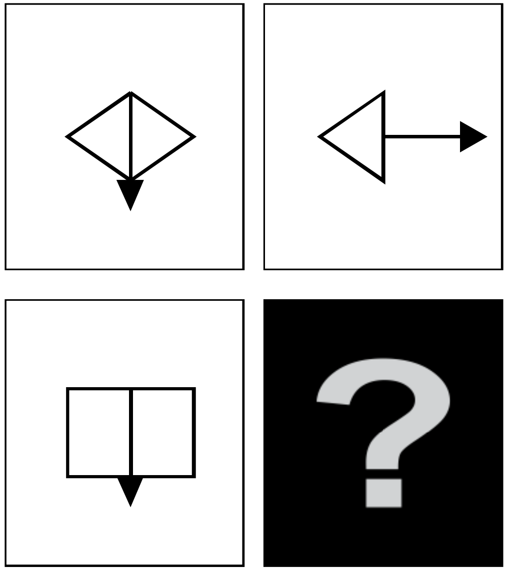


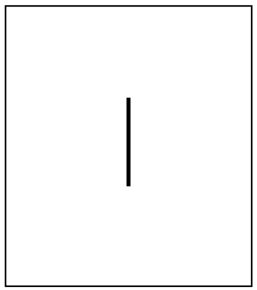


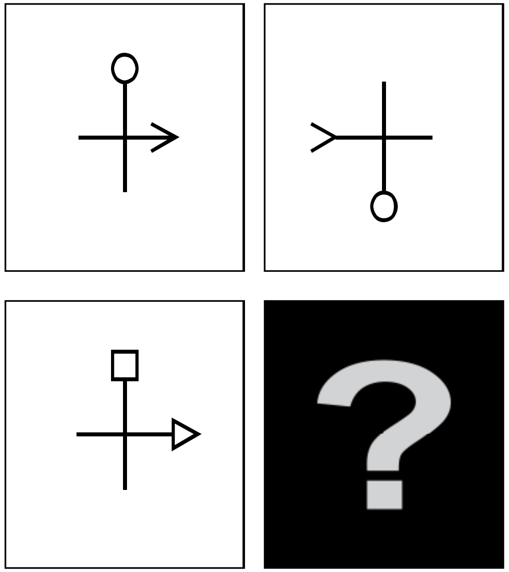


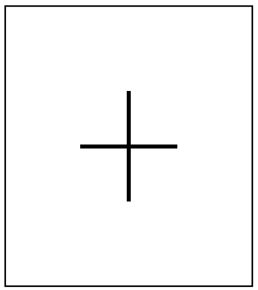


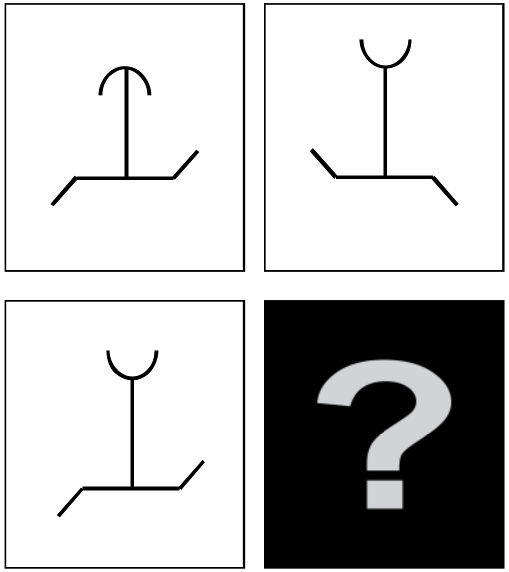


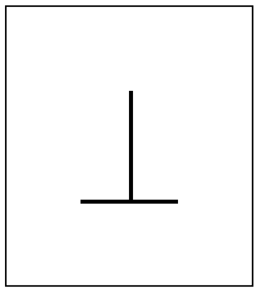




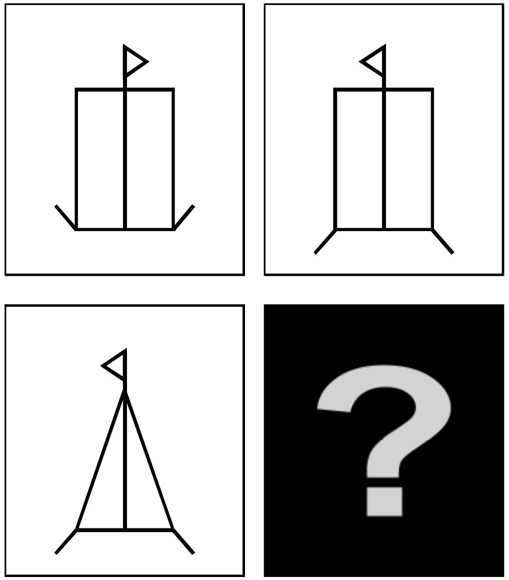


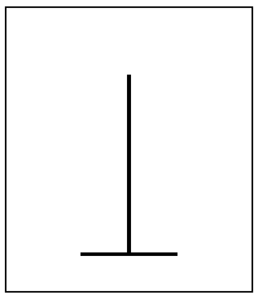


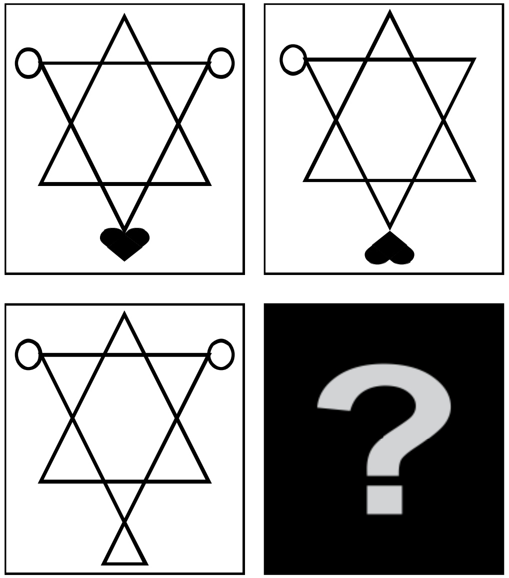


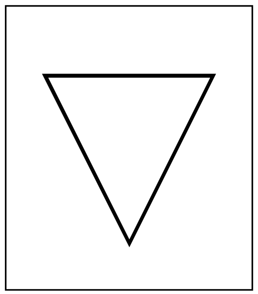


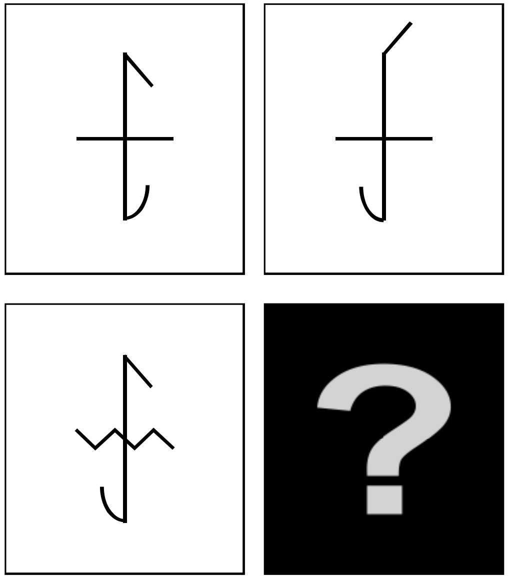
3C

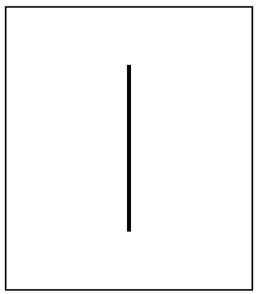


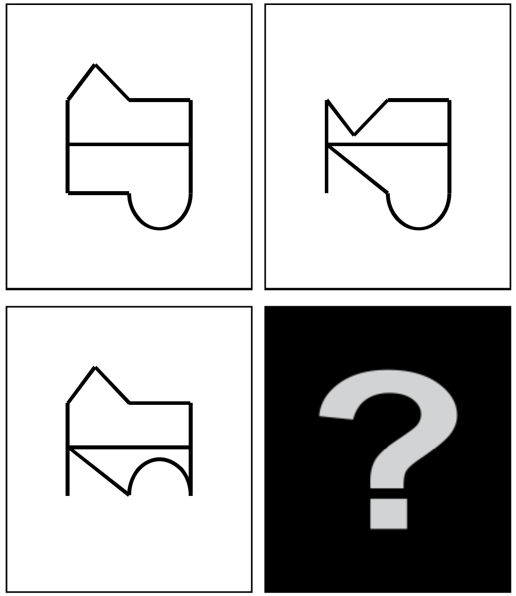


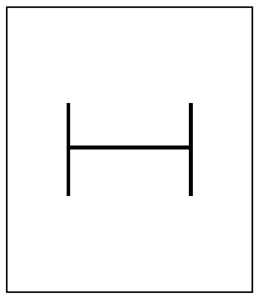


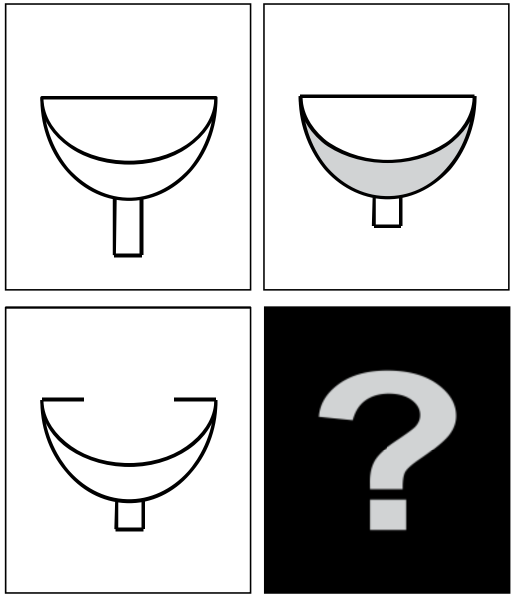


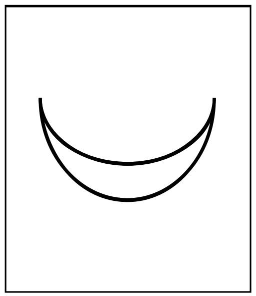


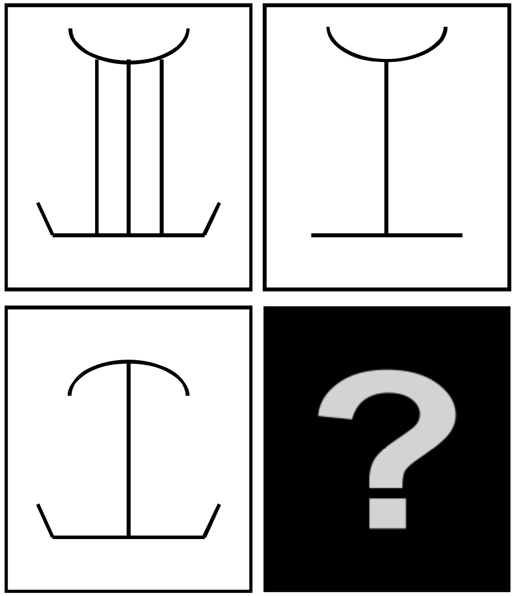


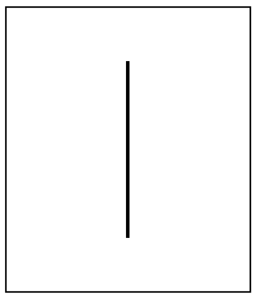


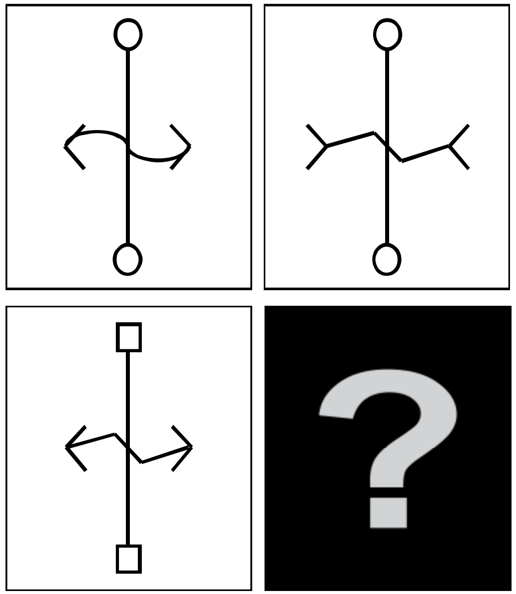


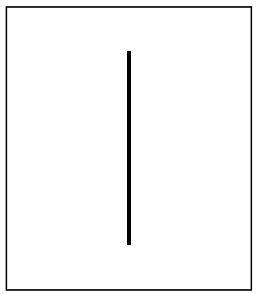


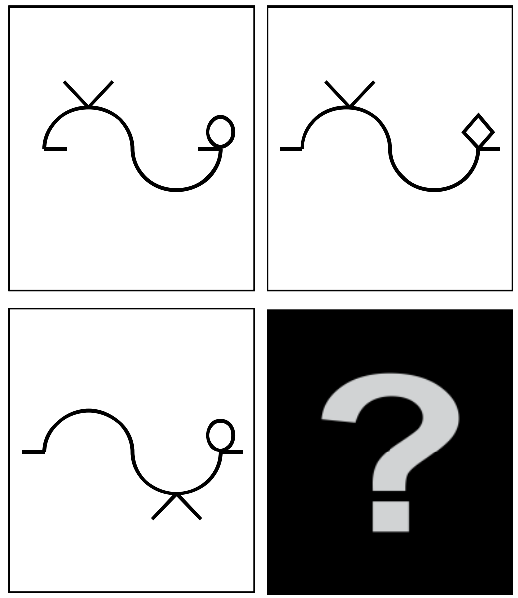


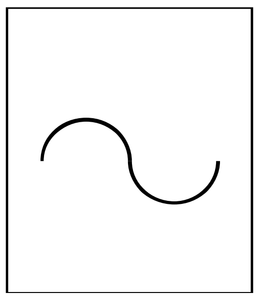


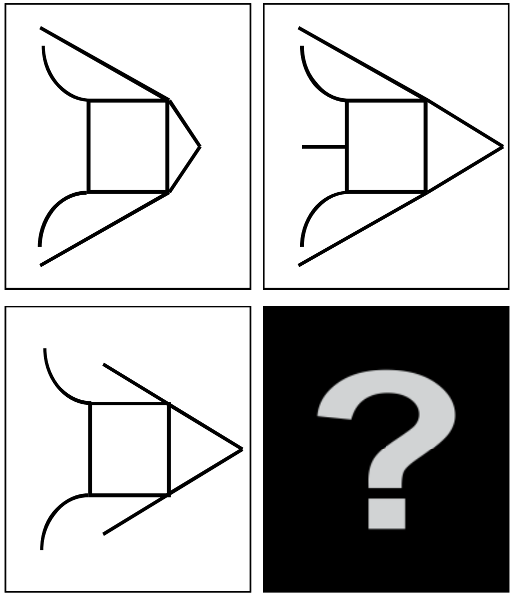


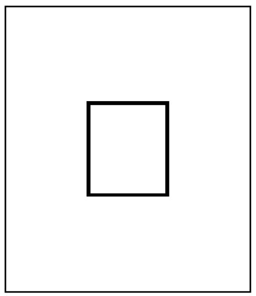


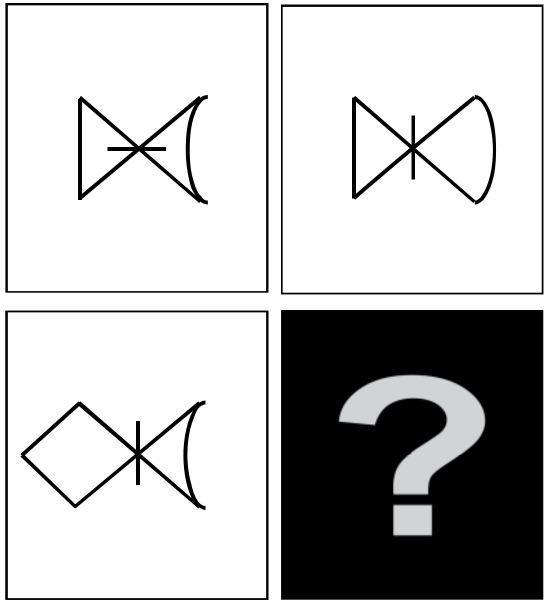


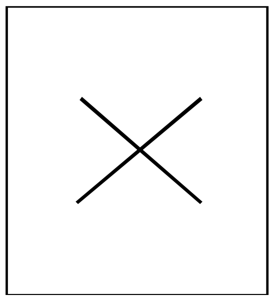




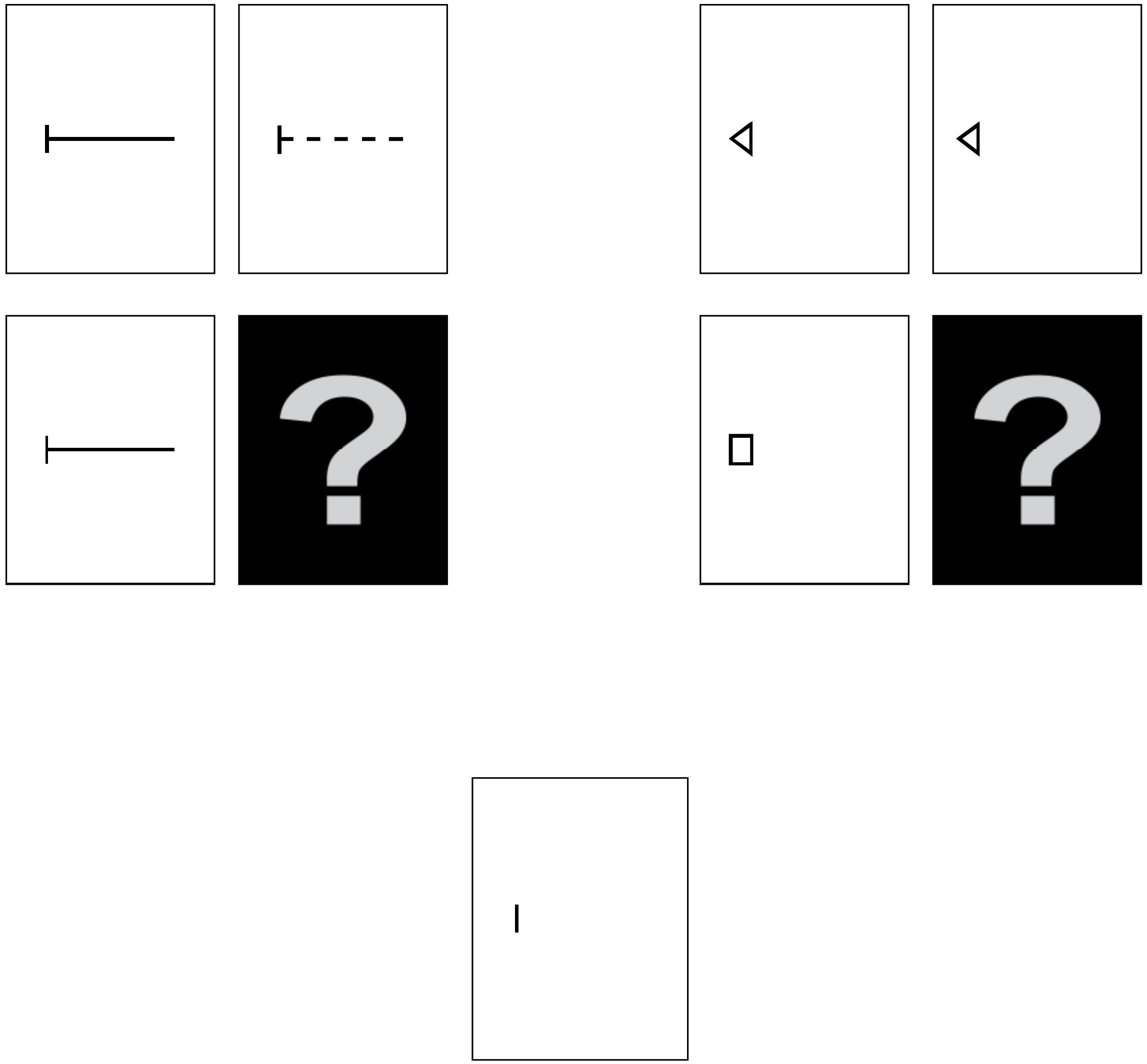


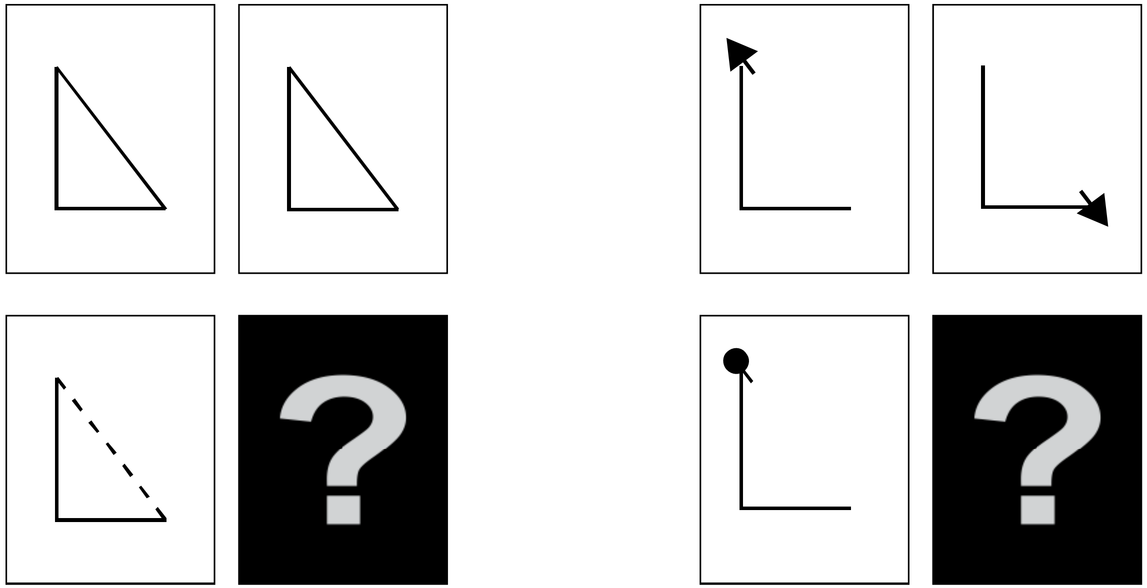


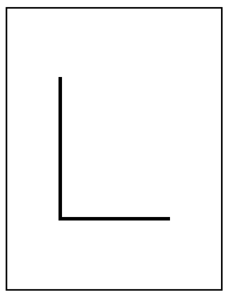


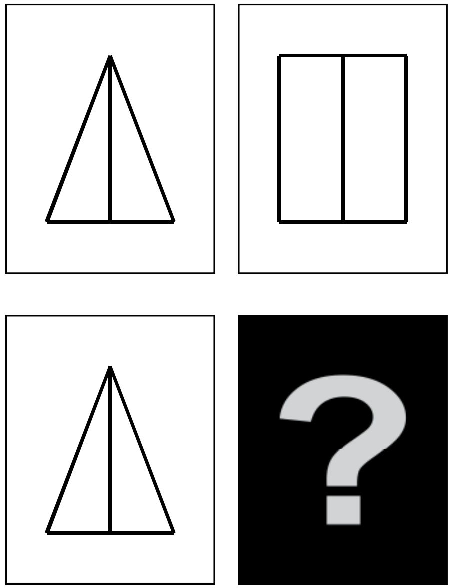


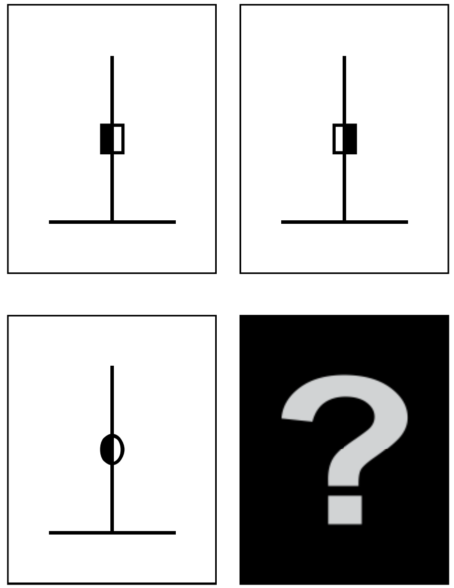
2S

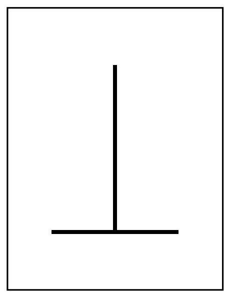


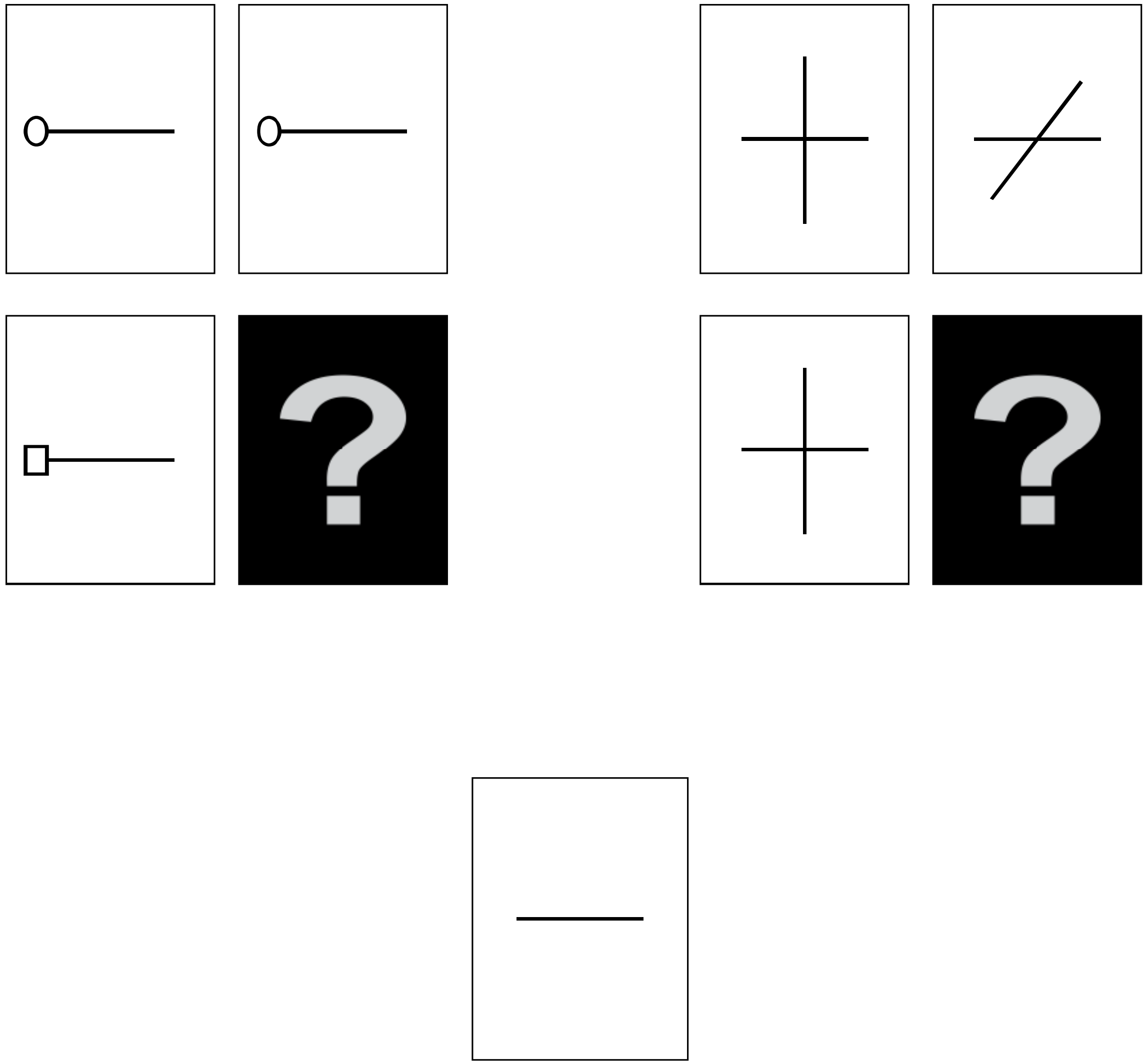


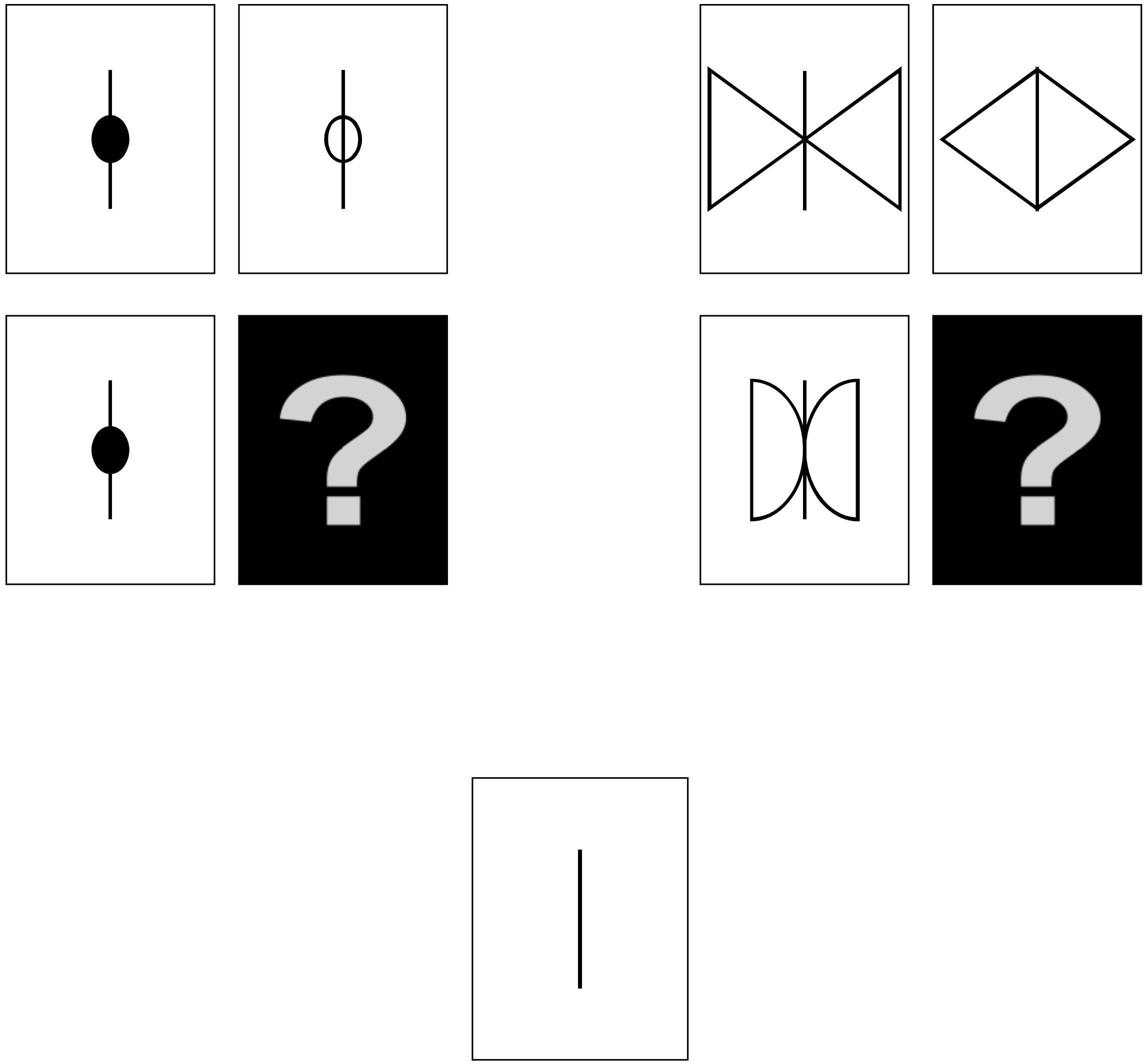


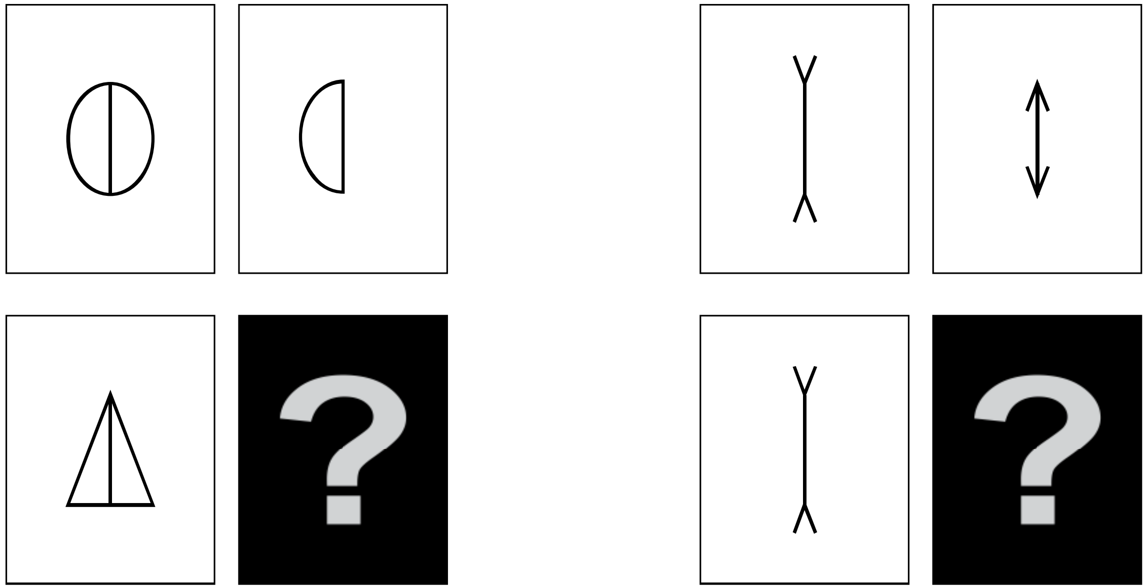


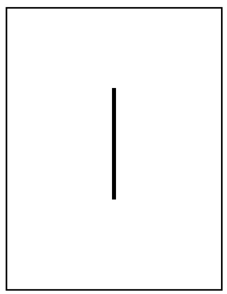


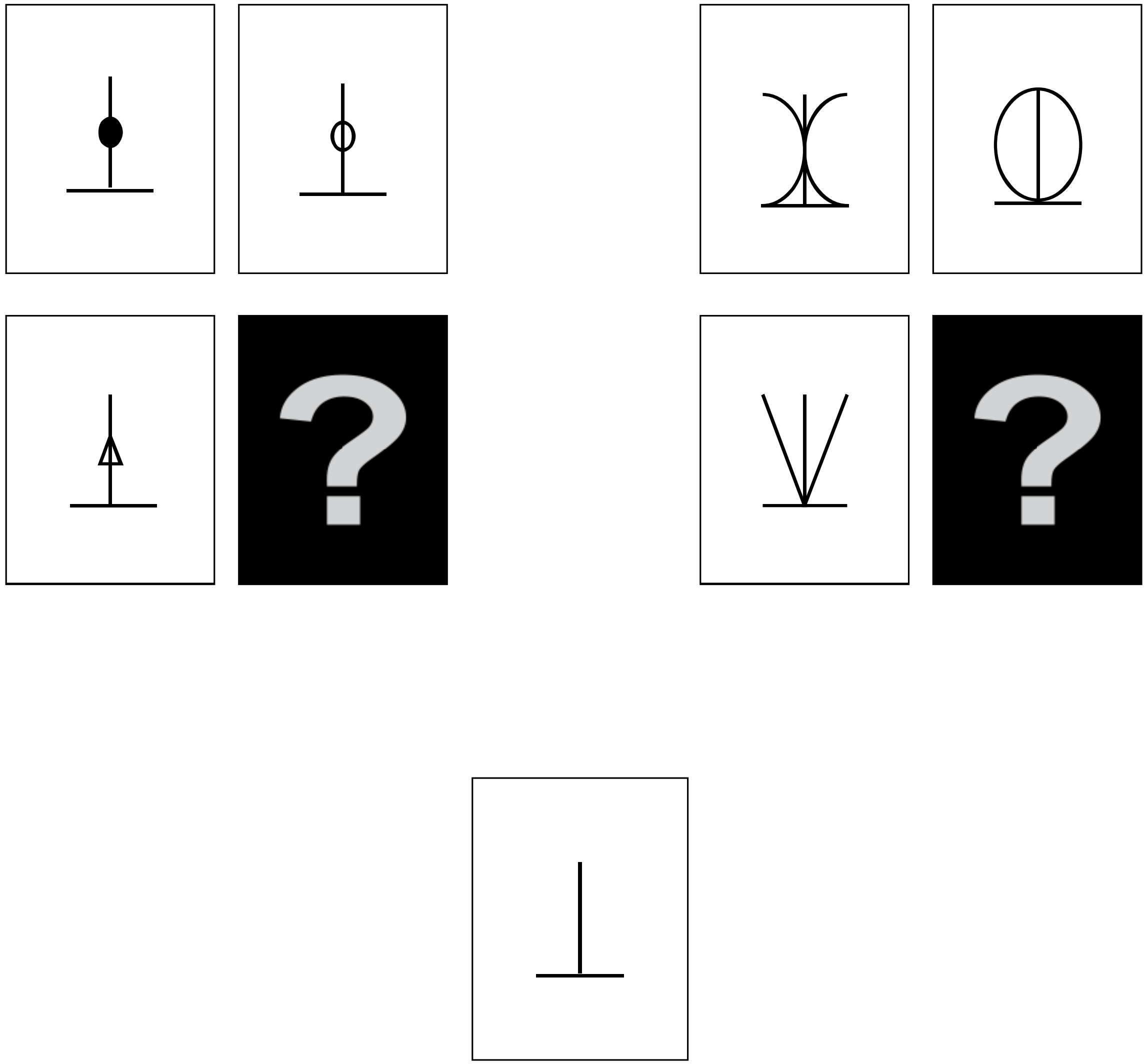




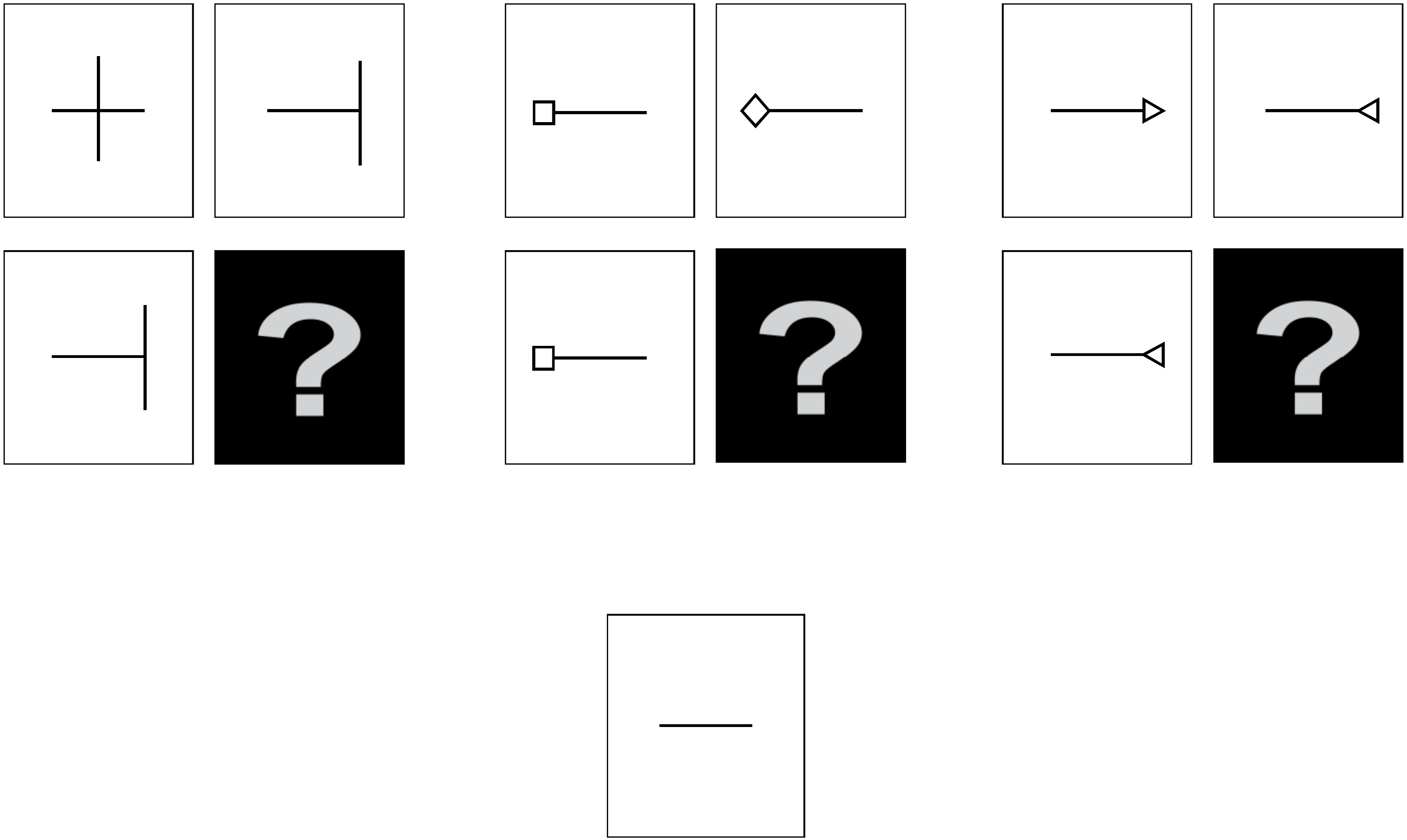


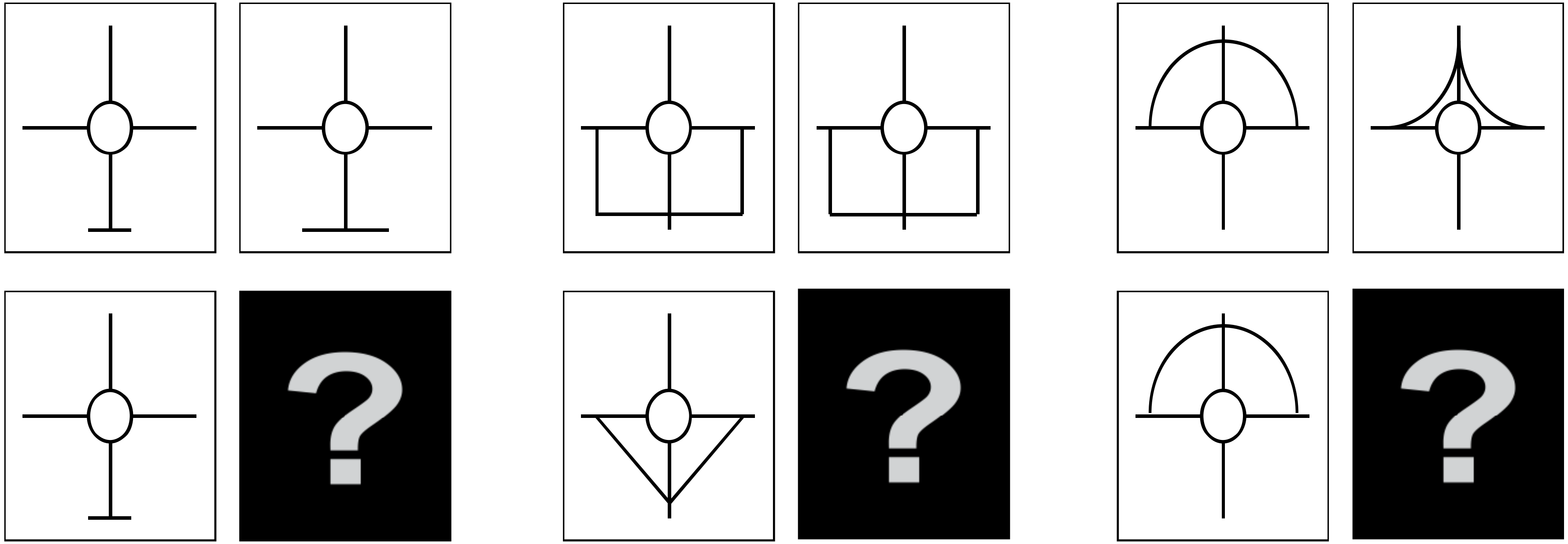


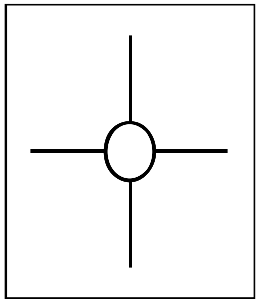


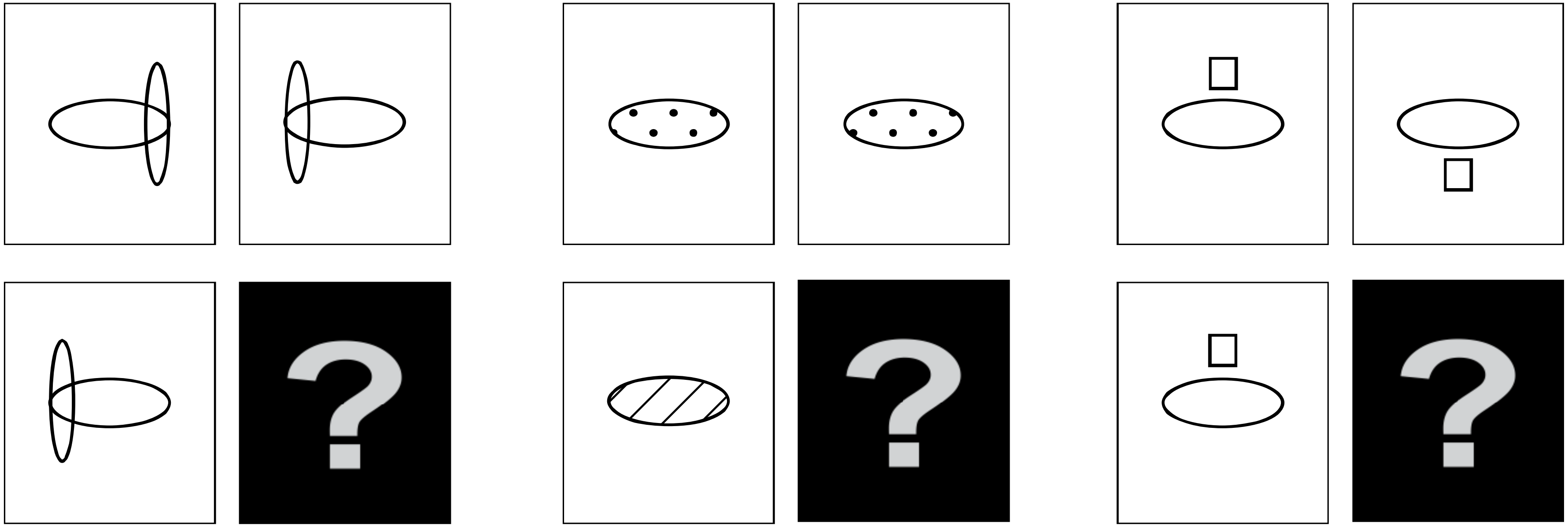


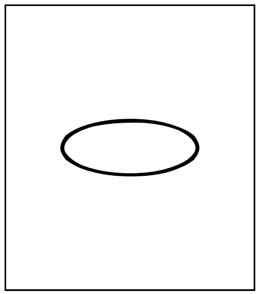
3S

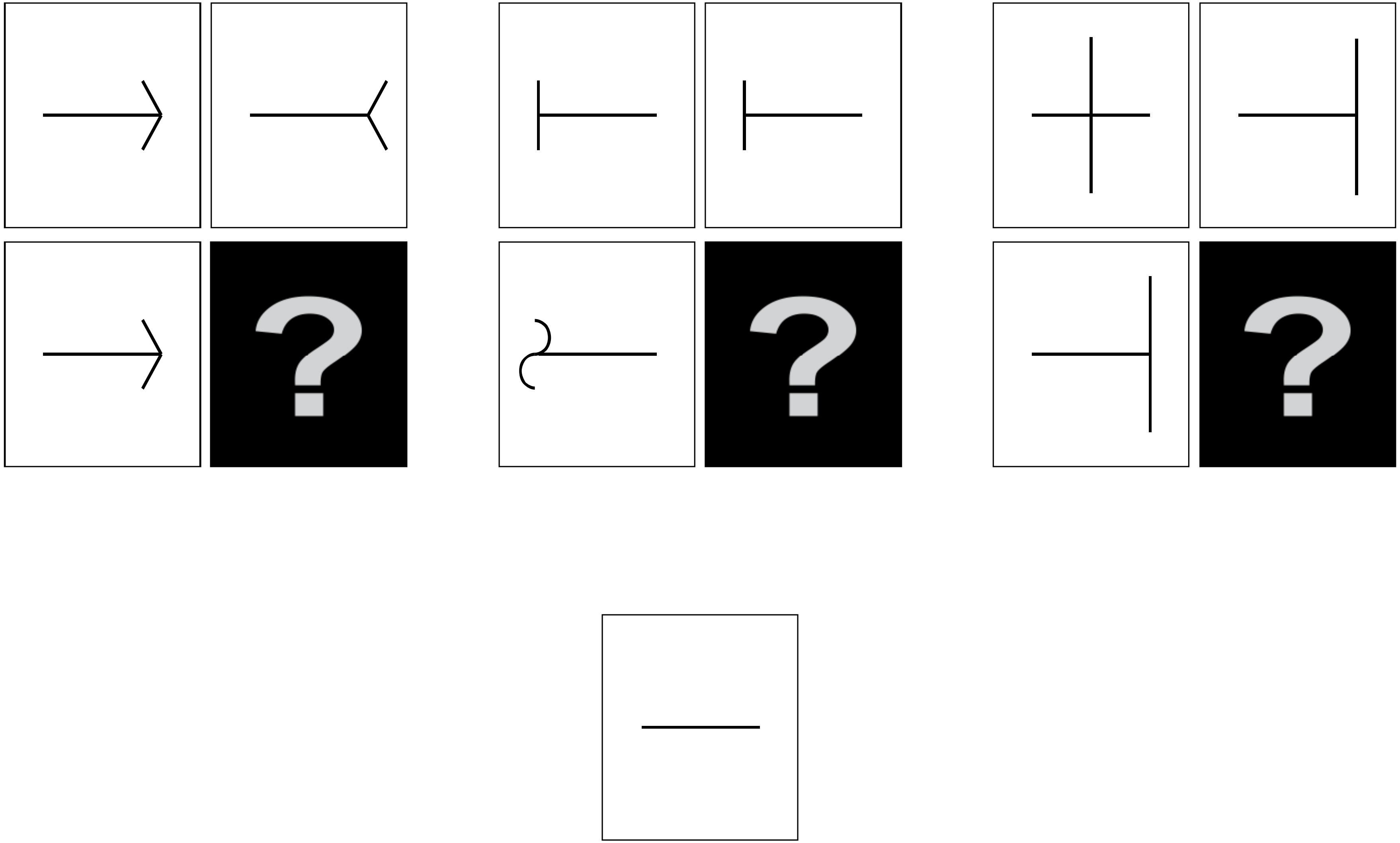


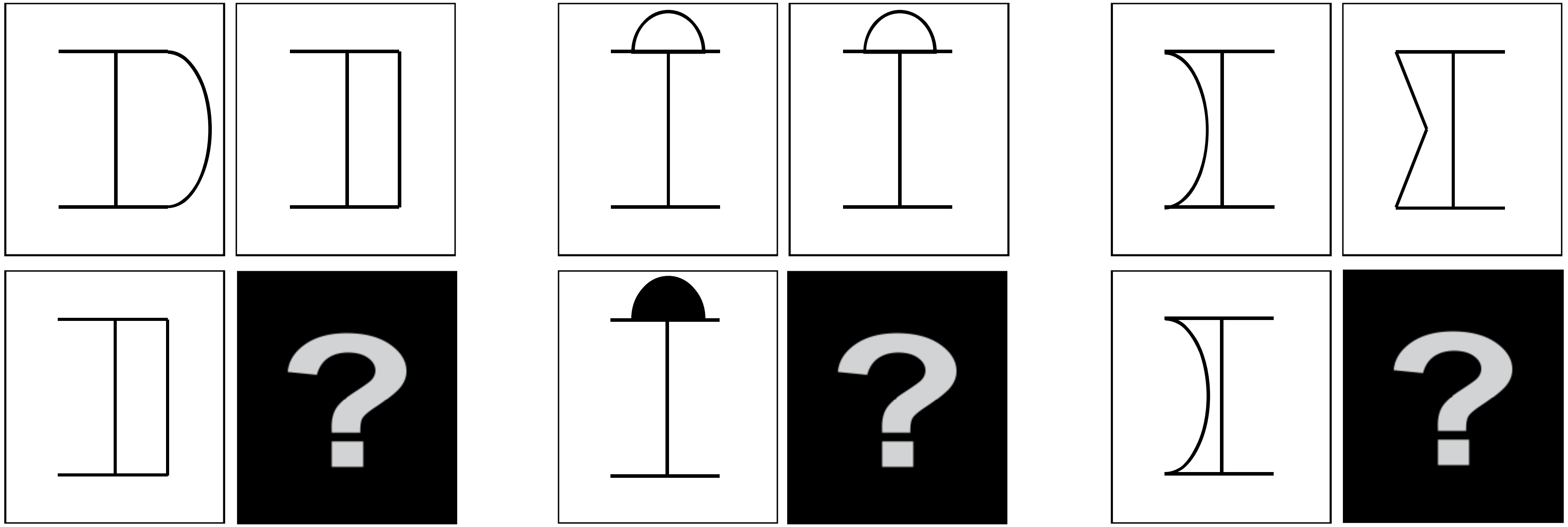


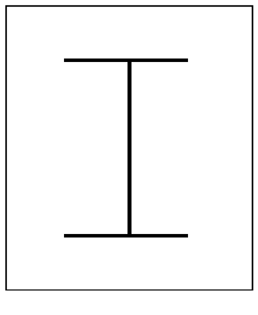


