The Development of Cognitive Flexibility in Early Childhood

Yesim Yavaslar Dogru

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Science
Department Psychology

May 2021

Abstract

The aim of this thesis was to better characterise how cognitive flexibility (CF) develops in early childhood. Cognitive flexibility has been typically characterised as overcoming perseveration between the ages of three and four years (Zelazo, 2006). Although this view has improved our understanding of the development of CF, more recent studies have suggested that this view may be too simplistic (Blakey, Visser, & Carroll, 2016). In this research, I aim to address two gaps with previous research to better characterise CF development in the early years. First, many existing studies have assessed CF with tasks that constrain the kinds of response children can produce on the tasks. Second, most prior research has only assessed CF in 3- to 4-year-olds as most CF tasks are too difficult for younger children. To address these gaps in the present work, I developed less constrained CF tasks that allowed for a greater variety of responses and that would be used in a younger and wider age range. To fully capture the varied responses that children could produce on the tasks, I applied statistical modelling techniques that were able to identify different performance patterns on the tasks. Finally, I was able to see how far any performance patterns generalised across tasks and replicated across studies. The results showed that CF has a more protracted developmental trajectory than previously thought that begins with 2-year-olds showing unsystematic mixed responding after the rule changes, some 3-year-olds showing perseveration and 4-year-olds showing successful switching. These three performance patterns appear to follow a developmental trajectory, and they were seen across two very different CF tasks. However, I found that the degree of perseveration in CF development is likely to change as a function of the task used to measure CF, suggesting this may be a less stable pattern of performance.

Acknowledgements

First, I would like to thank my wonderful supervisors *Dr Dan Carroll* and *Dr Emma Blakey* for being such great supervisors from the very beginning of my PhD. You have always been warmth, kind, considerate, and friendly. I have learnt so much in our supervision meetings, and have enjoyed so much in our gatherings in parks. You have always encouraged me with your endless patience and supported me for believing in myself. You are my academic role models for ever. Maybe beyond that, you have always touched my life with your unconditional positive regard and valuable thoughts. You are my supervisors, my friends, and 'my family' in Sheffield. As I always say, I am the luckiest PhD student ever as I have the greatest privilege of meeting you and working with you.

I would like to thank my examiners *Dr Danielle Matthews* and *Dr Nicolas Chevalier* for giving me the chance to talk about my research and for their valuable contributions to my thesis. It was an honour to discuss my research with you.

I would like to thank in millions to my partner, *Onur* for always being there for me no matter when I ask or what I ask. Sharing life with you is the most valuable thing that has ever happened to me. Can't wait to live many others together. I would not make it without your endless support, love, and caring.

I would like to thank my dearest family, *Ilknur Yavaslar*, *Iskender Yavaslar*, *Yeliz Burgac,* and *Pamuk* for always being there, supporting and listening me even if we were thousand miles apart. And thank you for my lovely little niece, *Gunes* for choosing us as your family.

I would like to thank *Merve* for being my best friend and for witnessing every moment of life since the beginning of primary school to the PhD years and I'm sure to the way beyond that. I'm also thankful to my lovely friends *Chen* and *Sema* for their infinite support from many miles away.

I would like to thank my dearest friend *Enja* for her warm presence and for being an excellent coffee date. You've always supported me with your calm personality, lovely smile and your thoughtful messages. You've 'literally' made Sheffield a home for me which is away from home. I'm also grateful for my lovely friend *Paulina* for always supporting me from the beginning of my PhD and for initiating the great adventures we had together in many parts of the UK. You will always be my hero for being there for me with a single call.

It was such a pleasure to be a member of Sheffield Cognitive Development Lab and to work with its lovely members *Steph*, *Ella*, *Burcu,* and *Emelie*. It was great to have delightful conversations about research in children and to share many experiences about PhD life. I am also thankful for my lovely friends in Cathedral Court, *Eric, Stan, Laura, Mate, Ali,* and *Fidan* for their armful warmth, support, and welcome. Also, a special thanks for *Claudia* for always listening and supporting me with a great smile on her face and for being a lovely 'auntie'.

Finally, all the children, parents, and nurseries, thank you for participating in my research. I couldn't make this research without your kind support.

Financial Support

This PhD research was supported by the Turkish Ministry of National Education. In addition, I am grateful to the Experimental Psychology Society (EPS) and the Sheffield Undergraduate Research Experience scheme for funding two summer research project placements awarded to Emelie Chard to assist me with data collection. I am also grateful to the Society for Research in Child Development for awarding me with a travel grant that allowed me to attend to the Society for Research in Child Development conference in 2019 to present my PhD research.

Declaration

I confirm that this thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means. This work has not been previously presented for an award at this, or any other, university. The experiments in this thesis were designed and conducted by myself and my supervisors Dr Dan Carroll and Dr Emma Blakey. In addition, Emelie Chard helped with data collection for study two (presented in Chapter three).

Experimental studies in Chapter two and in Chapter three of this thesis were presented at the following conferences:

Conference Talks

Yavaslar Dogru, Y., Carroll, D. J., & Blakey, E. (2020, July). *Cognitive flexibility in 2- to 4-year-olds: Insights from a new fine-grained switching task.* Postgraduate Psychology Virtual Conference, Sheffield, UK.