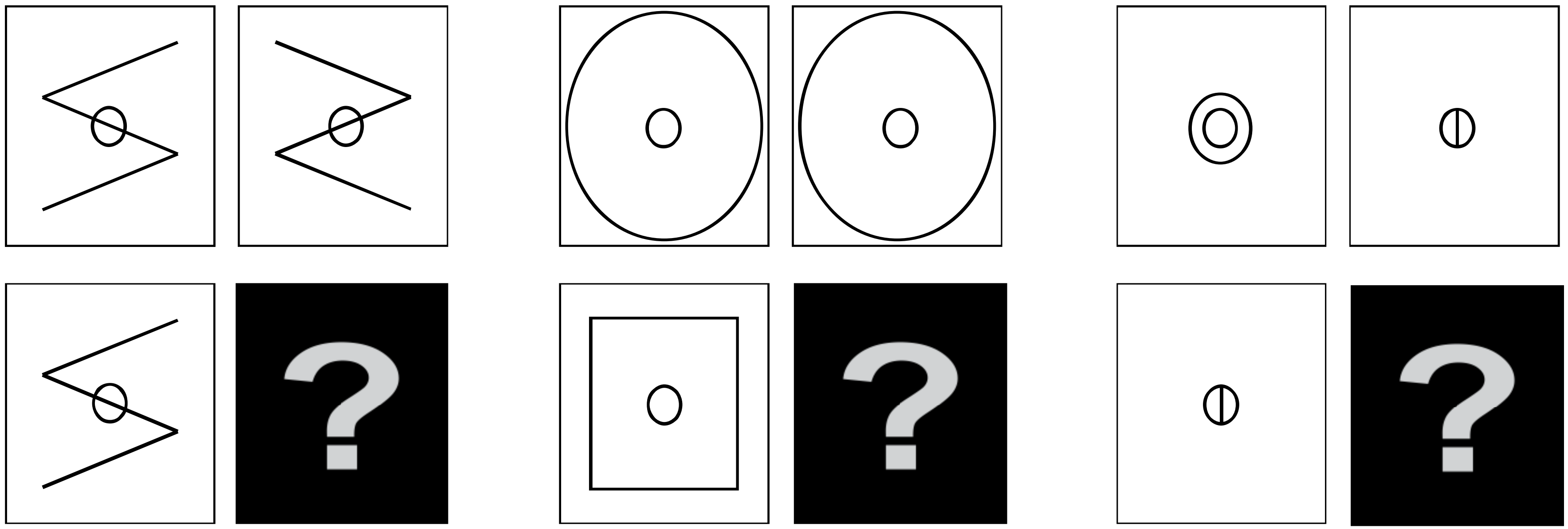


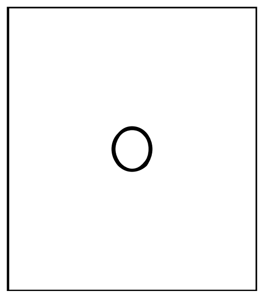


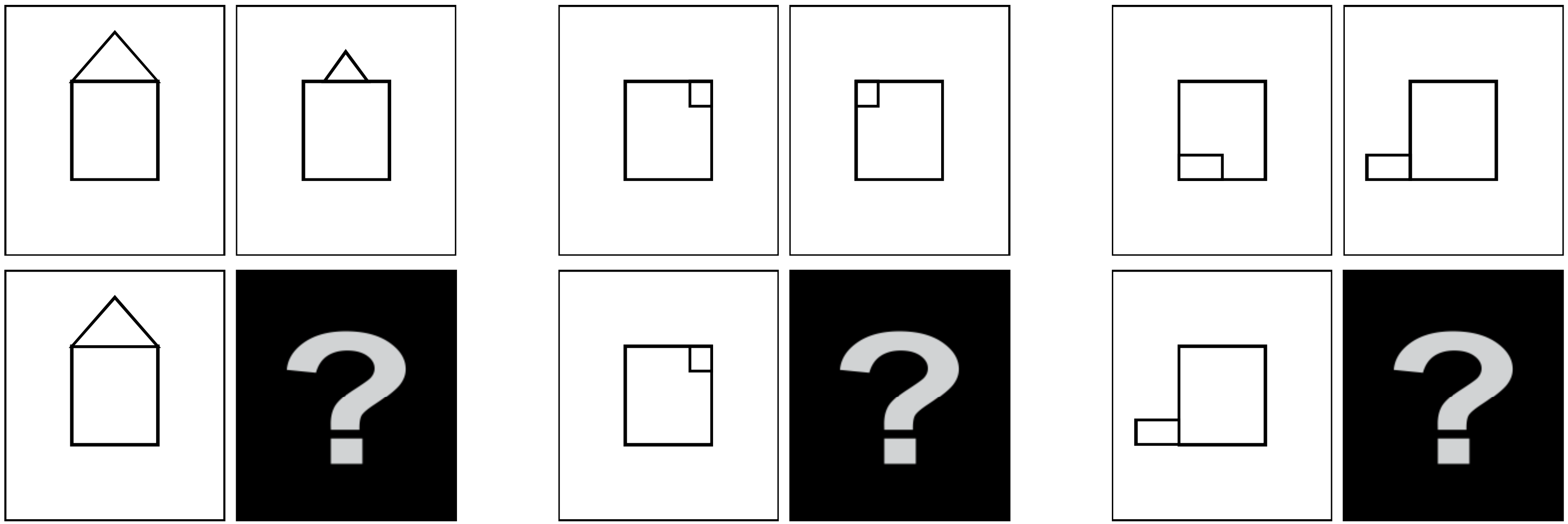


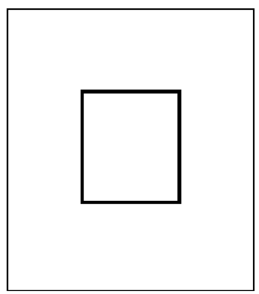


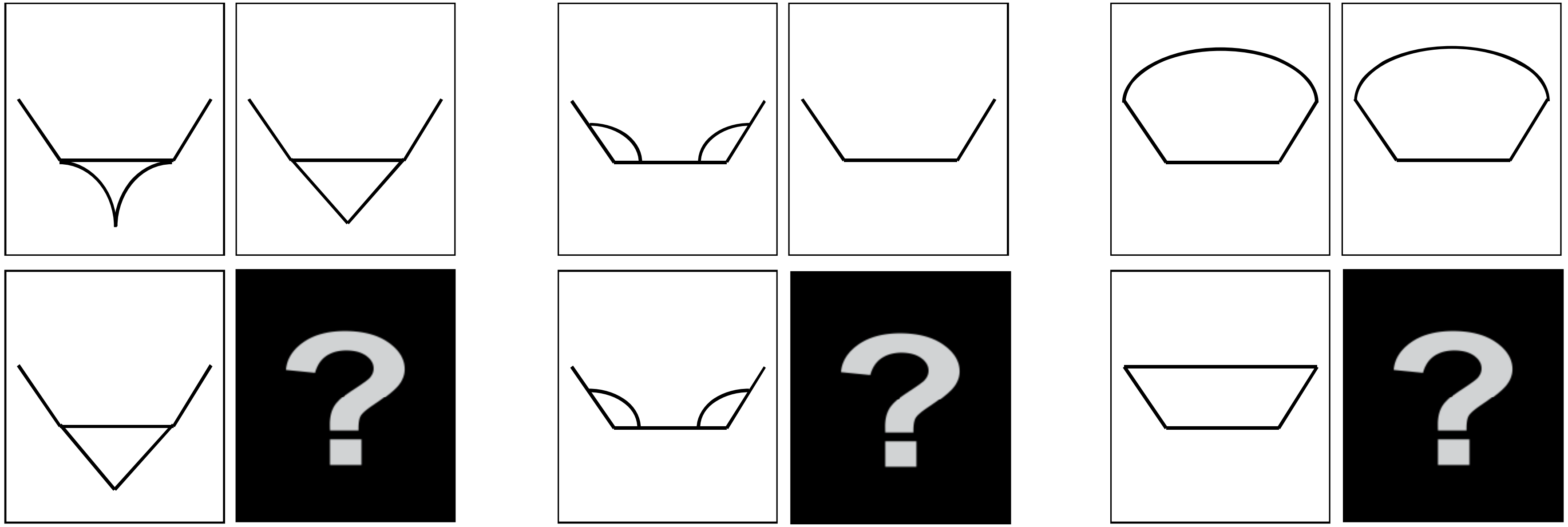


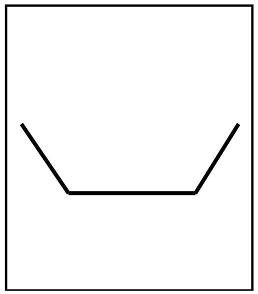


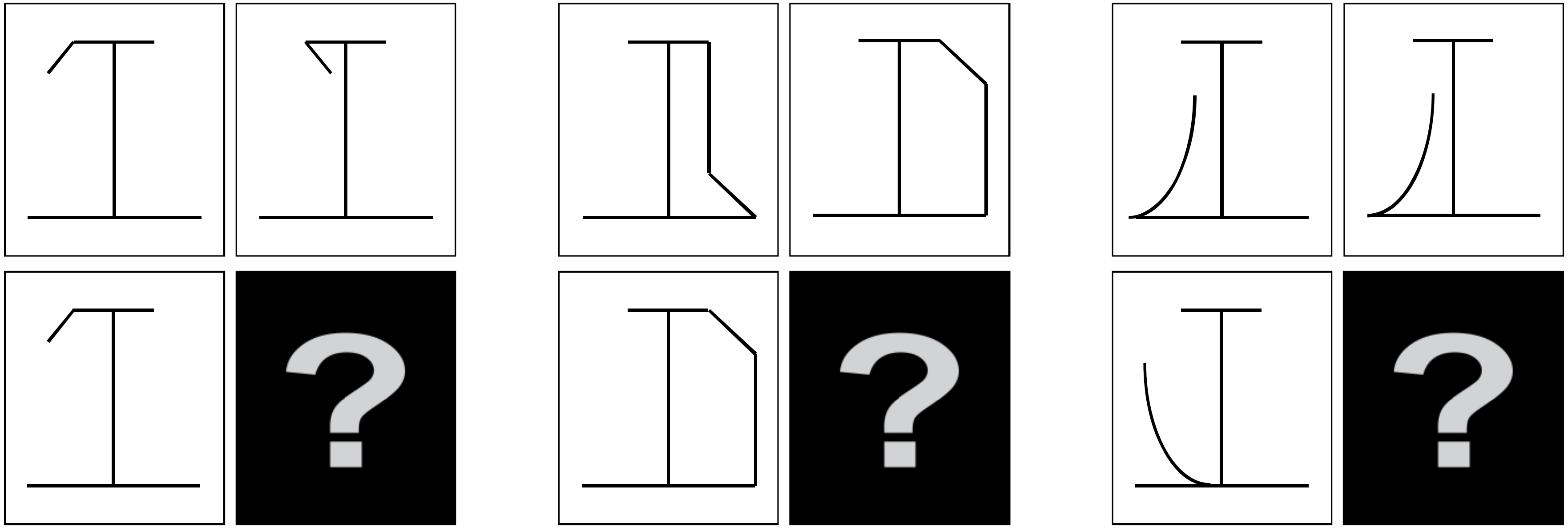


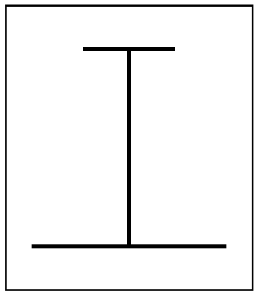


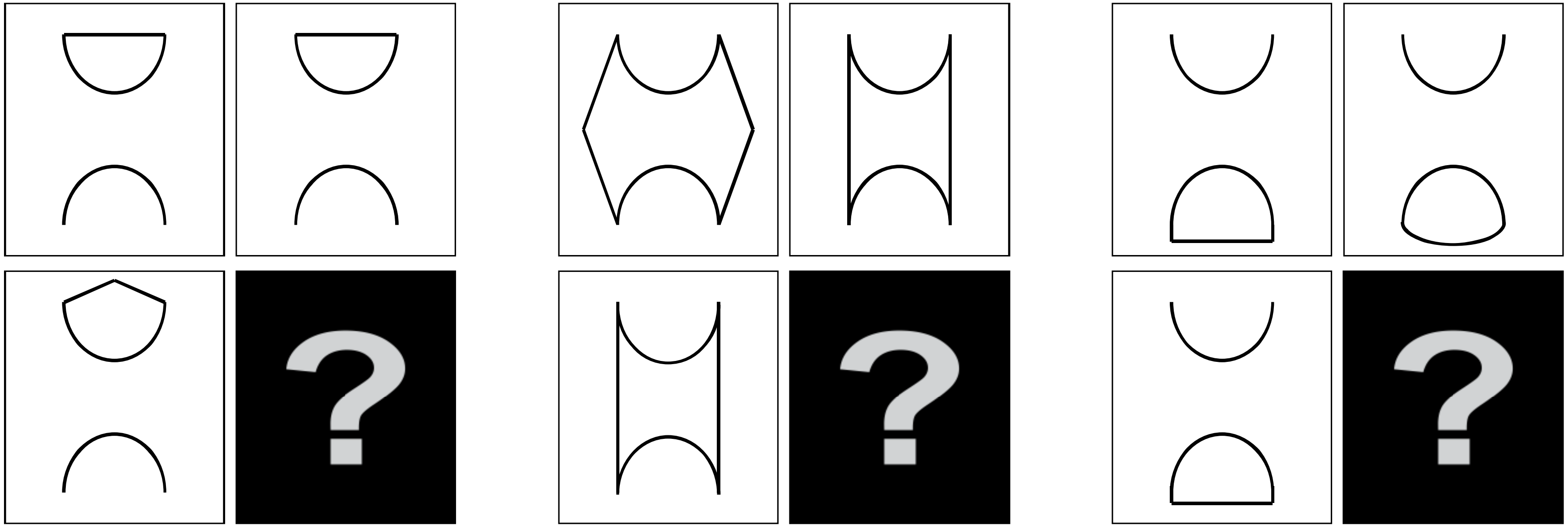


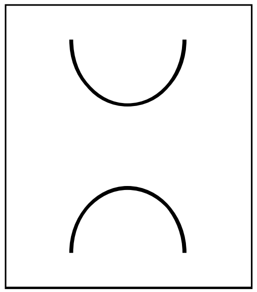






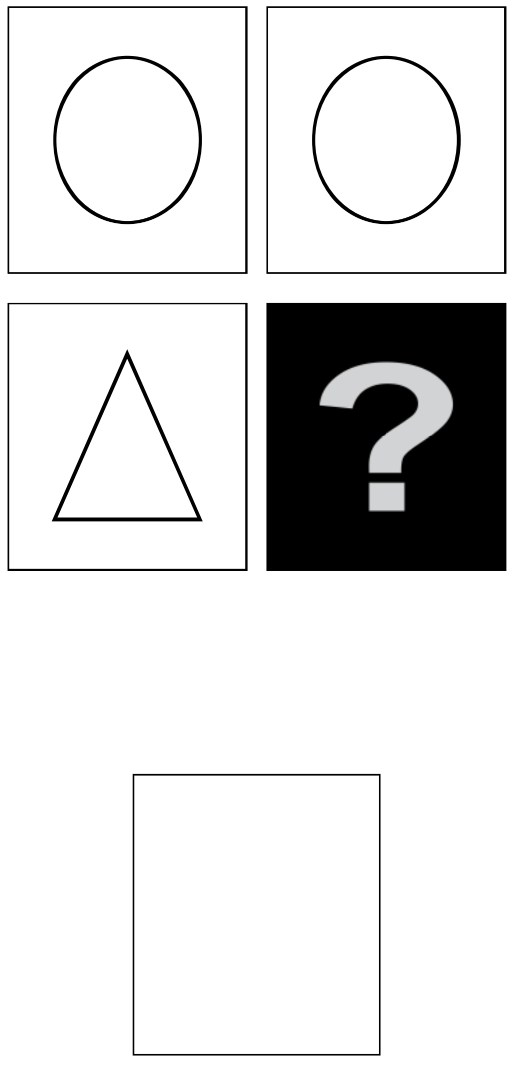


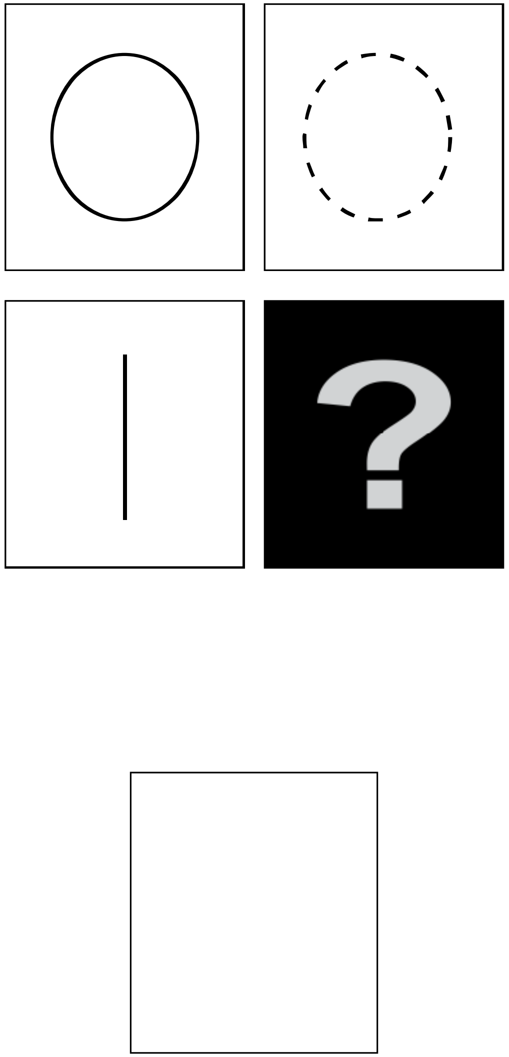


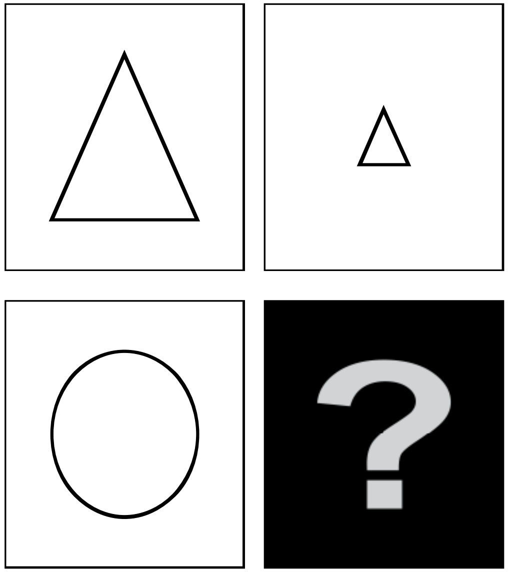


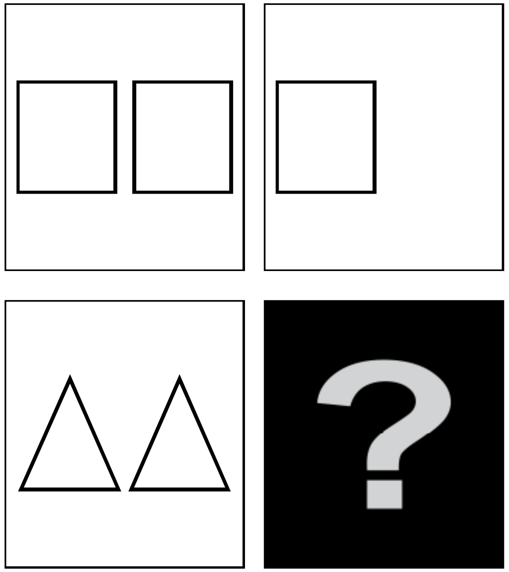
A2

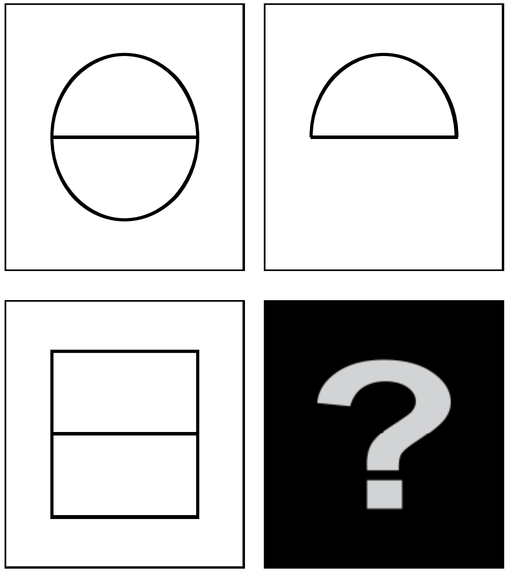
ID: ______________________

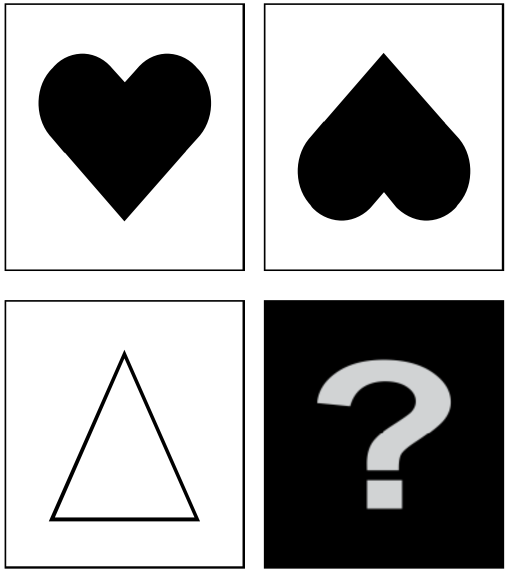


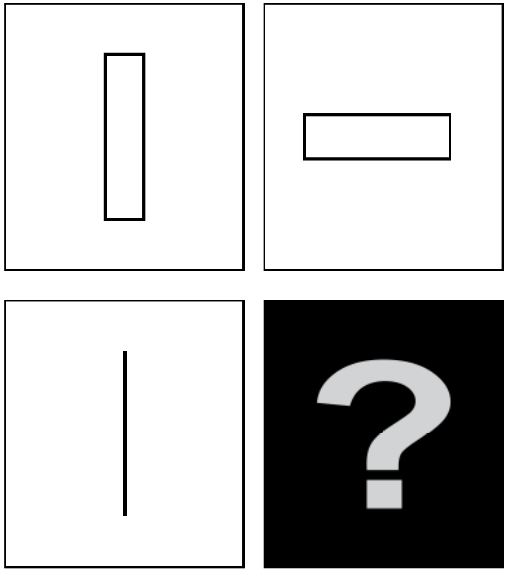


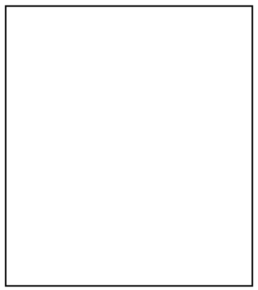




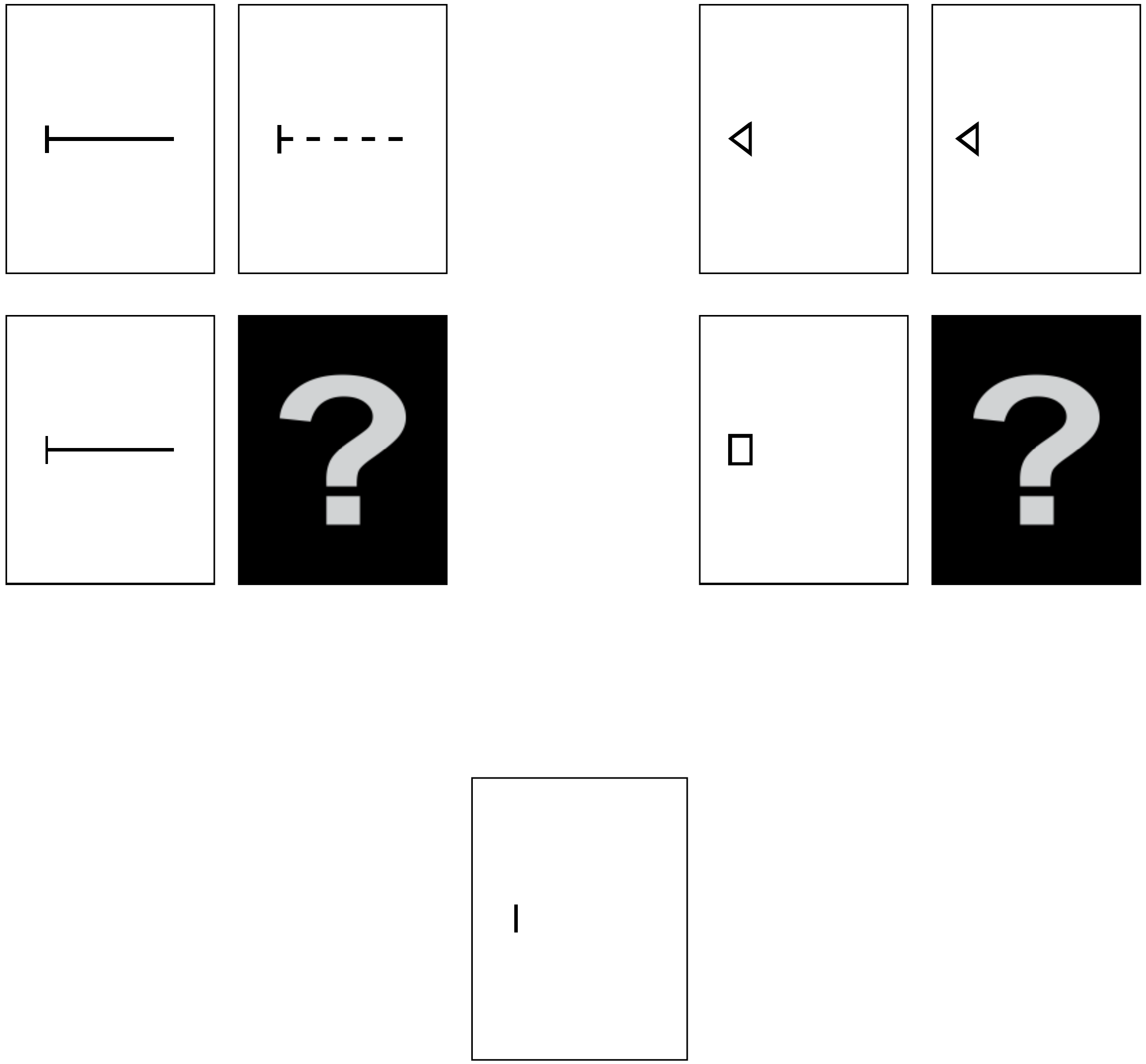


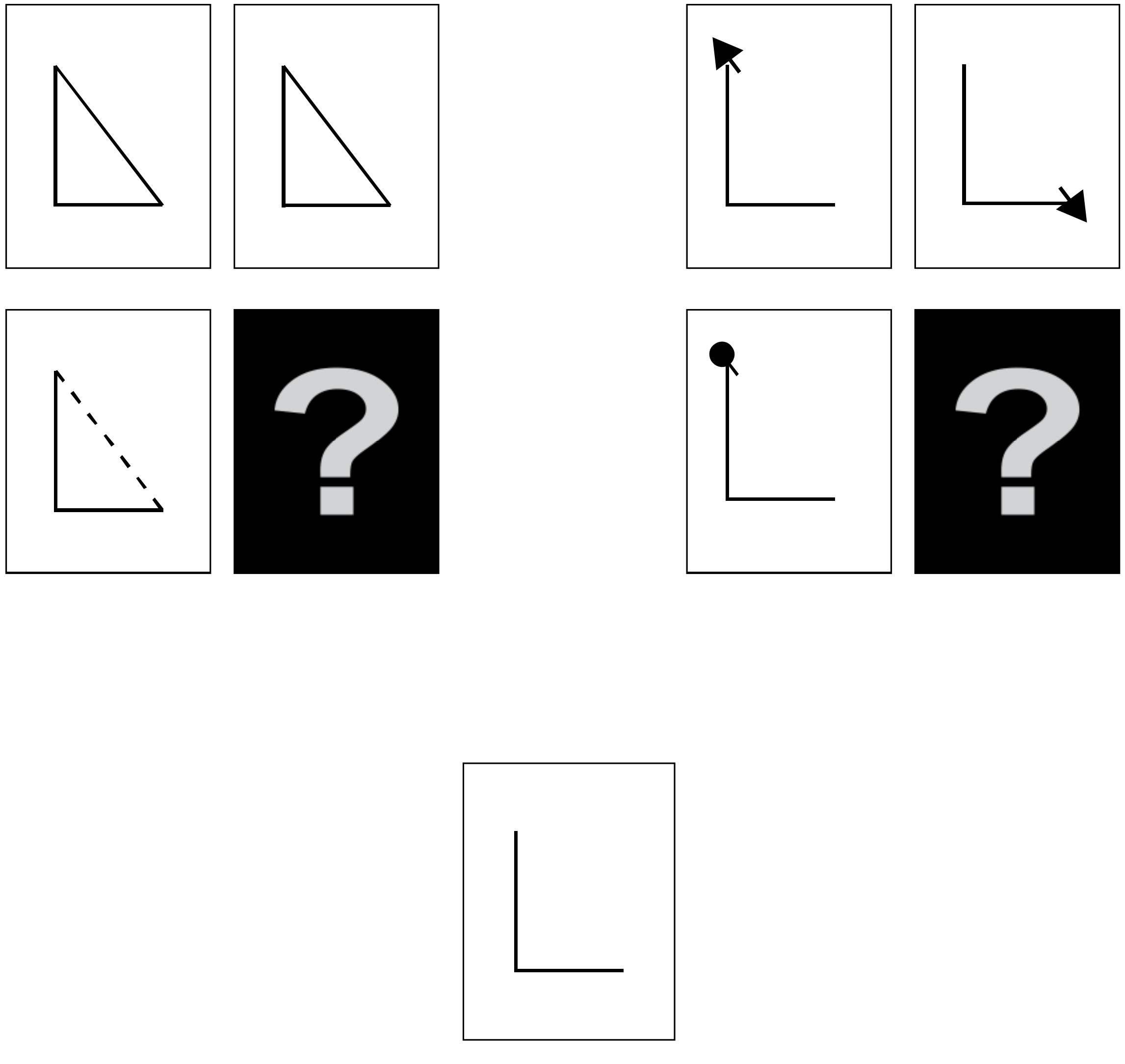


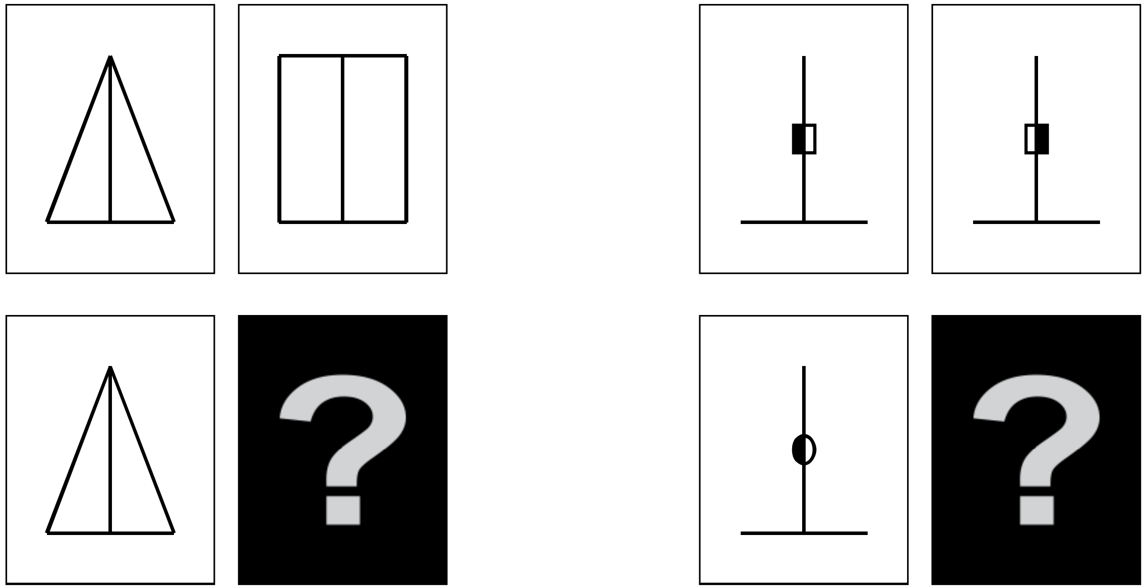


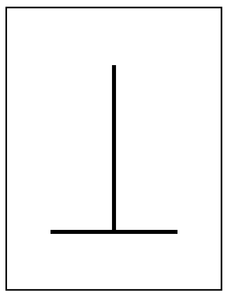


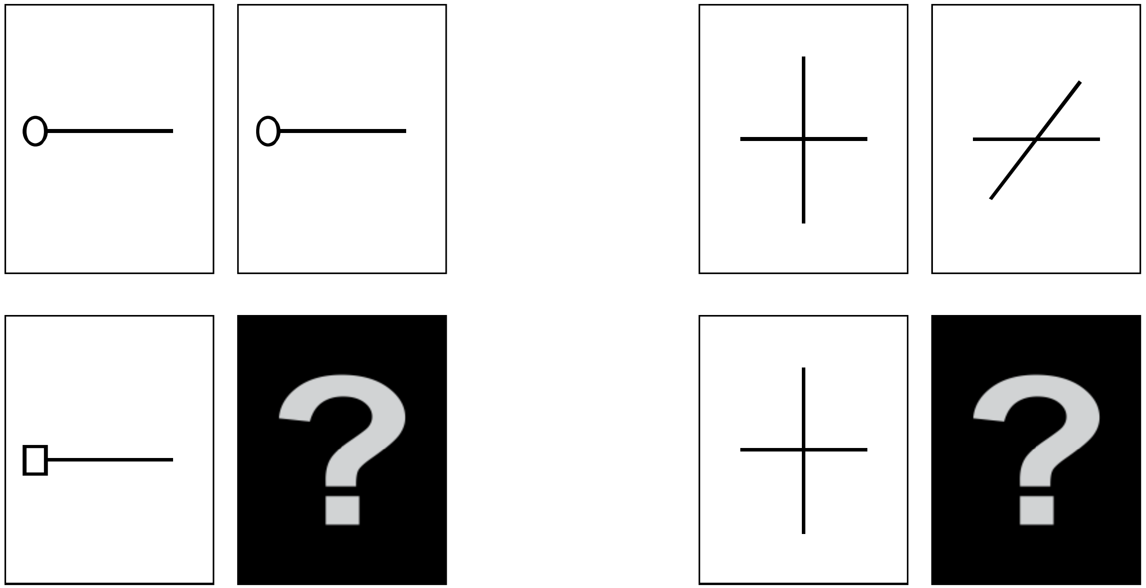
2S

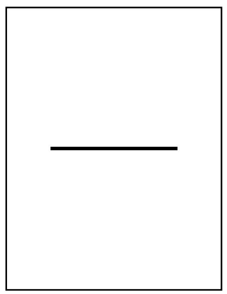


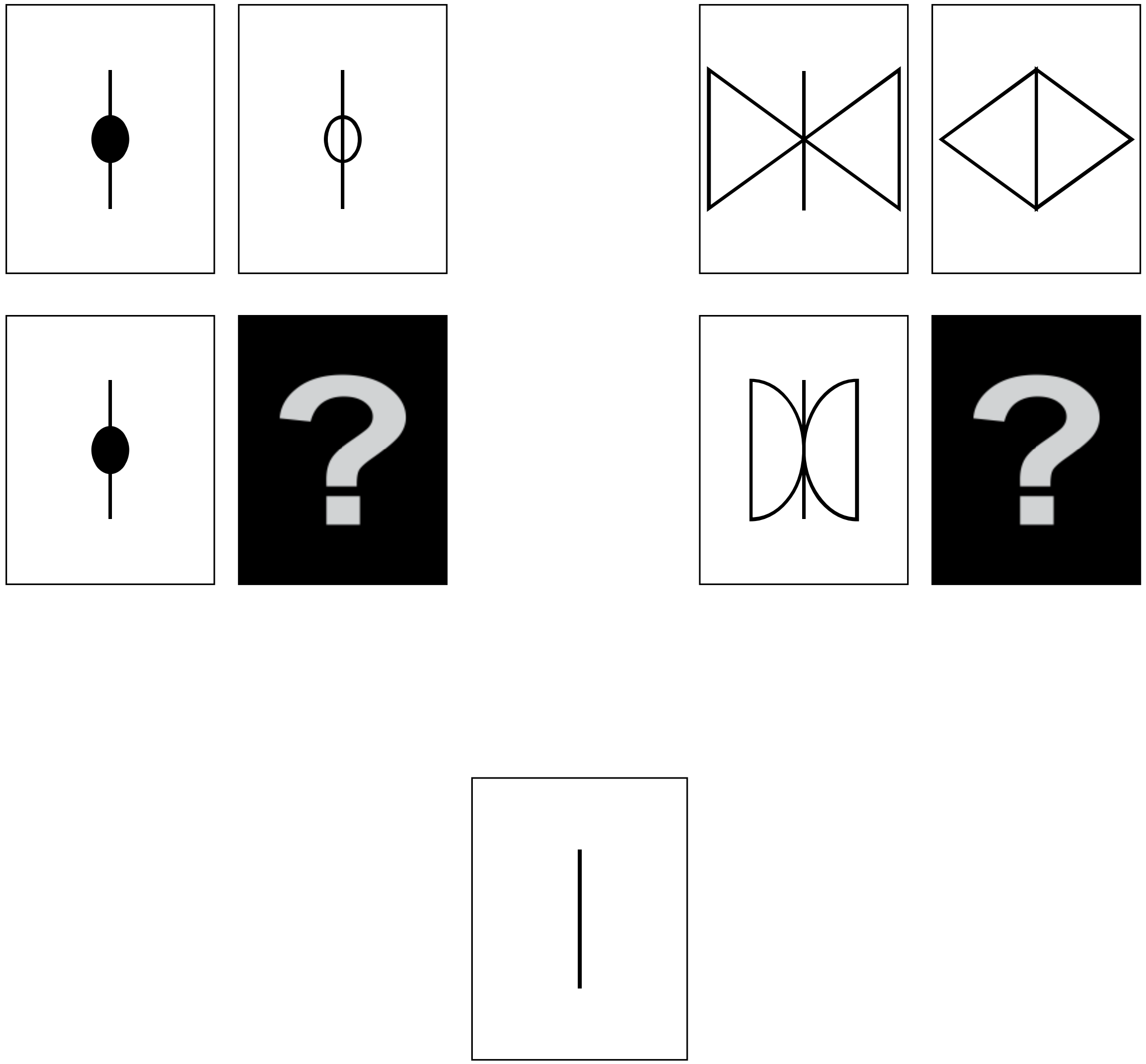


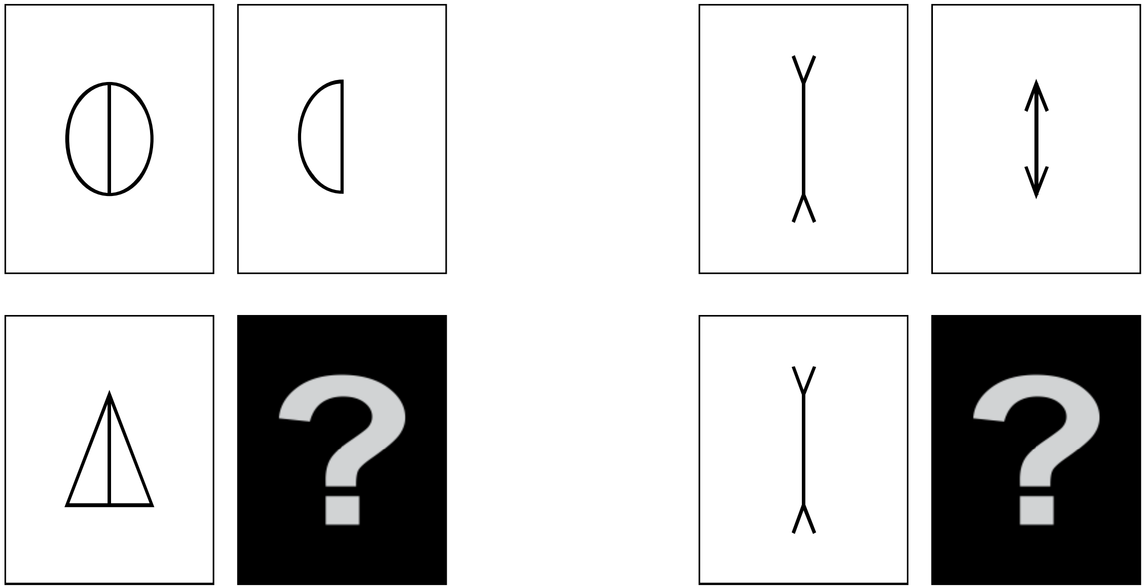


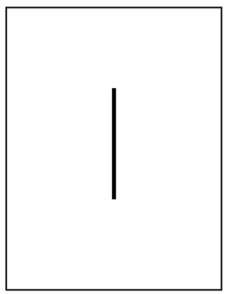


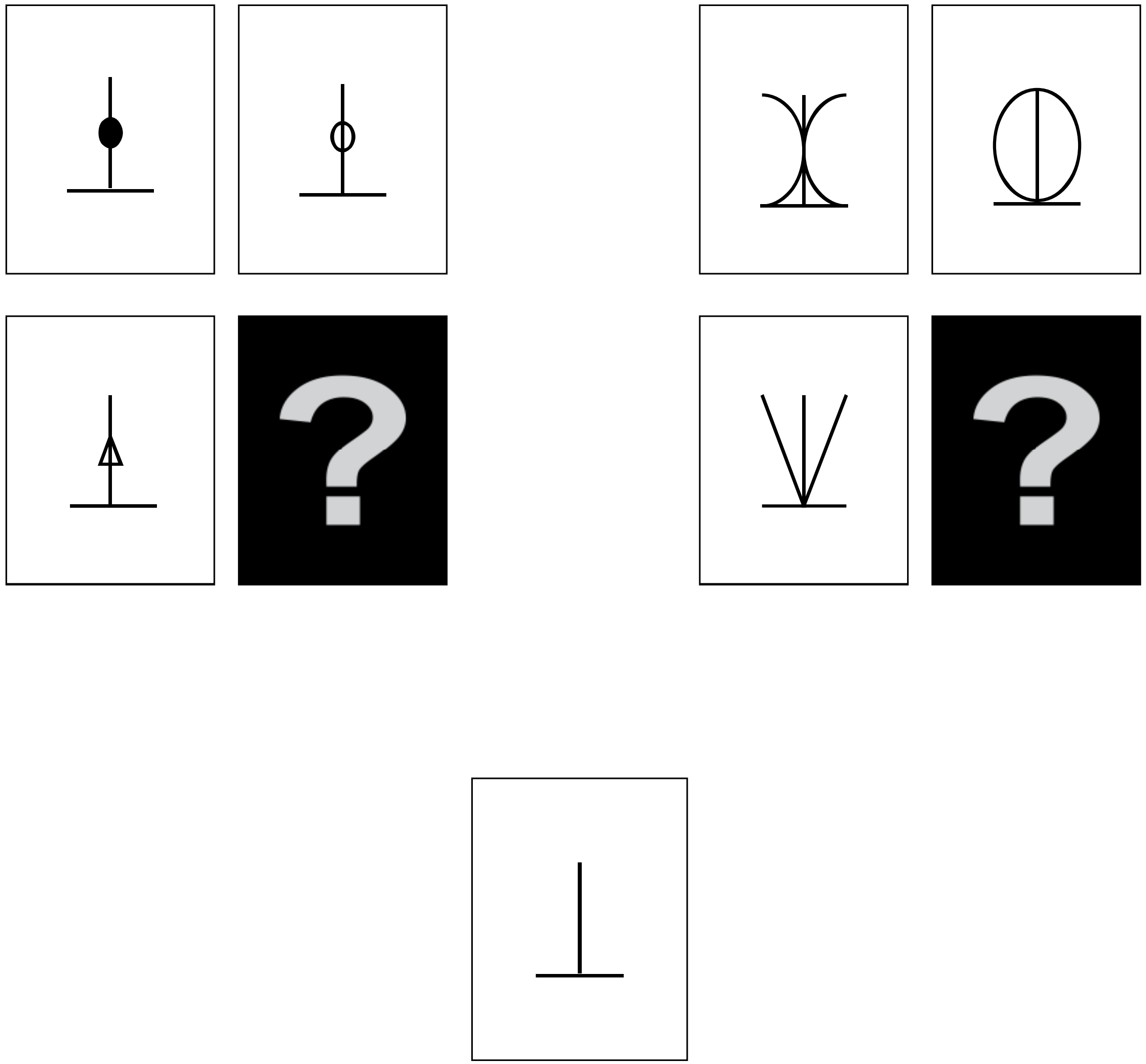




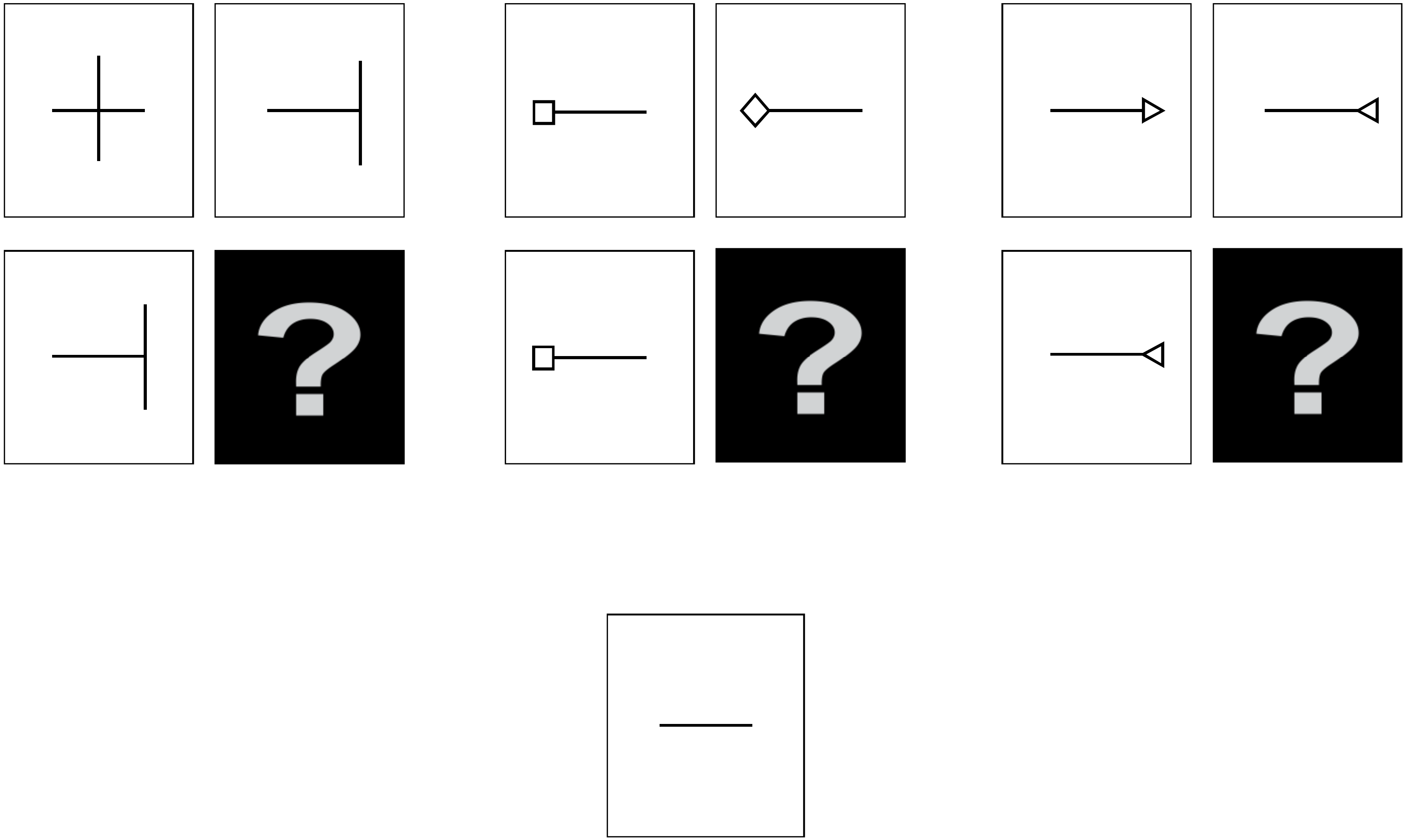


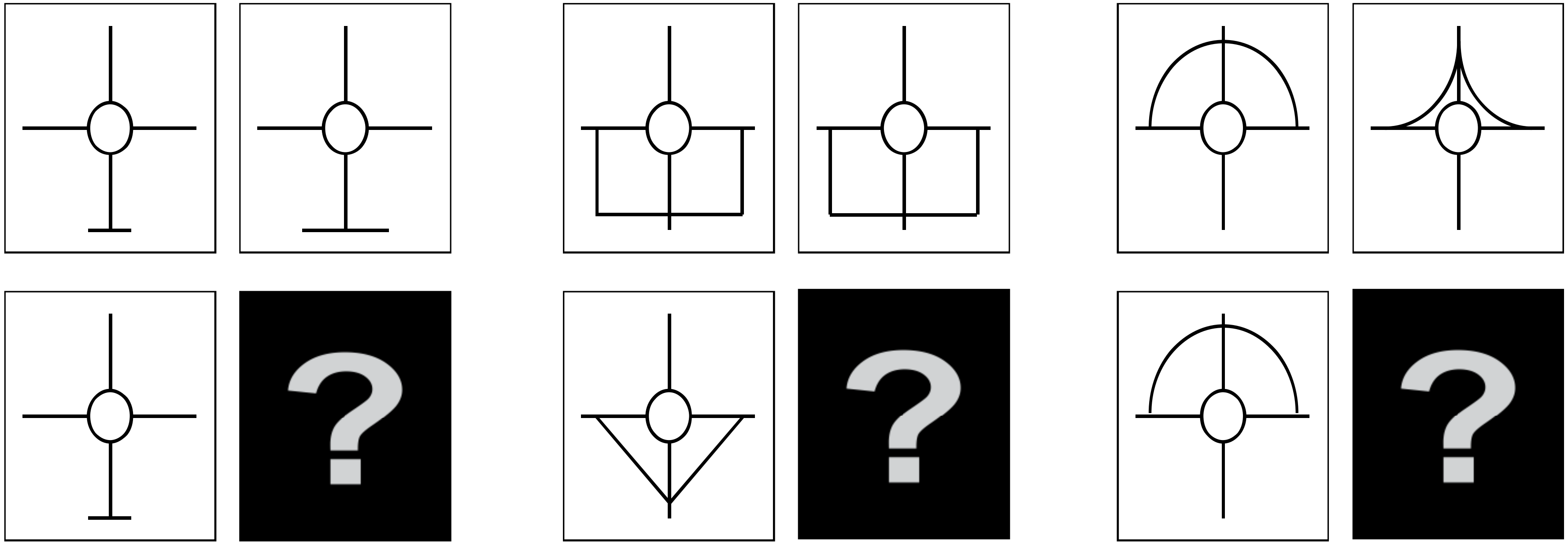


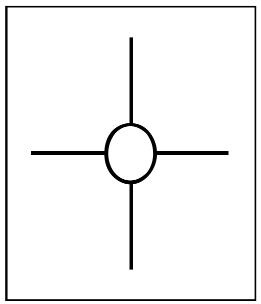


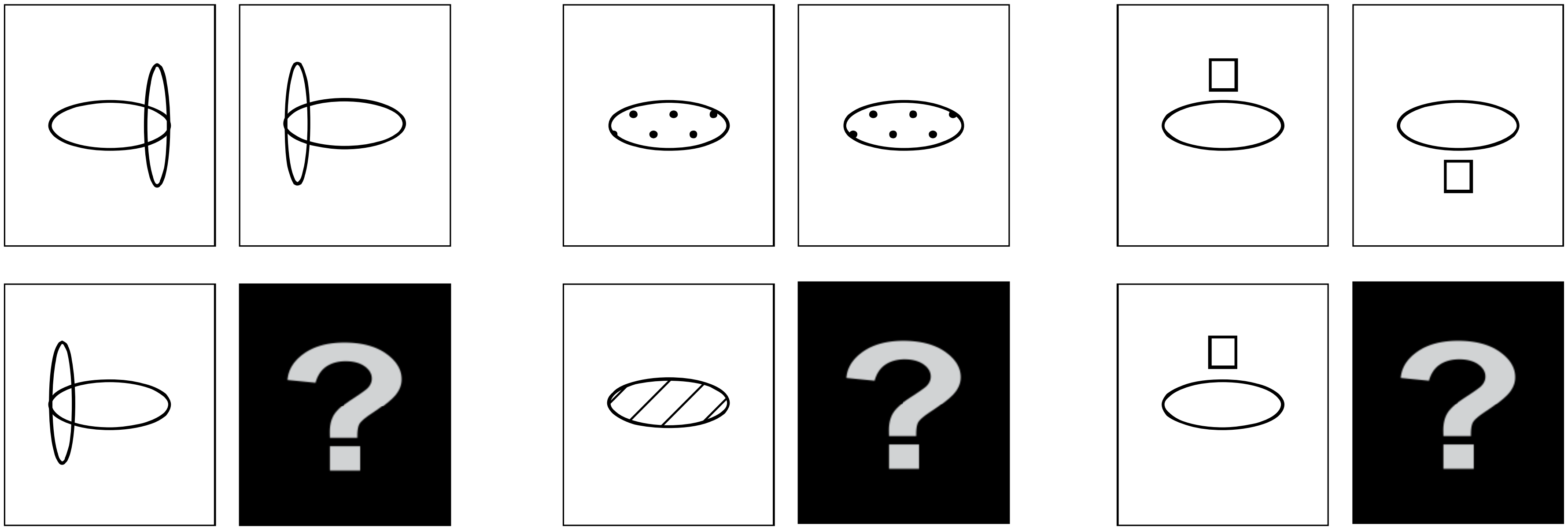


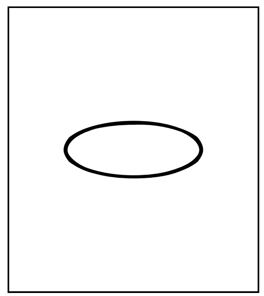
3S

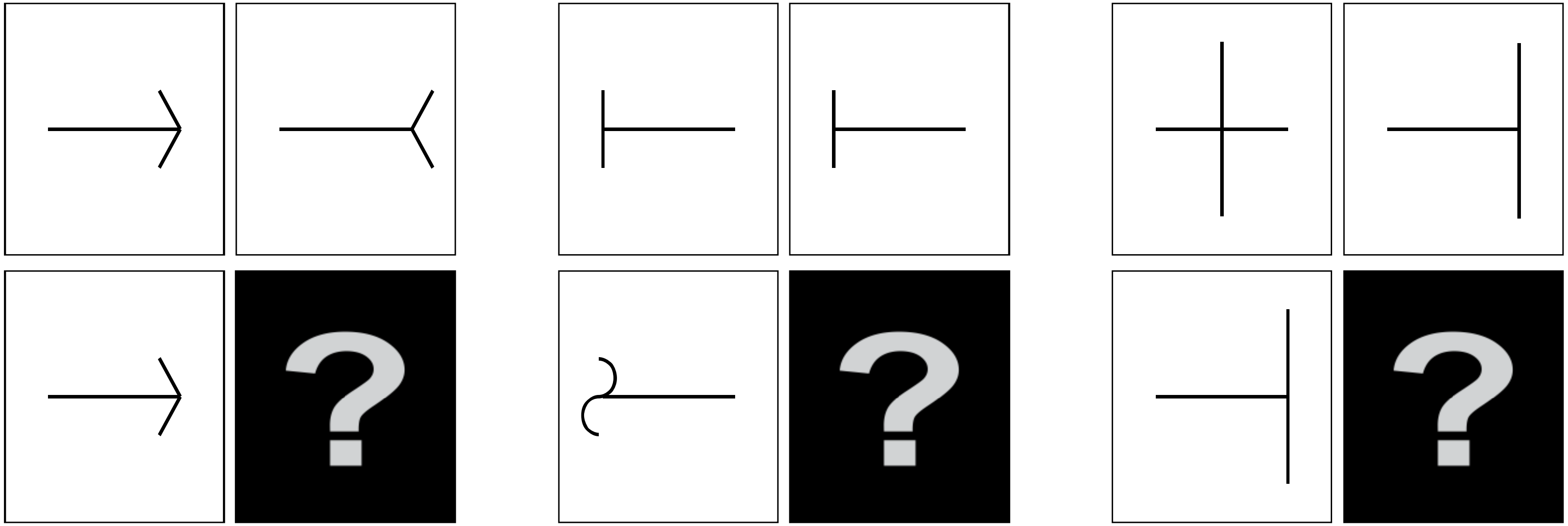


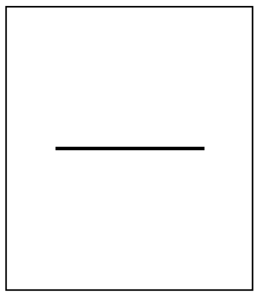


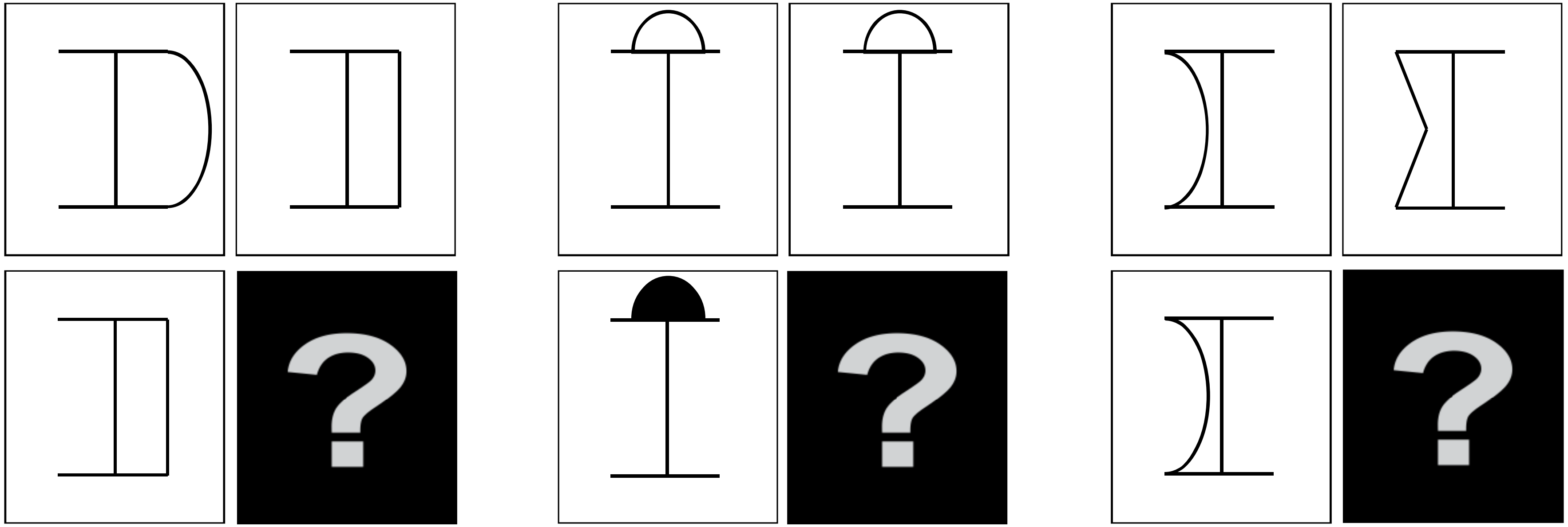


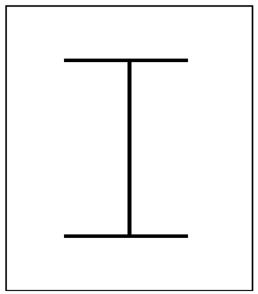


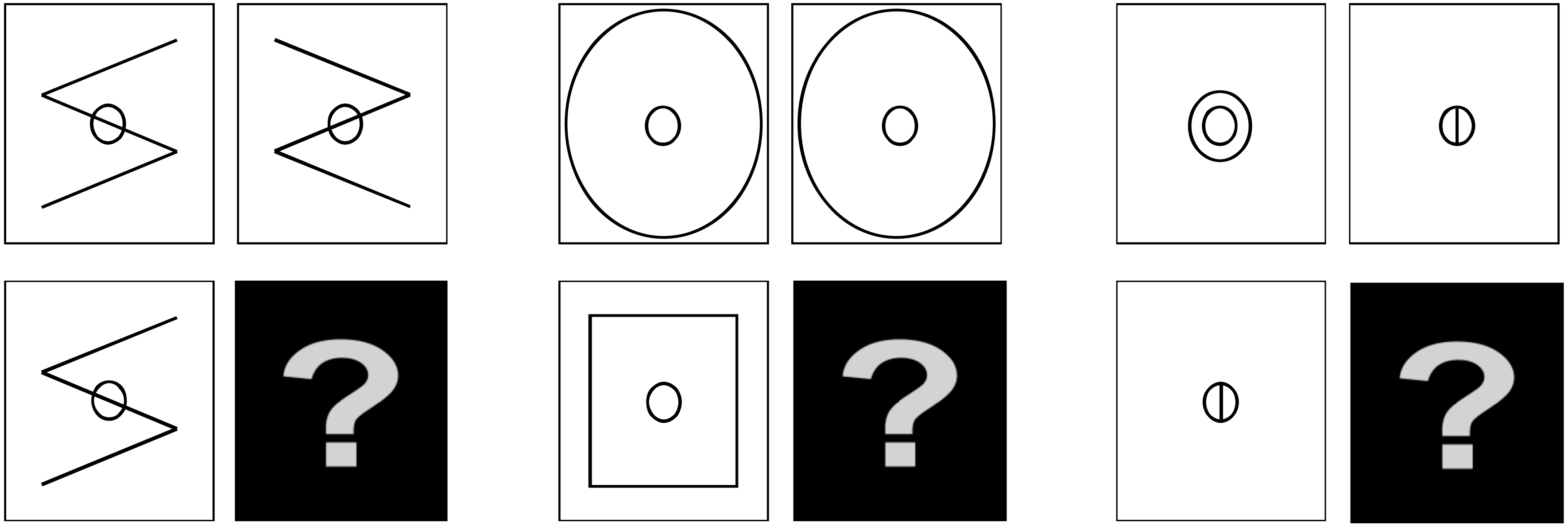


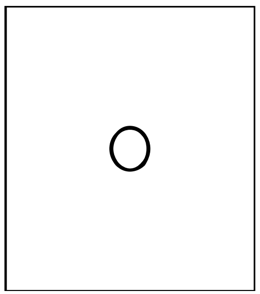


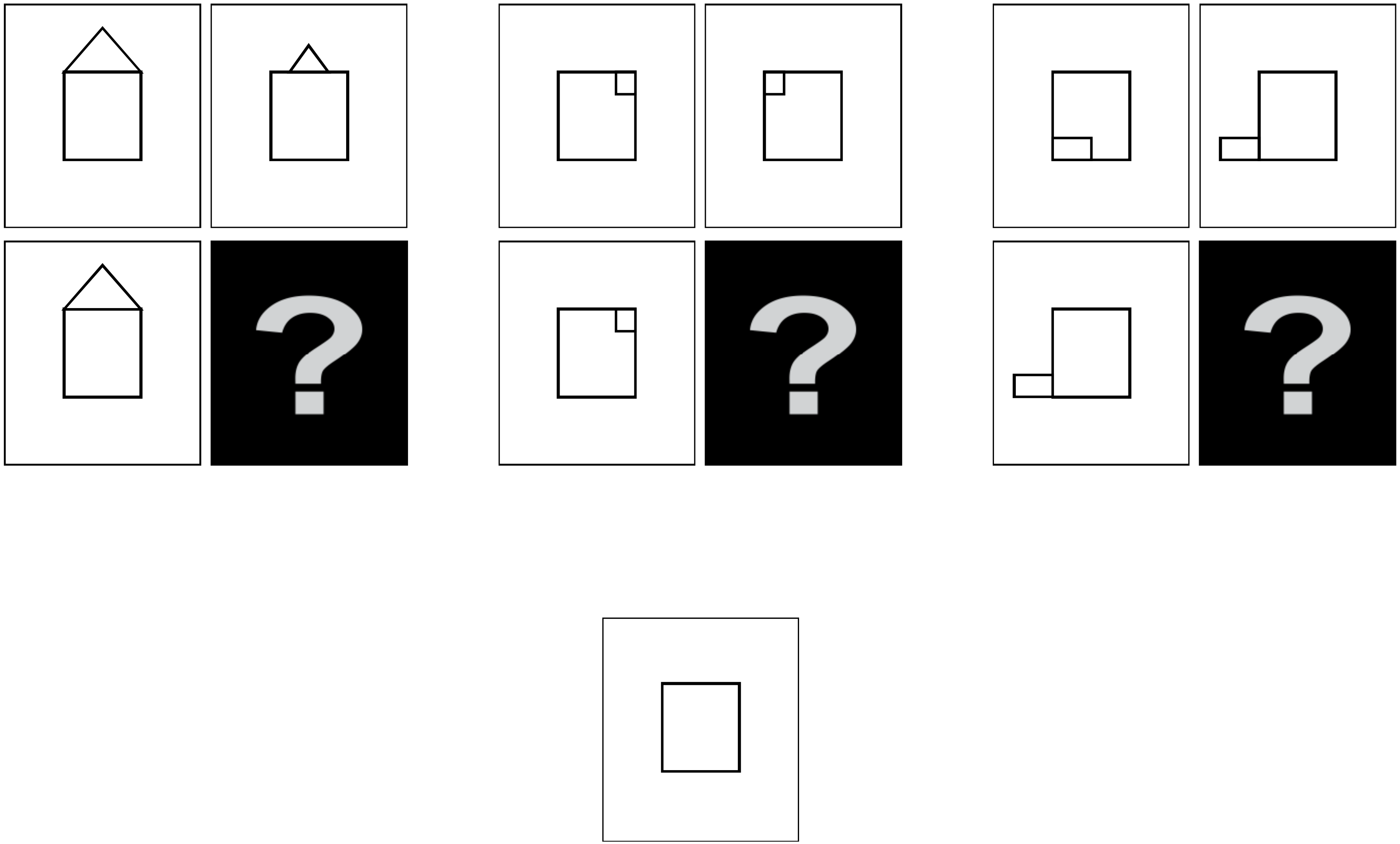


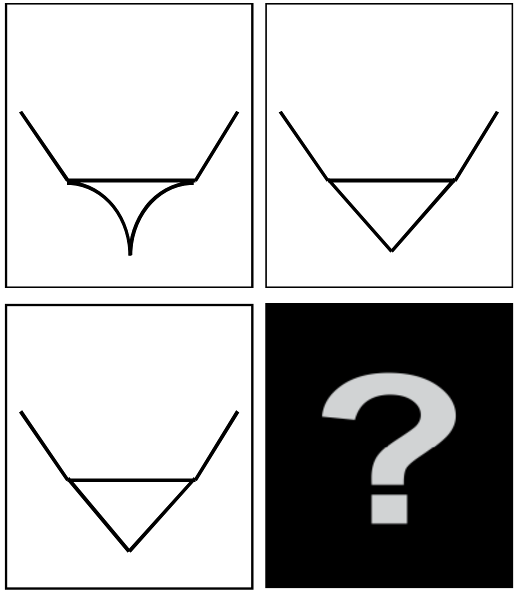
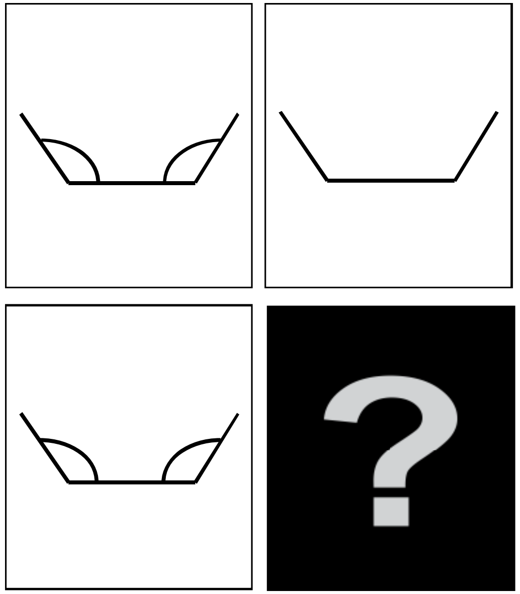
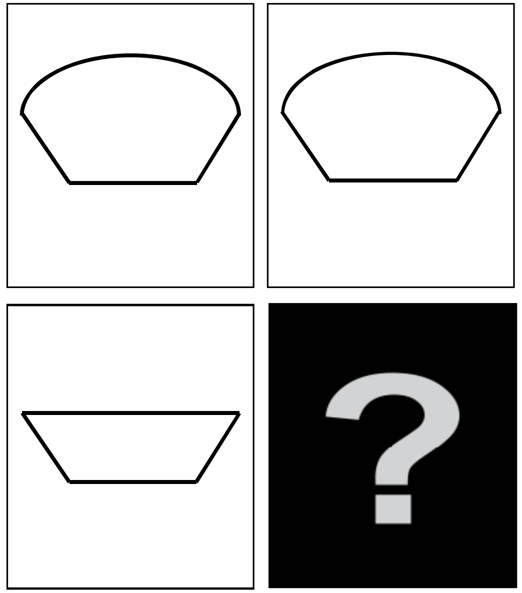
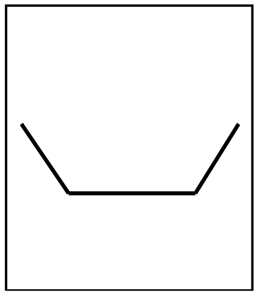