Yavaslar Dogru, Y., Carroll, D. J., & Blakey, E. (2019, September). *Cognitive flexibility in early childhood: Insights from a new fine-grained switching task.* British Psychological Society (BPS)'s Cognitive Psychology and Developmental Psychology Joint Conference, Stoke on Trent, UK.

Yavaslar Dogru, Y., Carroll, D. J., & Blakey, E. (2018, December). *How does working memory contribute to children's flexible cognition?* Greater Yorkshire Memory Meeting (GYMM), Sheffield, UK.

Yavaslar Dogru, Y., Carroll, D. J., & Blakey, E. (2018, May). *How do children learn to flexibly adjust their behaviour?* Postgraduate Psychology Conference, University of Sheffield, UK.

Conference Poster Presentations

Yavaslar Dogru, Y., Carroll, D. J., & Blakey, E. (2021, April). *Are patterns of cognitive flexibility development generalizable, or task specific?* Society for Research in Child Development (SRCD) Virtual Conference, USA.

Yavaslar Dogru, Y., Carroll, D. J., & Blakey, E. (2019, March). *The emergence of cognitive flexibility: How do children develop flexible behaviour?* Society for Research in Child Development (SRCD), Baltimore, MD, USA.

Table of Contents

List of Tables

List of Figures

Chapter One

The Development of Cognitive Flexibility in Early Childhood

This thesis investigates how cognitive flexibility (CF) develops in early childhood. The review of the literature begins with a brief overview of what executive functions are, and continues to charting the development of CF from infancy to school age. In this thesis, I consider CF to be an emergent skill, with its development made possible by improvements in working memory and inhibitory control. This chapter will discuss how executive functions contribute to CF, as well as the theoretical accounts that have been suggested in explaining CF development. The chapter will end with two important directions that should be addressed by future research if we are to better understand how CF develops.

1.1. Executive Functions

Executive functions are generally defined as a group of correlated cognitive skills that allow us to control our behaviour (Garon, Bryson, & Smith, 2008). Executive functions enable us to maintain information, to suppress internal and external distractions not relevant to the task, and to flexibly adjust our behaviour according to our environment and our goals (Garon, Smith, & Bryson, 2014). For example, if we were trying to accomplish a task in the face of distractions, executive functions would help us ignore or suppress those distractors, and resume where we were in order to complete a particular task (Hughes & Graham, 2002). Executive functions are especially important in new situations where habitual behaviours we are used to performing are inappropriate according to the current task. As an illustration of this, executive functions are necessary in the classic ‘Simon Says’ task where participants need to behave according to the instructions of the experimenter, but crucially, only when those instructions are preceded by the phrase “*Simon says”* (e.g., “Simon says, put your arms up”) (Carlson, 2005). In this task, the automatic, prepotent behaviour would be to perform the actions immediately when the experimenter gives the instructions. To be able to succeed on this task, however, children need to (i) maintain the task rule that the action should only be performed when there is a *Simon says* preceding it, and to (ii) suppress performing the behaviour automatically when there is no *Simon says* beforehand (Carlson, 2005). This may require combining different executive functions, including working memory (to maintain the rule), inhibitory control (to suppress the automatic tendency to do the action regardless of whether the phrase starts with Simon Says) and cognitive flexibility (to flexibility adjust behaviour according to whether the phrase starts with Simon Says, or not). Together, these three skills make up what most researchers regard as executive function.

In adults, executive functions are considered to have three distinct but moderately correlated cognitive skills: inhibitory control, working memory, and cognitive flexibility (Miyake et al., 2000). Miyake and colleagues (2000) tested adults on five different tasks that tap executive functions, and performed Structural Equation Modelling to identify the latent factors that best described adults’ performance across those tasks. They found three distinct latent factors. By examining the demands on the various tasks, the researchers concluded that the three factors reflected inhibitory control (enabling us to suppress automatic or task-irrelevant information), working memory (enabling us to maintain or remember information), and cognitive flexibility (enabling us to update our behaviour in response to changes in task goals) (Miyake et al., 2000).

However, the distinct structure of executive functions is less apparent in early childhood (Hendry, Jones, & Charman, 2016). Specifically, when studies have been conducted with children following a similar design to Miyake and colleagues (2000), typically only one or two latent factors are reported. For example, Wiebe and colleagues assessed performance across a set of executive function tasks that tap on working memory and inhibitory control in 3- to 6-year-olds (Wiebe, Espy, & Charak, 2008), as well as only in 3-year-olds (Wiebe et al., 2011). However, in both studies, only a single factor was identified. When more recent studies have assessed children's performance across a series of tasks intended to measure working memory, inhibitory control and CF in 3- to 5-year-olds, they identified only two factors (which they interpreted as reflecting working memory and inhibitory control: (Lerner & Lonigan, 2014; Monette, Bigras, Lafrenière, 2015). This leads to the suggestion that CF may not be a single executive function itself, but may instead emerge from developments in working memory and inhibitory control.

Executive functions are shown to develop rapidly in preschool years (Garon et al., 2008). Improvements in executive functions around this time have been found to predict a number of advantageous developmental outcomes for children, making it an important time to understand the development of executive functions (Diamond, 2013). Executive functions at preschool age have predictive relations with children's school success, cognitive development, and socio-emotional development during preschool age and beyond (Ackerman & Friedman-Krauss, 2017). For example, several studies have shown that executive functions in general predict children's school success (Blair & Razza, 2007; Bull, Espy, Wiebe, Sheffield, & Nelson, 2011; Clark, Sheffield, Wiebe, & Espy, 2013). In particular, inhibitory control was found to be positively associated with Maths ability and vocabulary development (Birgisdottir, Gestsdottir, & Thorsdottir, 2015; Lonigan, Allan, Goodrich, Farrington, & Phillips, 2017), and negatively associated with demonstrating externalizing behaviours (van Dijk et al., 2017). Working memory was found to be positively related to better Maths ability as well as better letter identification when children start to school (Miller, Muller, Giesbrecht, Carpendale, & Kerns, 2013; Welsh, Nix, Blair, Bierman, & Nelson, 2010). Cognitive flexibility was positively associated with more sophisticated theory of mind skills (Marcovitch et al., 2015) and more advanced emotional understanding (Wang, Liu, & Feng, 2021). This evidence highlights the importance of understanding how executive functions develop during the preschool period. This understanding will in turn allow us to better understand and support the means by which children achieve a diverse range of successful outcomes later in life.

1.2. Cognitive Flexibility

Cognitive flexibility can be defined as a skill through which we can update our thoughts or our actions in response to changes in task goals or in the environment (Chevalier et al., 2012). There is increasing evidence to suggest that CF should be thought of as an emergent skill supported from the developments in working memory and inhibitory control (Carroll, Blakey, & FitzGibbon, 2016; Chevalier et al., 2012), rather than as a distinct executive function itself. Conceptually, in order to flexibly update our behaviour to reflect (say) changes in task rules, it is necessary both to maintain multiple rules in mind, and to suppress rules that had been applied previously but are no longer relevant (Ackerman & Friedman-Krauss, 2017). To illustrate with a developmental example, when children first start at school, they may make the mistake of calling their teacher ‘mum’, particularly if they tend to do that at home when they need help. In order to flexibly learn to call their teacher the appropriate name, they would need to: i) maintain awareness that they are currently in another environment and their mum is not present, and ii) supress any tendency to say ‘mum’ when they need to attract a helpful adult’s attention (a behaviour which may have been relevant that morning, but is no longer relevant in the classroom environment). Therefore, in this thesis, CF is considered to be an emergent skill which is supported by working memory and inhibitory control (Carroll et al., 2016; Chevalier et al., 2012).

Cognitive flexibility is typically assessed in children by using rule-switching tasks (Zelazo, 2006). These tasks typically have two phases in which children first perform a task according to one particular rule, after the rule changes, they perform the task according to another rule. The core demand of rule-switching tasks is the requirement to update behaviour according to a change in rule (Carroll et al., 2016). For example, on the classic Dimensional Change Card Sort Task (DCCS; Zelazo, 2006), preschoolers need to sort cards with coloured pictures into one of the two trays, first based on a particular rule in the pre-switch phase (e.g., shape) and then based on a different rule in the post-switch phase (e.g., colour). The cards to be sorted have *both* shape and colour information (e.g., a blue heart). Therefore on any given trial, children could potentially sort by either shape or colour, as the to-be-sorted cards can be sorted according to either rule. This creates a *response conflict* after a rule change. Response conflict occurs when the stimulus to be sorted has properties that prompt both the previous rule and the current rule. In other words, response conflict occurs when the perceptual properties of the stimulus trigger multiple competing rules on the task. Therefore, in order to switch rules successfully, children need to resolve this response conflict – they are thought to do this by maintaining the current sorting rule in mind, and by attending to only the rule-related dimension of the cards for a given trial. This example indicates how successful performance on a CF task necessitates the interplay of different cognitive skills, in particular working memory and inhibitory control (Cragg & Chevalier, 2012). Many studies have shown that preschoolers tend to be able to successfully resolve this response conflict, and thus switch rules successfully, by around the age of four years (Buss & Spencer, 2014; Doebel & Zelazo, 2015).

1.3. The Development of Cognitive Flexibility in Early Childhood

Cognitive flexibility is important because it is a clear early example of children showing systematic, goal-oriented behaviour. But it is not the first example of children demonstrating goal-oriented behaviour. Instead, early goal-oriented behaviour begins to emerge in infancy, as evidenced by paradigms that index controlled attention (Ellis & Oakes, 2006; Hendry et al., 2016). Therefore, in this section, I will first give a broad overview of the development of CF and its precursors from infancy to school age. I will then go into more detail to discuss what prior research has found at each stage of development, starting with how key competences to controlled behaviour begin to develop in infancy, through to precursors to CF in toddlerhood. I will then explain how key developments in CF arise during the preschool years, and up to school age.