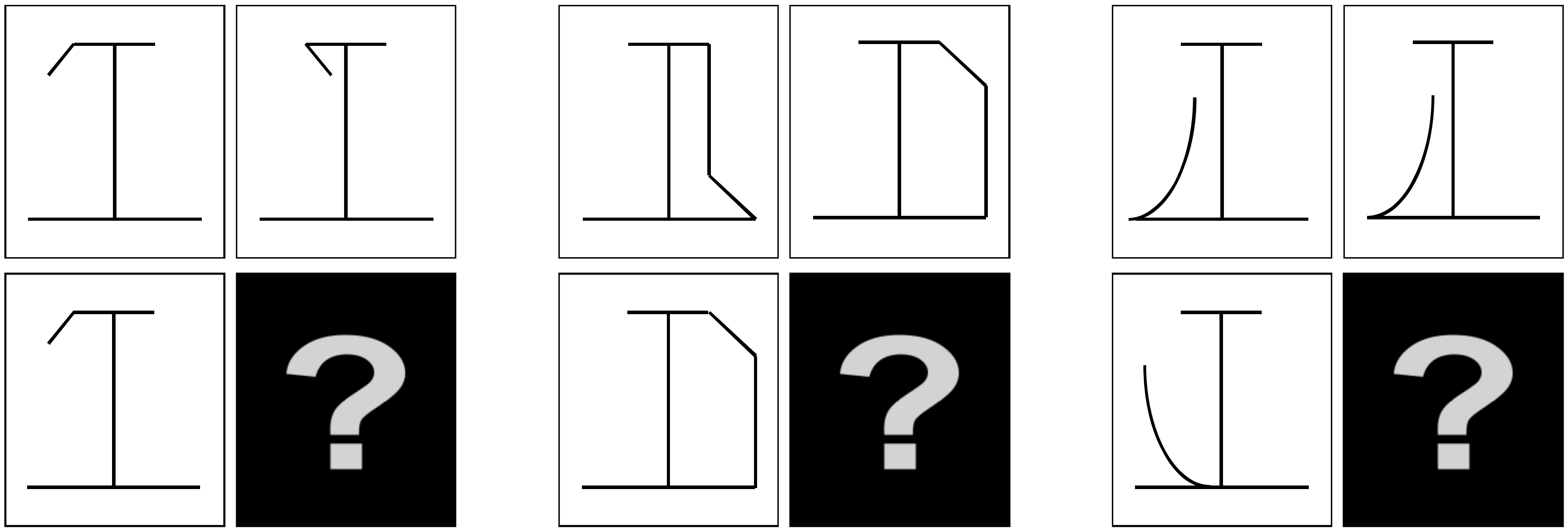


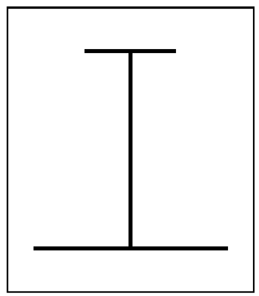


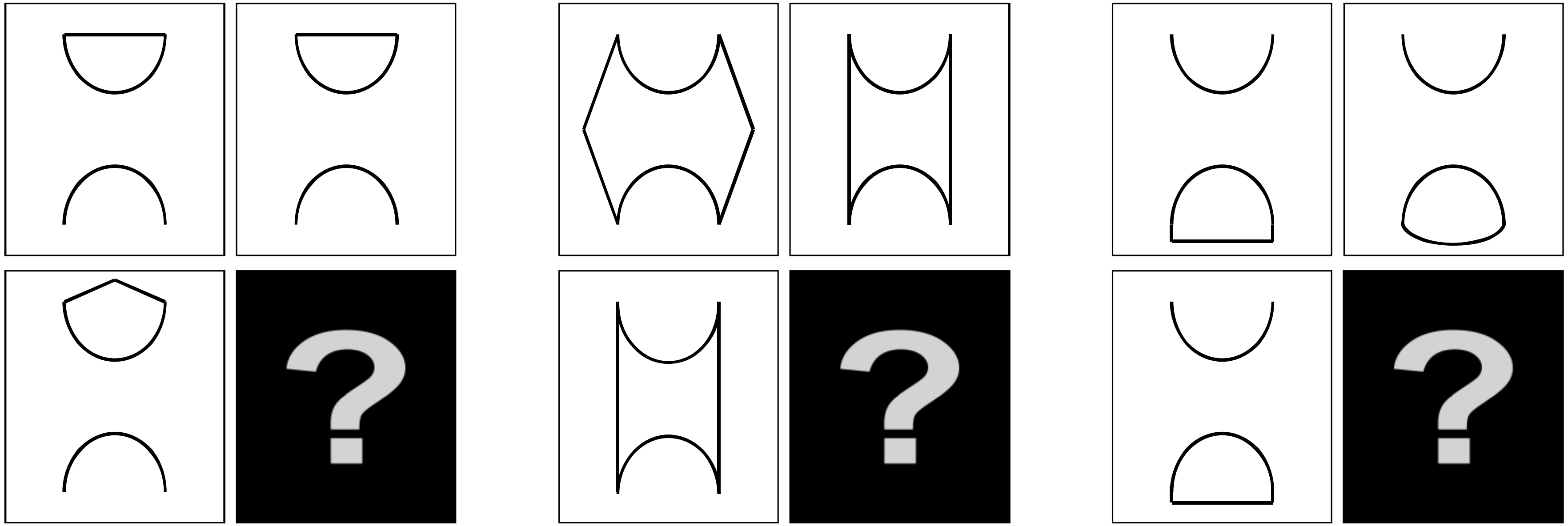


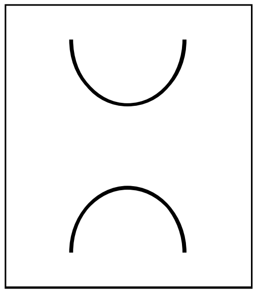


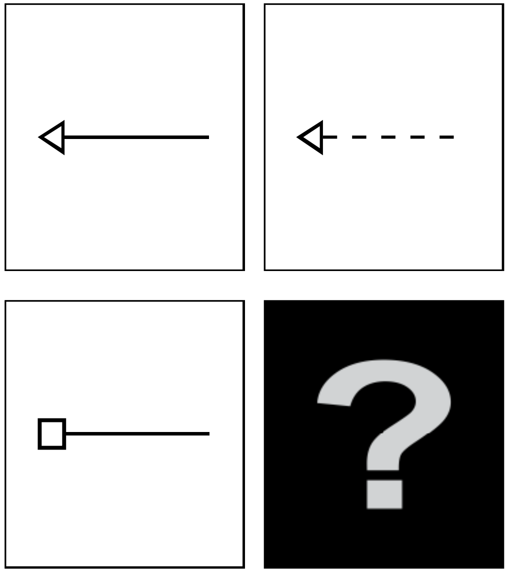


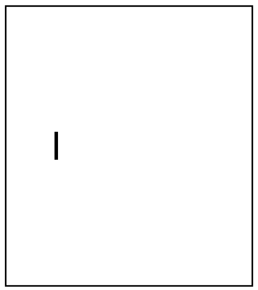


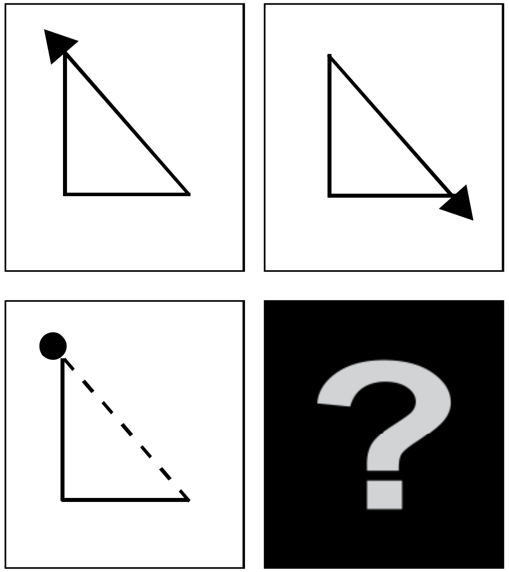


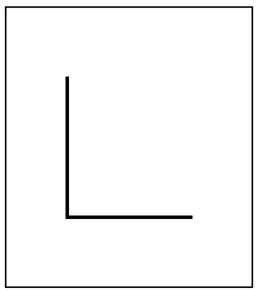


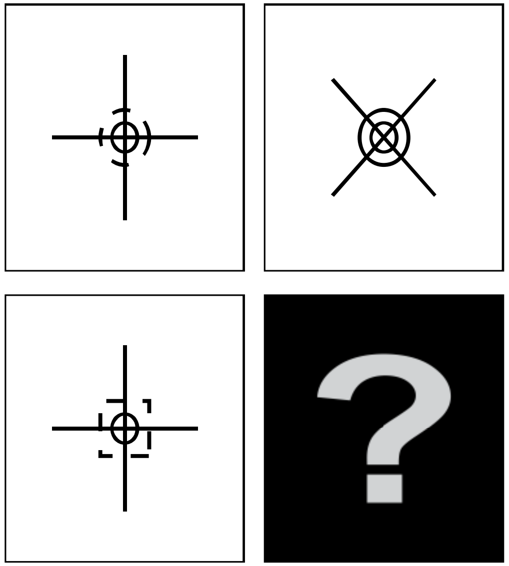
2C

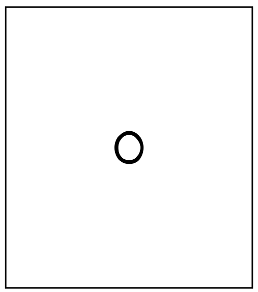


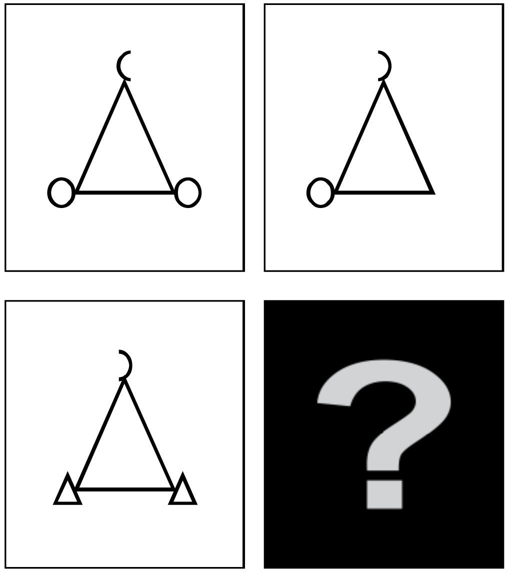


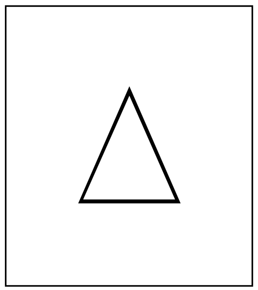


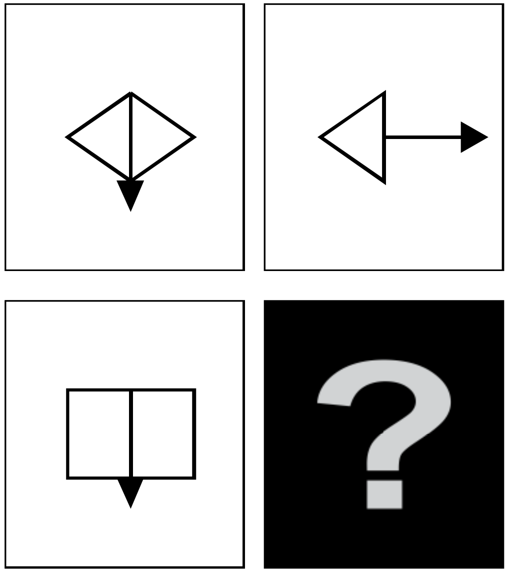


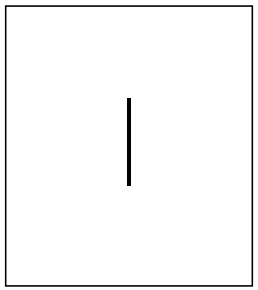


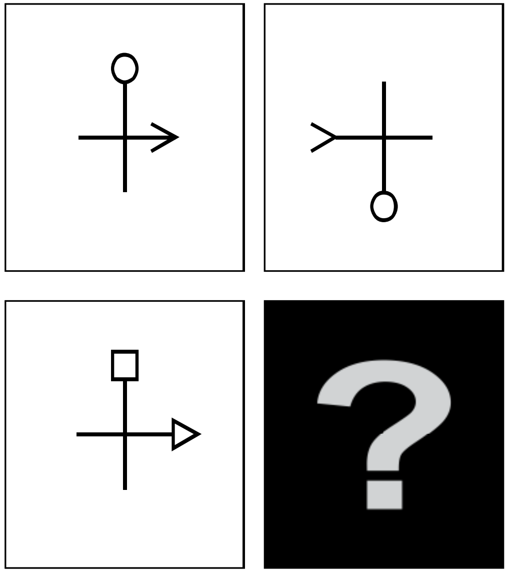


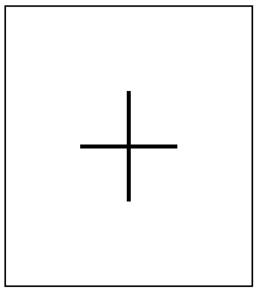


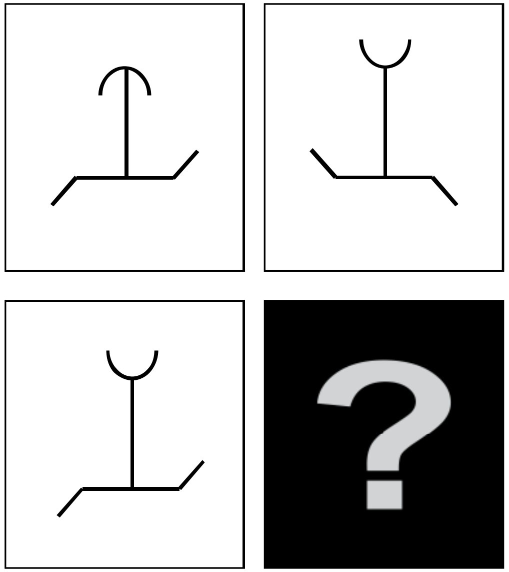


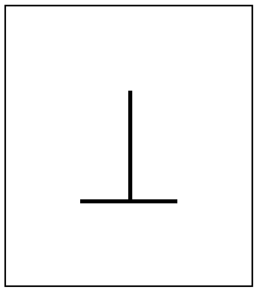




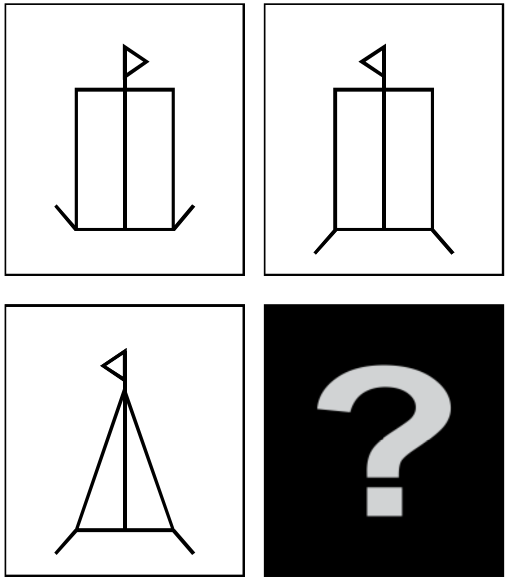


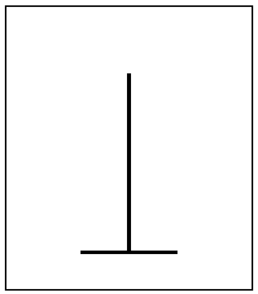


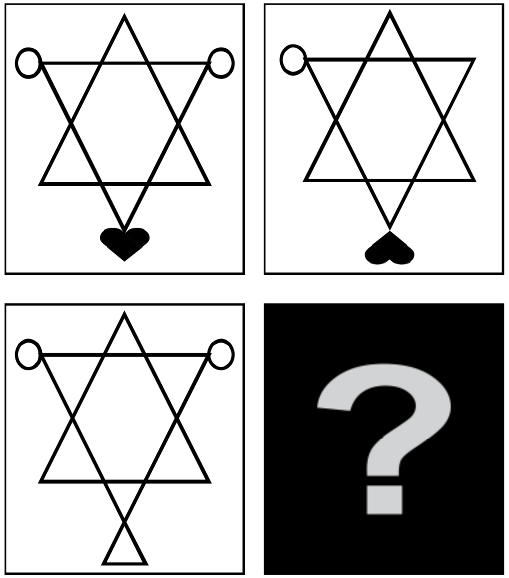


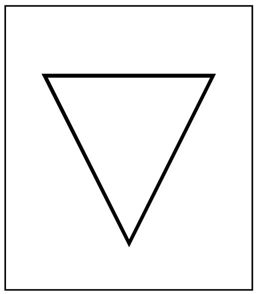


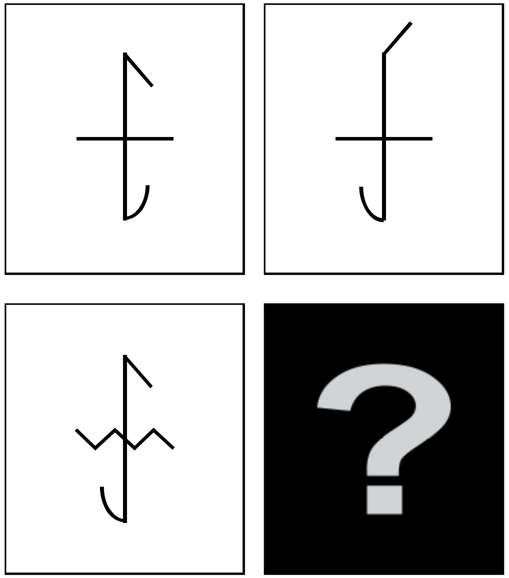
3C

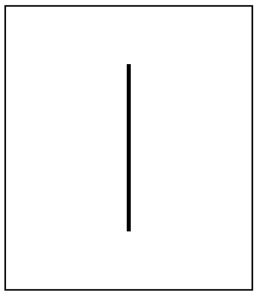


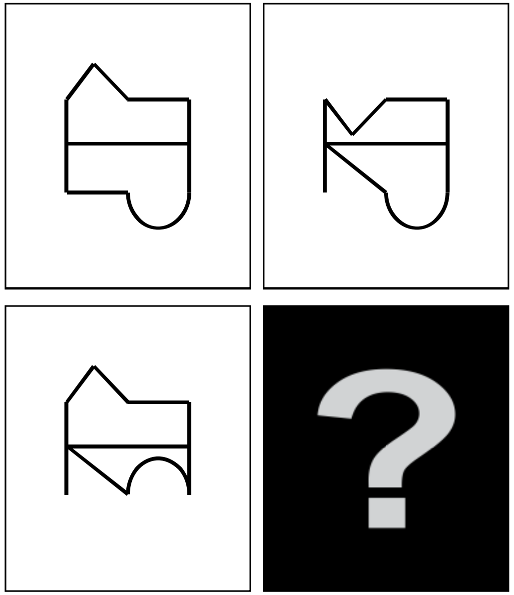


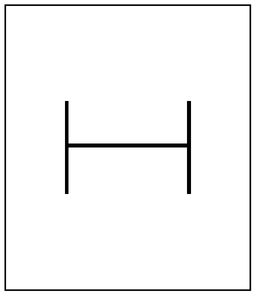


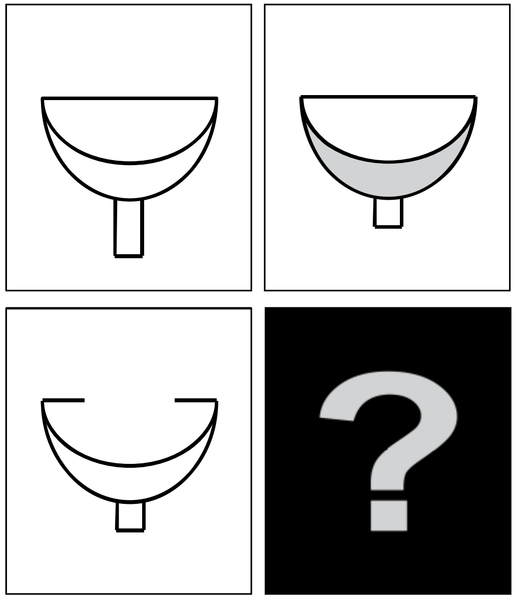


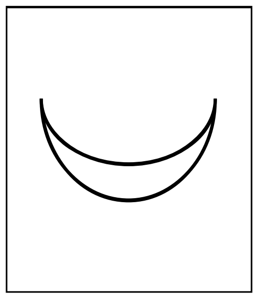


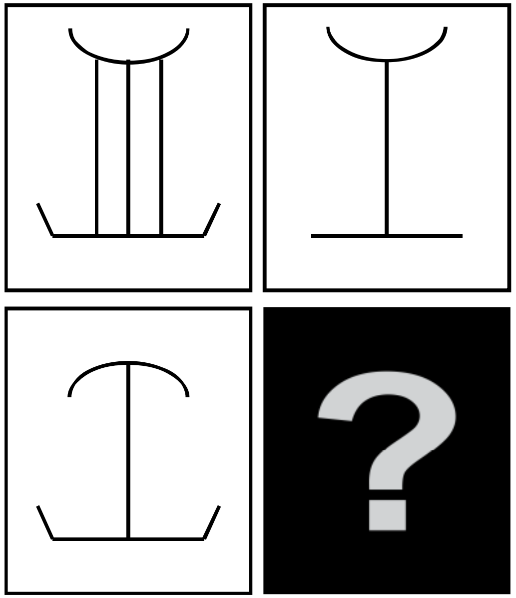


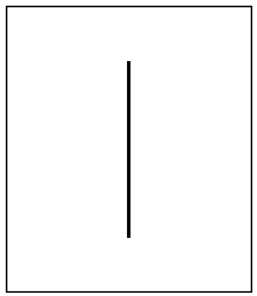


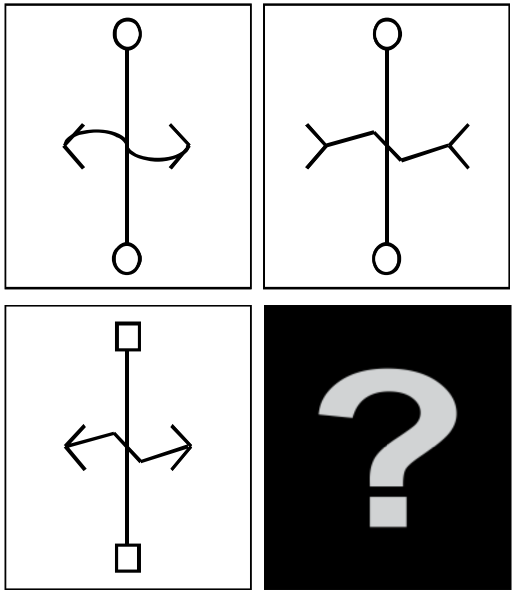


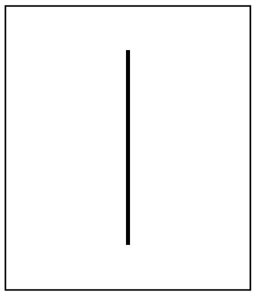


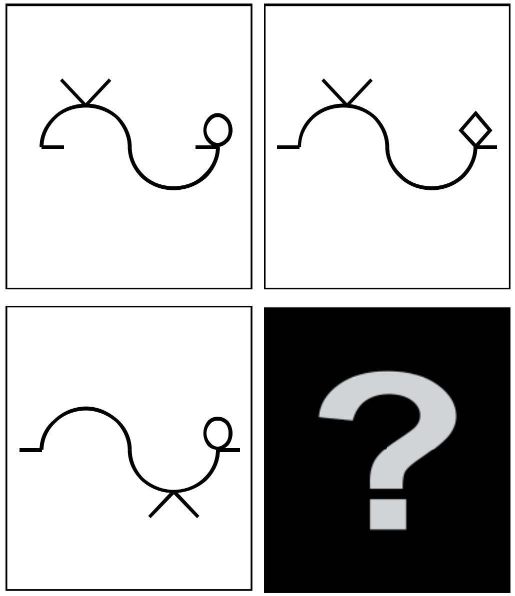


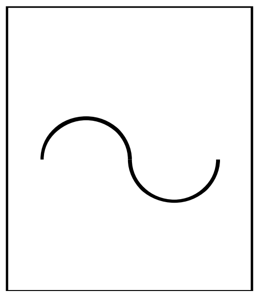


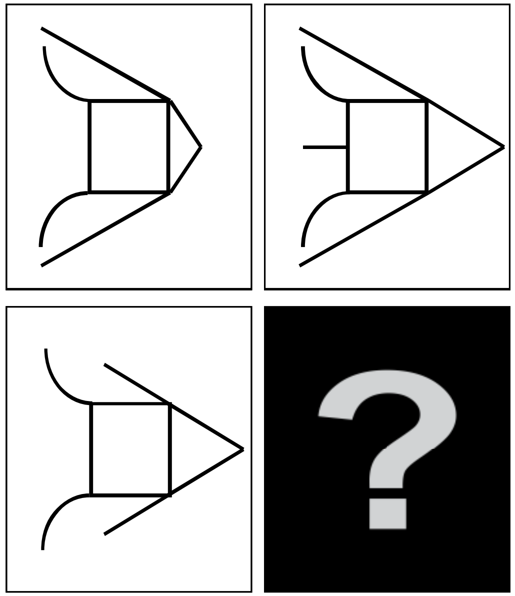


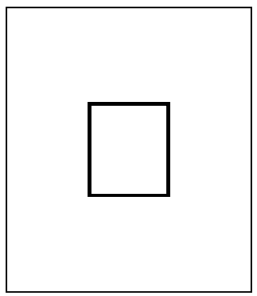


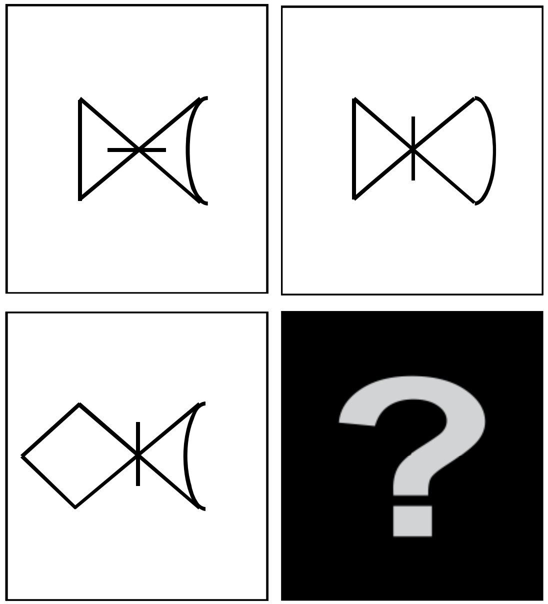


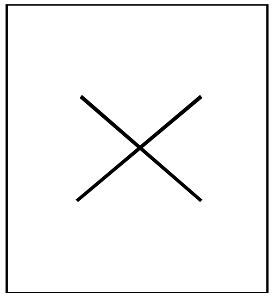






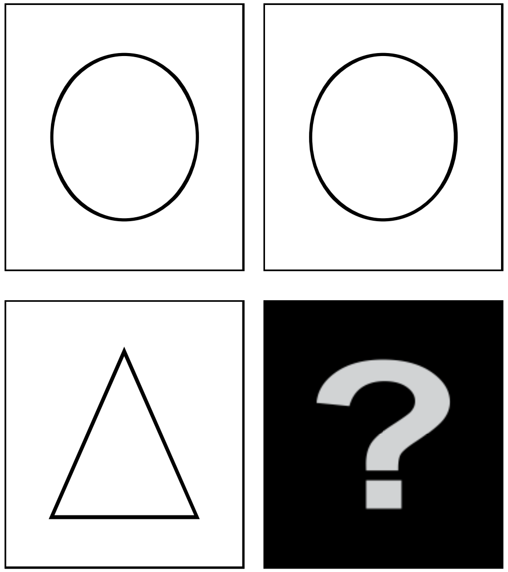


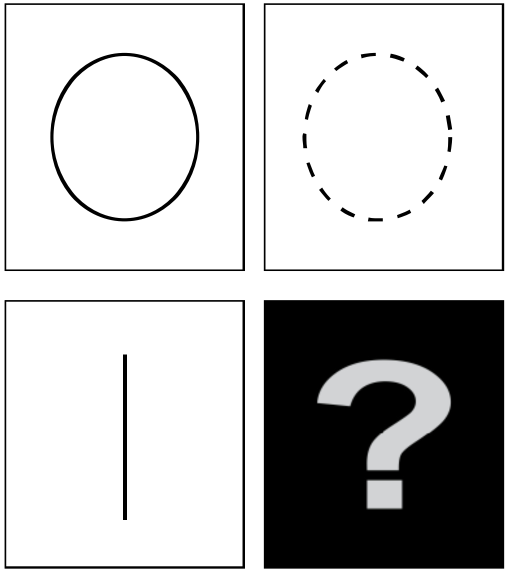


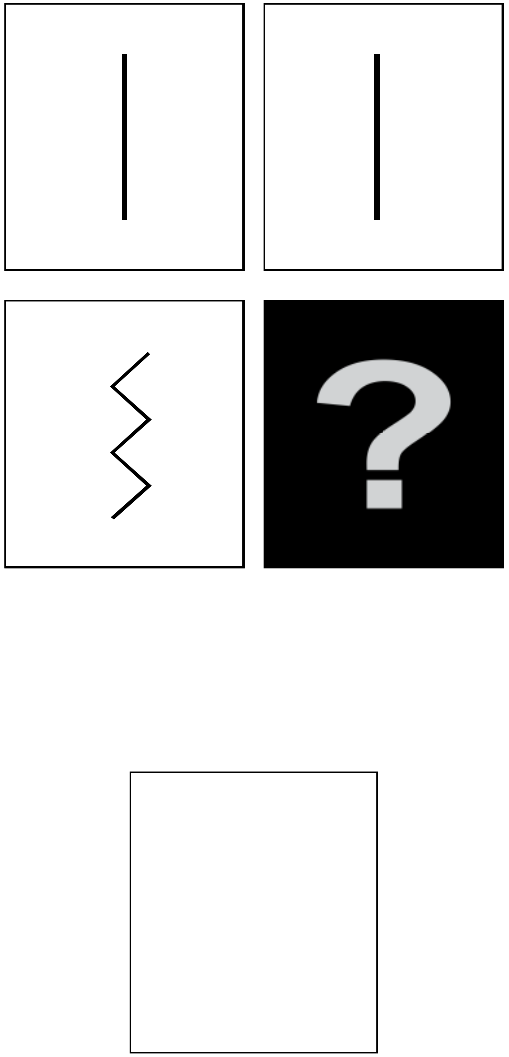


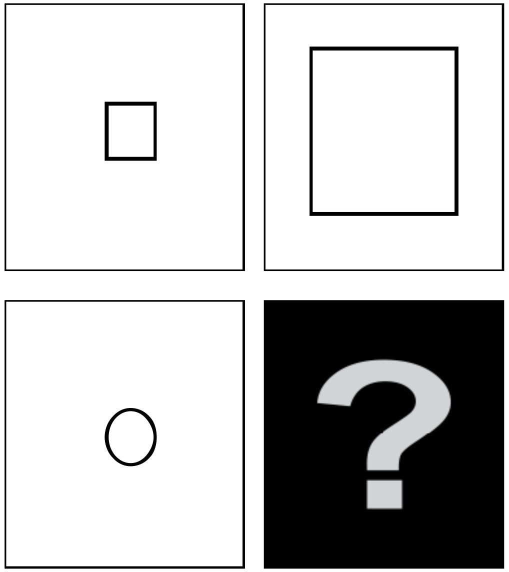
B1

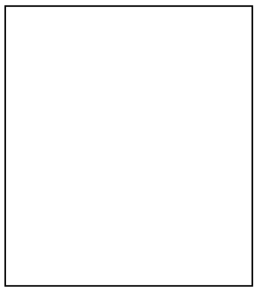
ID: ________________________

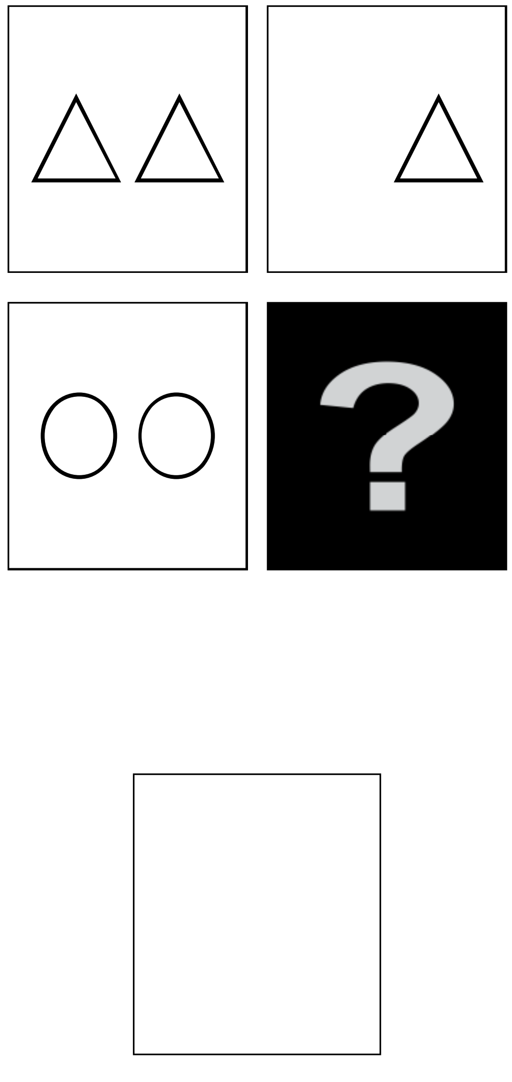


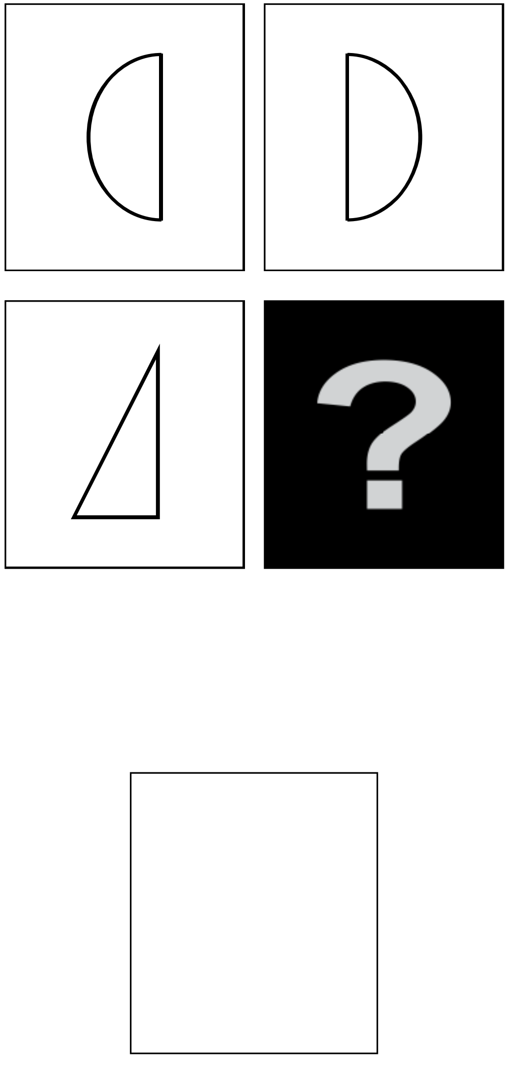


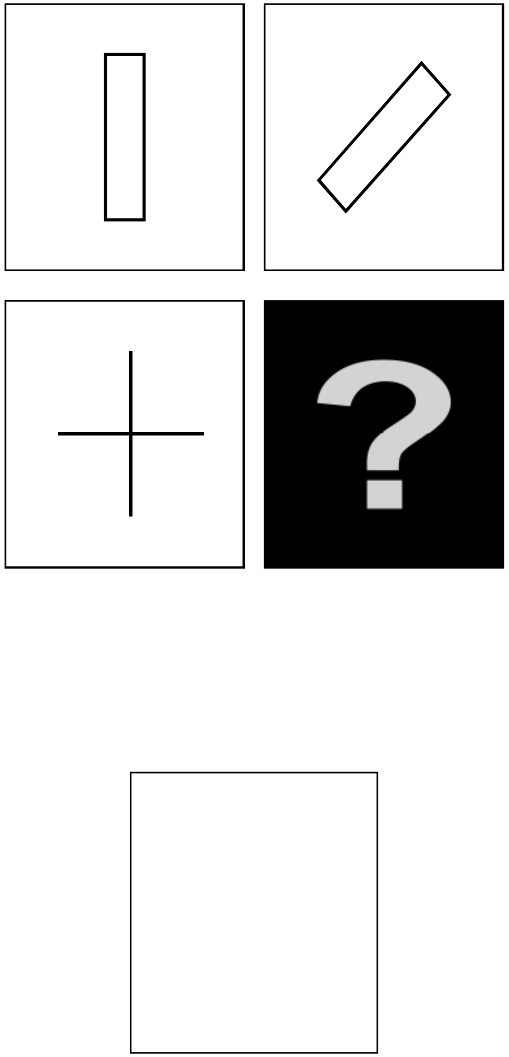




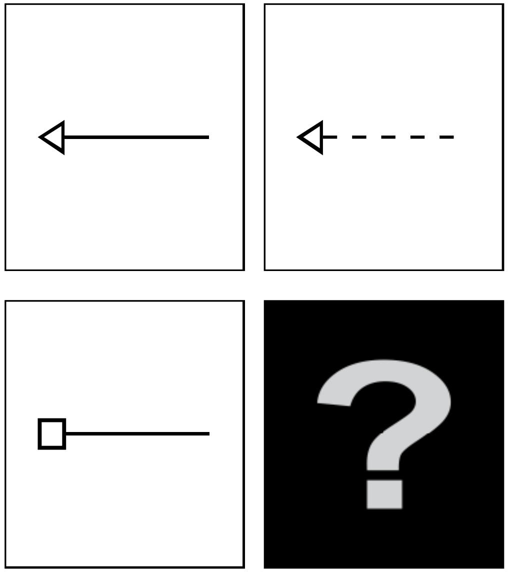


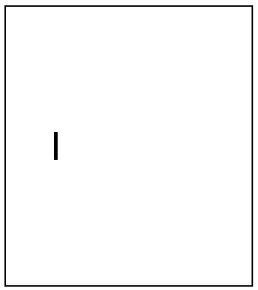


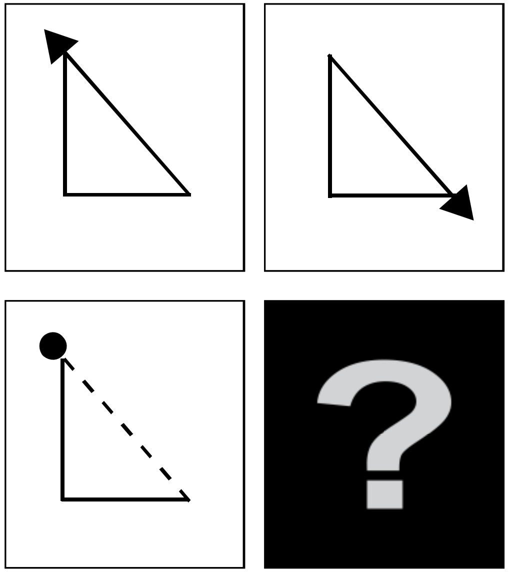


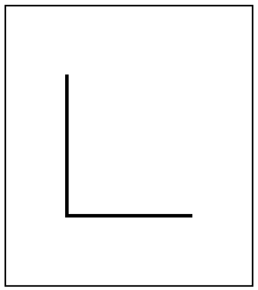


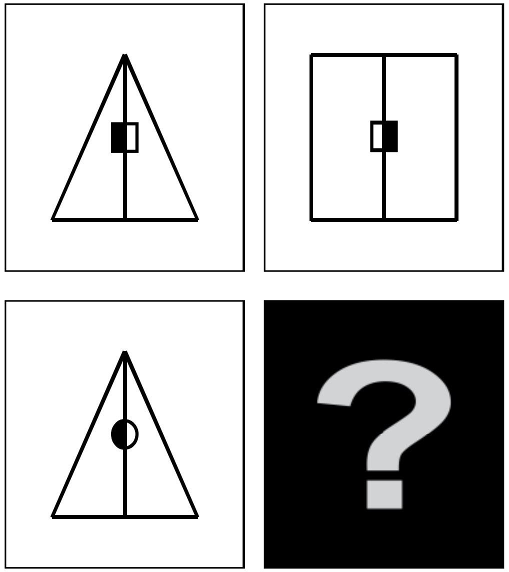
2C

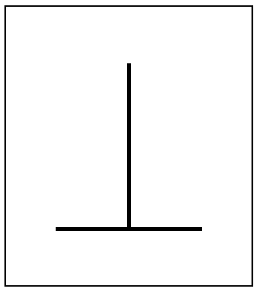


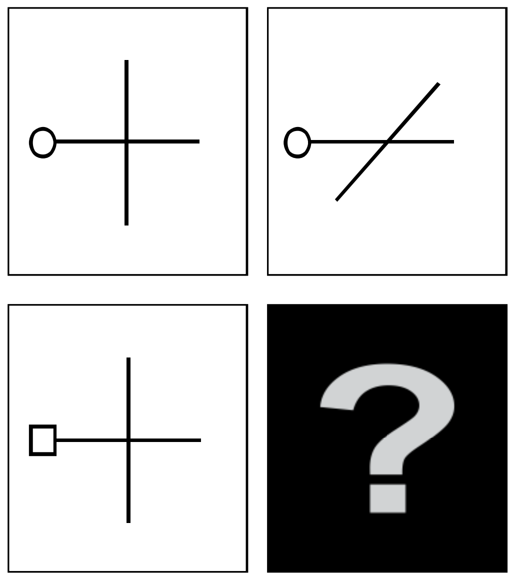


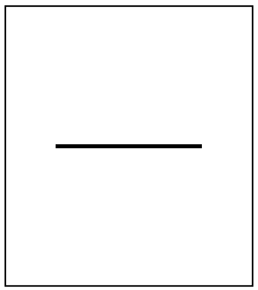


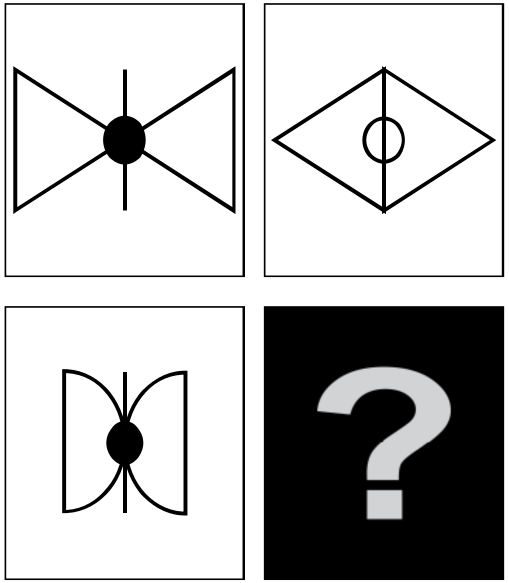


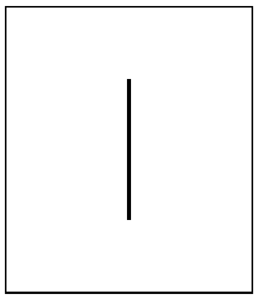


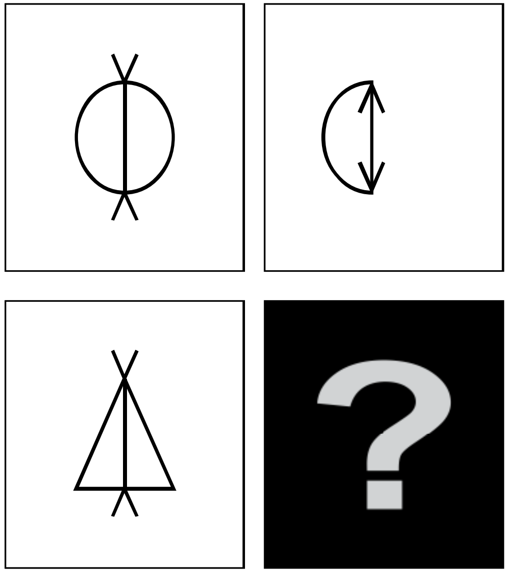


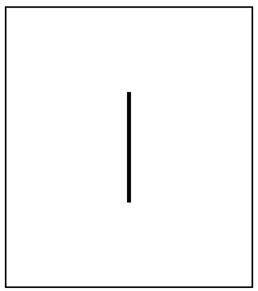


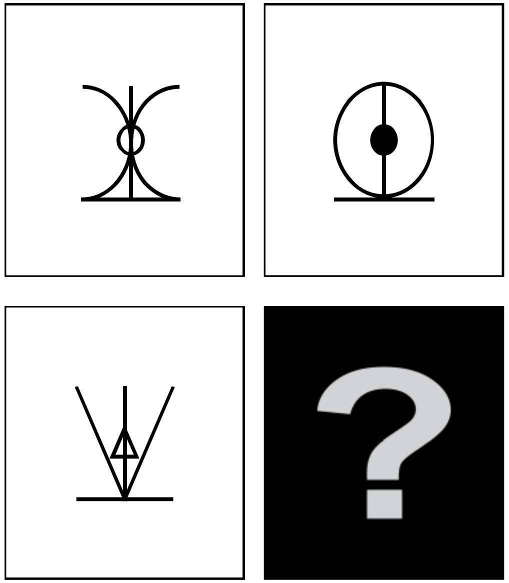


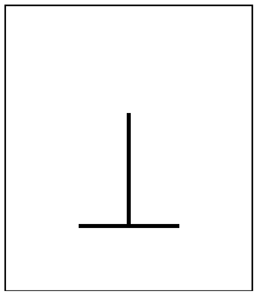




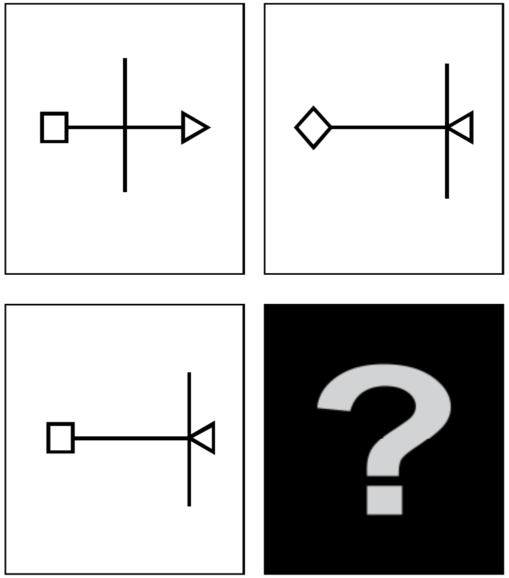


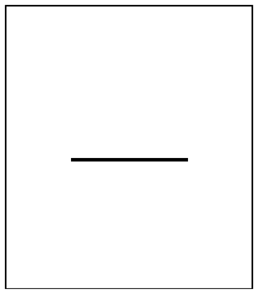


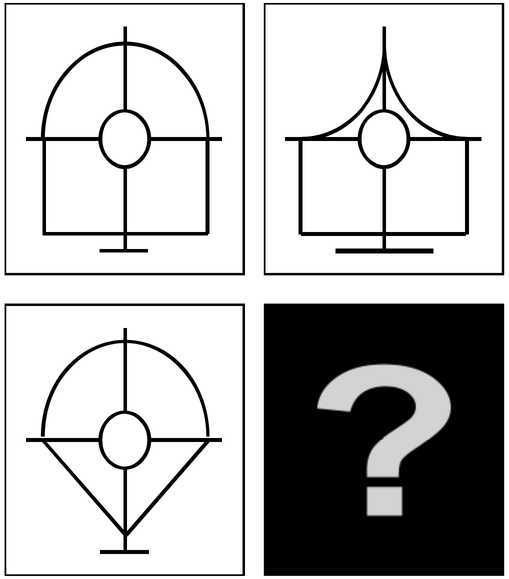


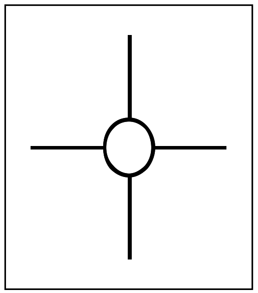


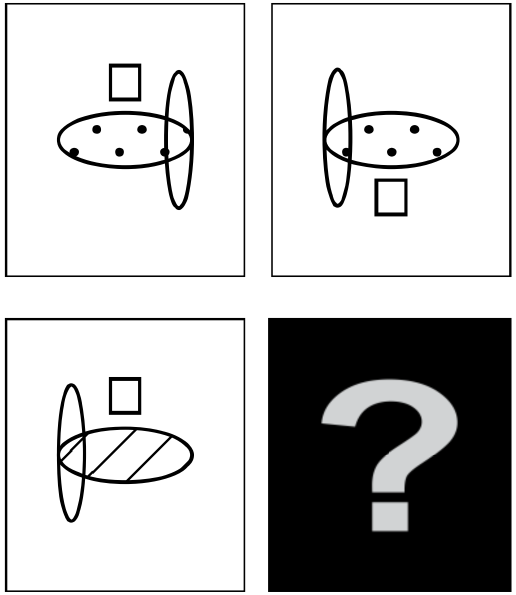
3C

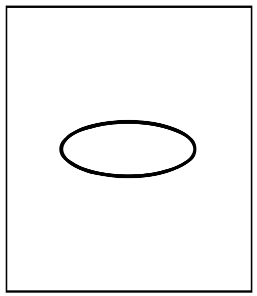


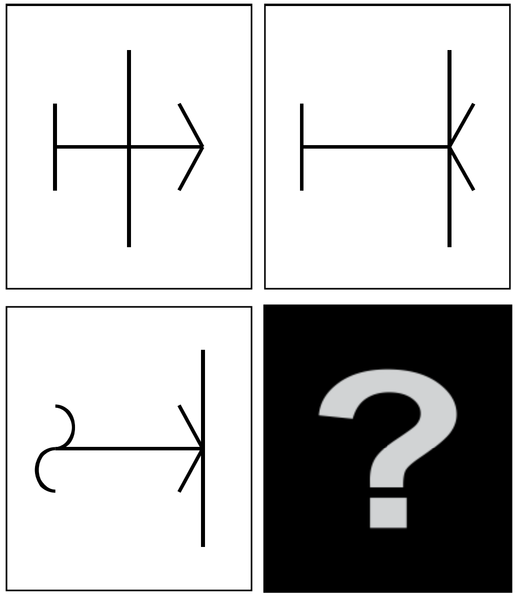


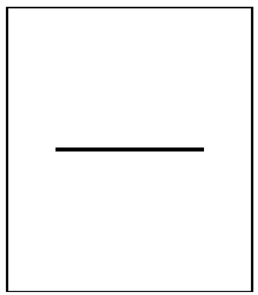


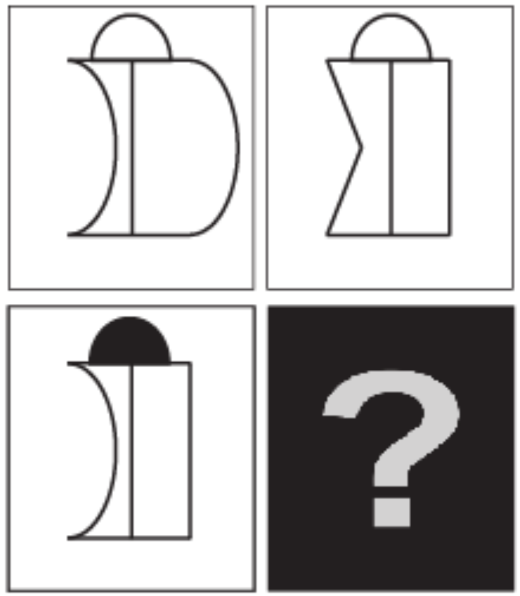


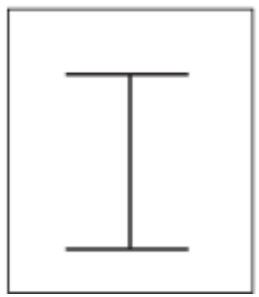


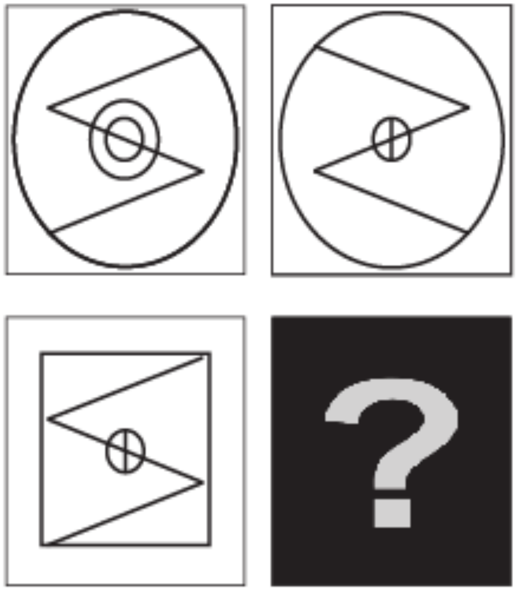


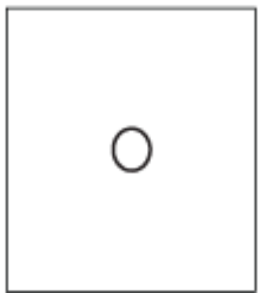


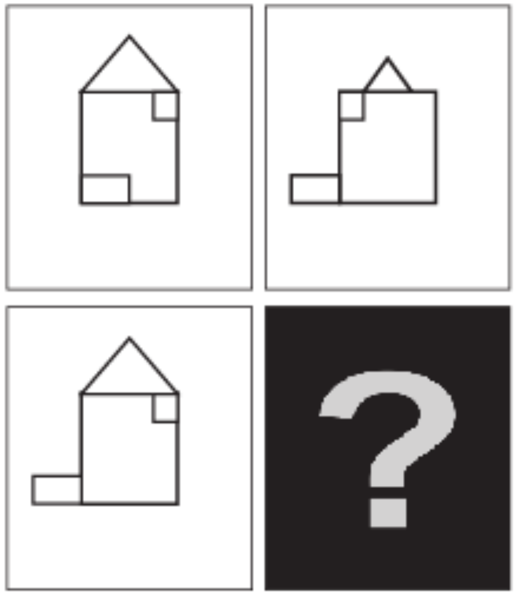


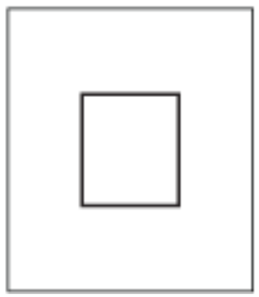


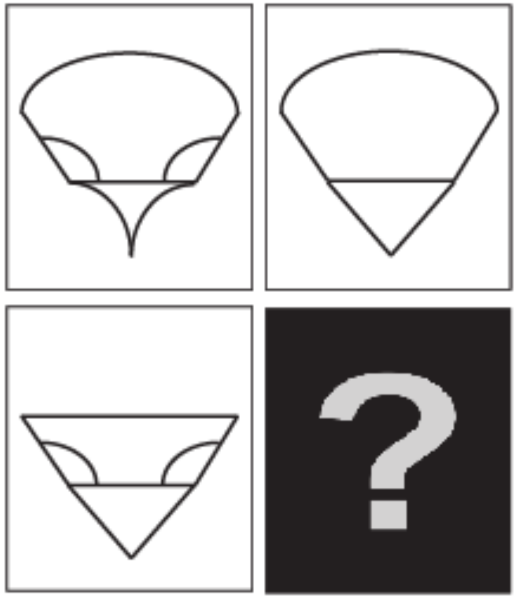


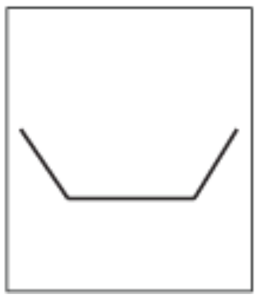


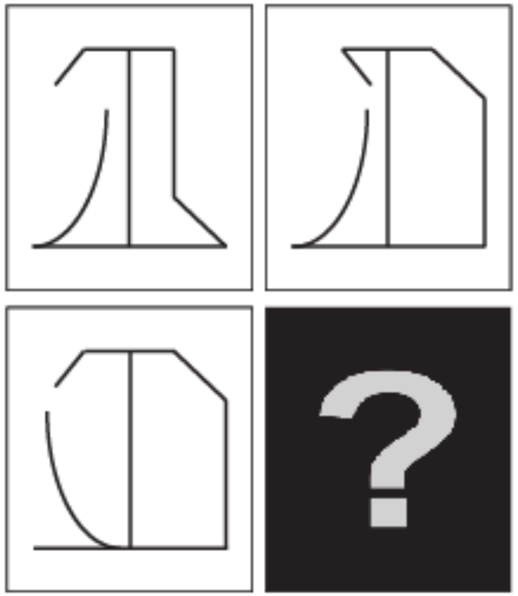


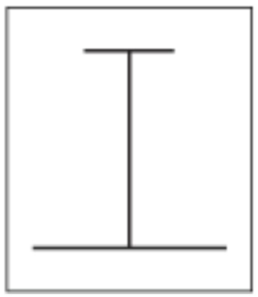


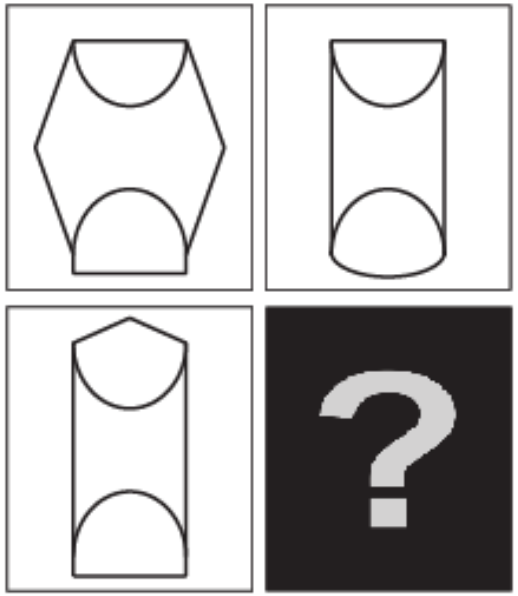


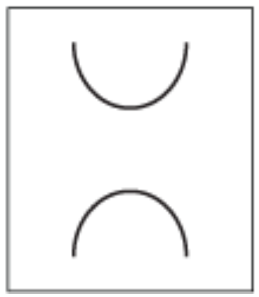




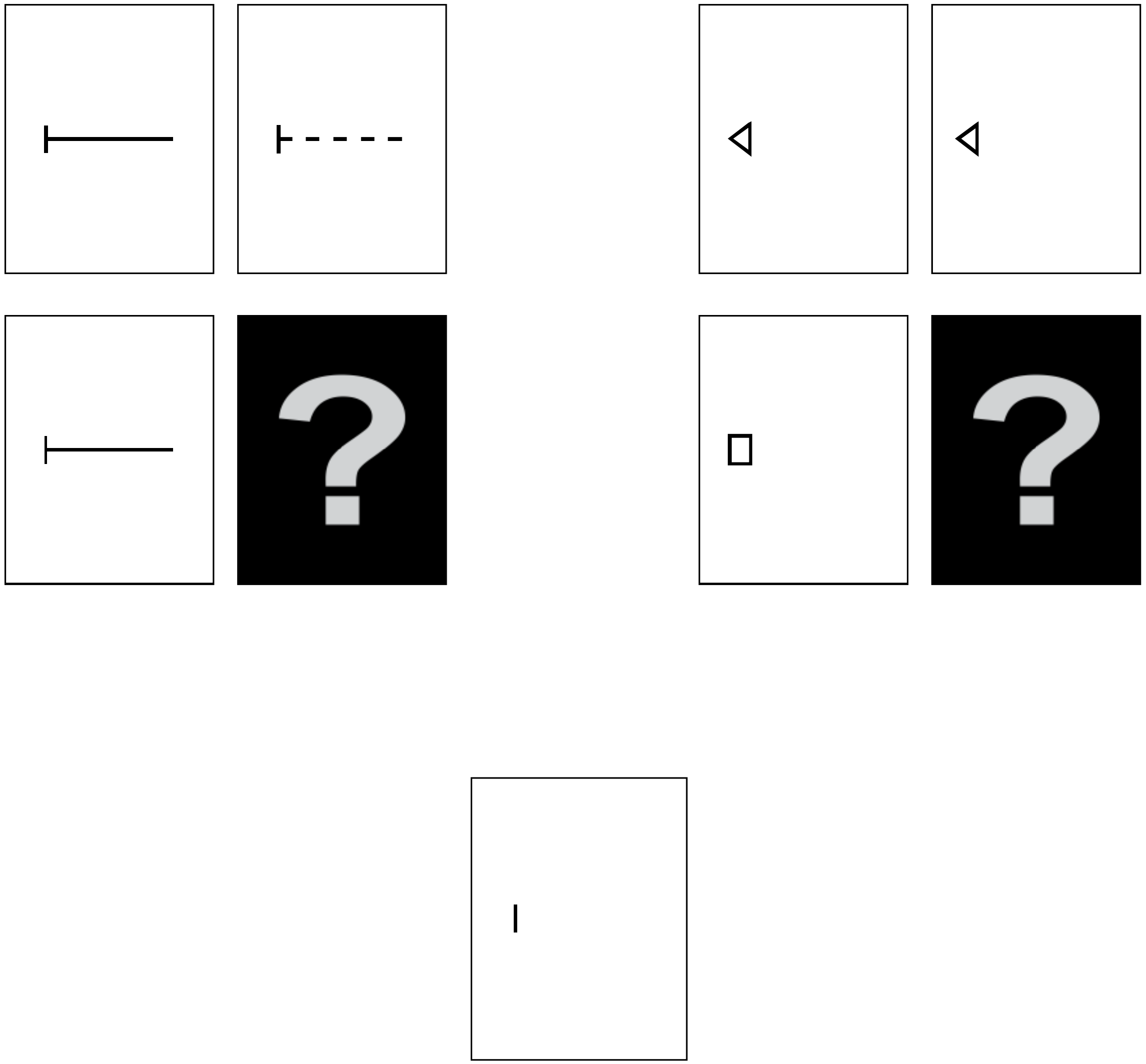


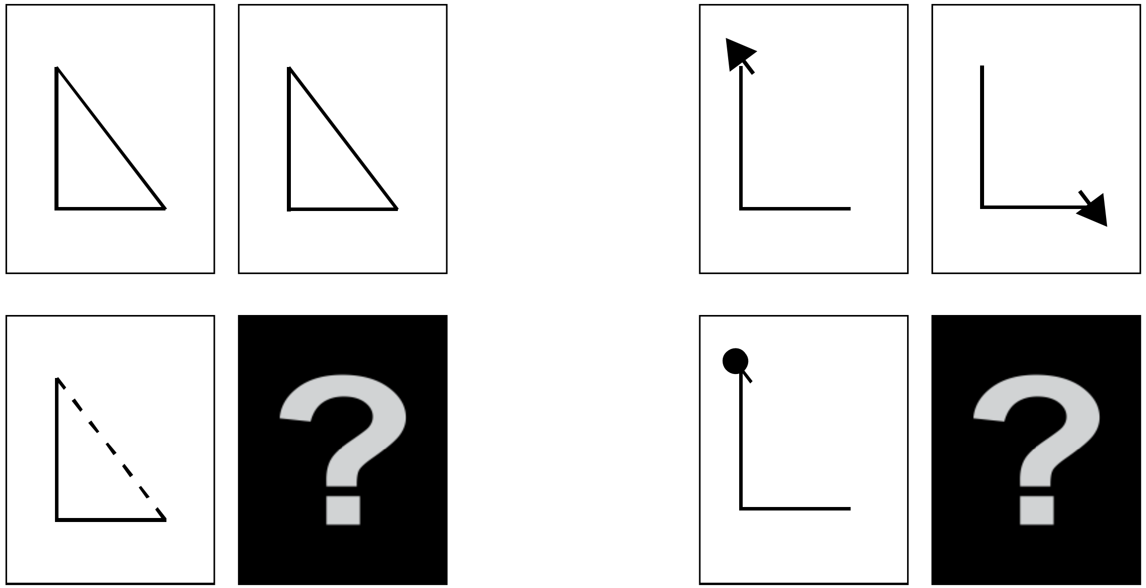


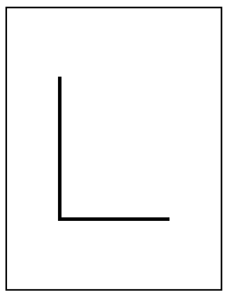


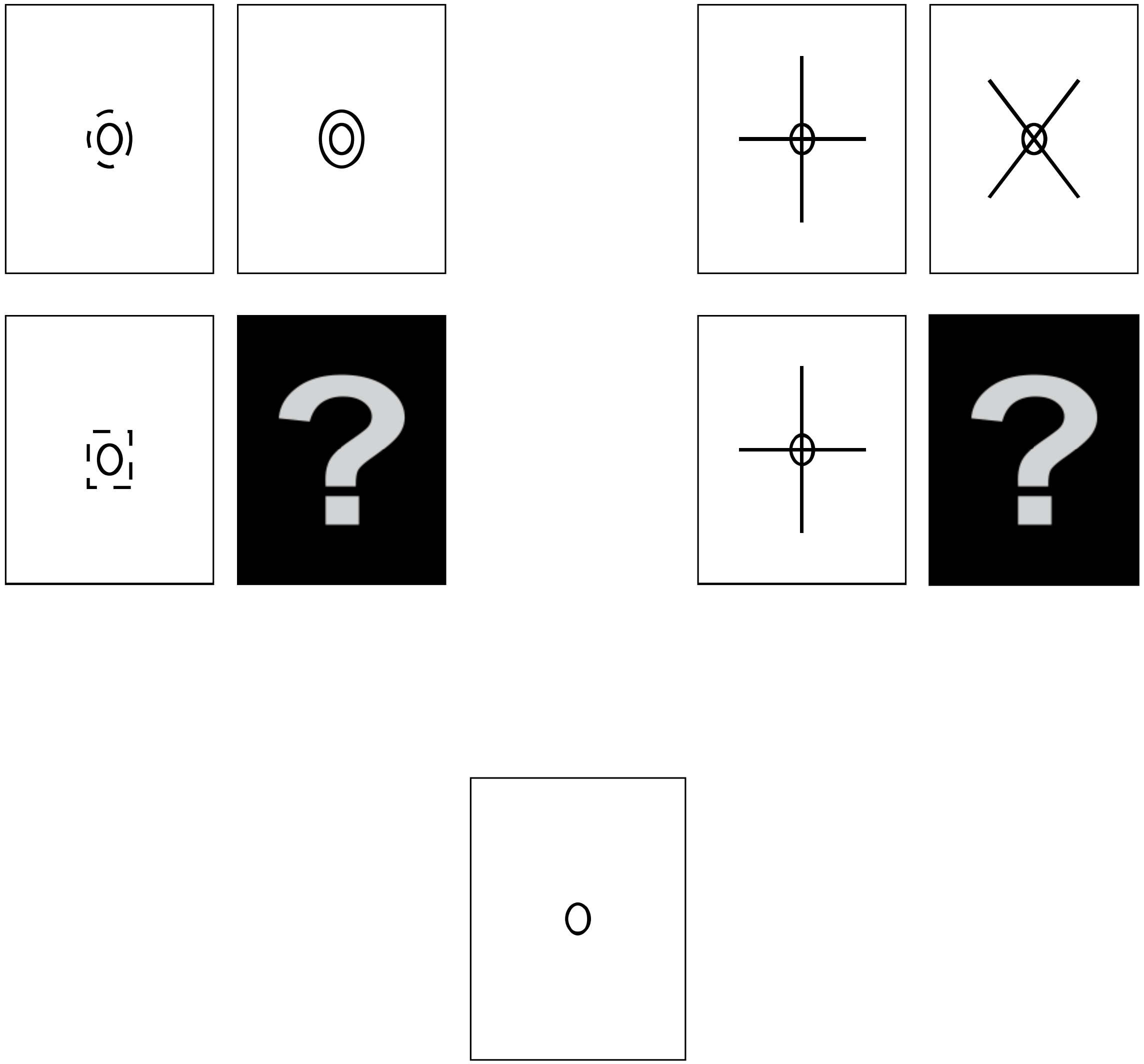


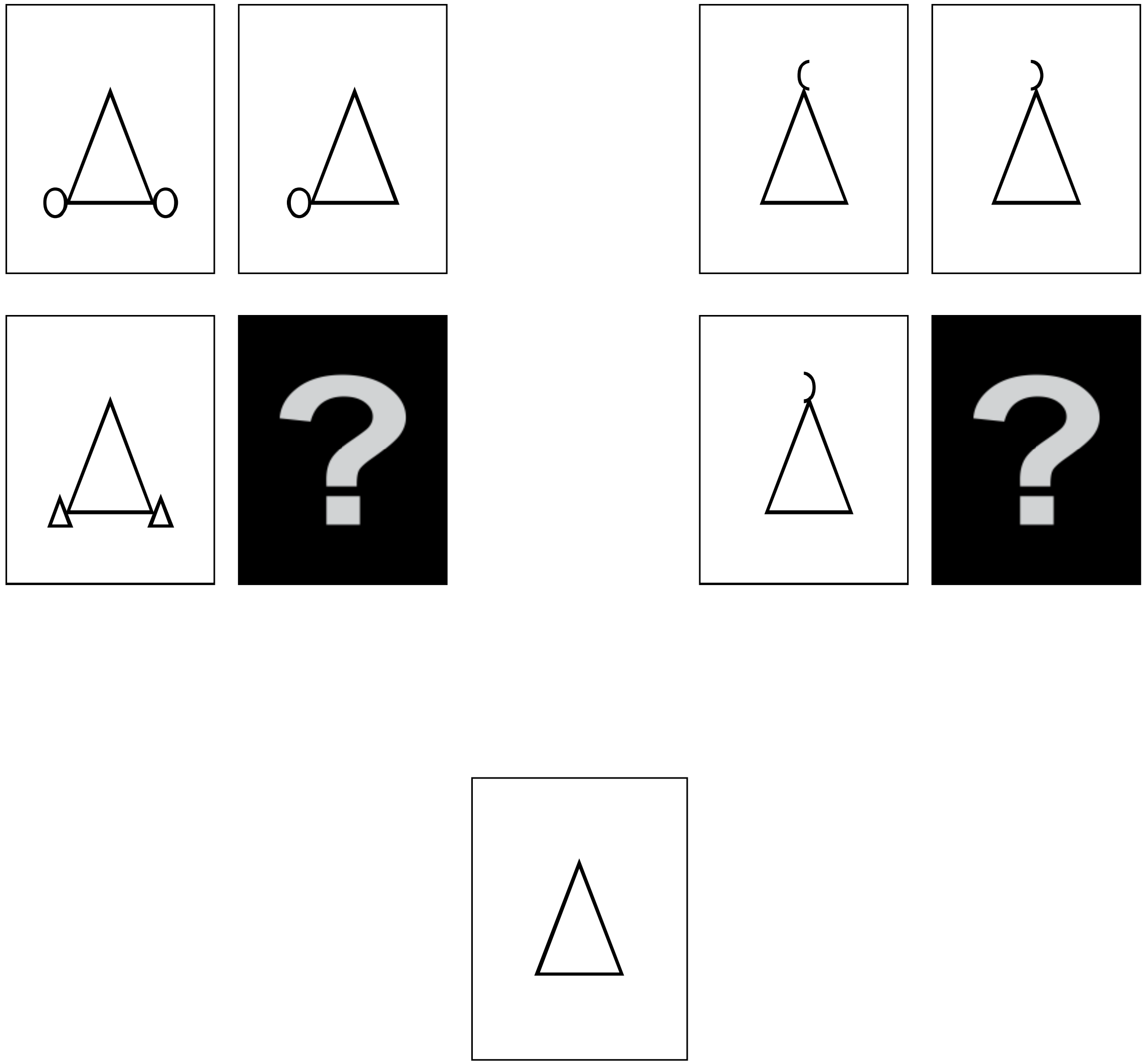
2S

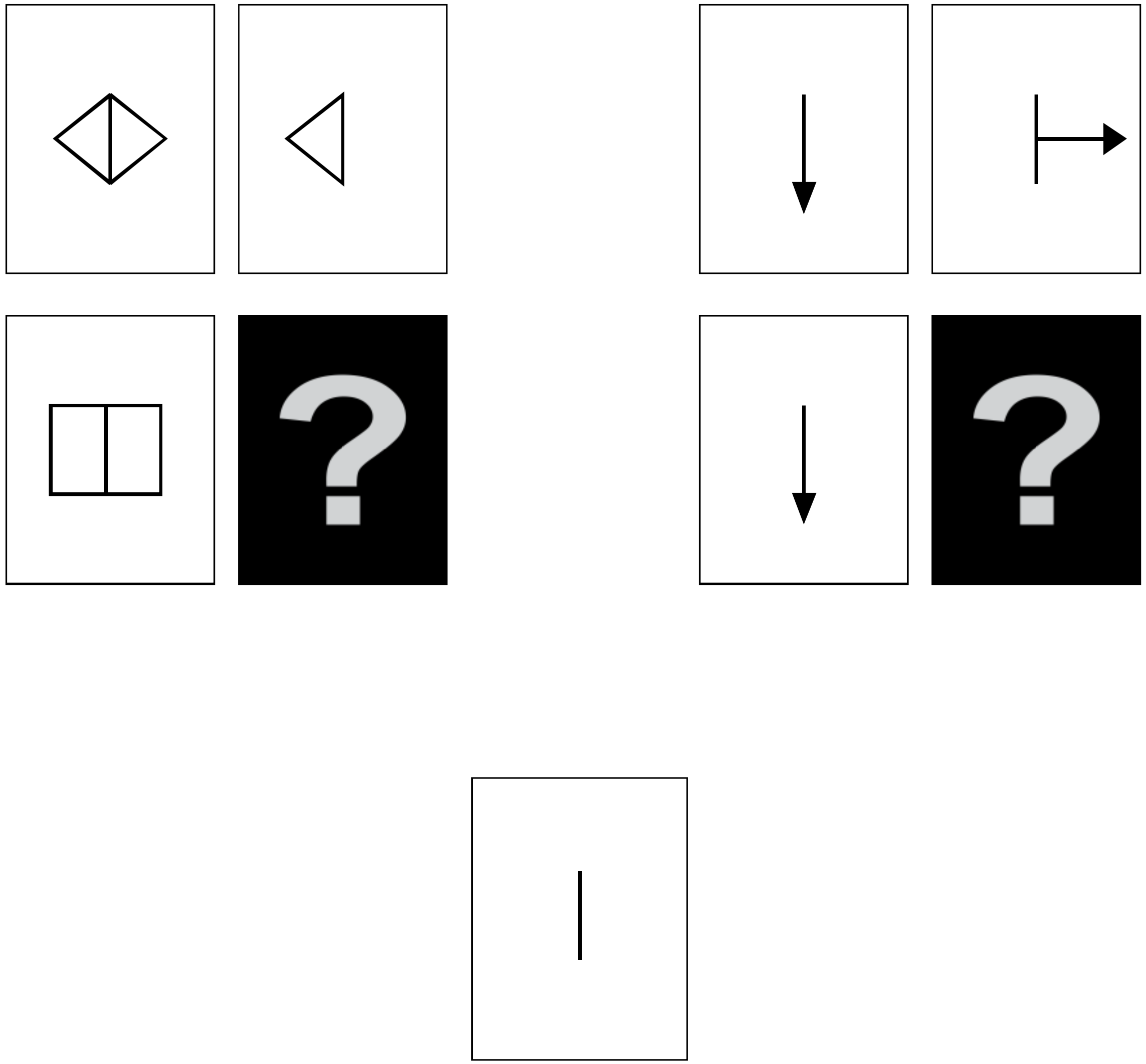


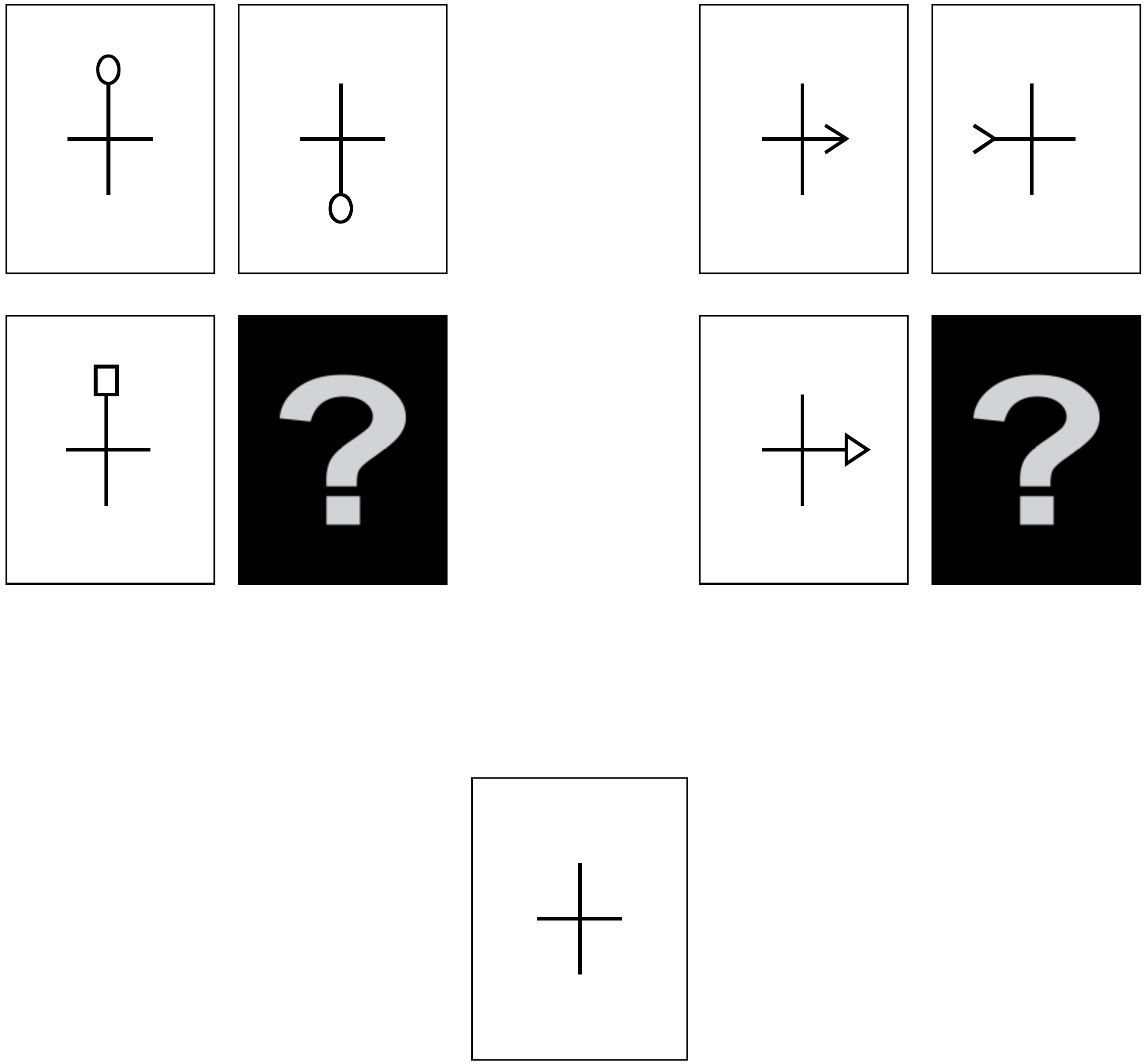


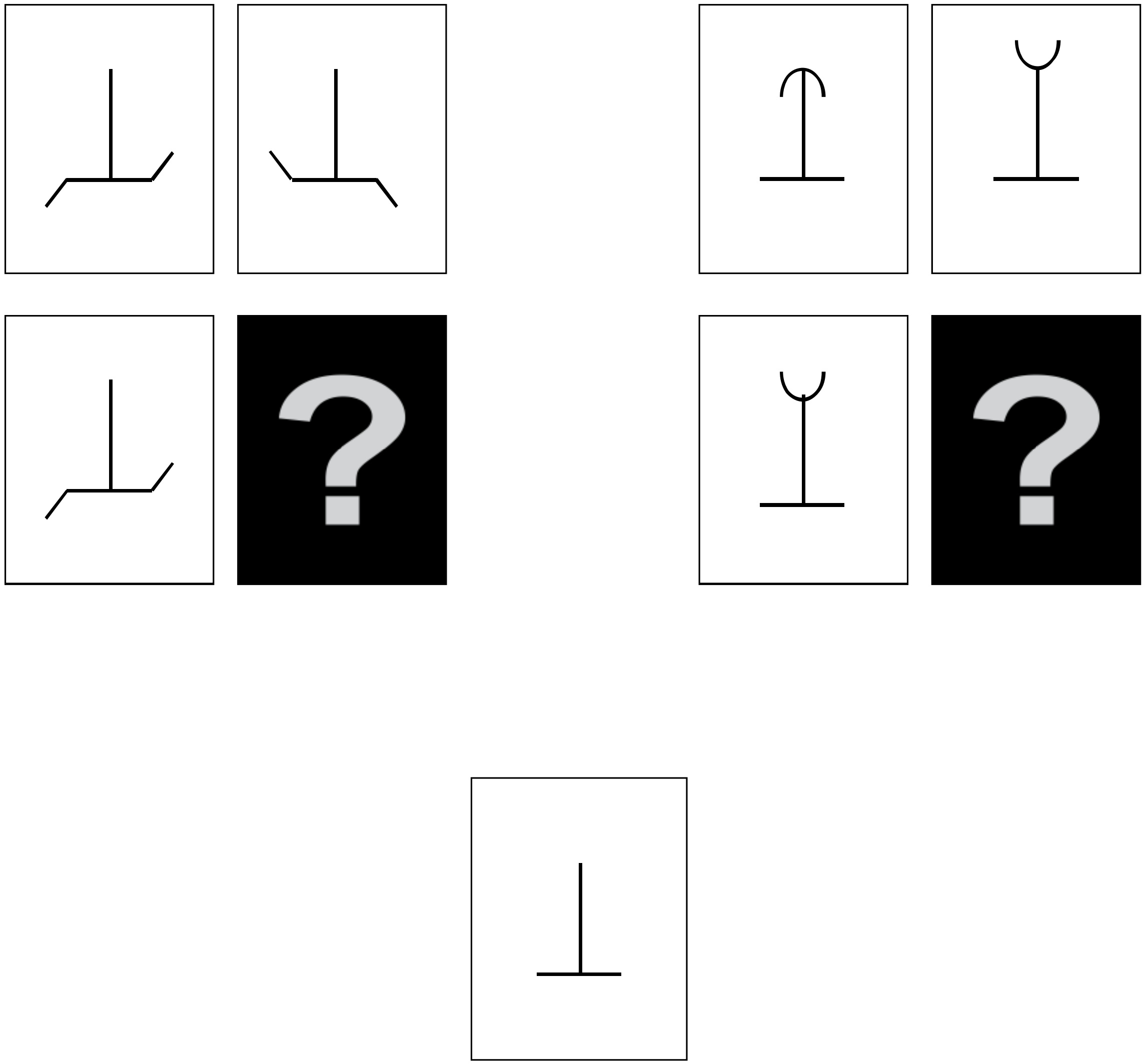




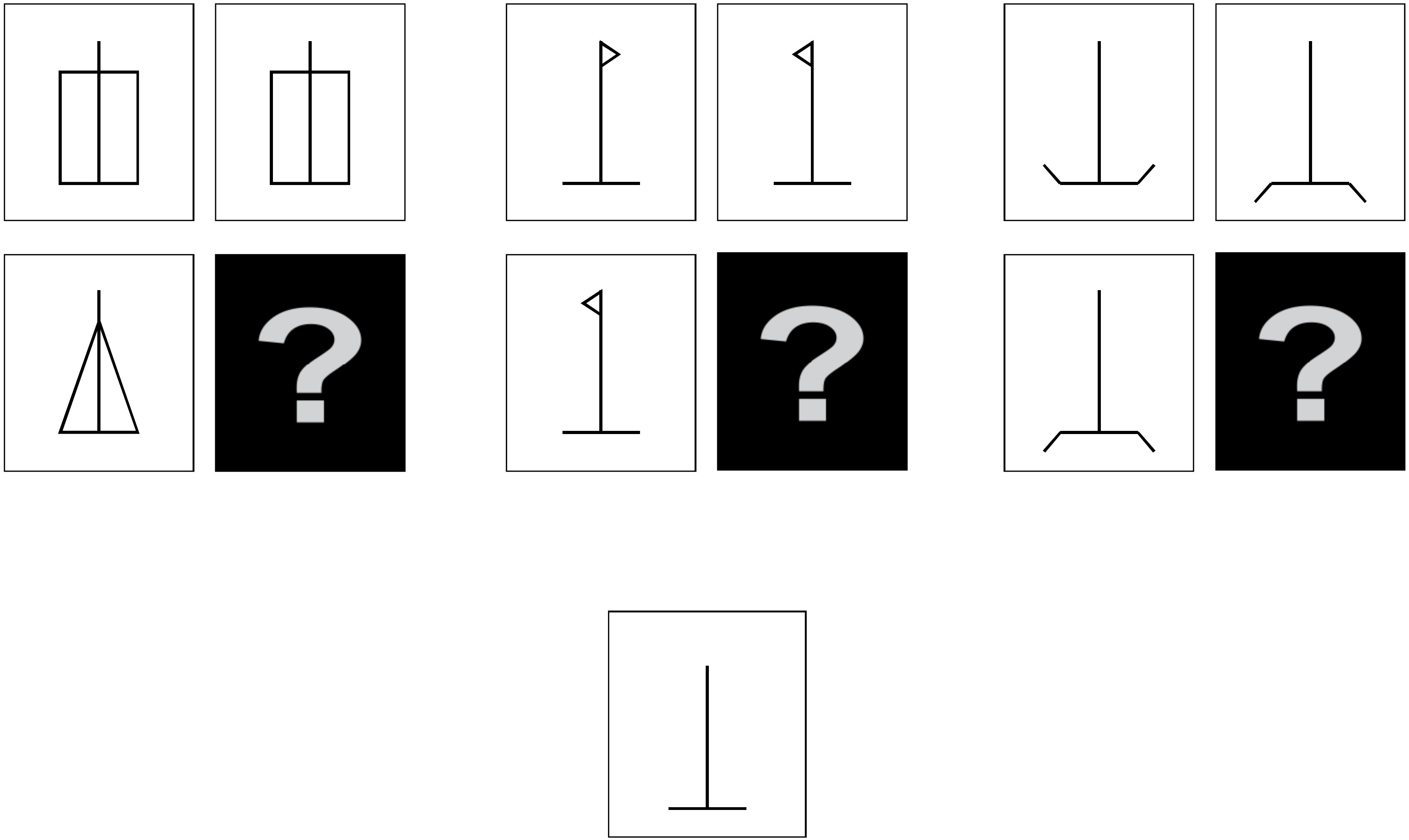


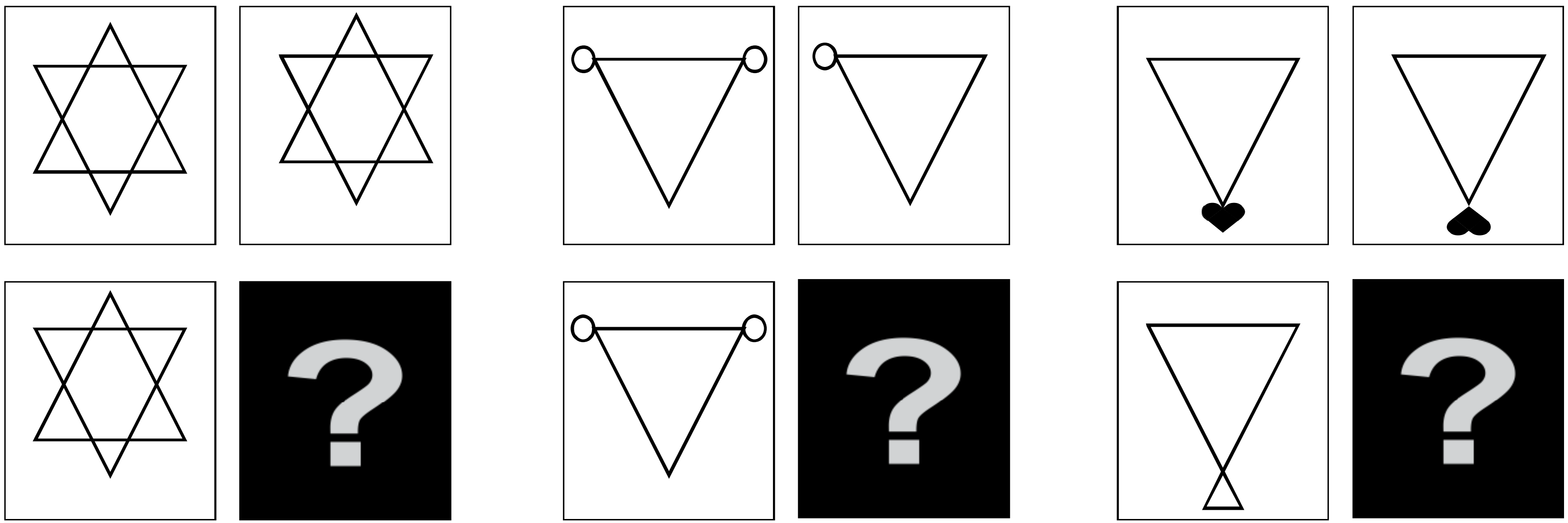


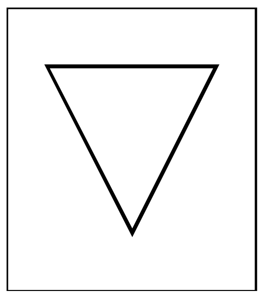


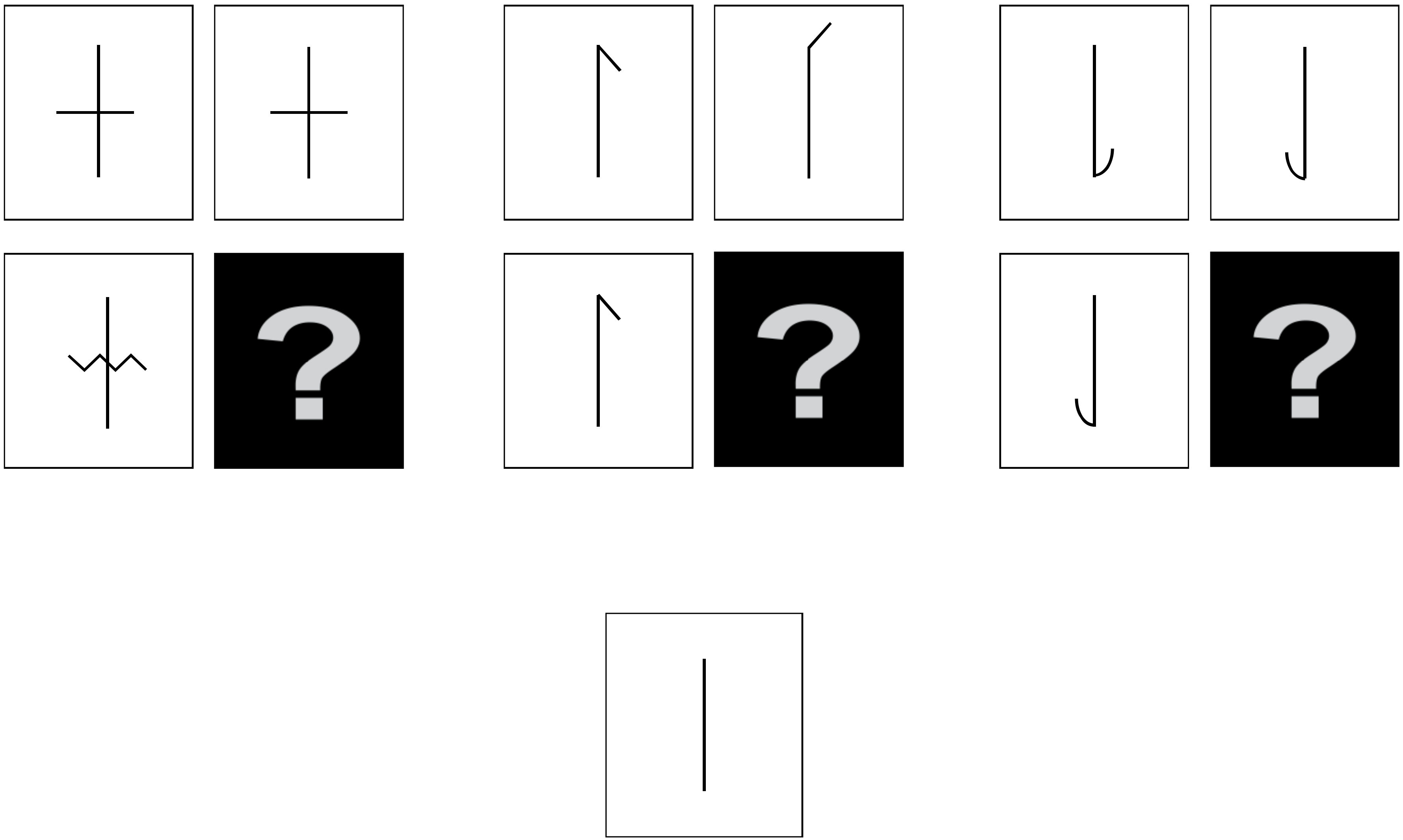


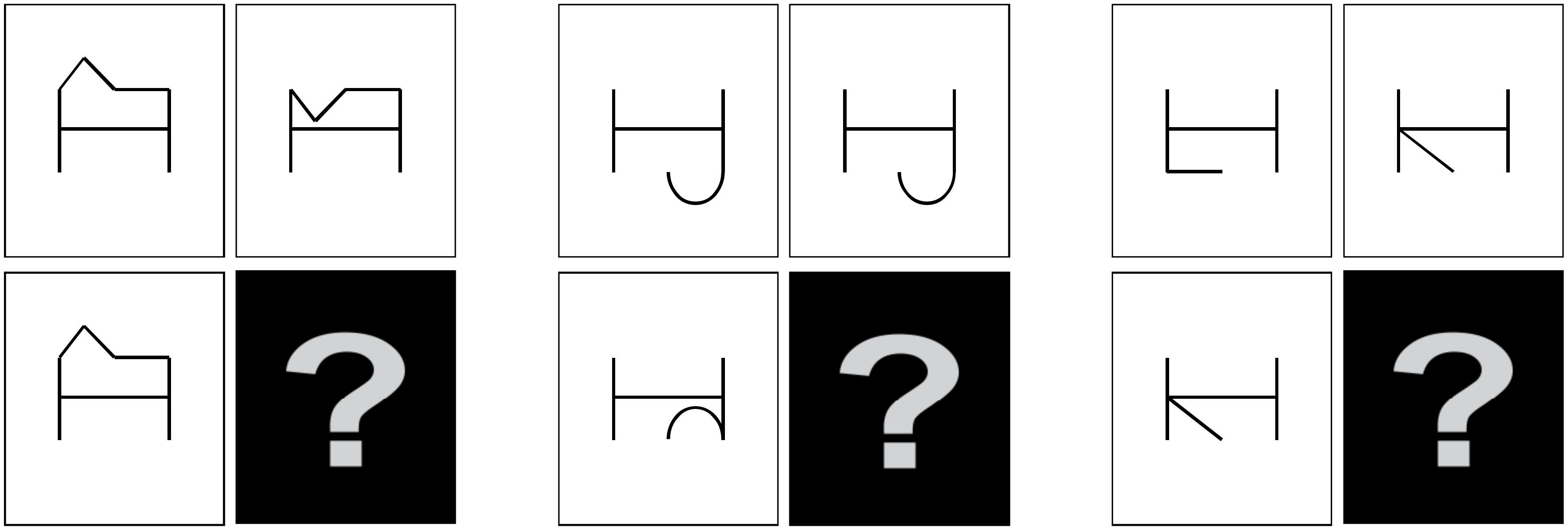
3S

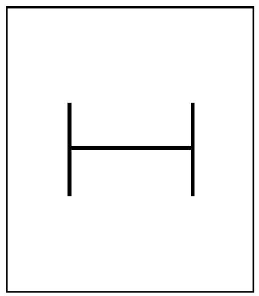


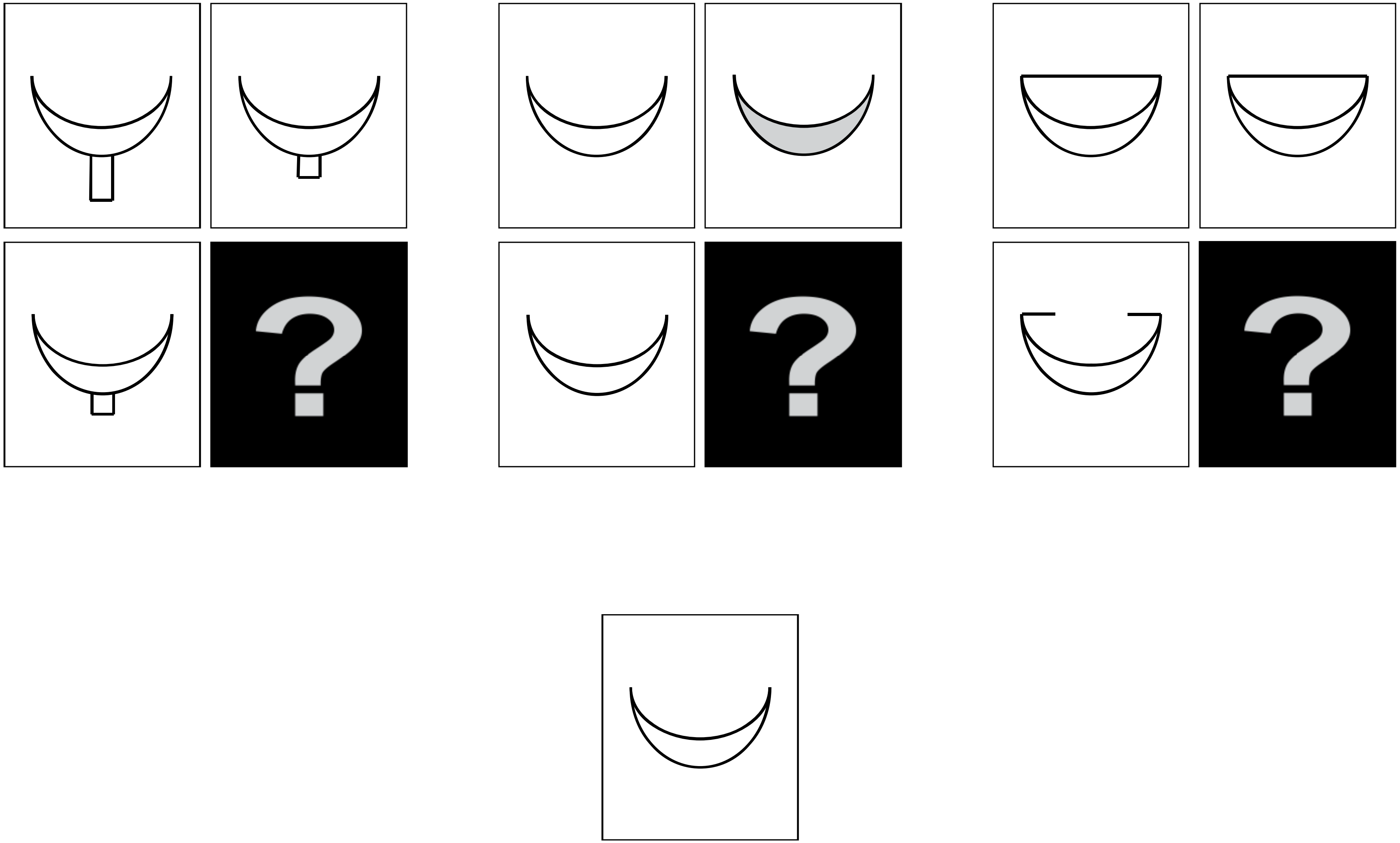


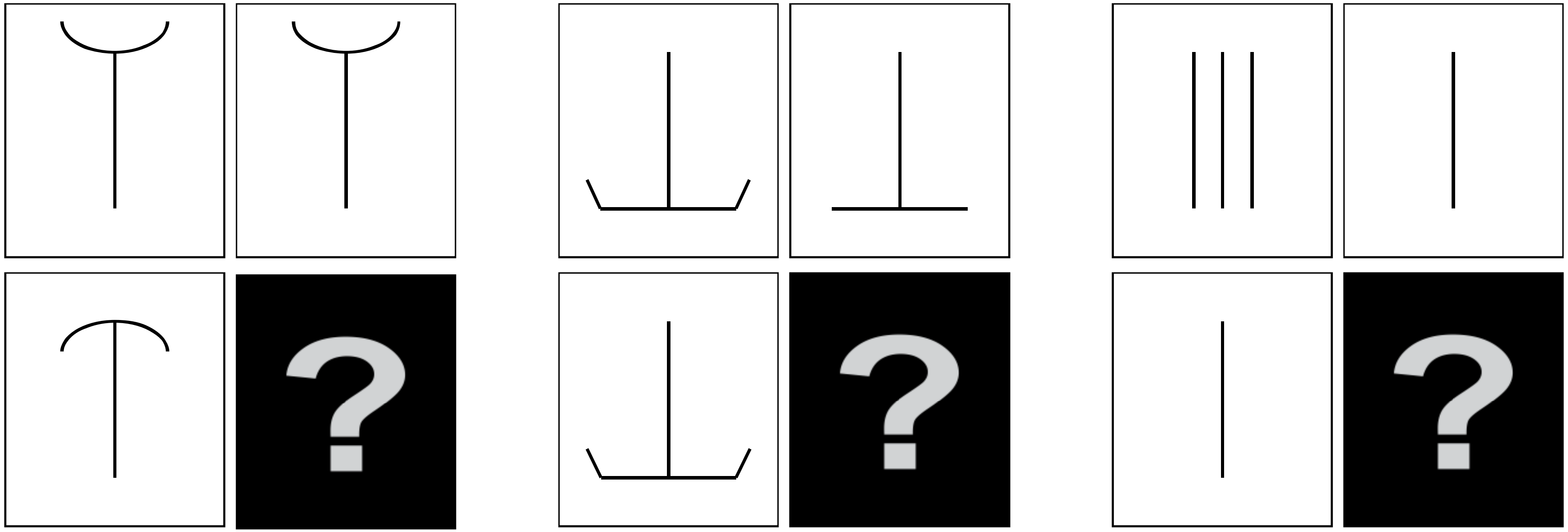


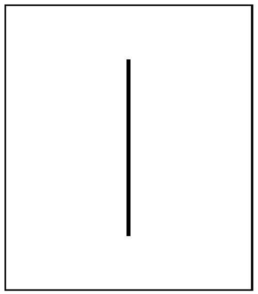


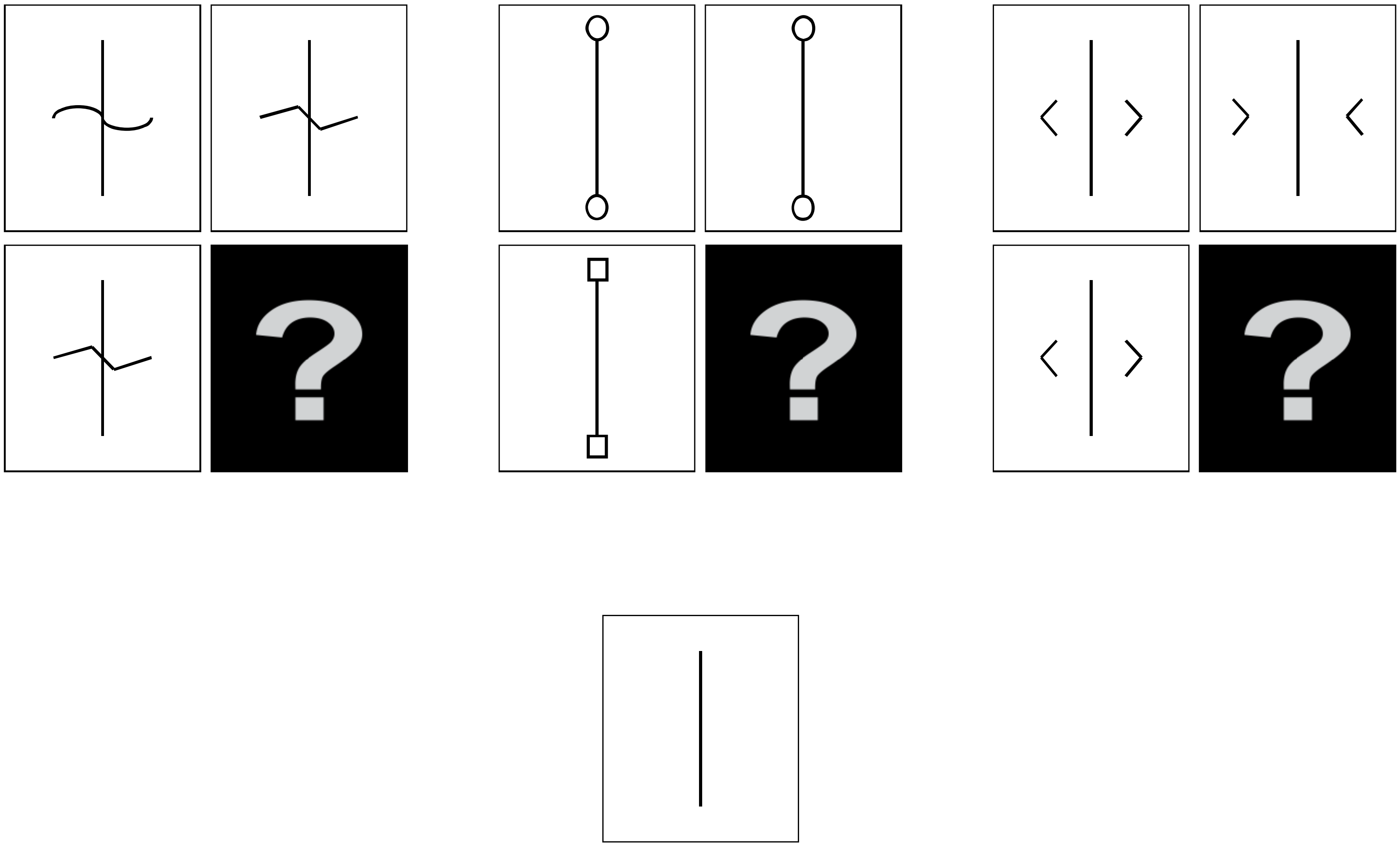


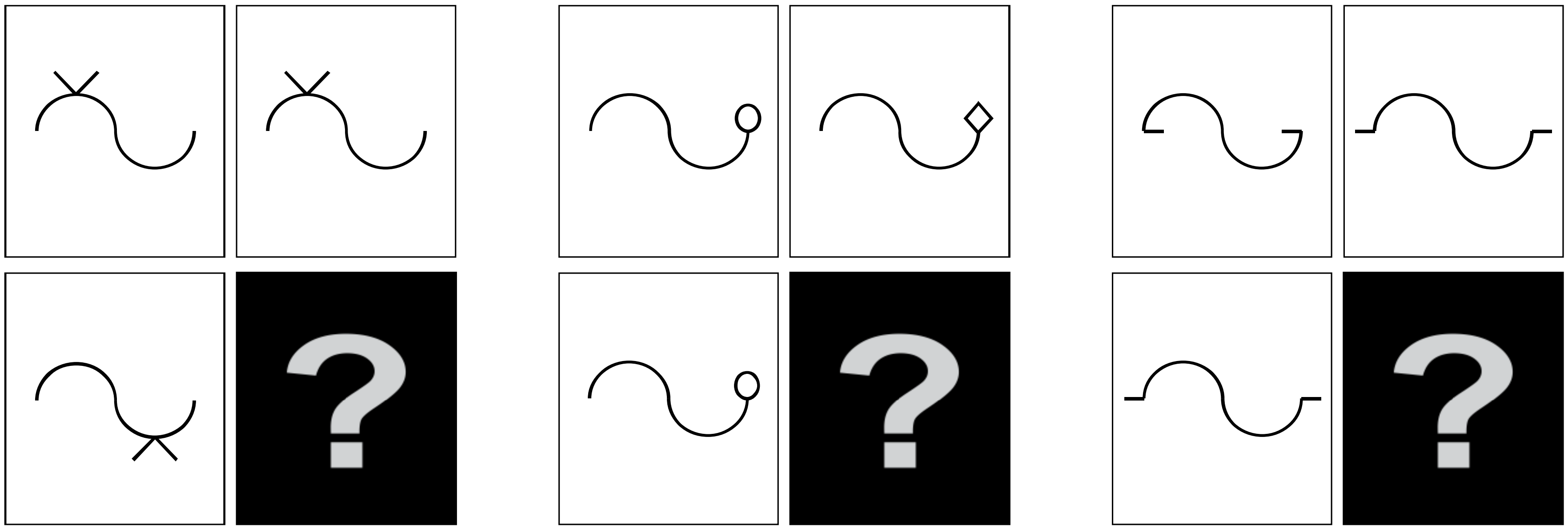


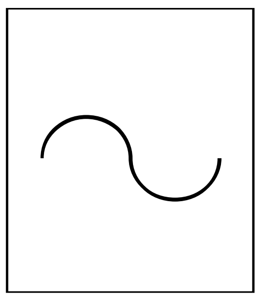


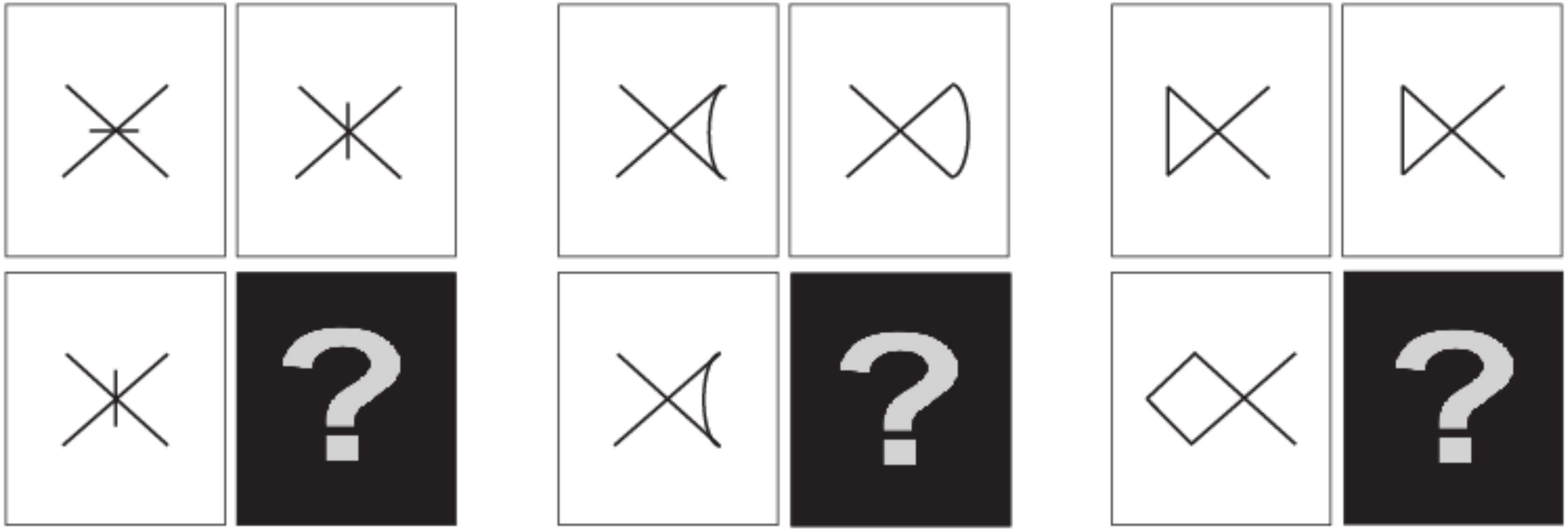


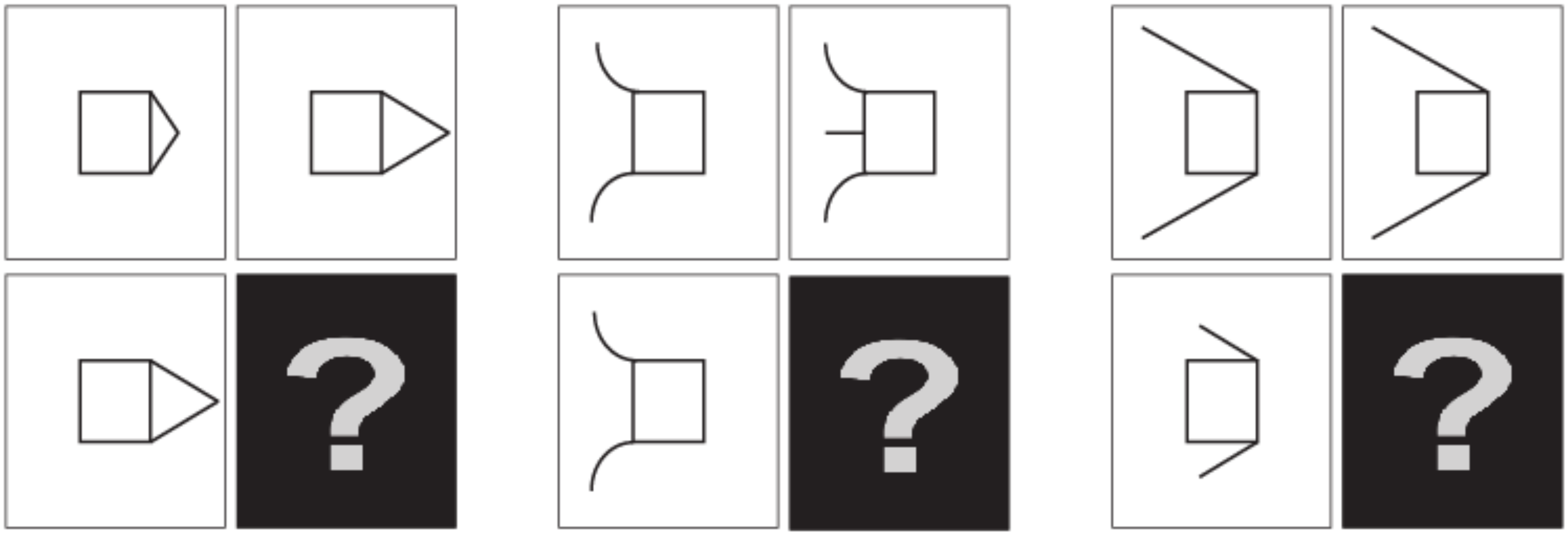


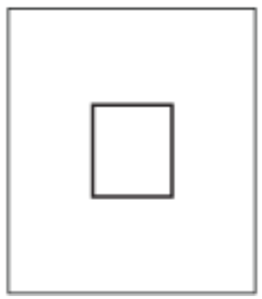






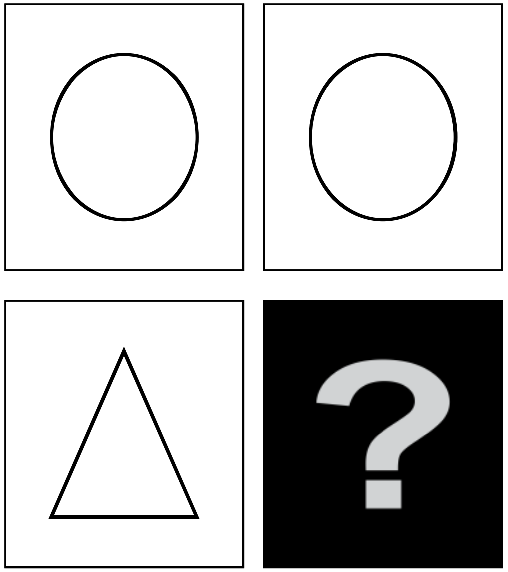


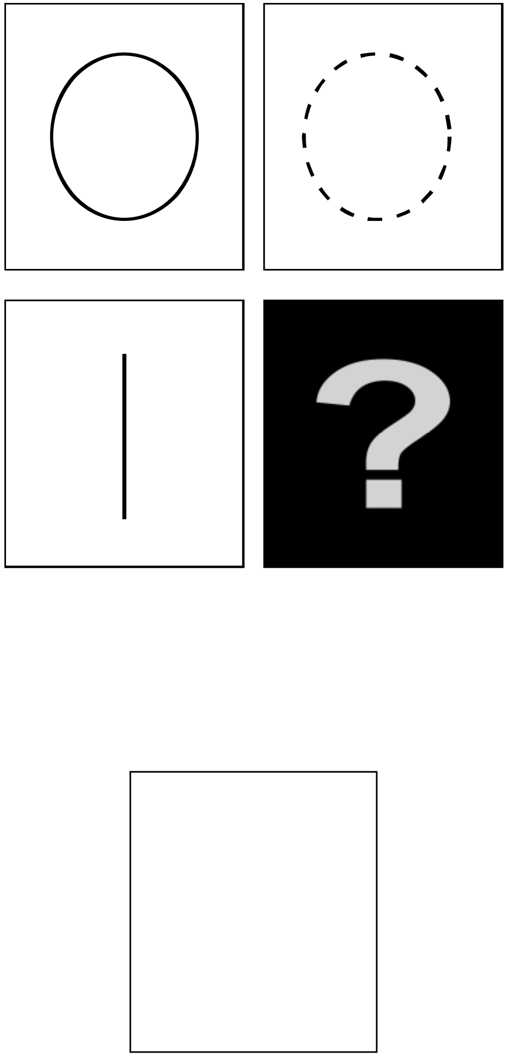


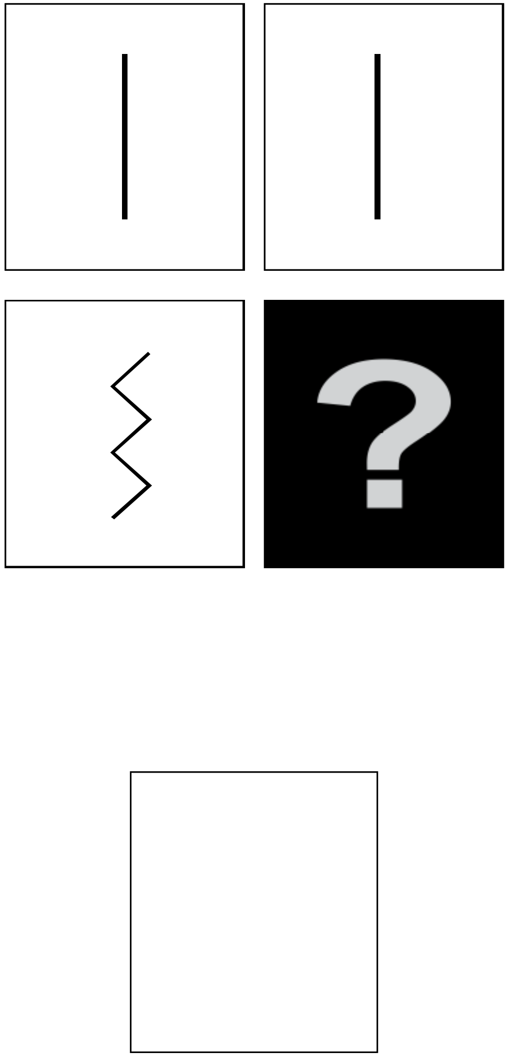


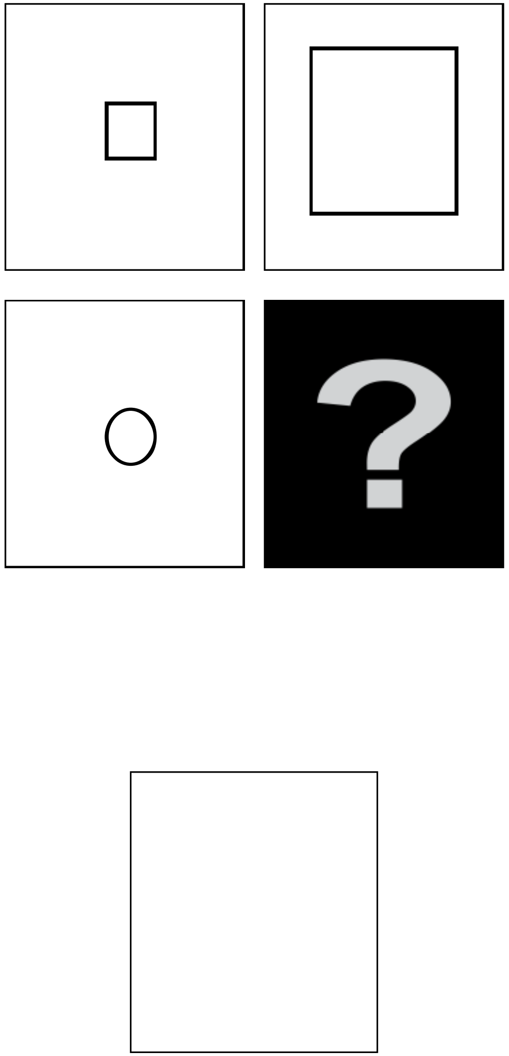
B2

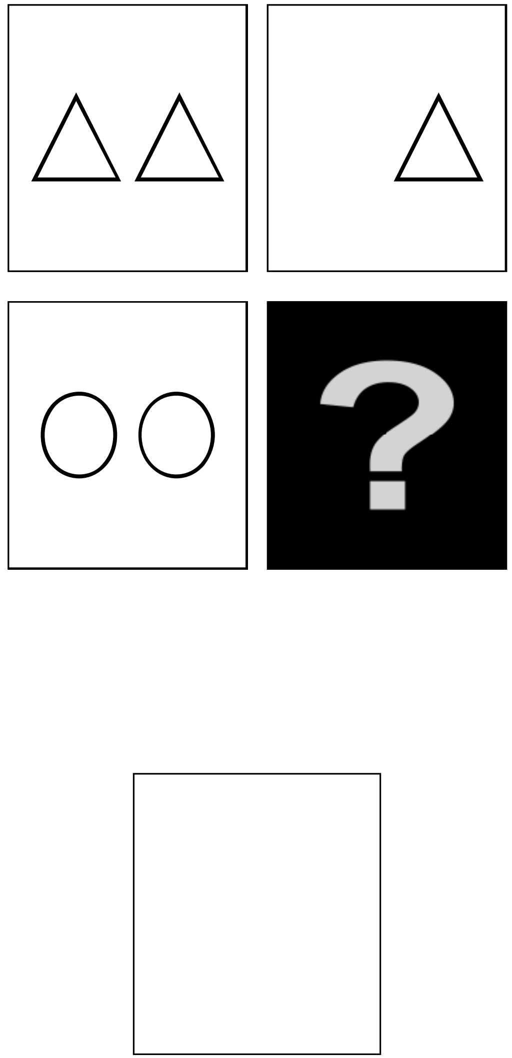
ID: ________________________

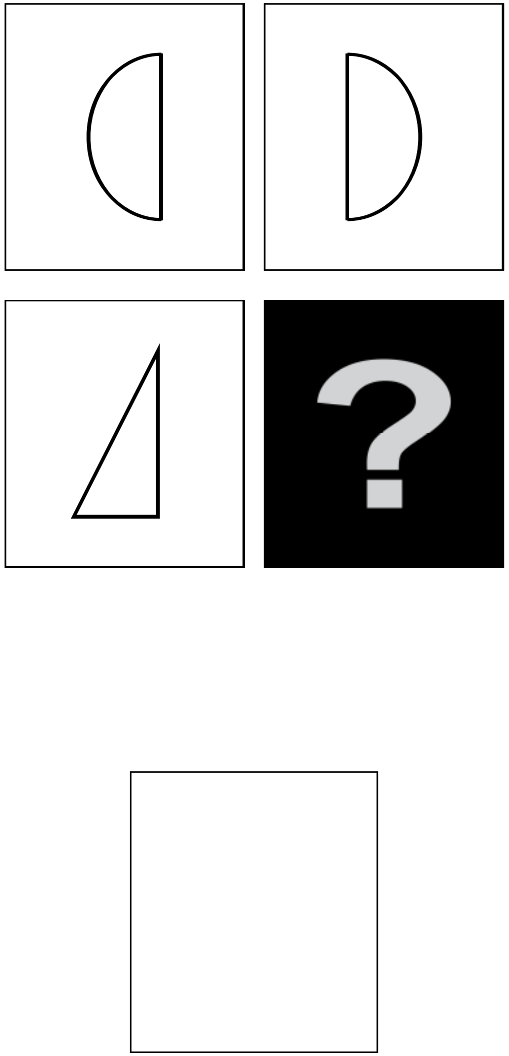


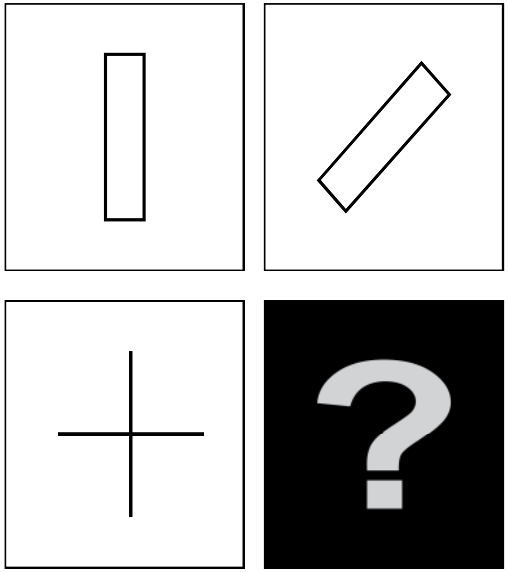




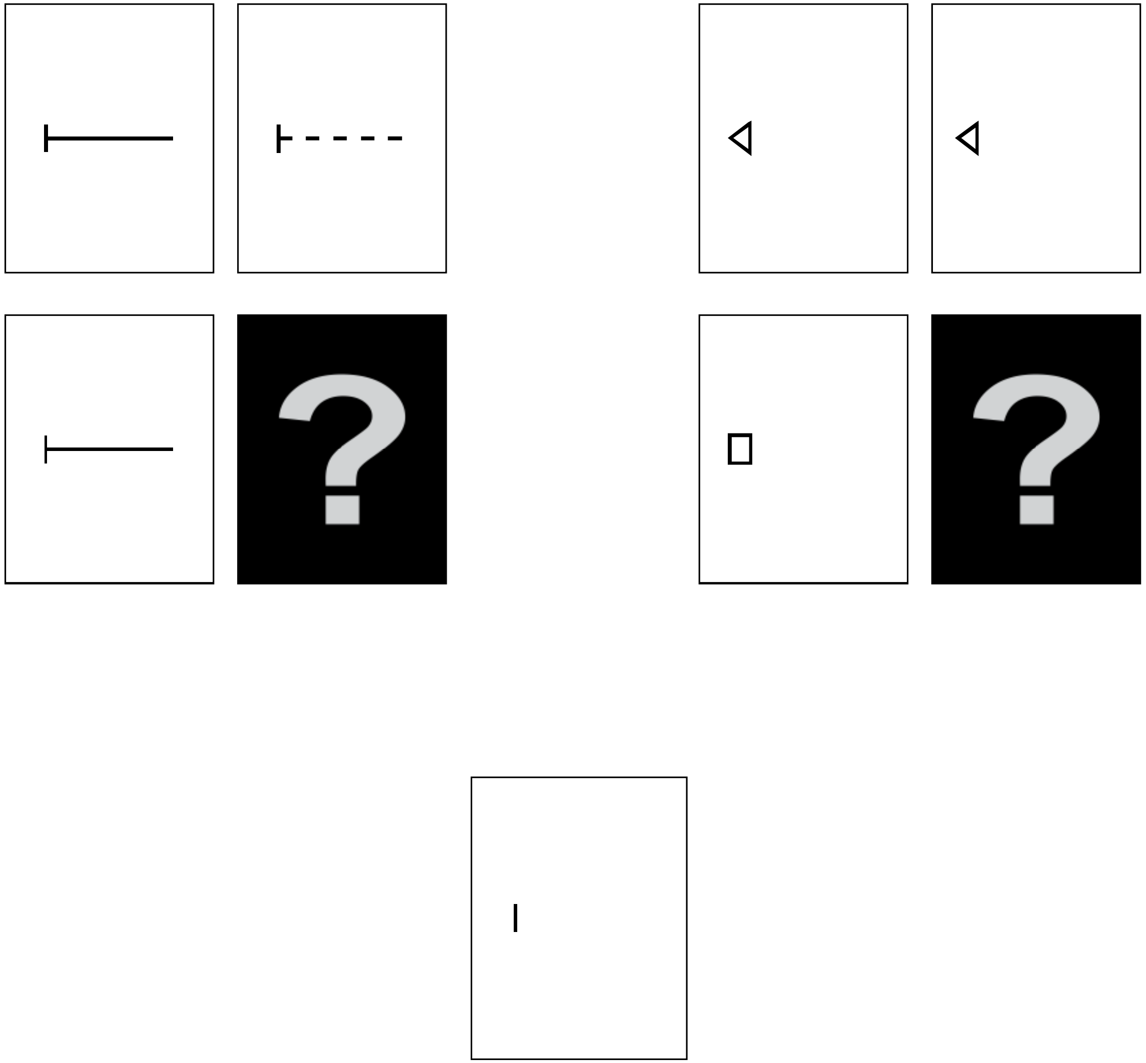


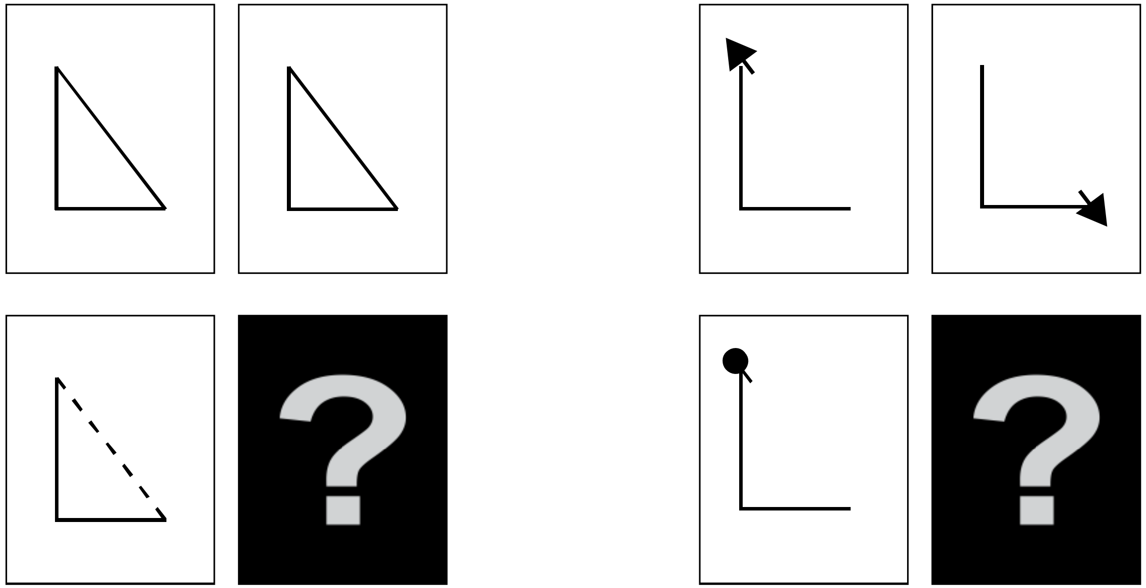


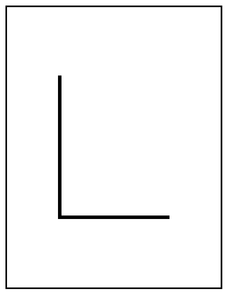


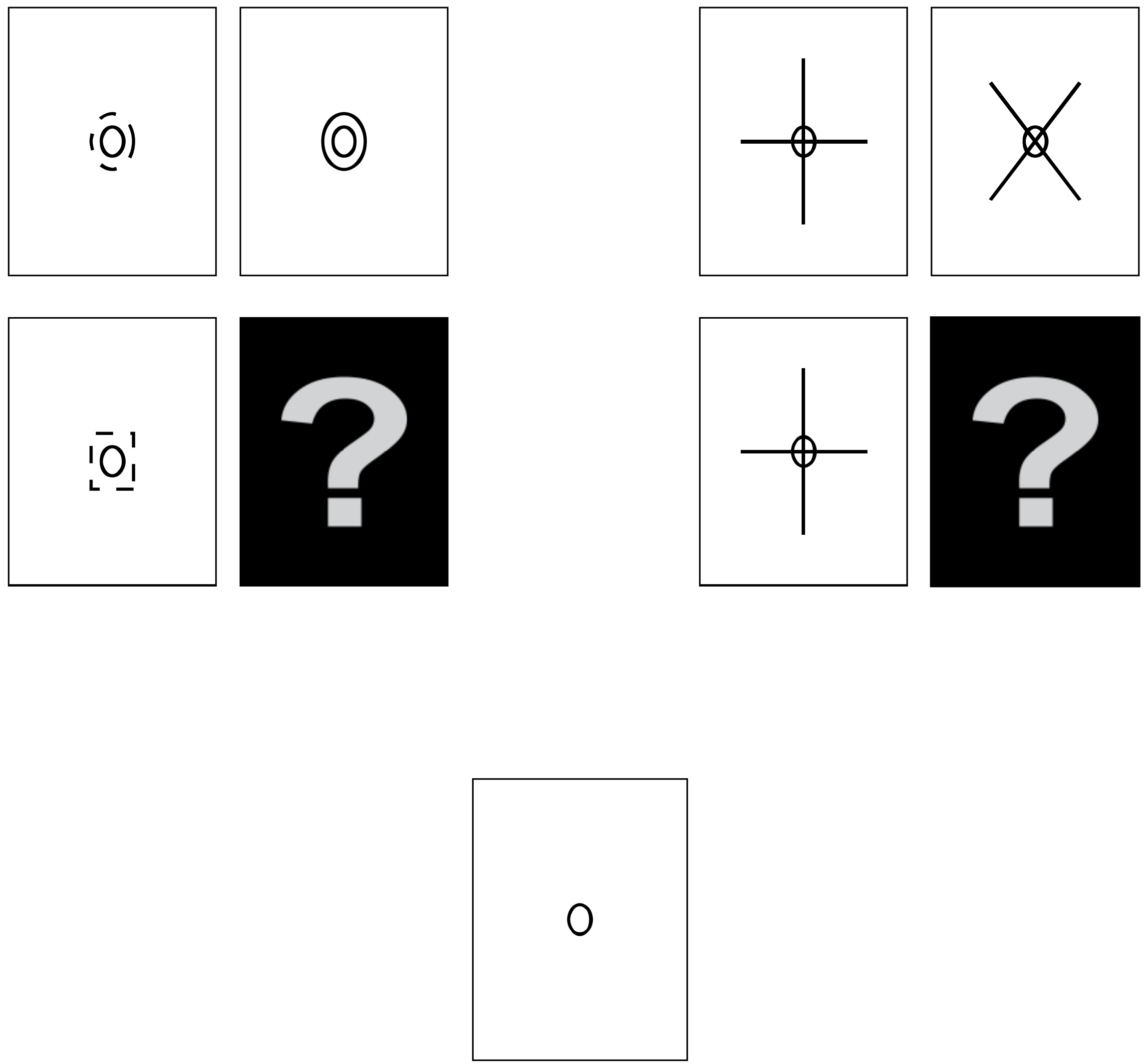


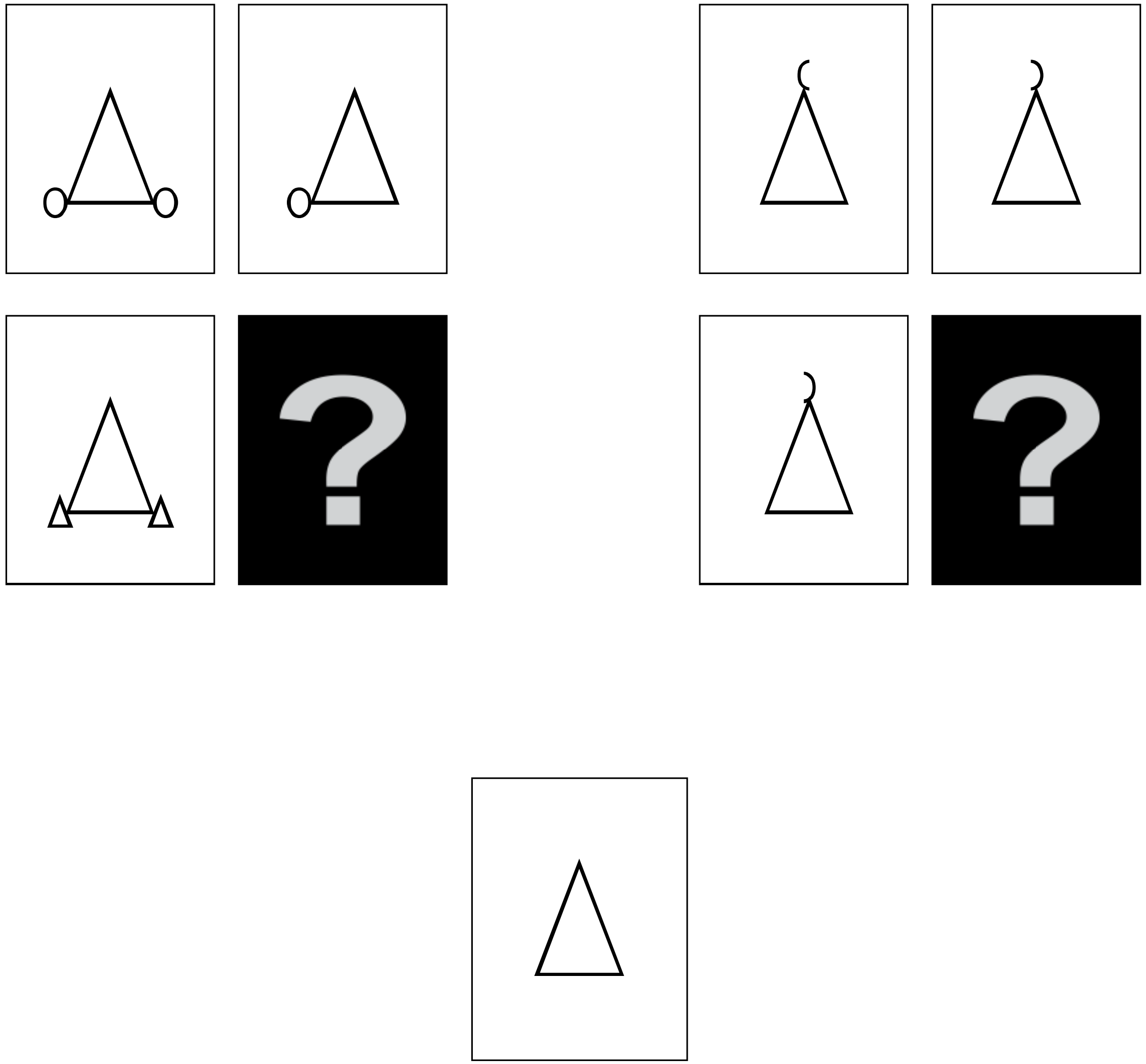
2S

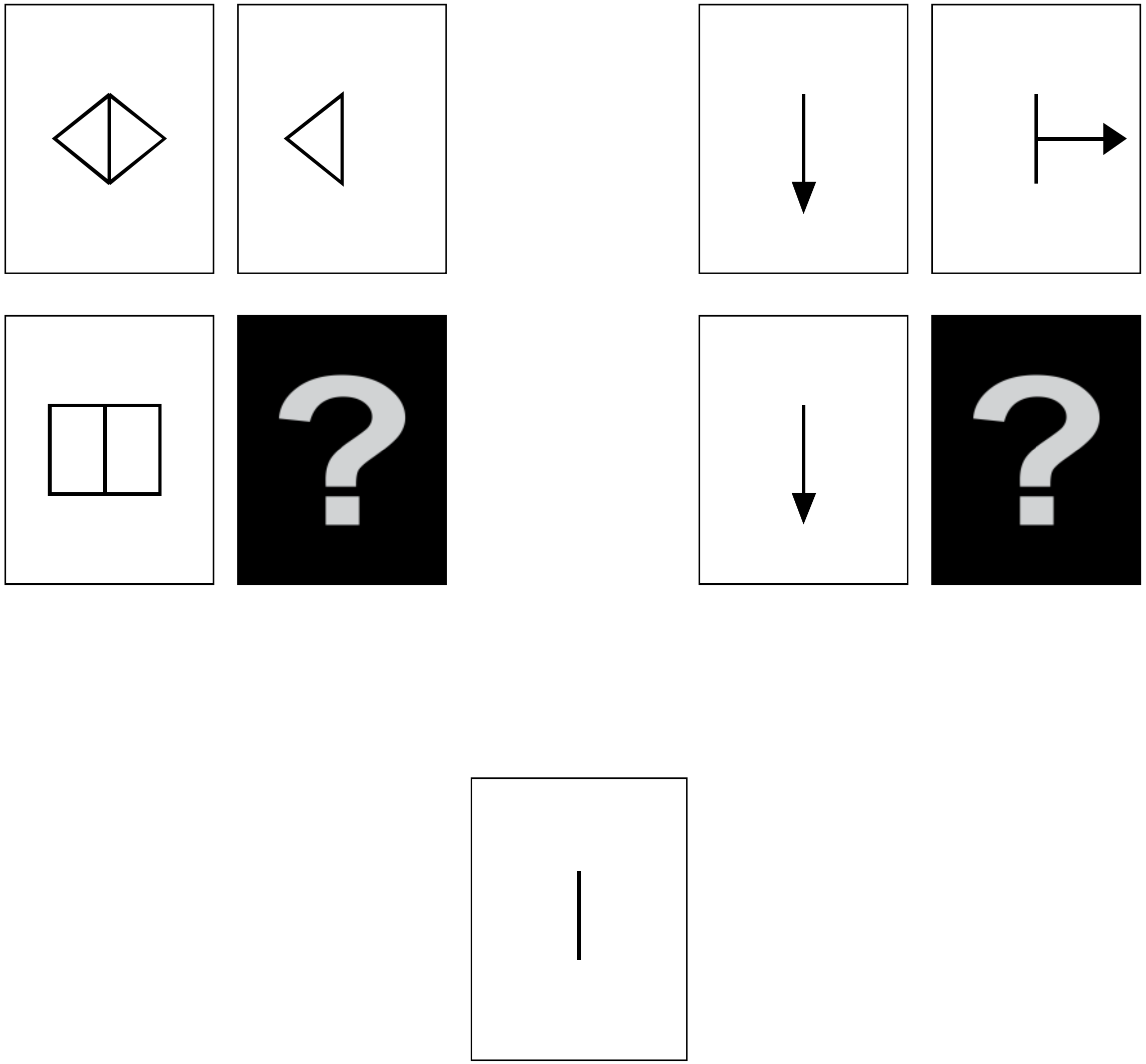


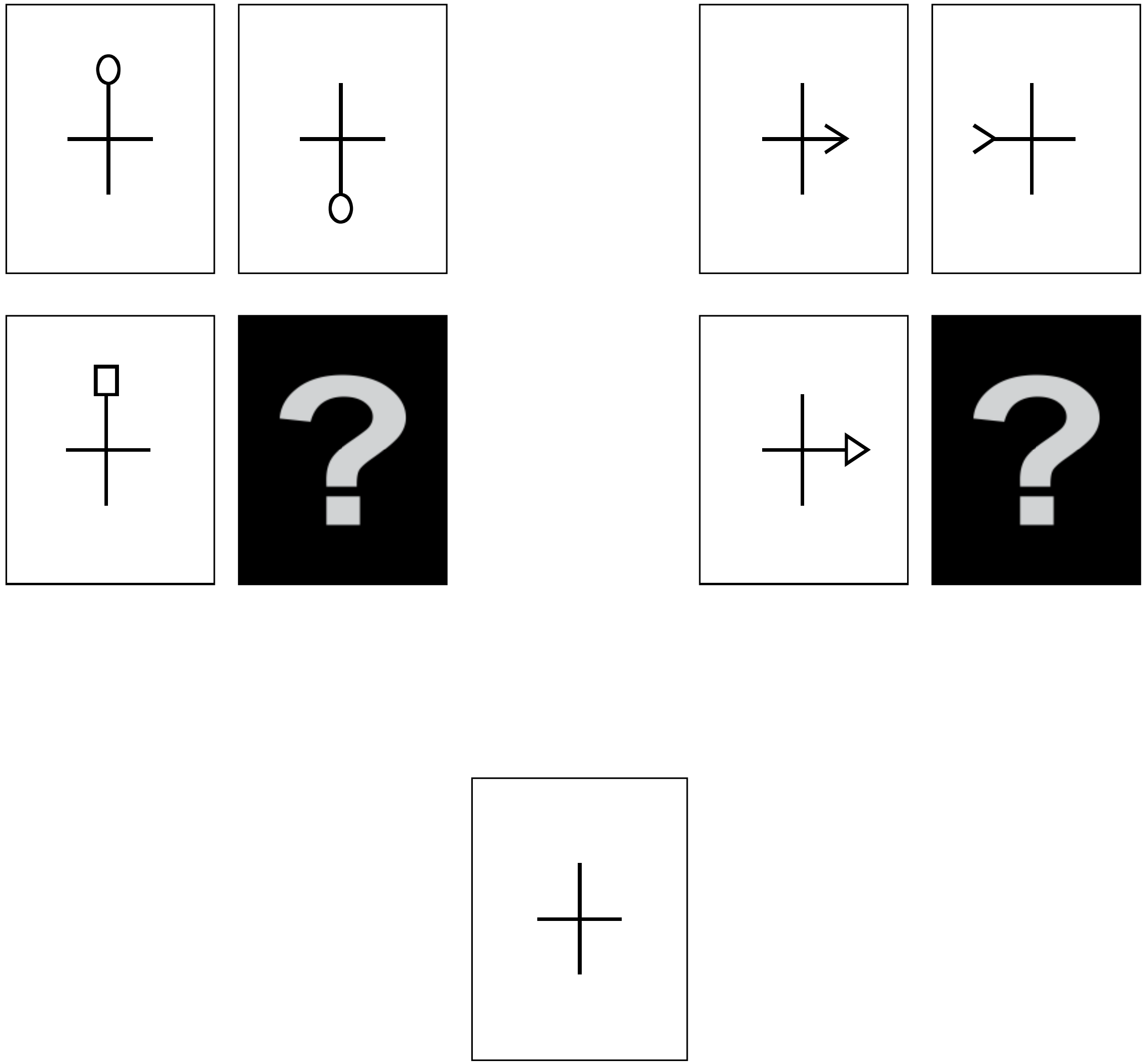


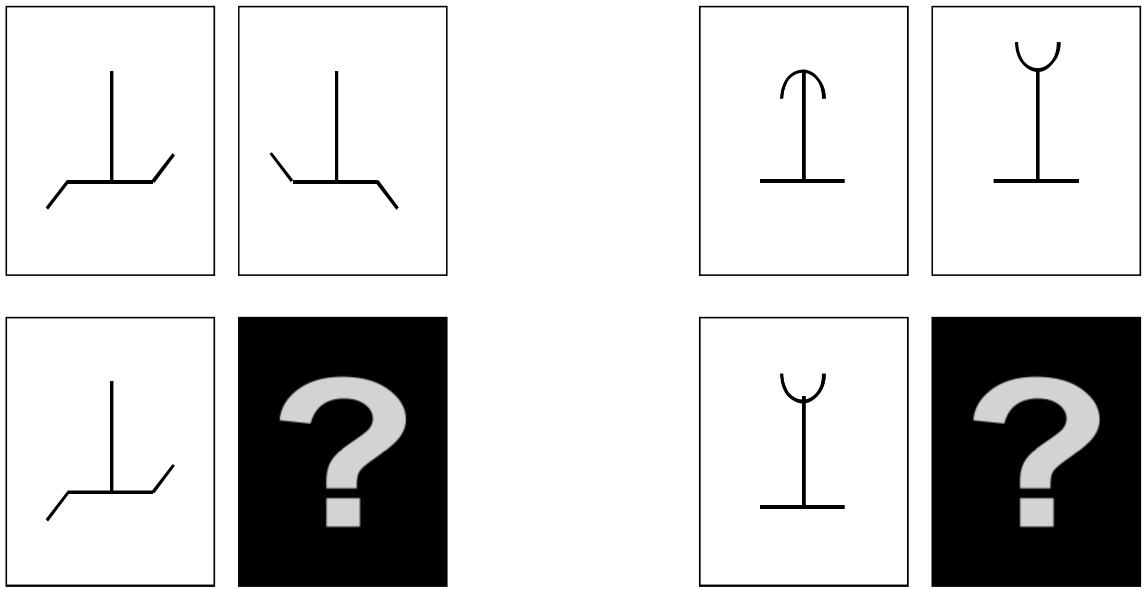


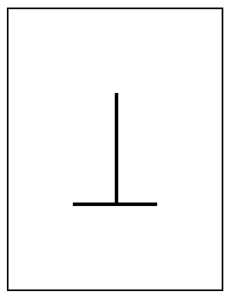




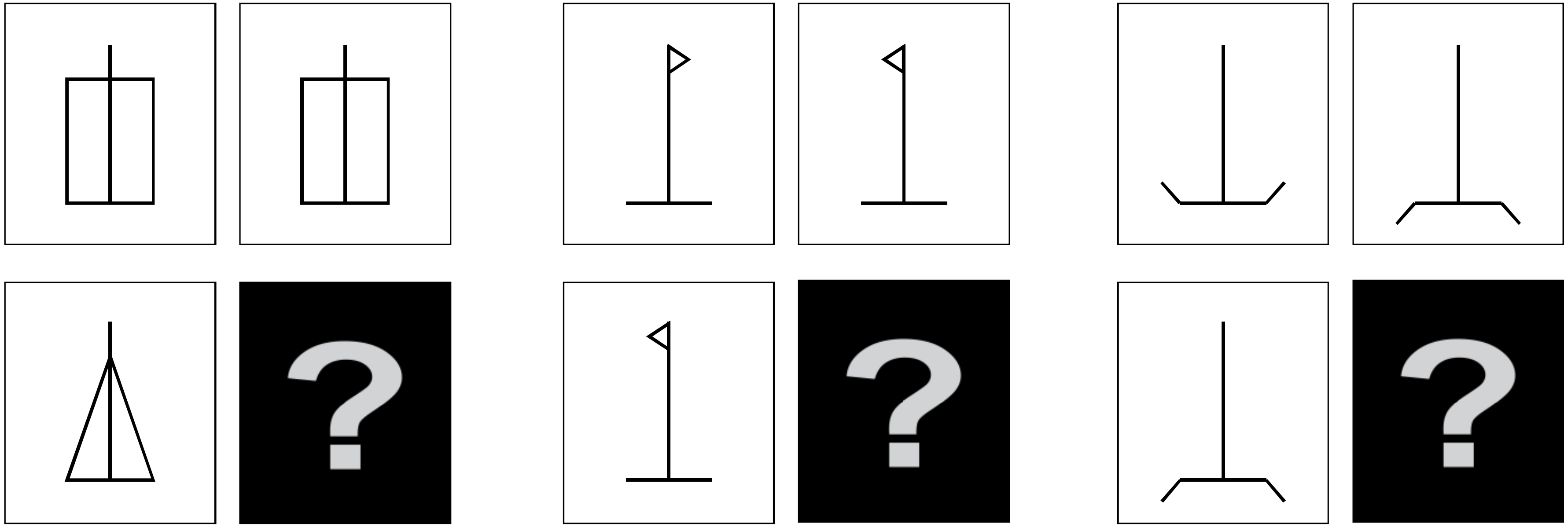


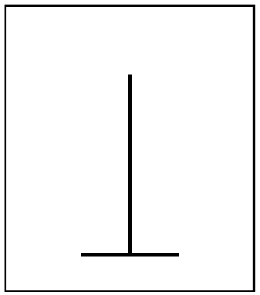


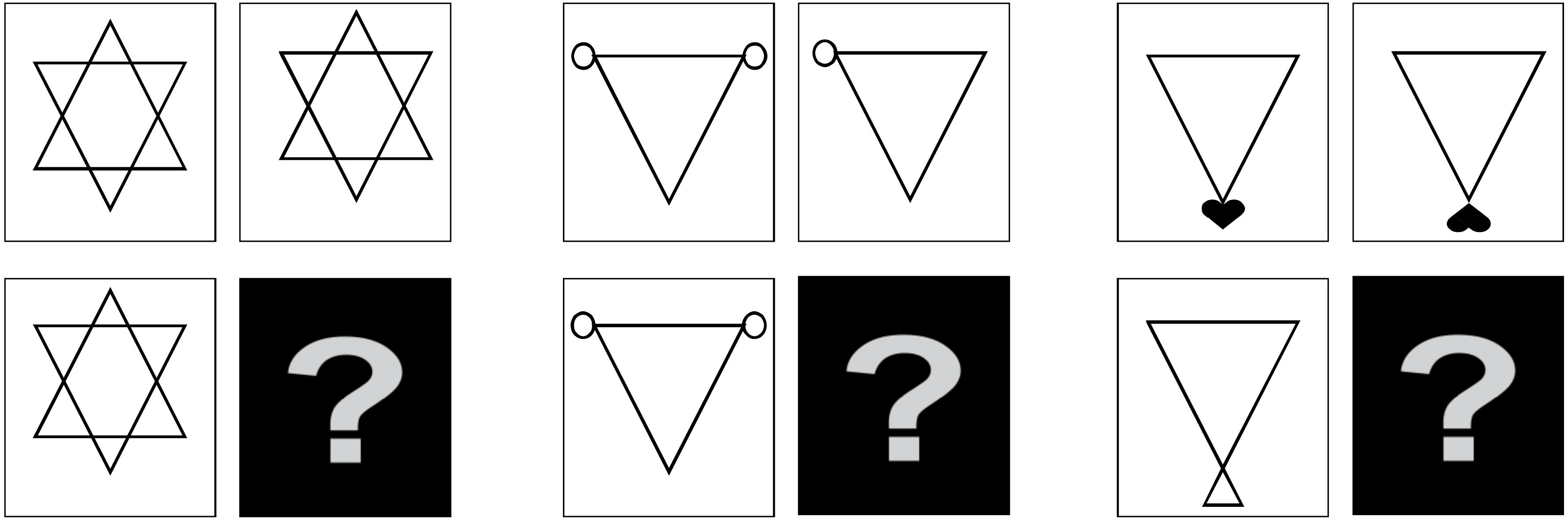


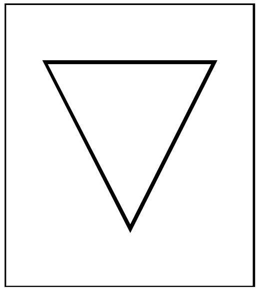


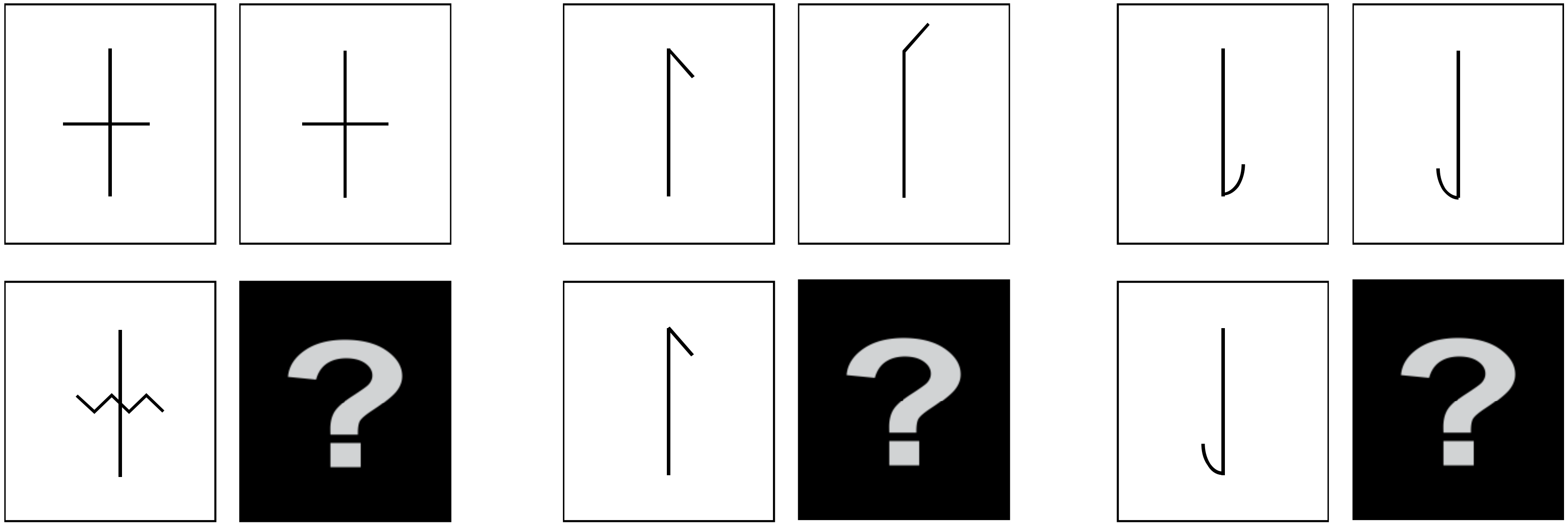
3S

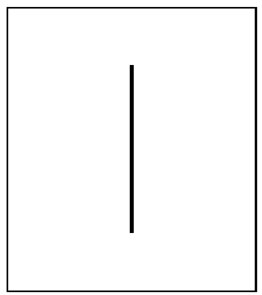


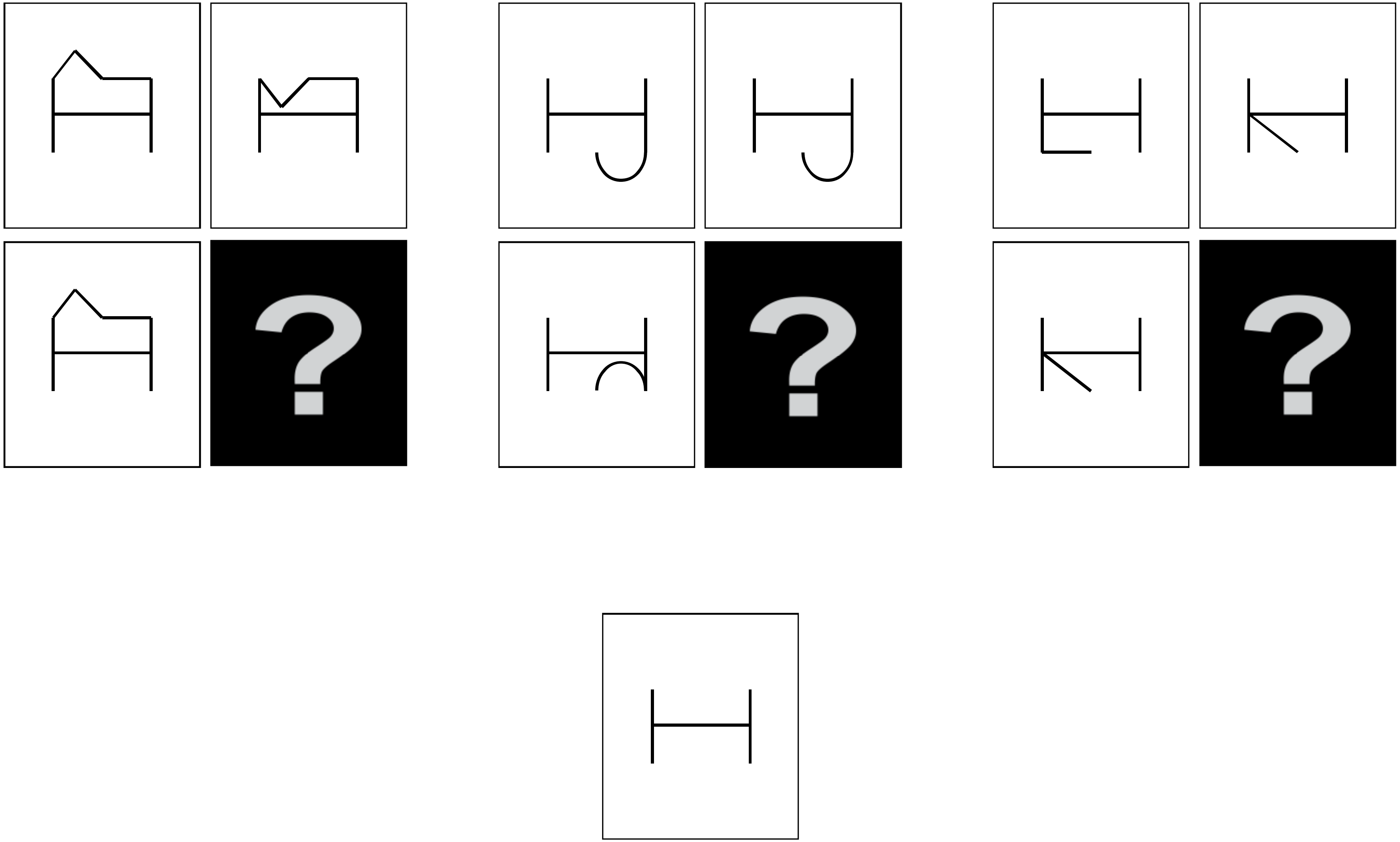


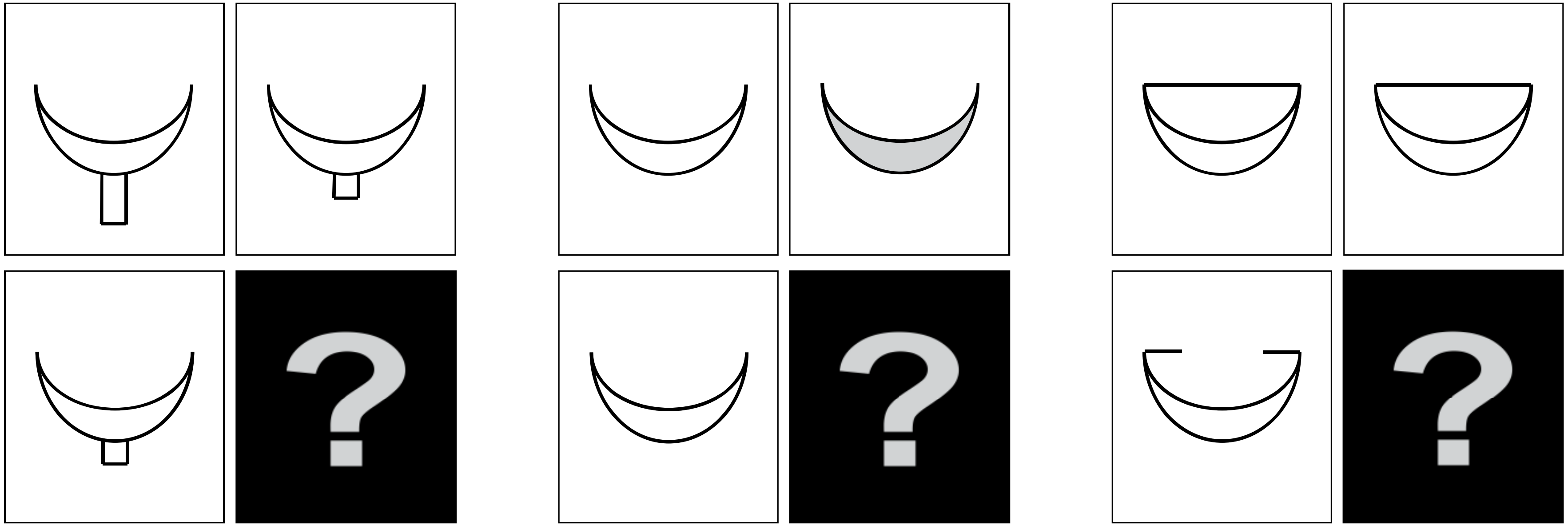


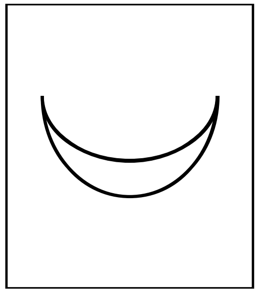


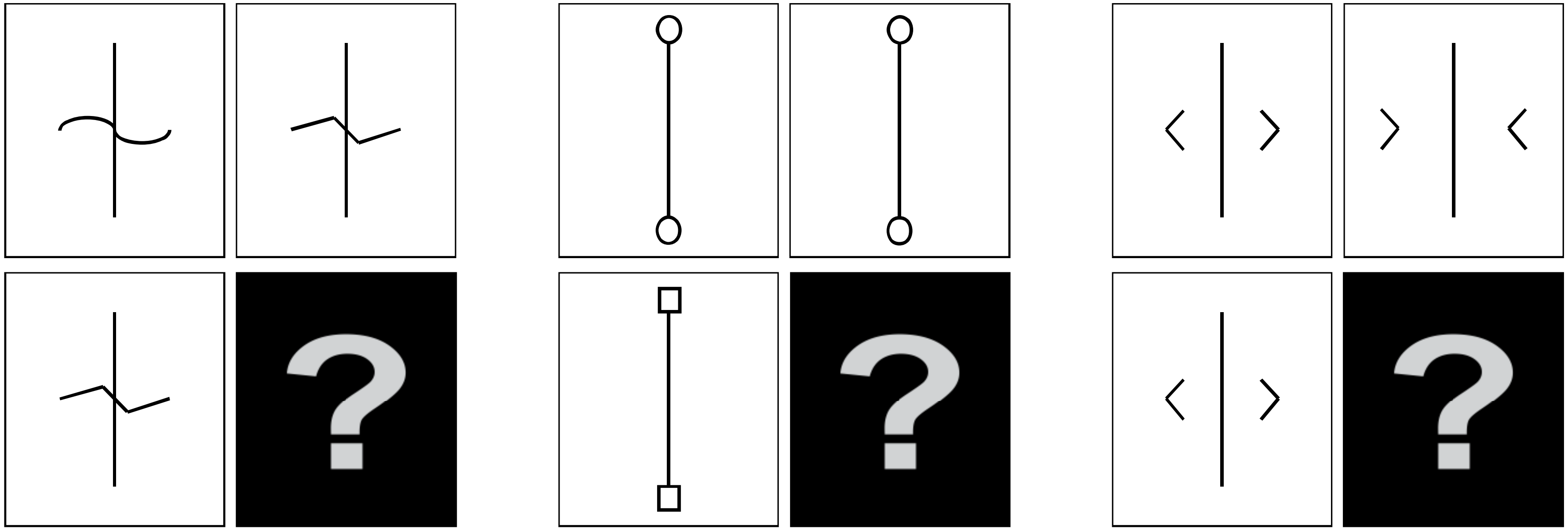


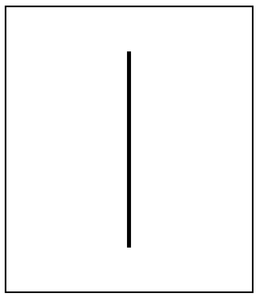


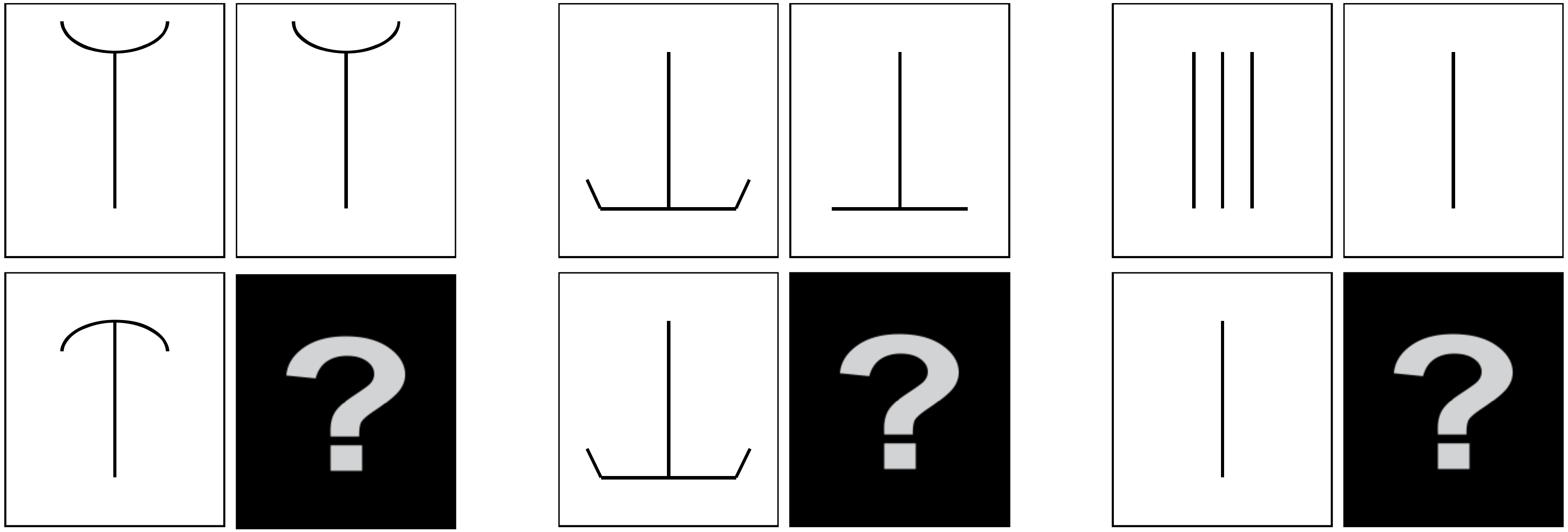


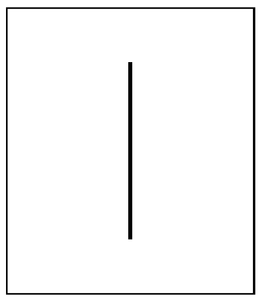


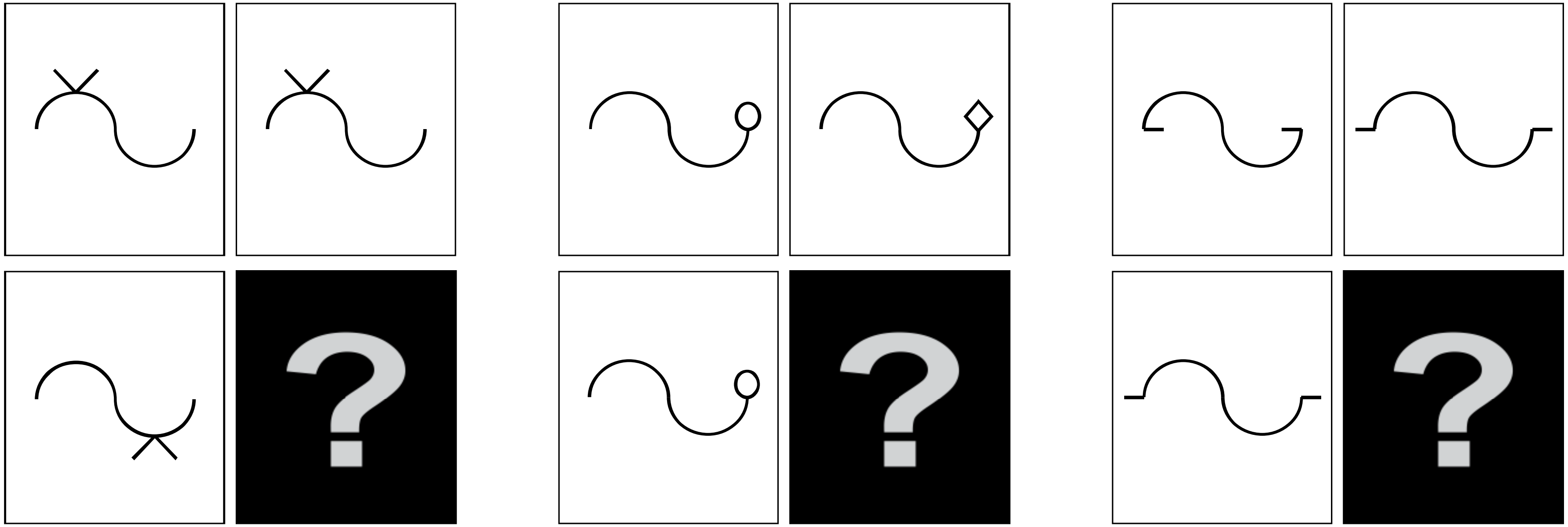


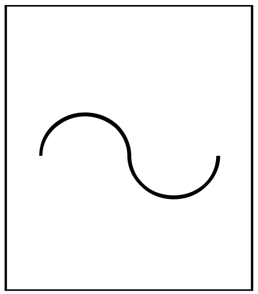


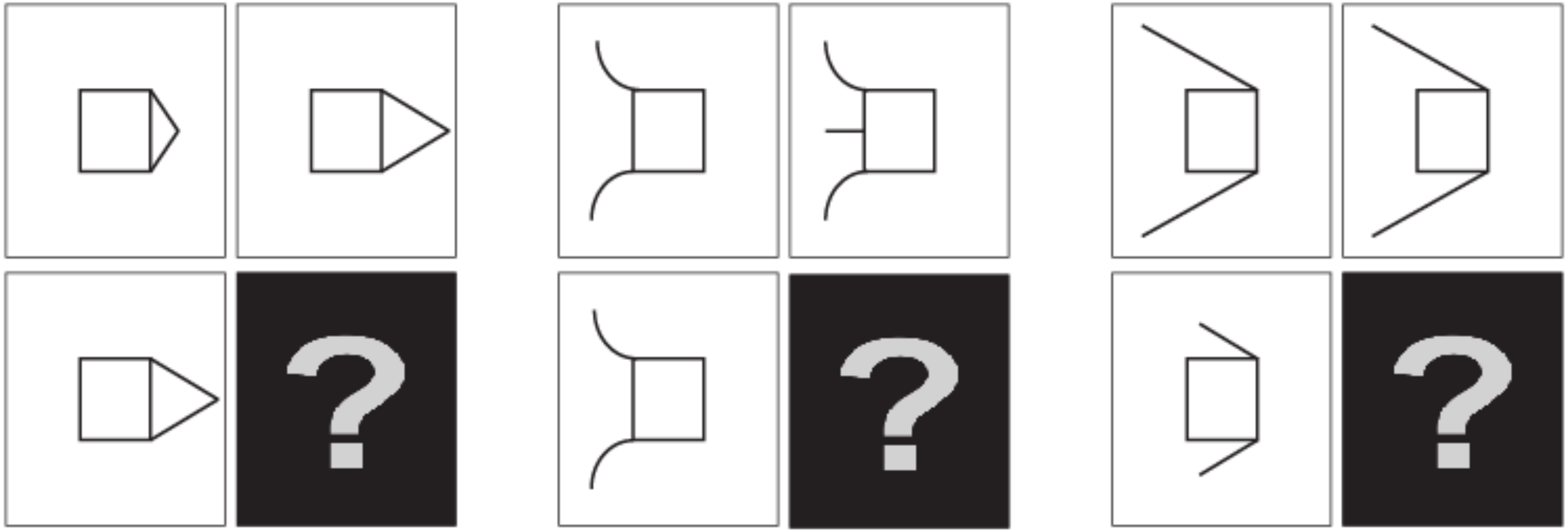


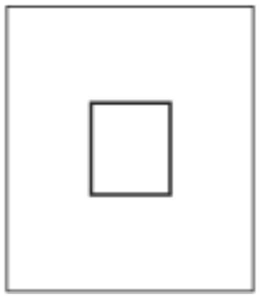


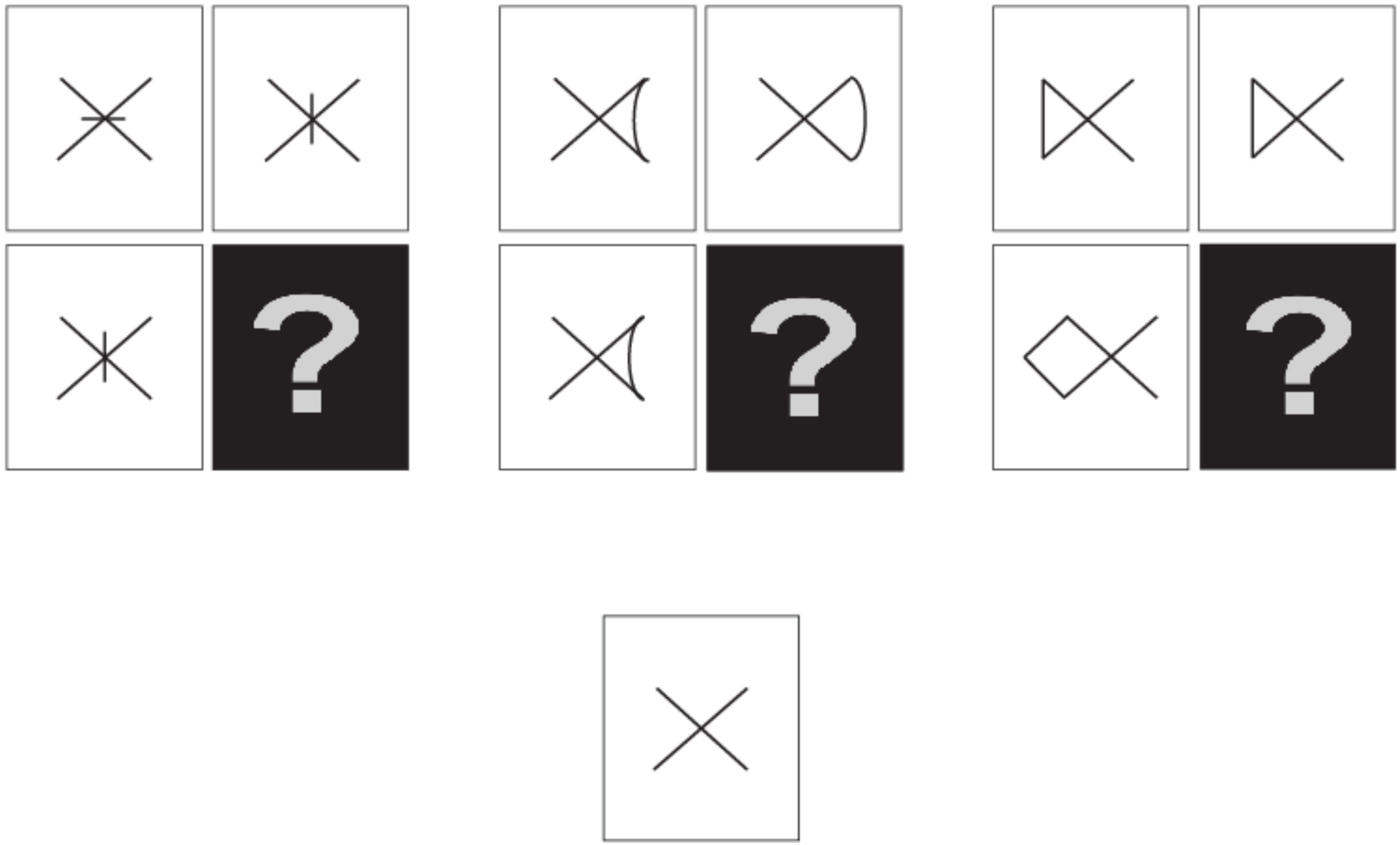




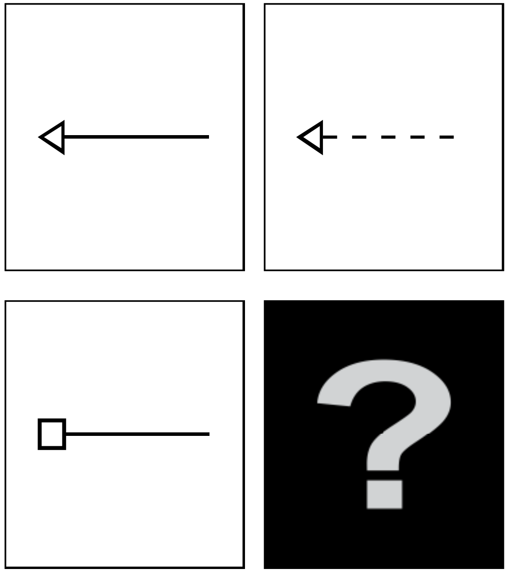


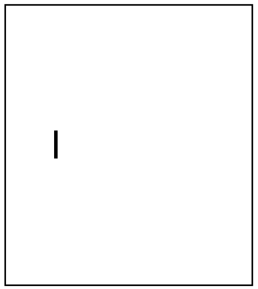


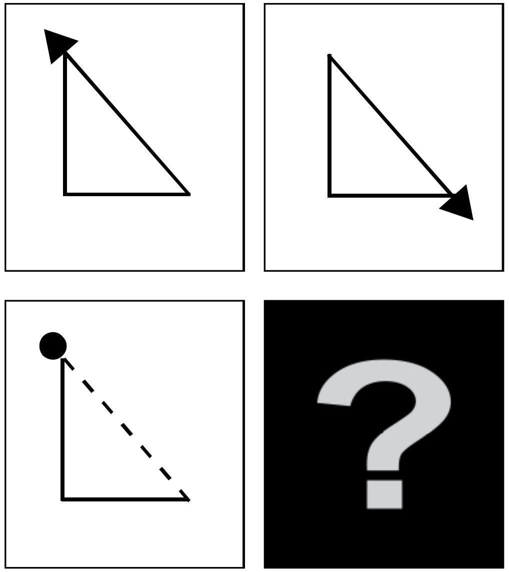


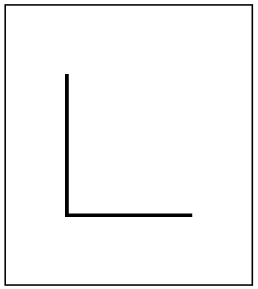


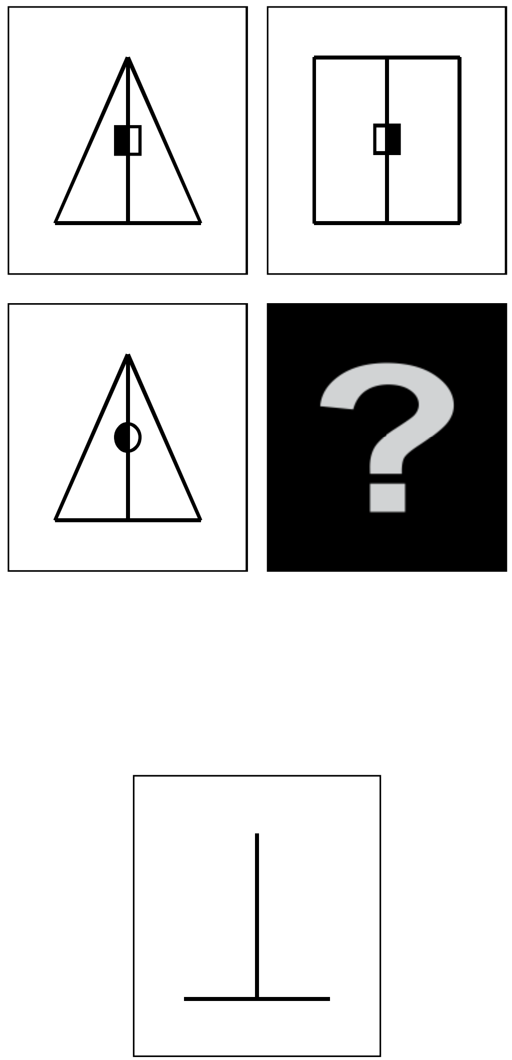
2C

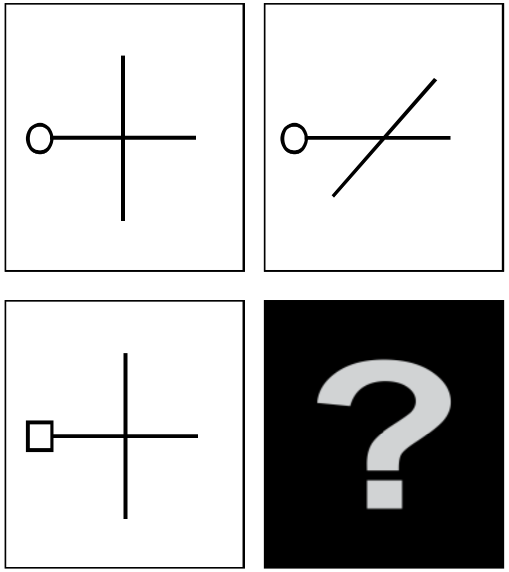


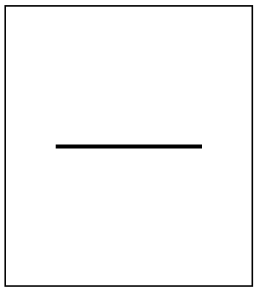


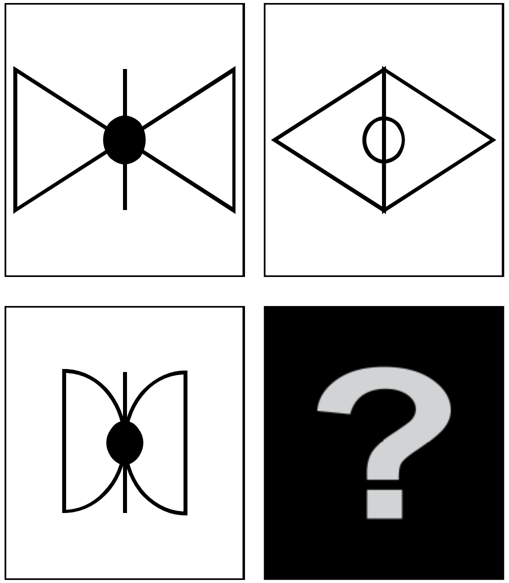


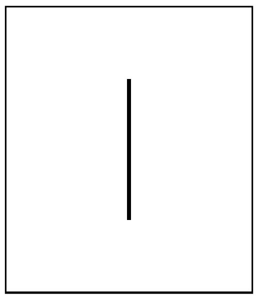


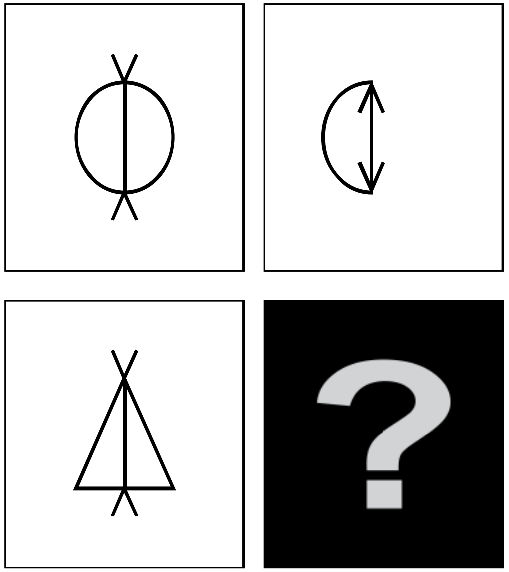


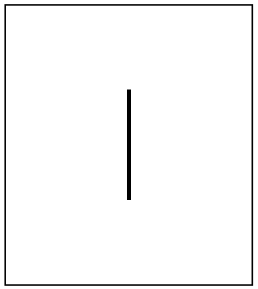


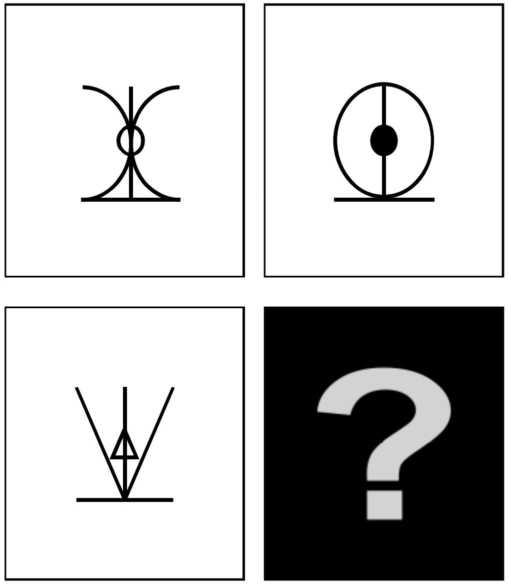


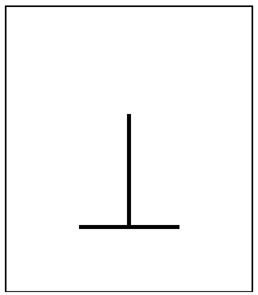




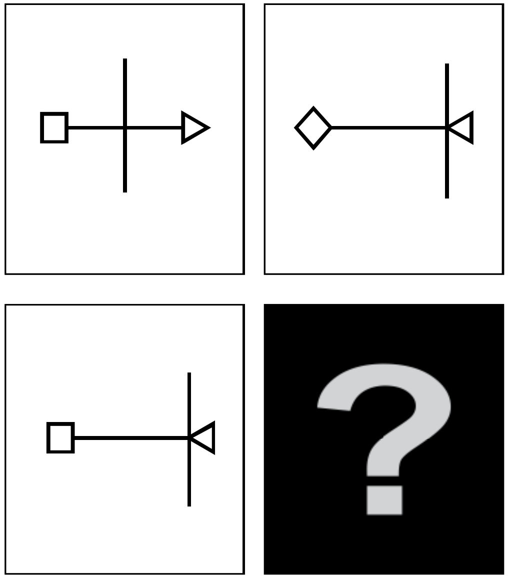


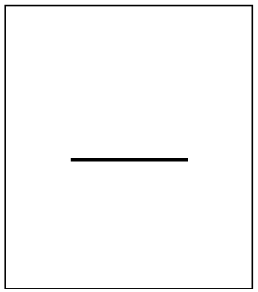


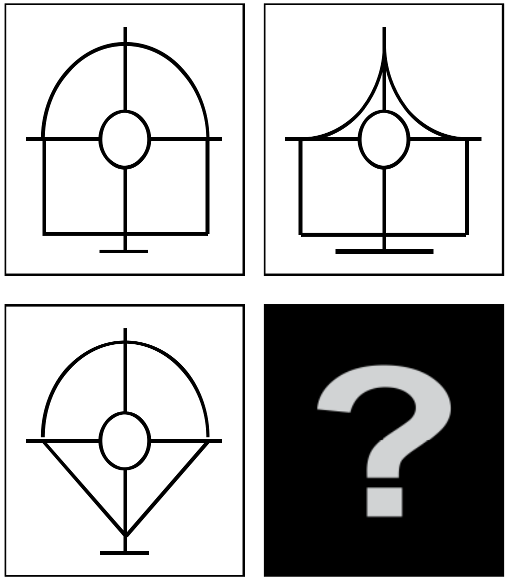


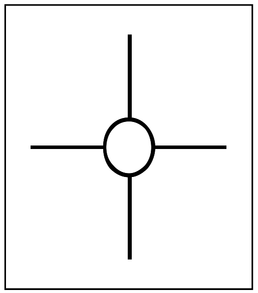


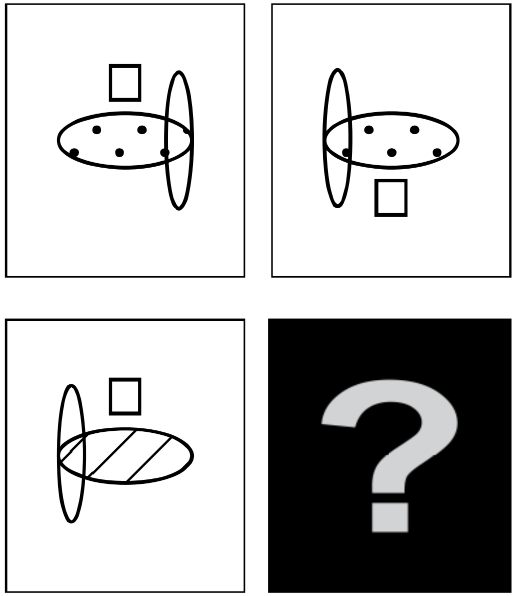
3C

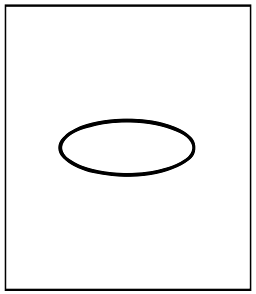


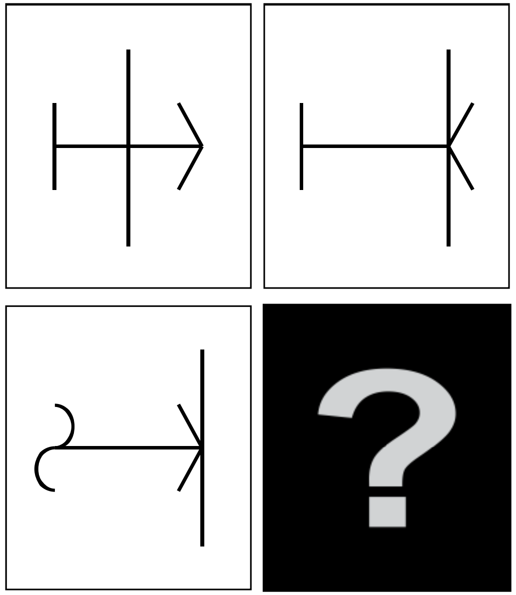


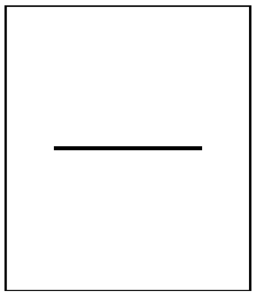


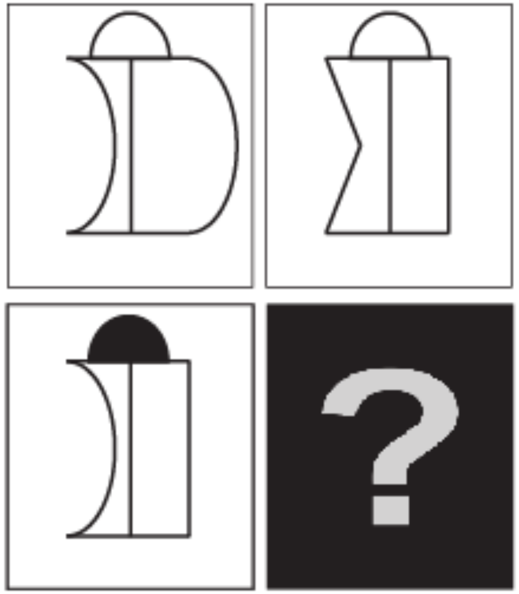


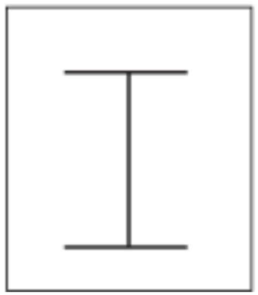


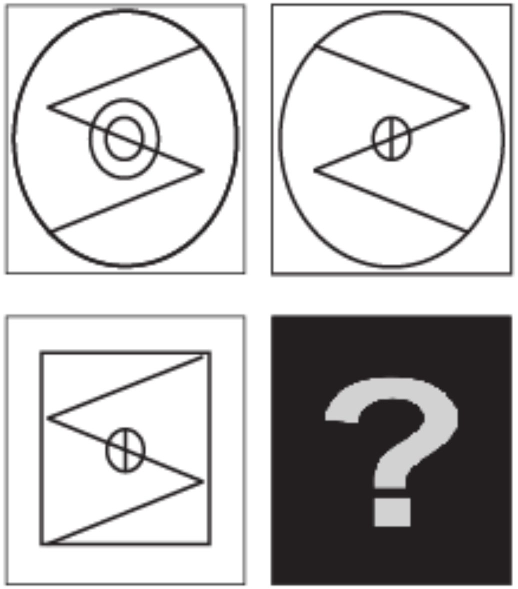


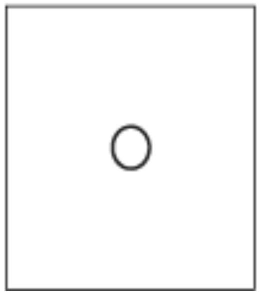


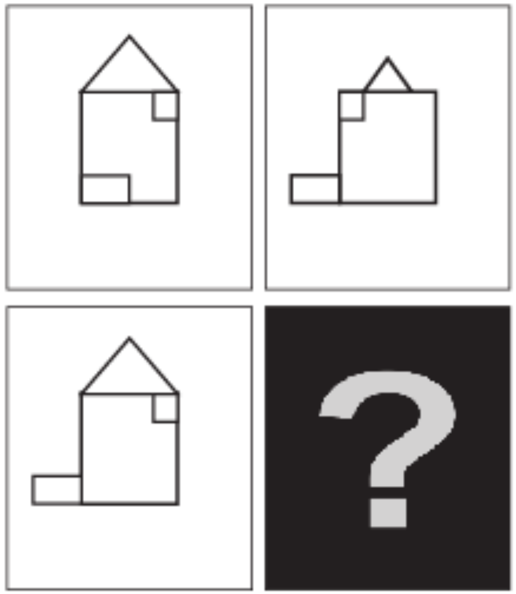


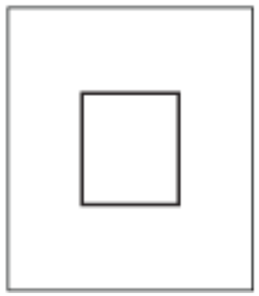