One of the key developments that occurs in the first few years of life is the emergence of goal-oriented behaviour. Evidence of goal-oriented behaviour can be seen from as early as four months of age, when infants start to control their looking. This can be considered an *implicit* goal-oriented behaviour, because it is not completely clear whether infants have a clear goal in their mind they try to achieve, such as reaching a toy. Then, between infancy and the toddler period, children progress from demonstrating implicit forms of goal-oriented behaviour to toddlers beginning to demonstrate early examples of explicit goal-oriented behaviour. The difference between implicit and explicit goal-oriented behaviour is that while implicit goal-oriented behaviour implies more automatic and spontaneous actions, explicit goal-oriented behaviour implies more deliberate and reasoned actions (Evans & Stanovich, 2013; Phipps, Hagger, & Hamilton, 2019). In other words, although both forms underpin goal-directedness, implicit goal-oriented behaviour is more fragile and more prone to changing moment to moment as attention is captured by different objects in the tasks or in the environment compared to explicit goal-oriented behaviour which is more strongly purposeful (Jennings, 2004). While there has been very limited executive function research done in toddlers, the few studies that have been done demonstrate that toddlers start to show explicit goal-oriented behaviour at around the age of two years by being able to follow a simple rule deliberately, and to then reverse that rule (e.g., sorting coloured cubes into *same* coloured boxes, followed by sorting the same cubes into *different* coloured boxes). Then, during the preschool years, children become able to resolve response conflict when sorting objects. Preschoolers demonstrate clear explicit goal-oriented behaviour, by being able to consistently switch their behaviour according to different rules (e.g., sorting colourful shapes based first on their shape, and then on their colour). Finally, children at early school age start to show more advanced forms of explicit goal-oriented behaviour – notably, by being able to switch their behaviour multiple times, and according to multiple rules, back and forth on a trial-by-trial basis (e.g., colour, shape, colour). In the sections that follow, these key developments on the road to CF will be explored in more detail.

1.3.1. Precursors to cognitive flexibility in infancy

The beginnings of implicit goal-oriented behaviour start to emerge in infancy (Hendry et al., 2016; Wiebe, Lukowski, & Bauer, 2010). Throughout this period, infants start to show increasing control over their looking and reaching. These developments are considered important precursors to CF because they show infants' control of their attention in response to external factors (Hendry et al., 2016). Because infants have limited verbal and motor abilities, it is difficult to definitively assess goal-oriented behaviour in infancy. Typical CF paradigms are obviously unsuitable for infants and toddlers, as they usually have prohibitively high language and motor demands. Therefore, to assess goal-oriented behaviour in infancy, researchers have usually used paradigms that examine infants' looking. These paradigms have generally examined infants' ability to shift focus of attention between objects that are located at different sides of a screen (Johnson, Posner & Rothbart, 1991). Infants became able to consistently shift their attention between objects at different sides of screen around four months of age (Hendry et al., 2016).

Control of looking can be considered as one of the earliest examples of *implicit* goal-oriented behaviour. As an example, Fixation-Shift tasks examined whether infants can engage in a shift of visual attention when an object shown at the centre of a screen disappears (Atkinson, Hood, Wattam-Bell, & Braddick, 1992). In these tasks, infants are presented with an attractive object at the centre of the screen and they are encouraged to attend to it. Then different other objects are presented at the sides of the screen *either* in the presence of the central object *or* in the absence of the central object. The measure of interest is how well infants can shift attention to the objects at the sides of the screen when the object at the centre stays on the screen *or* when it disappears. Until four months of age, infants are usually not able to shift their attention to the objects at the sides of the screen as they fixate their attention on the object at the centre of the screen (Hendry et al., 2016; Kulke, Atkinson, & Braddick, 2015). However, by around four months of age, infants could shift attention from the centre of the screen to the objects at the sides of the screen, especially when the object at the centre disappears while the other objects are presented at the sides of the screen. In addition, the time taken to disengage attention from the central object significantly decreases as infants get older (Hunnius, Geuze, & van Geert, 2006). Attentional control scores of infants were associated with better performance at later ages on different executive function tasks, including CF (Holmboe, Fearon, Csibra, Tucker, & Johnson, 2008).

Another example of an early form of goal-oriented behaviour is the control of reaching, which emerges at around eight months of age (Hendry et al., 2016). Control of reaching can be considered as a very early example of *explicit* goal-oriented behaviour, because infants are assumed to have a clear goal in these tasks (such as wanting to retrieve a hidden toy in the A-not-B task; Diamond 1985; Piaget, 1954). Being able to control reaching is an important precursor to CF, since it reflects the beginning of directing actions and attention in service of a goal (Clearfield, Diedrich, Smith & Thelen, 2006). A classic example is seen in the A-not-B task, in which an infant observes an experimenter place a toy in one of two boxes (Diamond, 1985). After two seconds, the infant is allowed to freely reach the toy. After hiding the toy in the same initial box (location A) for several trials, the toy is then placed in the other box (location B). After two seconds, the child is allowed to reach the toy. Twelve-month-olds tend to correctly reach to the new location to retrieve the toy. However, eight-month-olds typically continue to look for the toy in the initial "A" location. In other words, they make a perseverative error. Eight-month-olds' perseverative responses on the task have sometimes been characterised as executive in nature, reflecting difficulties in updating the representation where the toy is located, or difficulties in inhibiting attention to the previous location (Diamond, 1985).

When we examine infants' behaviour on the A-not-B task longitudinally, a striking trajectory is seen (Clearfield et al., 2006; Clearfield & Niman, 2012). For these longitudinal studies, a slightly different variant of the A-not-B task was used. That is, the experimenter placed two toys into a box in front of the child (toy A and toy B). The experimenter then waved toy A for a while until the infant's attention was captured. After that, the experimenter brought toy A closer to the infant (compared to toy B), by placing the toy on the front side of the display box. On the final step, the experimenter moved the display box forward so that the child could reach for the toys. This procedure of waving the toy A and allowing for the infant to reach the toys was repeated for three trials. After that, the important switch trial happens: the experimenter waved toy B and brought it closer to the infant (compared to toy A), by placing the toy on the front side of the display box. The infant then was allowed for the reach the toys. When the performance of infants was examined between the ages of five and 12 months, an interesting developmental pattern was found on the switch trials: Twelve-month-olds reached for toy B, thus switching away from reaching to toy A; 8-month-olds made a perseverative response, by continuing to reach for toy A; but 5-months-olds sometimes reached for toy A, and sometimes reached for toy B. They thus performed unsystematically, showing neither the perseverative response pattern of 8-month-olds, nor the successful switching of the 12-month olds.

At first glance, it might appear that the performance of 5-month-olds was better than the performance of 8-month-olds, because they made fewer perseverative responses. This is indeed a surprising and noteworthy finding, since it appears to contradict the basic expectation of that the performance of older children is better than the performance of younger children. However, although the 5-month-olds made fewer perseverative responses on B trials, it would be misleading to claim that they were actually performing better. Rather, the youngest infants' performance was qualitatively poorer, because they performed unsystematically, sometimes making a perseverative response and sometimes making a switching response. In contrast, the numerically lower number of correct responses from the 8-month-olds was arguably *qualitatively* better. Even though these children made more perseverative responses, their behaviour was stable and systematic, in a way that the 5-month-olds' responses were not. This more stable performance was found to be supported by developments in working memory (Clearfield, Dineva, Smith, Diedrich, & Thelen, 2009). Five- to six-month-olds had lower working memory scores than eight-months-olds. Their poorer working memory may mean that they had no reliable mental representation of which toys the experimenter had previously indicated, and they therefore selected between the two toys in a haphazard way. In contrast, eight-months-olds' better working memory likely meant that they had a strong mental representation for the initially waved toy that was shown to them multiple times. Therefore, the perseverative responses of the eight-months-olds may arise from their better working memory, meaning that they had strong representations of earlier events – and that their responses were based on this representation.

Another example of abilities that support goal-oriented behaviour is seen at 20-months-olds who start to show the ability to ignore task-irrelevant distractors (Wiebe et al., 2010). This ability reflects infants' growing ability to suppress information that is not related to their current task goal. For example, Wiebe and colleagues (2010) examined children's ability to copy sequences of actions performed by an experimenter. In the initial phase of the experiment, children watched the experimenter perform two different three-step actions. For example, to complete the action of *making a toy fish jump*, the experimenter i) set up a seesaw, ii) put the fish on the seesaw, and iii) pressed one end of the seesaw to make the fish jump. Then, to complete the action of *making a toy dog dance*, the experimenter i) set up a peg top, ii) put the dog on it, and iii) spin the peg top to make the dog dance. After that, children were presented with all the toys necessary to imitate first action (the fish and the seesaw), as well as only one of the toys necessary to imitate the second action (only the peg top). Therefore, children could only fully imitate the first action accurately. In this case, the peg top is a distracting toy. The results showed that around 20 months of age, children were found to accurately imitate what the experimenter did by ignoring the distractor toy (Wiebe et al., 2010).

1.3.2. The development of cognitive flexibility in toddlerhood

It should be noted that there is a shortage of cognitive research with 2-year-olds – perhaps for methodological reasons, since this age group is too old for many infant paradigms, but too young for many preschooler paradigms. This shortage limits our understanding of how CF develops at this age (Carroll et al., 2016). However, although they are few in number, those studies still suggest that there are important developments happening during toddlerhood regarding the development of goal-oriented behaviour, including becoming able to act according to a simple explicit rule, and becoming able to update behaviour in the presence of distractions.

It is during the toddler period that children begin to show early forms of explicit goal-oriented behaviour, perhaps made possible by improvements in language, memory and motor abilities. One of the first milestones on the road to CF in toddlerhood is that children can consistently follow an explicit simple rule (Wiebe & Bauer, 2005). This is an important ability, as it forms a baseline of stable goal-oriented behaviour which children can later learn to flexibly update, in response to different rules (Muller, Dick, Gela, Overton, & Zelazo, 2006). For example, when asked, 2-year-olds can put the toys into a bucket of other toys (see Johansson, Marciszko, Brocki, & Bohlin, 2016). Two-year-olds can also switch their behaviour in response to changing explicit rules – though notably, only as long as there is not a response conflict to be resolved. In other words, they can switch from following one rule to following another, as long as the stimuli to be sorted are quite simple. For example, on the Reverse Categorization task, 2-year-olds were asked to sort different coloured cubes into different coloured boxes, first in response to a straightforward matching rule (e.g., put the yellow cubes into the yellow box, and the blue cubes into the blue box), and then in response to the reverse rule (e.g., put the yellow cubes into the blue box, and the blue cubes into the yellow box) (Carlson, Mandell, & Williams, 2004). Two year-olds were successful both at initially placing the cubes into the relevant boxes, as well as with placing the cubes into the opposite boxes after the rule changes. This research suggests that 2-year-olds can, in some circumstances, flexibly switch their behaviour.