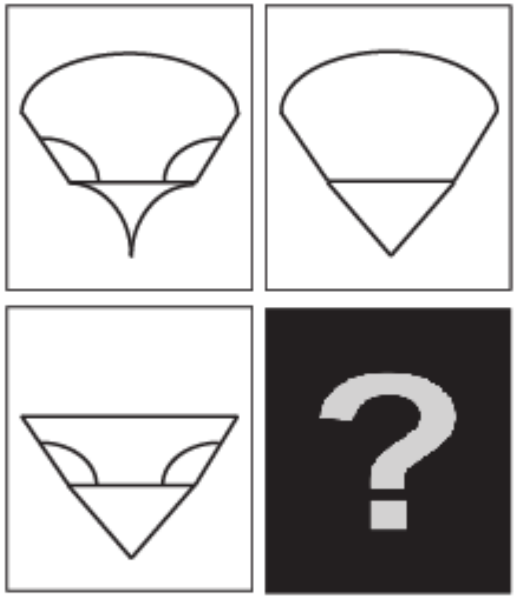


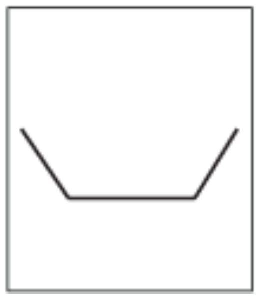


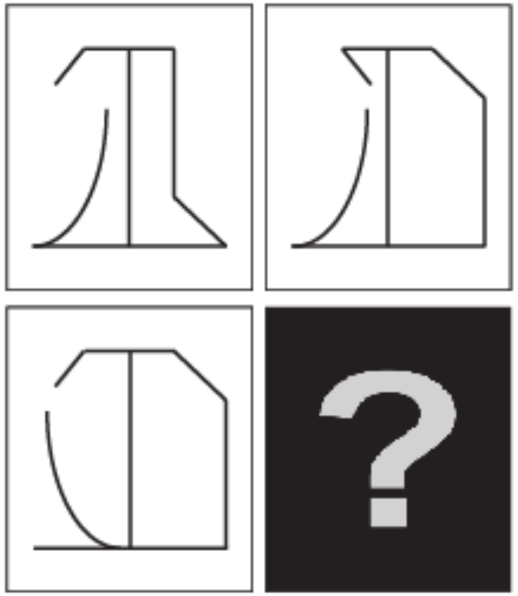


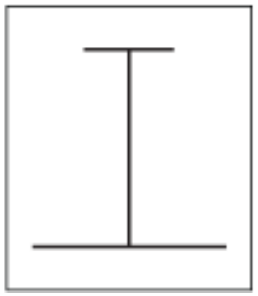


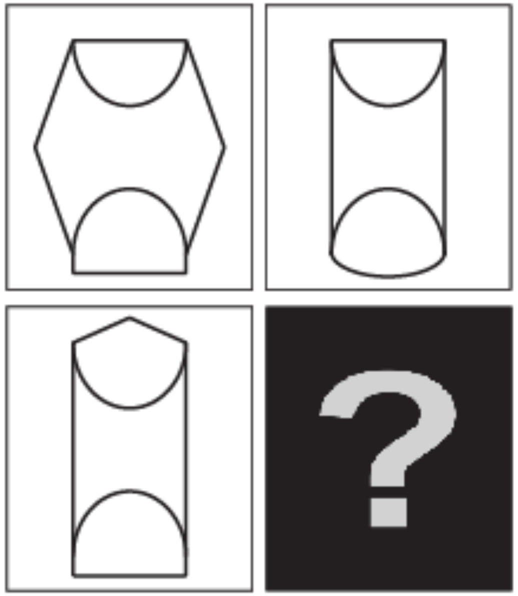


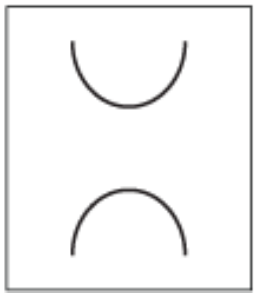












APPENDIX E TASKSET INSTRUCTIONS: A1 AND B1

SET A1/ B1

Trials

1-feature items: 2 practice, 5 items

Combined: 2 practice, 15 items (5x2-feat, 10x3-feat)

Segmented: 2 practice, 15 items (5x2-feat, 10x3-feat)

1 Feature Practice Instructions

In this task you will be solving some puzzles like this (POINT). To work out the answer, imagine what should go in the box with the question mark (POINT), then draw it in here (POINT).

Practice 1:

- Let me explain how the puzzle works.
- Here you see some boxes (POINT).
- When we look across the top row, notice what has happened. The same must happen in the bottom boxes.
- In this box (POINT to top left-hand box) there is a circle.
- In this box (POINT to top right-hand box) there is a circle too.
- In this box (POINT bottom left-hand box) there is a triangle.
- In this box (POINT to bottom right-hand box with question mark) something is missing. What shape goes in this box (POINT to question mark box) to complete the pattern?

Feedback:

IF THE CHILD KNOWS LET THEM DRAW THE SHAPE IN THE BOX BELOW.

- Use this box below (POINT to box below) to draw your answer, putting it in the correct place in the box.

IF THE CHILD DOESN'T KNOW THE ANSWER, EXPLAIN FURTHER.

- So here (POINT to the question mark box) you would put a triangle – then we would have two circles on top and two triangles on the bottom. Draw the triangle in the bottom box, putting it in the correct place in the box (POINT TO ANSWER BOX).

- You don't need to draw carefully – just well enough for us to know what you mean.
- Now we are going to try another one.

Practice 2:

- In this box (POINT to top left-hand box) there is a circle.
- In this box (POINT to top right-hand box) there is a circle too, except it has dotted line.
- In this box (POINT bottom left-hand box) there is a straight line.
- In this box (POINT to bottom right-hand box with question mark) something is missing. What shape goes in this box (POINT to question mark box) to complete the pattern?

Feedback:

IF THE CHILD KNOWS LET THEM DRAW THE SHAPE IN THE BOX BELOW.

- Use this box below (POINT to box below) to draw your answer.

IF THE CHILD DOESN'T KNOW THE ANSWER, EXPLAIN FURTHER.

- So here (POINT to the question mark box) you would put a dotted line – then we would have a solid circle changing to a dotted circle and a solid line changing to a dotted line. Draw the dotted line in the bottom box (POINT TO ANSWER BOX).

Start Task items:

- Well done. Now that you know what you have to do, we are going to do some more. Try your best for each one. Remember you don't need to draw carefully, just try to draw the shapes as well as you can. Even if you are not sure, it is good to draw something – you never know, you might be right! If you make a mistake and want to start again, just draw another box next to this one (POINT TO ANSWER BOX).
- Allow the child to do the 5 1-feature items.
- Well done, you did so well!!

2-Feature Instructions – Combined

Practice 1:

- In this next set of puzzles, you will see more shapes.
- This time there are two parts to the shapes that you must draw.

- In the answer box we have already drawn part of the answer to help you
- First look at the little shapes on the left. In this box (POINT to top left-hand box) there is a triangle on the left.
- In this box (POINT to top right-hand box) there is a triangle on the left too.
- In this box (POINT bottom left-hand box), there is a square on the left, so what would you expect here (POINT to question mark box)? Yes, a square on the left. So draw this in the box below.
- Now look at the lines. In this box (POINT to top left-hand box) there is a solid line but here (POINT to top right-hand box) the line has changed – can you see the change?

- In this box (POINT bottom left-hand box) we have a solid line.
- What would you expect here (POINT to question mark box) to complete the pattern?
- Remember, use the part that has already been drawn to help you.

<u>Feedback:</u>

IF THE CHILD KNOWS LET THEM DRAW THE SHAPE IN THE BOX BELOW.

- Use this box below (POINT to box below) to draw your answer.

IF THE CHILD DOESN'T KNOW THE ANSWER, EXPLAIN FURTHER.

- When we look across the top row, what has happened? The same must happen in the bottom boxes (POINT to bottom boxes).
- First look at the little shape on the left. In the top row, the little shape does not change. So in the bottom row the little shape should not change, and so we would draw another square in here (LET CHILD DRAW A SQUARE IN THE ANSWER BOX).
- Now look at the horizontal line. In the top row the line changes from a solid line to a dotted line. In the bottom row there is a solid line here, so what do you think should go here? (LET CHILD DRAW DOTTED/DASHED LINE IN ANSWER BOX). Do you understand? (IF NOT EXPLAIN TO THE CHILD AGAIN).

<u>Practice 2:</u>

- Now we are going to try another practice.
- Again, there will be two parts to the answer that you must draw.
- In the answer box we have already drawn some of the lines to help you.
- First, look at the small black shape. In this box (POINT to top left-hand box) there is a small black triangle at the top (POINT TO SMALL TRIANGLE).
- In this box (POINT to top right-hand box) the small black triangle is at the bottom– can you see?
- In this box we have a small black circle at the top, so what would you expect in here (POINT to question mark box) to complete the pattern?
- Now look at the big triangle. In the top row (POINT to top row) the big triangle has all solid lines in both boxes.
- We must use the information from in the pattern to help us complete it (POINT to question mark box).
- So, in this box there is a dotted line connecting the big triangle together (POINT bottom left-hand box).

- What would you expect here (POINT to question mark box) to complete the pattern?
- Remember, use the shape in the answer box to draw your answer.

Feedback:

IF THE CHILD KNOWS LET THEM DRAW THE SHAPE IN THE BOX BELOW.

- Use this box below (POINT to box below) to draw your answer.

IF THE CHILD DOESN'T KNOW THE ANSWER, EXPLAIN FURTHER.

- First look at the small black shapes. Here (POINT top left-hand box) we have a small black triangle at the top. In this box (POINT to top right-hand box) the small black triangle is at the bottom. If we look at the bottom box, we have a small black circle at the top. Thinking about what happened in the top row, where would you draw the circle in this box? LET CHILD DRAW A CIRCLE IN THE ANSWER BOX) Now look at the big triangle. In this box the triangle has solid lines. In this box the triangle stays the same - it also has solid lines. At the bottom, the triangle has a dotted line. So what type of line should connect the triangle together here? (LET CHILD DRAW DOTTED/DASHED LINE IN ANSWER BOX). Do you understand? (IF NOT EXPLAIN TO THE CHILD AGAIN).

Start of Task items:

- Well done. Now that you know what you have to do, we are going to do some more. Try your best for each one. Remember you don't need to draw carefully – just well enough for us to know what you mean. Even if you are not sure, it is good to draw something – you never know, you might be right! If you make a mistake and want to start again, just draw another box next to this one (POINT TO ANSWER BOX). The puzzles will get a little harder – sometimes you might even see three shapes!!! Part of the answer will be always be drawn in this box (POINT TO ANSWER BOX) to help you out ☺
- Are you ready?

2-Feature Instructions – Segmented

Practice 1

- Great, you are doing so well!
- This time, the task will be similar but with one change. Instead of just one pattern, you will see two patterns (POINT). Each pattern shows you one part of the answer you should draw in the box below. Just like before, there will already be part of the answer drawn in to help you.
- First pay attention to the pattern at the left (POINT).
- You see these boxes (POINT to top row) have a line. The line on the top row changes from a solid line to a dotted line. What do you think happens in the bottom row?
- If child answers correctly: Draw this into the answer box (ALLOW TO DRAW IN THE ANSWER BOX)

- Now look to the pattern on the right.
- This box has a triangle and the other box on the top-right hand side also has a triangle. Can you see? On the bottom row we have a square.
- So, what shape do you think we should draw in the answer box to complete the pattern? (POINT to the answer box)
- Draw the answer in this box (ALLOW TO DRAW IN THE ANSWER BOX)

Practice 2:

- Now we are going to try another practice. You will see two patterns again (POINT). Just like before, there is already part of the answer drawn in to help you.
- First pay attention to the pattern at the left (POINT).
- You see these boxes (POINT to top row) have a triangle with all solid lines. The triangle on the bottom row has a dotted line. What do you think happens in the bottom row?
- If child answers correctly: Draw this into the answer box (ALLOW TO DRAW IN THE ANSWER BOX)
- Now look to the pattern on the right.
- This box has a small black triangle at the top and the other box (POINT) has the small black triangle at the bottom. Can you see? On the bottom row we have a small black circle at the top.
- So, what do you think we should draw in the answer box to complete the pattern? (POINT to the answer box)
- Draw the answer in this box (ALLOW TO DRAW IN THE ANSWER BOX)

Feedback:

- Well done. Now that you know what you have to do, we are going to do some more. Try your best for each one. Remember you don't need to draw carefully – just well enough for us to know what you mean. Even if you are not sure, it is good to draw something – you never know, you might be right! If you make a mistake and want to start again, just draw another box next to this one (POINT TO ANSWER BOX).
- The puzzles will get a little harder – sometimes you might even see three shapes!!! Part of the answer will be always be drawn in this box (POINT TO ANSWER BOX) to help you out ☺
- Are you ready?

APPENDIX F TASKSET INSTRUCTIONS: A2 AND B2

SET A 2/B 2

Trials

1-feature items: 2 practice, 5 items

Combined: 2 practice, 15 items (5x2-feat, 10x3-feat)

Segmented: 2 practice, 15 items (5x2-feat, 10x3-feat)

1 Feature Practice Instructions

In this task you will be solving some puzzles like this (POINT). To work out the answer, imagine what should go in the box with the question mark (POINT), then draw it in here (POINT).

Practice 1:

- Let me explain how the puzzle works.
- Here you see some boxes (POINT).
- When we look across the top row, notice what has happened. The same must happen in the bottom boxes.
- In this box (POINT to top left-hand box) there is a circle.
- In this box (POINT to top right-hand box) there is a circle too.
- In this box (POINT bottom left-hand box) there is a triangle.
- In this box (POINT to bottom right-hand box with question mark) something is missing. What shape goes in this box (POINT to question mark box) to complete the pattern?

Feedback:

IF THE CHILD KNOWS LET THEM DRAW THE SHAPE IN THE BOX BELOW.

- Use this box below (POINT to box below) to draw your answer, putting it in the correct place in the box.

IF THE CHILD DOESN'T KNOW THE ANSWER, EXPLAIN FURTHER.

- So here (POINT to the question mark box) you would put a triangle – then we would have two circles on top and two triangles on the bottom. Draw the triangle in the bottom box, putting it in the correct place in the box (POINT TO ANSWER BOX).

- You don’t need to draw carefully – just well enough for us to know what you mean.
- Now we are going to try another one.

Practice 2:

- In this box (POINT to top left-hand box) there is a circle.
- In this box (POINT to top right-hand box) there is a circle too, except it has dotted line.
- In this box (POINT bottom left-hand box) there is a straight line.
- In this box (POINT to bottom right-hand box with question mark) something is missing. What shape goes in this box (POINT to question mark box) to complete the pattern?

Feedback:

IF THE CHILD KNOWS LET THEM DRAW THE SHAPE IN THE BOX BELOW.

- Use this box below (POINT to box below) to draw your answer.

IF THE CHILD DOESN’T KNOW THE ANSWER, EXPLAIN FURTHER.

- So here (POINT to the question mark box) you would put a dotted line – then we would have a solid circle changing to a dotted circle and a solid line changing to a dotted line. Draw the dotted line in the bottom box (POINT TO ANSWER BOX).

Start Task items:

- Well done. Now that you know what you have to do, we are going to do some more. Try your best for each one. Remember you don’t need to draw carefully, just try to draw the shapes as well as you can. Even if you are not sure, it is good to draw something – you never know, you might be right! If you make a mistake and want to start again, just draw another box next to this one (POINT TO ANSWER BOX).
- Allow the child to do the 5 1-feature items.
- Well done, you did so well!!

2-Feature Instructions – **Segmented**

Practice 1

- Great, you are doing so well!
- This time, the task will be similar but with one change. Instead of just one pattern, you will see two patterns (POINT). Each pattern shows you one part of the answer you should draw in the box below. Just like before, there will already be part of the answer drawn in to help you.
- First pay attention to the pattern at the left (POINT).
- You see these boxes (POINT to top row) have a line. The line on the top row changes from a solid line to a dotted line. What do you think happens in the bottom row?

- If child answers correctly: Draw this into the answer box (ALLOW TO DRAW IN THE ANSWER BOX)
- Now look to the pattern on the right.
- This box has a triangle and the other box on the top-right hand side also has a triangle. Can you see? On the bottom row we have a square.
- So, what shape do you think we should draw in the answer box to complete the pattern? (POINT to the answer box)
- Draw the answer in this box (ALLOW TO DRAW IN THE ANSWER BOX)

<u>Practice 2:</u>