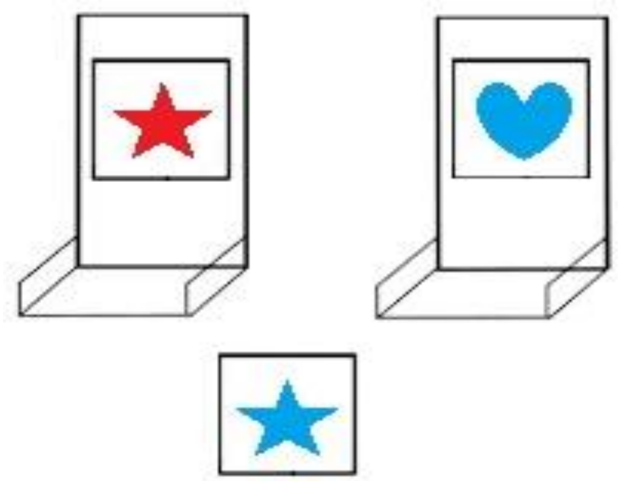
Recent research has suggested that 2.5-year-olds become able to update their behaviour in the presence of distracting, task-irrelevant information (Blakey, Visser, & Carroll, 2016). Though importantly, children at this age still cannot update their behaviour when there is a response conflict. For example, Blakey and colleagues (2016) tested 2- to 4-year-olds' CF using the Switching, Inhibition, and Flexibility Task (SwIFT). This is a computerised rule-switching task that requires children to pick one of the two colourful shapes that matches a prompt image on the relevant rule for a given trial (typically either shape or colour, with the rule changing midway through the task). Two separate versions of the SwIFT were used to assess CF: children either had to switch rules in the presence of a response conflict, or in the presence of distracting information (without any response conflict). Specifically, in the *conflicting condition*, children were asked to switch rules when there is a conflict to be solved. This means that, as in the standard DCCS task, after the rule changed, children could continue to sort stimuli by the previous rule. On the other hand, in the *distracting condition*, children were asked to switch rules when there is a need to ignore task-irrelevant information. This means that after the rule changed, children could not continue to sort stimuli by using the previous rule, since the stimuli to be sorted did not involve any information relating to that dimension; instead, they had to ignore distracting information (such as pattern). It was found that while 2-year-olds were, unsurprisingly, not able to switch rules while there is a response conflict to be solved, they were able to switch rules successfully in the distracting condition, where they had to overcome task-irrelevant distracting information. This was the first research to show that 2-year-olds *could* show explicit forms of CF on a switching task. Switching in the presence of distracting information constitutes an early example of explicit CF (Blakey et al., 2016).

1.3.3. The development of cognitive flexibility in preschool age

This ability to flexibly update responses in line with changing rules develops more fully during the preschool period. A key milestone in the development of CF during this time is that children can flexibly switch their behaviour when there is a response conflict to be resolved. The ability is usually assessed by tasks that ask children to switch from sorting bivalent stimuli by one dimension (e.g., shape) to sorting them by the other dimension (e.g., colour) (Zelazo, 2006; Espy, 1997). These tasks require children to shift their attention within an object, by first selectively attending to one dimension of a stimulus (e.g., the shape dimension of a coloured shape), and then after a rule change, redirecting their attention to the other dimension of that stimulus (e.g., the colour dimension of a coloured shape). Because the stimuli to be sorted contains both shape and colour information, during the post-switch trials, both rules will be triggered; and children must resolve this conflict in order to respond correctly after the rule changes. Children may resolve the response conflict by maintaining the currently relevant sorting rule in mind, and/or by focusing on the relevant dimension of the stimulus by inhibiting the dimension that is not currently relevant. Children tend to be able to do this at around four years of age (Doebel & Zelazo, 2015).

The most prevalent task to assess CF in preschoolers is the Dimensional Change Card Sort Task (DCCS; Zelazo, 2006; Zelazo, Frye, & Rapus, 1996). In the DCCS, there are two trays with marker cards attached (e.g., a red star and a blue heart - see Figure 1). There are also test cards to be sorted. Children are asked to sort cards into trays in response to a rule first (e.g., shape), and then in response to another rule (e.g., colour). In the standard version of the DCCS, the cards relate one tray by its shape, and the other tray by its colour. The DCCS consists of a pre-switch phase and a post-switch phase, with the sorting rule changing between these phases. The experimenter first shows how to sort the cards into one of the trays based on the pre-switch rule (e.g., shape). The child is then given cards one by one and is instructed to sort them based on the pre-switch rule (e.g., shape). The child sorts the cards for eight trials during the pre-switch phase (e.g., according to the shape rule), then the rule changes, and the child sorts cards for eight trials during the post-switch phase (e.g., according to the colour rule) (Zelazo, 2006).

Figure 1. The Dimensional Change Card Sort Task (DCCS).



Note. Children are instructed to sort the test cards (e.g., blue star) into one of the two trays with target cards attached (e.g., a red star and a blue heart). Children sort the test cards in response to a rule for eight trials in the pre-switch phase (e.g., shape), and then they sort the test cards according to a different rule for eight trials in the post-switch phase (e.g., colour).

Preschoolers' performance on the DCCS generally follows a simple trajectory: Almost all preschoolers can sort the cards correctly according to the first rule. However, after a rule-change, most 3-year-olds cannot resolve the response conflict, and continue to sort by the previous rule. In other words, after the rule changes, 3-year-olds perseverate by sorting according to the first (no longer relevant) rule. In contrast, 4-year-olds can typically resolve the response conflict, and can sort correctly by the current rule (Zelazo, 2006). This developmental change from perseverative performance at around age three to flexible performance at around age four has been found in many studies, and has generally been thought as the key development in CF (Cepeda & Munakata, 2007; Diamond, Carlson, & Beck, 2005; Munakata, Snyder, & Chatham, 2012). This has led many researchers to examine the cognitive processes that led children to overcome perseveration (Buss & Spencer, 2014; Cragg & Chevalier, 2012; also see section 5).

1.3.4. Cognitive flexibility in the early school years

During the early school years, at around five years of age, children show increasingly more advanced forms of explicit goal-oriented behaviour. They become able to flexibly perform more than a single switch back and forth, between multiple rules. Performance on the standard DCCS task generally reaches ceiling between the ages of four and five. At around this time, children become able to stay focused on tasks for longer, and can understand more complex task instructions than preschool children (Cragg & Chevalier, 2012). Therefore, to examine CF in school-age children, researchers use more advanced paradigms that require children to switch back and forth between multiple rules (e.g., sort by colour, shape, shape, colour, shape). At this stage, it is possible to use additional dependent variables to index CF. Thus, in addition to counting the number of correct switches children make, researchers can also use reaction time – and reaction-time bases calculations – as indices of different aspects of CF (Best, Miller, & Jones, 2009).

Cognitive flexibility from five years and above tends to be assessed by examining switch costs and mixing costs (Chevalier & Blaye, 2009; Cragg & Nation, 2009). Switch costs can be defined as the cost of switching from one rule to another (e.g., from shape to colour). Switch costs are usually measured by subtracting the reaction time on non-switch trials (e.g., sorting by pattern and then again sorting by pattern) from the reaction time on switch trials (e.g., sorting by pattern and then sorting by colour). Mixing costs can be defined as the cost of having to maintain the possibility of switching. Mixing costs are usually measured by subtracting the reaction time on non-switch trials in pure blocks (i.e. trials where no switch is possible) from the reaction time on non-switch trials in mixed blocks (i.e. trials where switching was possible, but did not happen). Between the ages of five and 11 years, these costs tend to decrease (Chevalier & Blaye, 2009; Davidson, Amso, Anderson, & Diamond, 2006).

Cognitive flexibility development in the school age period is thought to be enabled by improvements in executive functions (Cragg & Chevalier, 2012; Cragg & Nation, 2009). Researchers have noted that the development of CF in 5- to 6-year-olds is partly related to the improvements in inhibitory control – the ability to inhibit interference from non-relevant information (Chevalier, Blaye, Dufau, & Lucenet, 2010). For example, Chevalier and colleagues (2010) assessed 5- to 6-year-olds' CF on a variant of the DCCS in which children need to engage in multiple switches between two different rules (e.g., shape and colour). Children were presented with a prompt stimulus and two response stimuli, one of which matched the prompt stimulus by its shape, and one of which matched the prompt stimulus by its colour. The shape and the colour aspects of the prompt stimulus were separated in order to measure children's looking time to these relevant and non-relevant aspects by eye-tracking. Five-year-olds spent longer time looking at the non-relevant dimension of the stimulus compared to 6-year-olds. This suggests that failure to switch rules tends to be related to the inability to suppress attention from the non-relevant aspects of the stimuli (Chevalier et al., 2010).

The development of CF during early school age is also associated with improvements in working memory – the ability to represent and maintain task-related information (Blaye & Chevalier, 2011). For example, Chevalier and colleagues (2009) assessed how 5- to 6-year-olds switch rules in the presence of an arbitrary task cue on a variant of the DCCS. In this research, in the mixed block, children were presented with a border that covers the stimulus being sorted to nominate one rule, and they were presented with no border covering the stimulus being sorted to nominate the other rule (e.g., an arrow around the stimulus means the corresponding rule is *colour*, while no arrow around the stimulus means the corresponding rule is *shape*). In order to succeed on the task, children need to represent the association between the arrow and the corresponding rule (and the association between no arrow and the corresponding rule), as well as they need to maintain that association during the task. Results showed that 5- to 6-year-olds performed poorly at representing task cues with corresponding task rules compared to their performance on the standard version of the task where rules were transparent (Chevalier & Blaye, 2009). However, the following experiment in the same study showed that the role of arbitrary task rule on switching accuracy reduce with age as children's working memory scores increased between the ages of seven and nine years (Chevalier & Blaye, 2009).

1.4. The Role of Executive Functions in the Development of Cognitive Flexibility

Cognitive flexibility has been regarded as an emergent skill that is dependent on working memory and inhibitory control (Carroll et al., 2016; Chevalier et al., 2012). These executive functions are thought to contribute to developmental changes in CF, because they help us maintain information related to current goals, as well as help us suppress our automatic or habitual behaviours (Cragg & Chevalier, 2012; Kirkham et al., 2003; Morton & Munakata, 2002). At an empirical level, researchers have used correlational studies or have adopted individual difference approaches to understand the role of executive functions in the development of CF. Those who have performed correlational studies suggested that both working memory and inhibitory control are positively correlated with CF. For example, preschoolers' performance on a working memory span task was positively correlated with their performance on the DCCS and on the Shape School task (Blackwell, Cepeda, & Munakata, 2009; Espy, Bull, Martin, & Stroup, 2006). Likewise, preschoolers' performance on different inhibitory control tasks was positively associated with their performance on a switching task that required children to sort animals according to different categories (Memisevic & Biscevic, 2018).