- Now we are going to try another practice. You will see two patterns again (POINT). Just like before, there is already part of the answer drawn in to help you.
- First pay attention to the pattern at the left (POINT).
- You see these boxes (POINT to top row) have a triangle with all solid lines. The triangle on the bottom row has a a dotted line. What do you think happens in the bottom row?
- If child answers correctly: Draw this into the answer box (ALLOW TO DRAW IN THE ANSWER BOX)
- Now look to the pattern on the right.
- This box has a small black triangle at the top and the other box (POINT) has the small black triangle at the bottom. Can you see? On the bottom row we have a small black circle at the top.
- So, what do you think we should draw in the answer box to complete the pattern? (POINT to the answer box)
- Draw the answer in this box (ALLOW TO DRAW IN THE ANSWER BOX)

<u>Feedback:</u>

- Well done. Now that you know what you have to do, we are going to do some more. Try your best for each one. Remember you don't need to draw carefully – just well enough for us to know what you mean. Even if you are not sure, it is good to draw something – you never know, you might be right! If you make a mistake and want to start again, just draw another box next to this one (POINT TO ANSWER BOX).
- The puzzles will get a little harder – sometimes you might even see three shapes!!! Part of the answer will be always be drawn in this box (POINT TO ANSWER BOX) to help you out ☺
- Are you ready?

<u>2-Feature Instructions – **Combined**</u>

<u>Practice 1:</u>

- In this next set of puzzles, you will see more shapes.
- This time there are two parts to the shapes that you must draw.
- In the answer box we have already drawn part of the answer to help you

- First look at the little shapes on the left. In this box (POINT to top left-hand box) there is a triangle on the left.
- In this box (POINT to top right-hand box) there is a triangle on the left too.
- In this box (POINT bottom left-hand box), there is a square on the left, so what would you expect here (POINT to question mark box)? Yes, a square on the left. So draw this in the box below.
- Now look at the lines. In this box (POINT to top left-hand box) there is a solid line but here (POINT to top right-hand box) the line has changed – can you see the change?

- In this box (POINT bottom left-hand box) we have a solid line.
- What would you expect here (POINT to question mark box) to complete the pattern?
- Remember, use the part that has already been drawn to help you.

Feedback:

IF THE CHILD KNOWS LET THEM DRAW THE SHAPE IN THE BOX BELOW.

- Use this box below (POINT to box below) to draw your answer.

IF THE CHILD DOESN'T KNOW THE ANSWER, EXPLAIN FURTHER.

- When we look across the top row, what has happened? The same must happen in the bottom boxes (POINT to bottom boxes).
- First look at the little shape on the left. In the top row, the little shape does not change. So in the bottom row the little shape should not change, and so we would draw another square in here (LET CHILD DRAW A SQUARE IN THE ANSWER BOX).
- Now look at the horizontal line. In the top row the line changes from a solid line to a dotted line. In the bottom row there is a solid line here, so what do you think should go here? (LET CHILD DRAW DOTTED/DASHED LINE IN ANSWER BOX). Do you understand? (IF NOT EXPLAIN TO THE CHILD AGAIN).

Practice 2:

- Now we are going to try another practice.
- Again, there will be two parts to the answer that you must draw.
- In the answer box we have already drawn some of the lines to help you.
- First, look at the small black shape. In this box (POINT to top left-hand box) there is a small black triangle at the top (POINT TO SMALL TRIANGLE).
- In this box (POINT to top right-hand box) the small black triangle is at the bottom– can you see?
- In this box we have a small black circle at the top, so what would you expect in here (POINT to question mark box) to complete the pattern?
- Now look at the big triangle. In the top row (POINT to top row) the big triangle has all solid lines in both boxes.
- We must use the information from in the pattern to help us complete it (POINT to question mark box).
- So, in this box there is a dotted line connecting the big triangle together (POINT bottom left-hand box).
- What would you expect here (POINT to question mark box) to complete the pattern?

- Remember, use the shape in the answer box to draw your answer.

Feedback:

IF THE CHILD KNOWS LET THEM DRAW THE SHAPE IN THE BOX BELOW.

- Use this box below (POINT to box below) to draw your answer.

IF THE CHILD DOESN'T KNOW THE ANSWER, EXPLAIN FURTHER.

- First look at the small black shapes. Here (POINT top left-hand box) we have a small black triangle at the top. In this box (POINT to top right-hand box) the small black triangle is at the bottom. If we look at the bottom box, we have a small black circle at the top. Thinking about what happened in the top row, where would you draw the circle in this box? LET CHILD DRAW A CIRCLE IN THE ANSWER BOX) Now look at the big triangle. In this box the triangle has solid lines. In this box the triangle stays the same - it also has solid lines. At the bottom, the triangle has a dotted line. So what type of line should connect the triangle together here? (LET CHILD DRAW DOTTED/DASHED LINE IN ANSWER BOX). Do you understand? (IF NOT EXPLAIN TO THE CHILD AGAIN).

Start of Task items:

- Well done. Now that you know what you have to do, we are going to do some more. Try your best for each one. Remember you don't need to draw carefully – just well enough for us to know what you mean. Even if you are not sure, it is good to draw something – you never know, you might be right! If you make a mistake and want to start again, just draw another box next to this one (POINT TO ANSWER BOX). The puzzles will get a little harder – sometimes you might even see three shapes!!! Part of the answer will always be drawn in this box (POINT TO ANSWER BOX) to help you out ☺
- Are you ready?

APPENDIX G LETTER OF ETHICAL APPROVAL

Karen Douglas
Secretary

Dr J Holmes
MRC Cognition and Brain Sciences Unit
University of Cambridge

UNIVERSITY OF CAMBRIDGE

CAMBRIDGE
PSYCHOLOGY RESEARCH
ETHICS COMMITTEE

18 September 2018

Application No: PRE.2018.051

Dear Dr Holmes

Cognitive segmentation in children with low IQ

The Cambridge Psychology Research Ethics Committee has given ethical approval to your research project: 'Cognitive segmentation in children with low IQas set out in your application submitted on 30 April 2018.

The Committee attaches certain standard conditions to all ethical approvals. These are:

(a) that if the staff conducting the research should change, any new staff should read the application submitted to the Committee for ethical approval and this letter (and any subsequent letter concerning this application for ethical approval);
(b) that if the procedures used in the research project should change or the project itself should be changed, you should consider whether it is necessary to submit a further application for any modified or additional procedures to be approved;
(c) that if the employment or departmental affiliation of the staff should change, you should notify us of that fact.

Members of the Committee also ask that you inform them should you encounter any unexpected ethical issues.

Ethical approval will expire on 31 October 2021. If you require an extension, please submit an amendment request before the expiry date (guidance at www.bio.cam.ac.uk/psyres/amendments).

If you would let us know that you are able to accept these conditions, we will record that you have been given ethical approval.

Please note that there have been changes to the procedures regarding amendments. Full details are given on the REC website.

Yours sincerely

[Signature redacted]

K S Douglas

Cc Sinéad O'Brien

APPENDIX H STUDY PREREGISTRATION

ASPREDICTED

Wharton CREDIBILITY LAB
UNIVERSITY of PENNSYLVANIA

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Cognitive Segmentation in Children with low IQ (#13338)

Created: 08/14/2018 04:51 AM (PT)

This is an anonymized copy (without author names) of the pre-registration. It was created by the author(s) to use during peer-review.
A non-anonymized version (containing author names) should be made available by the authors when the work it supports is made public.

---

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Main hypothesis: Children with low IQ fail to use cognitive segmentation in the same way adults do on tests of fluid intelligence. Two additional questions will be asked: 1) At what age does the ability to cognitively segment develop? 2) What is the relationship between cognitive segmentation and other higher order cognitive processes?

3) Describe the key dependent variable(s) specifying how they will be measured.
The dependent variables will be the proportion of correct responses on a both a standard and modified matrix reasoning task.

4) How many and which conditions will participants be assigned to?
Sixty children from three different age groups will be recruited (N=180): 5-6 yrs, 7-8yrs, and 9-10yrs. Note that the same sample of children will be used to address all aims / questions. Question 1: A 2x2 design. Two within-subject conditions are task: Combined; and Separated modified matrix reasoning task. The Combined condition is similar to a traditional matrix reasoning task, except children are asked to draw the answer in a box provided, instead of choosing from an array of possible answers. In the Separated condition each matrix reasoning item is separated into its component parts. Children solve each of the parts and then draw the answer in the box provided. The between subjects' factor will be Group (high or low IQ), the number of participants in each of these groups will depend on recruitment. Question 2: The two within subject conditions from Question 1 above (task) will be included and performance will be correlated with age and / or compared across age groups. Question 3: The two within subject conditions from Question 1 above (task) will be included and performance will be correlated with measures of working memory, attention and other fluid reasoning measures. General linear modelling will be used to explore whether these abilities predict the ability to cognitively segment (proportion of correct responses on the two matrix reasoning tasks) and performance on fluid intelligence tests.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Do children with low IQ fail to segment in the same way as adults? General linear models will be used to predict proportion of correct responses from Condition (combined and separated) and IQ. When does the ability to cognitively segment develop? Correlation analysis will be used to check for the association between cognitive segmentation and age. Regression analysis will then be used to assess whether or not age predicts cognitive segmentation abilities. What is the relationship between cognitive segmentation and other cognitive processes? Correlation analysis will be used to check for the association between cognitive segmentation and higher order cognitive processes. Regression analysis will then be used to determine whether or not cognitive segmentation predicts higher order cognitive processes. Factor Analysis will be used to reduce to number of variables and determine the structure of the relationships between different cognitive processes and fluid intelligence. Latent variable analysis may be used to explore whether or not cognitive segmentation mediates relationships between higher order cognitive processes and fluid intelligence. Cluster analytic methods may be subsequently performed on the data to examine: 1) the presence of sub-groups within the data based on specific cognitive deficits or strengths; (2) whether sub-groups are different in terms of cognitive segmentation skills.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
A sample of typically developing children is required for this study therefore, participants with known neurological or genetic conditions will be excluded. Univariate outliers with scores more the 3.5 SD from the sample mean on each task will be identified and removed. Multivariate outliers will be identified using Mahalanobis D and removed.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
Sixty children from three different age groups will be recruited (N=180): 5-6 yrs, 7-8yrs, and 9-10yrs. Recruitment will stop at 180 participants but ultimately sample size will depend on access to participants via schools

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Version of AsPredicted Questions: 2.00

Available at https://aspredicted.org/BD3_3KW

Appendix I Published Manuscript: STXBP1-associaited neurodevelopmental disorder: a comparative study of behavioural characteristics

O'Brien *et al. Journal of Neurodevelopmental Disorders* (2019) 11:17
https://doi.org/10.1186/s11689-019-9278-9

Journal of Neurodevelopmental Disorders

RESEARCH Open Access

STXBP1-associated neurodevelopmental disorder: a comparative study of behavioural characteristics

Sinéad O'Brien[1], Elise Ng-Cordell[1,2], The DDD Study[3], Duncan E. Astle[1], Gaia Scerif[2] and Kate Baker[1,4*]

Abstract

Background: De novo loss of function mutations in *STXBP1* are a relatively common cause of epilepsy and intellectual disability (ID). However, little is known about the types and severities of behavioural features associated with this genetic diagnosis.

Methods: To address this, we collected systematic phenotyping data encompassing neurological, developmental, and behavioural characteristics. Participants were 14 individuals with *STXBP1*-associated neurodevelopmental disorder, ascertained from clinical genetics and neurology services UK-wide. Data was collected via standardised questionnaires administered to parents at home, supplemented by researcher observations. To isolate discriminating phenotypes, the *STXBP1* group was compared to 33 individuals with pathogenic mutations in other ID-associated genes (ID group). To account for the potential impact of global cognitive impairment, a secondary comparison was made to an ability-matched subset of the ID group (low-ability ID group).

Results: The *STXBP1* group demonstrated impairments across all assessed domains. In comparison to the ID group, the *STXBP1* group had more severe global adaptive impairments, fine motor difficulties, and hyperactivity. In comparison to the low-ability ID group, severity of receptive language and social impairments discriminated the *STXBP1* group. A striking feature of the *STXBP1* group, with reference to both comparison groups, was preservation of social motivation.

Conclusions: De novo mutations in *STXBP1* are associated with complex and variable neurodevelopmental impairments. Consistent features, which discriminate this disorder from other monogenic causes of ID, are severe language impairment and difficulties managing social interactions, despite strong social motivation. Future work could explore the physiological mechanisms linking motor, speech, and social development in this disorder. Understanding the developmental emergence of behavioural characteristics can help to focus clinical assessment and management after genetic diagnosis, with the long-term aim of improving outcomes for patients and families.

Keywords: STXBP1, Epilepsy, Intellectual disability, Language, Social

Background

It is increasingly possible to identify the genetic cause of severe neurodevelopmental disorders, with potential diagnostic yields now exceeding 50% [1]. For each rare genetic cause, there is an ongoing imperative to characterise associated medical and behavioural features in order to provide evidence-based prognostic advice, targeted family support, and ultimately personalised interventions.

Syntaxin-binding protein 1 (STXBP1, formally known as MUNC18-1) is part of the SEC1 family of membrane-trafficking proteins that interact with SNARE proteins to facilitate the docking of synaptic vesicles at presynaptic active zones to enable neurotransmission [2]. STXBP1 is expressed ubiquitously throughout the brain and neocortex during development and postnatal life. De novo *STXBP1* variants were initially identified in five patients as a cause of Ohtahara syndrome, also known as early infantile epileptic encephalopathy [3]. Thereafter, the epilepsy

* Correspondence: Kate.Baker@mrc-cbu.cam.ac.uk
[1]MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, UK
[4]Department of Medical Genetics, University of Cambridge, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0XY, UK
Full list of author information is available at the end of the article

spectrum associated with *STXBP1* variants expanded to include infantile spasms, West syndrome, Dravet syndrome, Lennox-Gastaut syndrome, and various other types of childhood-onset epilepsy [4–8]. The estimated prevalence of de novo *STXBP1* variants in severe childhood epilepsy is 3% [9]. In parallel, *STXBP1* variants have been discovered in individuals broadly ascertained for neurodevelopmental disorders, with and without epilepsy [7, 10–14]. The prevalence of *STXBP1* variants amongst individuals with unspecified developmental disorders is 0.25–0.5% [15, 16]. Given the 1% prevalence of intellectual disability (ID), the numbers of individuals diagnosed with *STXBP1*-associated neurodevelopmental disorder will rise substantially over the coming years.

Collation of clinical information has highlighted an association between *STXBP1* variants and complex long-term neurodevelopmental impairments, in patients with and without epilepsy. Existing case series suggest that there is considerable variability in the types and severities of neurodevelopmental problems associated with *STXBP1* variants. Reported problems include movement impairments such as ataxia and tremor, language delay and behavioural symptoms including hyperactivity, stereotypies, and autistic features [7, 9, 13, 17–20]. However, despite the increasing number of patients diagnosed with *STXBP1* variants, no study has carried out post-diagnostic, standardised assessments to systematically characterise developmental and behavioural aspects of the *STXBP1*-associated phenotype. These data are necessary to provide families with valid information about the types and severities of neurodevelopmental difficulties that can be expected for an individual diagnosed with an *STXBP1* variant. Hence, the primary aim of the current study was to describe in detail the neurodevelopmental phenotype of children and adolescents with *STXBP1* variants, ascertained as broadly as possible.

The secondary aim of the study was to relate neurodevelopmental characteristics associated with *STXBP1* variants to existing basic science literature, including animal models. STXBP1 plays a role in calcium-dependent, short-term synaptic facilitation, which may be particularly relevant to learning [21]. *STXBP1* loss of function mutations lead to broad impairments in synaptic physiology that nonetheless result in specific learning deficits and behavioural features. In cultured neurons from human embryonic stem cells, homozygous mutations impede synaptic transmission which results in neurodegeneration, whereas heterozygous mutations have a milder impact [21]. Homozygous knockout mice have shown that, despite normal brain assembly, lack of *STXBP1* expression inhibits pre-synaptic events which leads to synaptic degeneration, neuronal apoptosis, and early fatality [22–24]. In contrast, heterozygous *STXBP1* mouse models have reduced protein expression and stability, but structurally normal synaptogenesis. Mice demonstrate a seizure phenotype characterised by twitches, jerks, and jumps which often coincided with EEG spike-wave discharges [25]. Currently, there is no systematic, comparative data available on the human behavioural phenotype to verify whether behavioural results in *STXBP1* mouse models [26–28] recapitulate the human disorder. To achieve this, we have compared individuals with *STXBP1* variants to participants with equivalently severe ID, to determine which if any characteristics are consistently observed and are specific to this group. This comparison can add specificity to the prognostic information available to families and clinicians and set hypotheses for future investigation of molecular, neural, and cognitive mechanisms.

Methods

Participants and recruitment

Individuals who had previously been diagnosed with a pathogenic de novo variant in *STXBP1* were identified via clinical genetics services, paediatric neurology services, and clinical testing laboratories. Responsible clinicians were provided with a recruitment pack for each potential participant, and 14 families volunteered to participate in the study. Eight participants had been diagnosed with an *STXBP1* variant via the Deciphering Developmental Disorders research study [16], and the remainder had been diagnosed via a clinical testing pathway. For both testing pathways, pathogenicity of variants had been evaluated by the regional genetics service according to ACMG guidelines and local confirmation procedures. The same recruitment procedure was followed for the comparison sample of individuals with neurodevelopmental disorder associated with single nucleotide variants in other ID-associated genes. Participants were recruited with the assistance of 10 regional genetics centres.

Phenotyping assessments

All 14 participants were assessed in their homes. A standardised, study-specific, structured medical history interview was conducted, followed by the Vineland Adaptive Behaviour Scales, survey interview form [26]. Parents or carers were also given age-appropriate questionnaires to complete which included the Developmental Behavioural Checklist (DBC) [27], Conners-3 short-form [28], Social Responsiveness Scale (SRS-2) [29], and the Movement ABC Parent Checklist [30]. The Movement Assessment Battery for Children (MABC-2) was used to directly assess participants' manual dexterity, aiming and catching, and balance [30]. Detailed notes and video recordings were obtained in order to document qualitative observations of each participant.

Data completion rates

For descriptive reporting, data from all *STXBP1* participants are included ($n = 14$) (Additional file 1). For comparative analyses, analysis was restricted to a consistent *STXBP1* group for whom a complete questionnaire dataset was available ('*STXBP1* full dataset group', $n = 8$). Partial data collection was achieved for younger participants (protocol only appropriate for ages 4 and above) and for three participants without a fully completed parent questionnaire pack.

Analysis

Raw scores from questionnaire data were standardised to published normative data. Non-parametric comparisons were conducted throughout (Mann-Whitney U statistics).

Results

Demographics of *STXBP1* and comparison groups (Table 1)

The *STXBP1* group comprised 14 individuals with a confirmed pathogenic, de novo variant in *STXBP1*. The ID comparison group comprised 33 individuals with confirmed pathogenic or likely pathogenic de novo or inherited variants in other ID-associated genes (Additional file 2). The *STXBP1* full dataset and ID groups did not differ in age, gender or socioeconomic status (SES) distributions. On average, the *STXBP1* group had more severe global adaptive impairments than the ID comparison group (Mann-Whitney $U = 43.5$, $p = 0.004$), with more severe deficits in all adaptive behaviour subscales except for gross motor function (Table 2). To identify specific behavioural characteristics not explained by these global impairments, secondary analyses were carried out between the *STXBP1* full dataset group and a low-ability subset of the ID comparison group, selected to match the range of Vineland Adaptive Behaviour Composite scores. The *STXBP1* full dataset and low-ability ID comparison groups did not differ in age, gender, or SES representation.

STXBP1 group—medical, neurological, and neurodevelopmental characteristics

General health

Duration of pregnancy ranged from 32 to 42 weeks. No abnormal ultrasounds or threatened miscarriages were reported. However, five mothers reported persistent vomiting throughout pregnancy. There were concerns about foetal well-being during the birth of six of the participants. Four participants were born by Caesarean section. Only three cases subsequently spent time in SCBU (with none requiring neonatal intensive care). Three individuals were readmitted to hospital within the first month of life for investigation of suspected seizures.

A large proportion of parents (85%) reported feeding issues during infancy and early childhood, common problems being reflux, choking, and difficulty moving solids. Sleep difficulties were reported in four cases during infancy, characterised by problems falling and staying asleep. In two cases, sleep difficulties persisted into adolescence.

Parents also reported a range of sensory issues with regard to hearing, sight, and touch. Six parents reported sight issues including; ptosis, misshapen eye, difficulty focusing, excessive blinking, and long-sightedness. Three parents reported hearing issues: one participant suffered from recurrent ear infections, one had glue ear in both ears, and one participant required hearing aids. A diverse range of sensory issues involving touch and sound were reported in eight cases (3: touch, 3: sound, 2: touch and sound). Three participants enjoyed sensory feedback from sensory objects and people (e.g., hugging and rough play), whereas two did not. One child appeared to enjoy high-pitched sounds such as babies crying, whereas four parents noticed that their children were averse to certain loud sounds such as hoovers, fireworks, and weather. One of these children had received a diagnosis of hyperacusis.

Table 1 Demographic information

Demographic measure		Group			
		STXBP1 all (n = 14)	*STXBP1* full dataset (n = 8)	ID all (n = 33)	ID low ability (n = 8)
Age	Mean (SD)	9.93 (5.837)	13.63 (2.326)	13.91 (5.553)	14.88 (3.271)
	Range (years)	1–17	10–17	5–25	8–18
Gender	% females(n)	71.4 (10)	62.5 (5)	51.5 (17)	62.5 (5)
SES**	Median (SD, range)	6.5 (2.51, 2–10)	7 (1.96, 4–10)	6 (2.33, 1–10)	6.5 (1.98, 3–9)
Epilepsy (ever)	% (n)	92.9 (13)	87.5 (7)	39.4 (13)	50 (4)
Epilepsy (current)	% (n)	92.3 (12)	85 (6)	62 (8)	75 (3)
Epilepsy age of onset	Mean (SD)	49.67 (67.059)	84.14 (70.115)	70.73 (76.352)	12.67 (9.866)
	Range (months)	1–156	1–156	1–204	6–24
Global intellectual ability	Mean (SD)	45.86 (14.992)	41.0 (12.13)	58.3 (15.661)*	48.0 (9.592)
	Range	26–69	31–67	20–96	36–68

**SES calculated using 'English Indices of Deprivation' (2015)
*Comparison to *STXBP1* full dataset group, Mann-Whitney U test, $P < 0.05$

Table 2 Adaptive function

Vineland (standard score) Domains Subdomains		Groups			
		STXBP1 all (n = 14)	*STXBP1* full dataset (n = 8)	ID all (n = 33)	ID low ability (n = 8)
Communication	Mean (SD)	42.64 (11.5)	43.38 (12.1)	59.15 (18.6)*	47.50 (7.6)
	Range	30–67	33–67	21–100	38–61
Receptive	Mean (SD)	4.82 (2.7)	4.75 (2.6)	8.70 (3.1)*	7.13 (1.5)*
	Range	2–10	2–10	1–16	5–10
Expressive	Mean (SD)	3.55 (2.7)	3.38 (2.8)	6.88 (4.0)*	3.00 (1.9)
	Range	1–8	1–8	1–17	1–6
Written	Mean (SD)	5.55 (1.8)	5.38 (1.8)	7.97 (3.4)*	5.88 (1.0)
	Range	4–9	4–9	3–20	5–8
Daily living skills	Mean (SD)	45.57 (19.7)	38.38 (15.5)	57.36 (18.2)*	43.00 (13.3)
	Range	25–77	25–73	21–98	30–71
Personal	Mean (SD)	4.71 (4.5)	2.88 (3.8)	6.18 (4.3)*	2.75 (2.6)
	Range	1–12	1–12	1–19	1–8
Domestic	Mean (SD)	6.43 (4.3)	4.38 (3.3)	8.73 (3.8)*	5.63 (3.3)
	Range	2–13	2–12	3–17	3–13
Community	Mean (SD)	4.86 (3.9)	3.00 (1.9)	6.18 (3.6)*	3.88 (2.6)
	Range	1–12	1–7	1–13	1–9
Socialisation	Mean (SD)	52.29 (12.1)	47.75 (8.4)	63.30 (16.5)*	56.13 (9.2) *
	Range	34–75	40–67	20–108	46–75
Interpersonal	Mean (SD)	6.00 (3.3)	4.63 (2.1)	7.94 (3.0)*	6.63 (1.8)
	Range	2–13	2–8	1–15	4–10
Play and leisure	Mean (SD)	4.93 (3.4)	3.63 (3.2)	7.76 (3.6)*	5.00 (3.7)
	Range	1–11	1–11	1–15	1–11
Coping skills	Mean (SD)	6.79 (1.8)	6.00 (1.3)	9.30 (3.1)*	8.00 (1.3)*
	Range	4–9	5–8	6–20	6–10
Motor	Mean (SD)	60.43 (16.4)	63.25 (17.4)	69.03 (12.4)	62.63 (5.6)
	Range	37–97	40–97	56–107	56–72
Gross	Mean (SD)	8.50 (2.8)	9.63 (3.0)	9.58 (1.8)	9.00 (1.5)
	Range	5–15	6–15	7–16	7–11
Fine	Mean (SD)	8.29 (4.1)	8.00 (3.1)	10.09 (3.0)*	8.50 (0.9)
	Range	4–17	4–14	7–20	7–10

*Comparison to *STXBP1* full dataset group, Mann-Whitney U test, $P < 0.05$

Neurological histories

Thirteen *STXBP1* participants have a history of seizures, diagnosed as epilepsy by a paediatric neurologist. The age of seizure onset varied from just after birth to 13 years of age. Half of the *STXBP1* participants ($n = 7$) had a seizure in their first year of life. Four of these participants had a seizure within the first 2 months of life. For the remaining six individuals with epilepsy, the age of onset of seizures was broadly spread across childhood. Three participants experienced their first seizure during the peripubertal phase. The type of seizure also varied. One participant was diagnosed with infant encephalopathy, two had tonic-clonic seizures, and six experienced focal seizures. Seizure phenotypes were mixed or uncertain for the remaining participants. Epilepsy has been treated with a wide range of anti-epileptic drugs (AEDs), the most frequently used was levetiracetam ($n = 4$). Whilst seizure frequency was generally reduced on treatment, no participant was reported to be completely seizure free during the months prior to research assessment. One participant experienced infrequent seizures (approximately three between 5 and 14 years of age) but was not on medication. Four participants also experienced tremors in their hands, which were observed by the researcher during assessments.

Developmental milestones

Thirteen participants learned to sit independently from 7 to 42 months. Seven out of thirteen sat independently before 10 months, which is within the normal to mildly delayed range, whilst the remainder were clearly delayed. The youngest child in the sample (1-year-old) had extremely low muscle tone and was not able to sit independently at the time of assessment. Four children did not learn to crawl but 'bum-shuffled' or 'bunny hopped'. Nine participants were walking independently at the time of assessment. However, eight acquired this skill late: between 25 and 60 months old. One participant, aged 14, could walk a few steps but required constant support. Three participants between 1.9 and 6.8 years of age had not yet learned to walk.

Speech and language acquisition were delayed in all cases. Eight participants currently use some verbal communication, with age of first words being between 18 and 156 months (13 years). The age of participants not currently using verbal communication is between 1.1 and 15 years. Of the four participants who can form simple phrases or sentences, three of these acquired this skill late, after 5.5 years of age.

Current motor abilities

Vineland gross and fine motor subscale scores reflect the extent of impairment and range of motor abilities within the sample. Gross motor standardised scores ranged between 6 and 15: from severely impaired to average for age (maximum score 24, normative population mean = 15, SD = 3). The participant with the highest score could run, jump, hop, skip, and kick, throw, and catch a ball, whereas, at age 6, the participant with the score was unable to pull themselves to a standing position. Similarly, fine motor standard scores ranged from 4 to 17. Participants with greater fine motor abilities could draw shapes freehand and build three-dimensional structures whereas three participants were unable to turn pages or stack blocks.

The MABC-2 Parent Checklist provided a parent rating of their child's motor abilities in predictable and unpredictable environments. In addition, a 13-item subscale asks parents to indicate any non-motor factors that may affect their child's movement. Over 85% of parents in the STXBP1 group reported that being overactive, distractible, impulsive, disorganised, and hesitant contributed to their child's motor skills difficulties. In addition, 62.5% of children became anxious in stressful, movement-related situations. All participants attempted MABC-2 motor tasks but only one participant was able to complete the full protocol enabling standardised scoring.

Current adaptive and communication abilities

The global adaptive function was estimated via the Vineland Adaptive Behaviour Scales (parent interview) composite score. Seven participants had scores in the severe ID range and seven participants were in the mild to moderate ID range. Notably, three participants under 2 years of age scored in the mild ID range, which may indicate increasing diversion from developmental expectations with age.

Vineland Communication standard scores have a mean of 100 and a standard deviation of 15. Participants scored between 33 and 67 (i.e. moderately to severely impaired range). The *STXBP1* participants' receptive abilities ranged from an inability to listen and attend to speech to being able to recall and carry out instructions given 5 min previously. Expressive ability ranged from difficulty with prespeech expression to engaging in interactive speech.

Current emotional and behavioural difficulties

The Developmental Behavioural Checklist (DBC) was completed by parents to characterise the current behavioural and emotional difficulties they observe in their children (participants over the age of 5). Seven out of eight *STXBP1* participants had a DBC Total Problem Behaviour Score above the clinical cut-off (*T* score 46) indicative of the likely presence of psychopathology. Behaviours such as biting, spitting, ripping things, hitting, and running away from caregivers were common, and parents reported demanding situations and frustration. Two parents remarked that these behaviours had become increasingly difficult to manage following the onset of puberty. High levels of anxiety associated with specific causes such as fireworks, adverse weather, hoovers, lawnmowers, and the emotions of others were reported in five cases (4 females, 1 male). Six parents also reported that their child rarely cried and did not appear to feel pain.

Parents completed the Conners Short Form to assess the characteristics of ADHD and comorbid problems in the sample. The *STXBP1* group showed very elevated scores on the inattention, hyperactivity, learning problems, and peer relations subscales. Two participants had a *T* score in the normal range for executive functions and five participants scored within the normal range on the aggression subscale.

Current social functioning

Vineland socialisation subscale scores indicated that the *STXBP1* group demonstrate social adaptive impairments. The Social Responsiveness Scale (SRS-2) was used to characterise atypical social processes linked to autism spectrum features in the sample. Three participants had an SRS total score within the mild to moderately impaired range and five scored above 76, which is indicative of severe social impairment likely to meet criteria for an autism spectrum diagnosis. Participants' scores fell within the moderate to severe range on social awareness, social cognition, social communication and restricted interests and repetitive behaviours subscales. In contrast, scores on the social motivation subscale were notably preserved and within the normal range. Parents typically described a social profile characterised by self-confidence in social settings and an interest in engaging with peers and adults. The study researchers also observed this strong interest in social interaction, despite limited communication and social skills as assessed by Vineland scales.

Comparison of behavioural features in the *STXBP1* full dataset and ID groups

Motor function

There was no difference in gross motor skills between groups as assessed via Vineland Adaptive Behaviour Scales (Table 2) and Movement ABC (Table 3). However, the *STBXP1* group had poorer Vineland fine motor subscale scores when compared to the ID group ($U = 74.0$, $p = 0.053$). There was no difference between groups in how parents evaluated the contextual factors influencing children's motor performance.

Emotional and behavioural difficulties

The *STXBP1* and ID groups did not differ on DBC total problem behaviour scores or subscale scores (stratified for ID severity), indicating that emotional and behavioural difficulties in general are not more prevalent or severe in the *STXBP1* group than expected for intellectual disability due to other causes (Table 3). Anxiety and disruptive behaviours were prominent problems for some but not all individuals within the *STXBP1* group; however, group mean DBC anxiety subscale scores did not differ from the ID comparison group. On the Conners scales, the *STXBP1* group were rated as significantly more hyperactive/impulsive compared to the ID group ($U = 52.0$, $p = 0.026$). In summary, the *STXBP1* group demonstrated similar total problem behaviour scores to the ID comparison group, with variation within the group in the severity and types of problems reported for each individual. High hyperactivity/impulsivity scores were more consistent within the *STXBP1* group and on average were more severe than reported for individuals with other monogenic causes of ID.

Social functioning

There was a significant difference between *STXBP1* and the ID groups on Vineland Socialisation scores: the ID group on average display stronger skills on all three subscales—interpersonal relations, play and leisure, and coping skills. The groups did not differ on SRS total score or social awareness, social cognition, social communication and restricted interests and repetitive behaviours (RRB) subscales. However, the ID comparison group were more impaired on social motivation, compared to the *STXBP1* group who fell within the 'normal' range for this subscale ($U = 74.0$, $p = 0.06$). In summary, the *STXBP1* group demonstrated an atypical profile of social functioning in comparison to ID in general, with more severely impaired everyday social behaviour (on Vineland scales), equivalently impaired social cognition (on SRS), but preserved social motivation.

Comparison of behavioural features in the *STXBP1* full dataset and low-ability ID groups

To explore the specificity of behavioural characteristics associated with *STXBP1* mutation in more detail, secondary comparisons were made to a subset of the ID comparison group selected to match the range of global adaptive abilities within the *STXBP1* group. Although the two groups had comparable global adaptive function scores, communication impairments were more severe in the *STXBP1* group. The comparison group on average demonstrated stronger receptive than expressive abilities whereas the *STXBP1* group showed severe restriction of both receptive and expressive abilities (with significant difference between groups in receptive score: $U = 11.0$, $p = 0.025$). Groups did not differ in gross or fine motor abilities. Behavioural problems assessed via the DBC (total, subscales) did not differ between groups. The *STXBP1* group also did not differ from the low-ability ID group in Conners hyperactivity or impulsivity scores, indicating that although these difficulties are a consistent problem area for the *STXBP1* group, they are not more impaired in these domains than other individuals with equivalent global impairments. Comparison of Vineland scores indicated that everyday social behaviours were more impaired in the *STXBP1* group in comparison to the low-ability ID group (socialisation domain: $U = 11.0$, $p = 0.027$, coping: $U = 9.0$, $p = 0.014$). The analysis of SRS scores (see Fig. 1) indicated a stable pattern of results whether comparing to the whole sample ID group or low-ability ID group—total SRS scores did not differ, social cognitive scores did not differ, restricted and repetitive behaviour scores did not differ, but social motivation was significantly preserved in the *STXBP1* group ($U = 9.5$, $p = 0.018$).

Discussion

In this study, we systematically assessed the behavioural characteristics of children and adolescents with variants in *STXBP1*, aiming to improve prognostication and management for individuals with this rare genetic diagnosis. Participants in the *STXBP1* group demonstrated complex and persistent difficulties across multiple domains of everyday function and emotional-behavioural development. Consistent problems within the group included severe communication impairments and hyperactivity-impulsivity. Some participants were greatly troubled by anxiety

Table 3 Emotional and behavioural characteristics

Outcome measure		Groups		
		STXBP1 full dataset (*n* = 8)	ID all (*n* = 33)	ID low ability (*n* = 8)
MABC parent checklist				
Motor competence (static environments)	Mean (SD)	24.5 (12.3)	18.45 (9.2)	23.00 (5.6)
	Range	3–42	1–37	11–31
Motor competence (dynamic environments)	Mean (SD)	26.00 (11.9)	22.12 (8.8)	25.5 (6.8)
	Range	3–43	3–39	14–34
Total motor score	Mean (SD)	50.5 (23.6)	40.61 (17.2)	48.5 (10.9)
	Range	6–85	6–74	35–62
Conners (*T* score)				
Inattention	Mean (SD)	84.75 (7.5)	79.19 (12.0)	86.00 (5.3)
	Range	72–92	55–90	79–90
Hyperactivity	Mean (SD)	85.25 (6.5)	72.37 (15.0)*	77.25 (13.2)
	Range	73–90	40–90	56–90
Learning problems	Mean (SD)	86.38 (8.3)	79.44 (10.9)	86.13 (5.0)
	Range	66–90	57–90	78–90
Executive functions	Mean (SD)	76 (14.8)	64.56 (14.4)	66.25 (13.3)
	Range	54–90	43–90	48–83
Aggression	Mean (SD)	59.88 (17.4)	52.3 (10.5)	47.5 (4.8)
	Range	44–90	44–83	44–58
Peer relations	Mean (SD)	84.13 (7.2)	80.33 (15.0)	81.38 (16.1)
	Range	72–90	45–90	45–90
DBC (percentile/stratified by ID severity)				
Disruptive	Mean (SD)	68.25 (24.3)	50.91 (28.8)	54.00 (25.2)
	Range	28–90	4–98	26–90
Self-absorbed	Mean (SD)	73.75 (19.2)	70.12 (25.8)	77.75 (17.9)
	Range	34–94	10–100	40–100
Communication	Mean (SD)	75.5 (17.8)	76.42 (23.1)	78.00 (18.8)
	Range	48–96	6–100	40–92
Anxiety	Mean (SD)	42.5 (24.5)	58.3 (26.9)	66.25 (30.1)
	Range	12–76	10–100	10–96
Social relations	Mean (SD)	49.5 (22.8)	58.85 (28.0)	57.00 (27.3)
	Range	24–96	12–100	24–98
Total problems	Mean (SD)	73.50 (24.1)	63.94 (27.4)	67.25 (22.3)
	Range	30–94	0–100	38–98
SRS (*T* score)				
Social awareness	Mean (SD)	78.75 (5.9)	73.64 (11.2)	76.38 (12.3)
	Range	70–87	49–93	56–92
Social cognition	Mean (SD)	73.88 (11.5)	72.7 (11.0)	74.38 (6.6)
	Range	53–92	47–96	67–85
Social communication	Mean (SD)	76.25 (9.2)	72.7 (11.7)	80.88 (11.1)
	Range	66–90	50–96	67–96
Social motivation	Mean (SD)	58.00 (13.6)	64.52 (11.8) *	69.5 (13.6)*
	Range	45–89	41–95	58–95

Table 3 Emotional and behavioural characteristics *(Continued)*

Outcome measure		Groups		
		STXBP1 full dataset ($n = 8$)	ID all ($n = 33$)	ID low ability ($n = 8$)
RRB	Mean (SD)	79.00 (11.9)	78.36 (15.2)	87.38 (14.7)
	Range	62–92	48–108	72–108
SRS total score	Mean (SD)	77.75 (9.3)	75.12 (12.0)	81.75 (10.9)
	Range	66–94	49–98	70–98

[a]Comparison to *STXBP1* full dataset group, Mann-Whitney *U* test, $P < 0.05$

including specific phobia, but these symptoms varied across the group. We observed an additional contrast in some individuals between low social anxiety and high non-social anxiety. In view of moderate to severe ID, it is important to consider whether behavioural characteristics within this group differ from developmental expectations—we found that receptive language ability and socialisation skills (but not hyperactivity) discriminated the *STXBP1* and ID comparison groups, suggesting a specific contribution of aetiology to these features.

We were struck by a consistent profile of social behaviour within the *STXBP1* group. Participants showed great enjoyment of social interactions with family members and demonstrated reciprocal behaviours such as sharing and turn-taking. Researcher experience was that, from the outset of the testing sessions, participants wanted to engage with the researchers (a novel stranger within the participants' home) particularly in social activities such as drawing, dancing, singing, and playing games. Even participants as young as 2 years held eye contact and were very sociable. On the other hand, parental questionnaire ratings indicated limited understanding of social expectations and difficulties integrating into normal social settings. The observation of significantly stronger social motivation remained true, whether comparing to the whole ID sample or low-ability ID sample, indicating that this feature is not simply a correlate of ID severity. Although several individuals scored above cut-off on the SRS for a likely autism spectrum diagnosis, this profile of social behaviours would not be typical

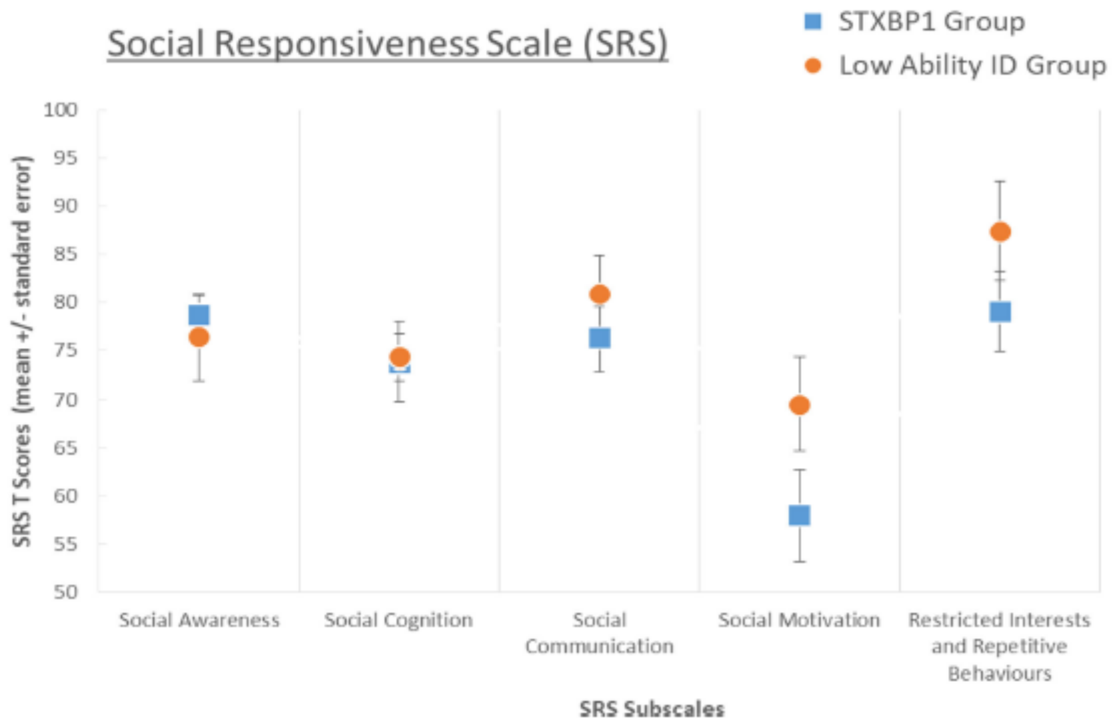


Fig. 1 Mean Social Responsiveness Scale subscale *T* scores in the STXBP1 and low-ability ID comparison groups

for an individual with autism. These findings support other research into the diverse social developmental trajectories contributing to autism-like impairments in individuals with genomic disorders [31–33].

Epilepsy prevalence within the *STXBP1* group was high, as previously reported, but the type, severity, age of onset, treatment sensitivity, and remission of seizures varied across participants. A relatively high prevalence of focal seizures was reported, in keeping with previous reports [7]. Further studies are needed to examine the contribution of subclinical seizure activity to developmental progress for individuals with *STXBP1* variants, including interactions between low-level sensory processing, cognitive biases, and the social environment to mediate anxieties or prosocial behaviour. In addition, comparison to other groups with similar epilepsy prevalence and seizure characteristics could tease apart aetiology-specific behavioural features from correlates of, for example, focal seizures.

A further goal of this research is to stimulate future investigation of the mechanisms linking presynaptic dysfunction to cognitive and social development, which may ultimately lead to aetiology-specific therapies. Toward this goal, we compared our observations of the *STXBP1* group to published behavioural phenotyping studies of mice with *STXBP1* mutations, to consider consistency across species. The behavioural phenotypes exhibited across different mice models include elevated anxiety, hyperactivity, and impairments in acquiring and maintaining spatial memory and reversal of previously learned strategies [25, 32, 34]. Importantly, not every aspect of learning and behaviour is impaired: *STXBP1* haploinsufficient mice demonstrate normal sociability, preference for social novelty, and normal profiles of attentional control without impulsivity [25]. In one study, social behaviours in *STXBP1* mouse models were examined by introducing a novel mouse into the chamber [27]. Regardless of specific *STXBP1* variant, heterozygous mice spent more time around the novel mouse and displayed normal sociability and preference for social novelty. These findings appear, at least on the surface, to be convergent with our observations of young people with *STXBP1* variants, and future research could explore whether similar neurodevelopmental mechanisms underpin social interactions across species.

In considering potential mechanisms underlying this and other aspects of the *STXBP1* phenotype, we have been struck by the phenotypic similarity with Angelman syndrome (AS). AS is characterised by developmental delay, seizure susceptibility including focal seizures, movement disorders, language deficits, impulsivity, short attention span, and specific phobias [33, 35–38]. The behavioural hallmarks of AS are hyperactivity and hypersociability [34]. Just as in *STXBP1*, individuals with AS display increased social motivation, prolonged social interest, and also excessive smiling and laughing [39]. Socialisation is believed to be underpinned by a network of brain regions including the amygdala, ventral striatum, orbital, and ventromedial regions of the prefrontal cortex. However, different regions and networks may have a greater role in specific aspects of sociability [40]. It has been hypothesised that reciprocal changes within striatal circuits give rise to the atypical social novelty profile associated with AS, consistent with experimental evidence for altered striatal dopamine balance in a mouse model [41, 42]. Examining the striatum and its dependent functional neural systems, in the context of the wider 'social' brain network, may provide a starting point for understanding the neural mechanisms driving the atypical social profile in *STXBP1*.

There are several important limitations to this study. The sample size is small and encompasses a wide age range. We aimed to ascertain participants as broadly as possible, from multiple regions of the UK and multiple medical specialties. In comparison to previous studies, the *STXBP1* participant group had a higher proportion of individuals ascertained via clinical genetics rather than paediatric neurology, which may provide a more comprehensive picture of the phenotypic spectrum or may under-represent individuals with early-onset epileptic encephalopathy. The data were collected via parent report, and it was not feasible to corroborate or elaborate via review of medical records. It was also not feasible for researchers collecting data to be blind to genetic diagnoses; however, our use of standardised parent-report questionnaire measures should mitigate potential bias of diagnosis-informed clinician ratings. Non-parametric statistical analyses are included to enhance our descriptive observations and provide an indication of which measures may discriminate between groups. Given the small sample size and large number of measures, adjusting the p value threshold for statistical significance to account for multiple comparisons is not practical, and results should be interpreted with due caution. The robustness of our findings will, we hope, be tested in future independent studies with pre-registered hypotheses and predictions building on our exploratory results. In view of small sample size, within-group analysis of behavioural variation was not justified (for example genotype-phenotype correlations, associations with epilepsy severity or age of onset), and these analyses would improve the prognostic utility of our results. A further limitation of the current study is that we have not directly assessed cognitive abilities in young people with *STXBP1* variants and are reliant on parent report of adaptive function and behavioural characteristics. To determine whether specific learning deficits characterise the disorder and contribute to behavioural characteristics, it is necessary to accurately measure relevant aspects of cognition in individuals with moderate to severe ID. However, assessment of cognitive function in individuals with AS or *STXBP1* is challenging due to low levels of understanding and hyperactivity. To achieve this, cognitive tasks must first be developed that assess the capacity for learning different skills (rather than simply documenting deficits), are intuitive, and are developmentally appropriate. In addition to behavioural methods, application of non-invasive physiological approaches such as wireless EEG in semi-naturalistic settings may be feasible and informative. These methods should be developed now and applied in future once a larger number of individuals have been diagnosed via genome-wide testing and can be followed from early childhood through to adult life.

Conclusion

This research presents the post-diagnostic evaluation of children and adolescents with *STXBP1*-associated neurodevelopmental disorder. Individuals with *STXBP1* variants have a wide range of complex adaptive, social and behavioural characteristics. Via comparison to individuals with other ID-related genetic diagnoses, we have highlighted a more limited set of characteristics which may discriminate individuals with *STXBP1* from individuals with other monogenic causes of neurodevelopmental disorder. Delineating this spectrum should assist in recognising this disorder and supporting families after diagnosis. Moreover, convergence between *STXBP1* phenotypes across species and across disorders (AS) provides a new starting point for understanding the molecular and neural mechanisms which underlie prosocial development.

Additional files

Additional file 1: STXBP1 Participant Information. (DOCX 17 kb)
Additional file 2: Gene groups and participant numbers recruited to the BINGO study. (DOCX 12 kb)

Abbreviations

ADHD: Attention deficit hyperactivity disorder; AED: Anti-epileptic drug; DBC: Developmental Behavioural Checklist; ID: Intellectual disability; MABC-2: The Movement Assessment Battery for Children 2; SCBU: Special care baby unit; SES: Socioeconomic status; STXBP1: Syntaxin-binding protein 1

Acknowledgements

We sincerely thank the study participants, their families, and carers for their contributions to this study. We are grateful to the Deciphering Developmental Disorders (DDD) study, the IMAGINE ID study, UK regional genetic centres, diagnostic laboratories, and referring clinicians for enabling us to publicise this study to the families who have taken part.

Authors' contributions

SOB and ENC acquired, analysed, and interpreted the data. SOB drafted the paper and all other authors substantially revised it. DEA, GS, and KB conceived and designed the study. The DDD study made a significant contribution to the genetic diagnosis of both study groups and facilitated recruitment to the study. All authors read and approved the final manuscript.

Funding

This study was funded by the Newlife Charity for Disabled Children and supported by the National Institute for Health Research and Medical Research Council UK. The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003], a parallel funding partnership between Wellcome and the Department of Health, and the Wellcome Sanger Institute [grant number WT098051]. The views expressed in this publication are those of the author(s) and not necessarily those of Wellcome or the Department of Health. The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network. This study makes use of DECIPHER (http://decipher.sanger.ac.uk), which is funded by the Wellcome Trust.

Availability of data and materials

All data analysed during this study are included in this published article (and its supplementary information files).

Ethics approval and consent to participate

This study was approved by the Cambridge Central Research Ethics Committee ('Phenotypes in Intellectual Disability', reference: 11/EE/0330). Parent or Carer gave written informed consent on behalf of participants under the age of 16. For participants aged 16 and over, lacking capacity to consent, a consultee was appointed. Prior to testing, research assistants ensured the project had been discussed with participants. The DDD study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC).

Consent for publication

Obtained within study participation consent.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, UK. [2]Department of Experimental Psychology, University of Oxford, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK. [3]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. [4]Department of Medical Genetics, University of Cambridge, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0XY, UK.

Received: 5 April 2019 Accepted: 22 July 2019
Published online: 06 August 2019

References

1. Gilissen C, Hehir-Kwa JY, Thung DT, Van De Vorst M, Van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511(7509):344–7 Nature Publishing Group.
2. Rizo J. Mechanism of neurotransmitter release coming into focus. Protein Sci. 2018;27(8):1364–91.
3. Saitsu H, Kato M, Mizuguchi T, Hamada K, Osaka H, Tohyama J, et al. De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. Nat Genet. 2008;40(6):782–8.
4. Carvill GL, Weckhuysen S, McMahon JM, Hartmann C, Møller RS, Hjalgrim H, et al. GABRA1 and STXBP1: novel genetic causes of Dravet syndrome. Neurology. 2014;82(14):1245–53.
5. Deprez L, Weckhuysen S, Holmgren P, Suls A, Van Dyck T, Goossens D, et al. Clinical spectrum of early-onset epileptic encephalopathies associated with STXBP1 mutations. Neurology. 2010;75(13):1159–65.
6. Di Meglio C, Lesca G, Villeneuve N, Lacoste C, Abidi A, Cacciagli P, et al. Epileptic patients with de novo STXBP1 mutations: key clinical features based on 24 cases. Epilepsia. 2015;56(12):1931–40.
7. Stamberger H, Nikanorova M, Accorsi P, Angriman M, Benkel-herrenbrueck I, Capovilla G, et al. STXBP1 encephalopathy: a neurodevelopmental disorder including epilepsy. Neurology. 2016;86:954–962.
8. Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, et al. De novo mutations in epileptic encephalopathies. Nature. 2013;501(7466):217–21 Nature Publishing Group.
9. Chan A, Bunga L, Lamoureux S, Faheem M, Drake J, Shlien A, et al. Germline and somatic mutations in STXBP1 with diverse neurodevelopmental phenotypes. Neurol Genet. 2017;3(6):e199.
10. Barcia G, Chemaly N, Gobin S, Milh M, Van Bogaert P, Barnerias C, et al. Early epileptic encephalopathies associated with STXBP1 mutations: could we

better delineate the phenotype? Eur J Med Genet. 2014;57(1):15–20 Elsevier Masson SAS.

11. Gburek-Augustat J, Beck-Woedl S, Tzschach A, Bauer P, Schoening M, Riess A. Epilepsy is not a mandatory feature of STXBP1 associated ataxia-tremor-retardation syndrome. Eur J Paediatr Neurol. 2016;20(4):661–5 Elsevier Ltd.
12. Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012;380(9854):1674–82.
13. Suri M, Evers JMG, Laskowski RA, O'Brien S, Baker K, Clayton-Smith J, et al. Protein structure and phenotypic analysis of pathogenic and population missense variants in STXBP1. Mol Genet Genomic Med. 2017;5(5):495–507.
14. Yuge K, Iwama K, Yonee C, Matsufuji M, Sano N, Saikusa T, et al. A novel STXBP1 mutation causes typical Rett syndrome in a Japanese girl. Brain Dev. 2018;40(6):493–7 The Japanese Society of Child Neurology.
15. McRae JF, Clayton S, Fitzgerald TW, et al. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542(7642):433–8.
16. Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, Van Kogelenberg M, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet. 2015;385(9975):1305–14.
17. Hamdan FF, Gauthier J, Dobrzeniecka S, Lortie A, Mottron L, Vanasse M, et al. Intellectual disability without epilepsy associated with STXBP1 disruption. Eur J Hum Genet. 2011;19(5):607–9.
18. Milh M, Villeneuve N, Chouchane M, Kaminska A, Laroche C, Barthez MA, et al. Epileptic and nonepileptic features in patients with early onset epileptic encephalopathy and STXBP1 mutations. Epilepsia. 2011;52(10):1828–34.
19. Mastrangelo M, Peron A, Spaccini L, Novara F, Scelsa B, Introvini P, et al. Neonatal suppression-burst without epileptic seizures: expanding the electroclinical phenotype of STXBP1-related, early-onset encephalopathy. Epileptic Disord. 2013;15(1):55–61.
20. Yamamoto T, Shimojima K, Yano T, Ueda Y, Takayama R, Ikeda H, et al. Loss-of-function mutations of STXBP1 in patients with epileptic encephalopathy. Brain Dev. 2016;38(3):280–4 The Japanese Society of Child Neurology.
21. Patzke C, Han Y, Covy J, Yi F, Maxeiner S, Wernig M, et al. Analysis of conditional heterozygous STXBP1 mutations in human neurons. J Clin Invest. 2015;125(9):3560–71.
22. Verhage M. Synaptic assembly of the brain in the absence of neurotransmitter secretion. Science. 2000;287(5454):864–9.
23. Toonen RFG, Verhage M. Munc18-1 in secretion: lonely Munc joins SNARE team and takes control. Trends Neurosci. 2007;30(11):564–72.
24. Grone BP, Marchese M, Hamling KR, Kumar MG, Krasniak CS, Sicca F, et al. Epilepsy, behavioral abnormalities, and physiological comorbidities in syntaxin-binding protein 1 (STXBP1) mutant zebrafish. PLoS One. 2016;11(3):1–25.
25. Kovačević J, Maroteaux G, Schut D, Loos M, Dubey M, Pitsch J, et al. Protein instability, haploinsufficiency, and cortical hyper-excitability underlie STXBP1 encephalopathy. Brain. 2018;141(5):1350–74.
26. Sparrow CB. The Vineland Adaptive behavior Scales-Second Edition-VBAS II. In: Major psychological assessment instruments, vol. 2. Needham Heights: Allyn & Bacon; 2005. p. 199–231.
27. Einfeld SL, Tonge BJ. Developmental Behavior Checklist. 2nd Edi ed. Australia: University of New South Wales and Monash University; 2002.
28. Conners CK. Conner's 3rd Edition Technical Manual. 3rd Edit ed. Toronto: MHS Assessments; 2008.
29. Constantino JN, Gruber CP. The Social Responsiveness Scale Manual, Second Edition (SRS-2). 2nd Edi ed. Torrance: WPS; 2012.
30. Henderson SE, Sugden DA, Barnett A. Movement ABC-2 Movement Assessment Battery for Children – Second Edition. 2nd Edi ed. Pearson Education Ltd: Minneapolis; 2013.
31. Richards C, Powis L, Moss J, Stinton C, Nelson L, Oliver C. Prospective study of autism phenomenology and the behavioural phenotype of Phelan-McDermid syndrome: comparison to fragile X syndrome, Down syndrome and idiopathic autism spectrum disorder. J Neurodev Disord. 2017;9(1):1–15.
32. Orock A, Logan S, Deak F. Molecular and Cellular Neuroscience Munc18-1 haploinsufficiency impairs learning and memory by reduced synaptic vesicular release in a model of Ohtahara syndrome. Mol Cell Neurosci 2018; 88:33–42. Elsevier.
33. Mabb AM, Judson MC, Zylka MJ, Philpot BD. Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. Trends Neurosci. 2011;34(6):293–303 Elsevier Ltd.
34. Hager T, Maroteaux G, Julsing J, Van VR, Stiedl O. Munc18-1 haploinsufficiency results in enhanced anxiety-like behavior as determined by heart rate responses in mice. Behav Brain Res. 2014;260:44–52 Elsevier B.V.
35. Horsler K, Oliver C. The behavioural phenotype of Angelman syndrome. J Intellect Disabil Res. 2006;50(1):33–53.
36. Bakke KA, Howlin P, Retterstøl L, Kanavin ØJ, Heiberg A, Nærland T. Effect of epilepsy on autism symptoms in Angelman syndrome. Mol Autism. 2018;9(1):1–8.
37. Pelc K, Cheron G, Dan B. Behavior and neuropsychiatry manifestations in Angelman syndrome. Neuropsychiatr Dis Treat. 2008;4(3):577–84.
38. Huang HS, Burns AJ, Nonneman RJ, Baker LK, Riddick NV, Nikolova VD, et al. Behavioral deficits in an Angelman syndrome model: effects of genetic background and age. Behav Brain Res. 2013;243(1):79–90 Elsevier B.V.
39. Oliver C, Woodcock K, Adams D. The importance of aetiology of intellectual disability. In: Grant G, Ramcharan P, Flynn M, Richardson M, editors. Learning disability: a life cycle approach to valuing people. Milton Keynes: Open University Press/Wiley; 2010. p. 135–46.
40. Chevallier C, Kohls G, Troiani V, Brodkin ES, Schultz RT. The social motivation theory of autism. Trends Cogn Sci. 2012;16(4):231–9.
41. Stoppel DC, Anderson MP. Hypersociability in the Angelman syndrome mouse model. Exp Neurol. 2017;293:137–43 Elsevier Inc.
42. Berrios J, Stamatakis AM, Kantak PA, Mcelligott ZA, Judson MC, Aita M, et al. Loss of UBE3A from TH-expressing neurons suppresses GABA co-release and enhances VTA-NAc optical self-stimulation. Nat Commun. 2016;7:1–8 Nature Publishing Group.

Publisher's Note

APPENDIX J MEDICAL HISTORY QUESTIONNAIRE

Medical Assessment

Date of assessment	
Date of birth	
Age at assessment	

PERSONAL DETAILS

Name:	**ID:**
Address:	
Telephone:	**Email:**
Parent(s)/Guardian(s) Names:	
Parent / Guardian Occupation(s):	

MEDICAL PROFESSIONALS

GP Name: **Address:**
Present illnesses for which the child is being treated
Prescription medication
Paediatrician Name: **Address:**
Neurologist Name:

Address:
Psychologist Name: **Address:**
Therapists Name: **Address:**
Other Medical Professionals:

MEDICAL HISTORY

Antenatal

Duration of pregnancy			
Any concerns about the baby during pregnancy?			
Any concerns about the mother during pregnancy?			
Any abnormalities on the ultrasound screening?			
Vomiting	**Bleeding**	**Infections**	**Threatened Miscarriage**
Smoking	**No. of cigarettes**	**Alcohol use**	
Other complications			
Maternal medications			

Birth/ Neonatal

Duration of Labour:		**Birth weight:**	
Labour type:	Spontaneous	Induced	
Type of Delivery:	Normal	Breach	Caesarean
Any concerns about the baby during delivery?			
Resuscitation? Apgars			
Was he/she in SCBU:	Jaundice	Infections	Other
Any unusual physical features			
Describe the first few days and weeks:			
Early feeding:			
Did you have any concerns from the start?			

Infancy

Difficult to calm or comfort:		
Colicky:	**Excessively irritable:**	**Floppy:**
Difficulty feeding:		
Disturbed sleep patterns (describe):		
Other:		

Current

Weight	
Height	

Genetic mutations

Childhood and Adolescence

	Medical Problems (hospitalisations, childhood diseases, allergies, eating problems, sleep disorders, other problems)	**Sensory Problems** (Vision or hearing problems)	**Educational Placement** (type of school, repeated years, special education services received)
Early Childhood (<3 yrs)			
Mid-Childhood (3-11)			
Adolescence			
Early Adulthood			
Later Adulthood			

Neurological Problems

Head Injury:		**Coma:**		**Convulsions with/out convulsions:**	
Epilepsy:					
Age of onset:			**Age stopped:**		
Type of seizure:	Generalised			Focal:	
How frequent are/ were the seizures?					

Scans

MRI	
Scan done:	When:
Hospital:	
EEG	
Done:	When:
Hospital:	

DEVELOPMENTAL MILESTONES

	Age	Early	Normal	Late
Sat without support				
Crawled				
Walked without assistance				
Spoke first words				
Said sentences				
Toilet trained				

CURRENT FUNCTION

Motor
Language
Cognitive
Social
Emotional-Behavioural

FAMILY HISTORY

Family Tree
For each of the following, please specify which relative *(parents, siblings, grandparents, aunts, uncles or cousins)* and which side of the family *(maternal or paternal)* has had a history of the problem or disorder.

Long-term medical illness	**Learning difficulties**
Depression, anxiety or other mental health problem	**Epilepsy**
Genetic Disorder	

Additional Notes: Interests