Those who performed individual difference studies suggest that working memory may allow children to have active representations of the post-switch rule that is strong enough for switching to the current rule (Marcovitch, Boseovski, Knapp, & Kane, 2010). For example, Marcovitch, Boseovski, and Knapp (2007) examined how working memory contributes to CF in preschoolers by varying the degree of the need for maintaining the rule in mind in order to switch rules. For this purpose, 4- and 5-year-olds performed on two different versions of the DCCS: one version involved to-be-sorted cards most of which were exactly the same with the target card ("mostly redundant" condition; Marcovitch et al., 2007). Therefore, while sorting those identical cards, children do not need to maintain the relevant rule in order to succeed on the task. The other version involved to-be-sorted cards most of which have information related to both rules thus they match the target card either by their colour or by their shape ("mostly conflicting" condition; Marcovitch et al., 2007). Therefore, while sorting conflicting cards, children need to maintain the relevant rule in mind strongly in order to sort the cards correctly and to succeed on the task (Marcovitch, et al., 2007). The results showed that children did better in "mostly conflicting" condition than in "mostly redundant" condition. The researchers suggested that sorting with mostly conflicting cards meant that children needed to maintain the relevant rule strongly in mind in order to sort the cards correctly. However, while sorting with mostly redundant cards, there was much less need for children to maintain the relevant rule in order to sort the cards correctly, because those cards were identical to the target cards – and would therefore be sorted correctly whether children were using the previous rule or the current post-switch rule. Because children sorting with mostly redundant cards encountered rule conflict less often than children sorting with mostly conflicting cards, they had less need to keep the relevant rule in mind, and thus they were less well able to switch rules on trials where they encountered rule conflict (Marcovitch et al., 2007). This shows how important working memory is – and in particular how important maintaining the currently relevant rule in mind is – in flexible switching behaviour.

Recent research has also indicated that working memory and inhibitory control contribute to some aspects of CF, by supporting multiple switches (that is, switching back and forth). For example, Chevalier and colleagues (2012) tested 3- to 5-year-olds on a version of the Shape School task in which children needed to make multiple switches between two rules. The results showed that neither working memory nor inhibitory control were related to switch costs. However, they were significantly related to mixing costs. That is, children who had better working memory and inhibitory control also had lower mixing costs. This suggests that executive functions can play a role in monitoring the need to switch, and in maintaining the currently available rules for a given task (Chevalier et al., 2012).

More recent research with preschoolers has suggested that working memory and inhibitory control may make quite distinct contributions to CF. For example, Blakey and colleagues (2016) tested 2- to 4-year-olds on different versions of the SwIFT that varied in terms of the type of the information presented to children: a Conflicting version and a Distracting version. In the Conflicting version, children needed to switch rules while resolving response conflict. In other words, after the rule changed, it was possible for children to keep sorting stimuli by the previous rule. For example, when the sorting rule changed from colour to shape, the incorrect response option on the sorting array still matches the prompt stimulus by the previous rule (e.g., it matches the prompt stimulus by colour). Conversely, in the Distracting version, children needed to switch rules while ignoring task-irrelevant, distracting information – but *without* any within-stimulus conflict. In other words, after the rule changed, it was not possible for children to keep sorting stimuli using the previous rule – since the to-be-sorted stimuli in the post-switch phase did not contain any information relating to that rule. For example, when the sorting rule changed from colour to shape, although the to-be-sorted stimuli have shape and colour information, the incorrect response option does not match the prompt stimulus by its colour. The results indicated that switching when there is a *response conflict* to be resolved was related to better working memory; whereas switching when there is a *distracting information* to be suppressed was related to better inhibitory control (Blakey et al., 2016). This suggests that working memory is likely to support CF through allowing children to remember the current rule in mind and overcome the previous rule; while inhibitory control is likely to support CF through allowing children to overcome task-irrelevant information (Blakey et al., 2016).

1.5. Theoretical Accounts of the Development of Cognitive Flexibility

There are two main theoretical accounts that are suggested to explain the development of CF. One, the Attentional Inertia account, argues that inhibitory control plays the crucial role; the other, the Graded Representations account, argues that working memory plays the crucial role. These accounts have generally focused on explaining the change from perseveration to switching in 3- to 4-year-olds on the DCCS. In other words, they focus on a single change in behaviour and at a very specific age. However, given the recent research suggesting that children younger than three years can show basic CF by overcoming distraction errors (Blakey et al., 2016, see section 3.2.), those theories may be less useful in explaining the development of CF outside of the age range of the 3- to 4-year-old shift in CF (Carroll et al., 2016). In addition, although these accounts have been informative about the development of CF in preschoolers, they are somewhat unfalsifiable – they make identical predictions for how children will behave, and differ solely in how they attribute a causal role. That is, preschoolers' perseveration could be explained by inhibitory control (that is, the inability to inhibit the previously relevant dimension) and also by working memory (that is, the inability to maintain and update the currently relevant sorting rule). Nevertheless, while these accounts do not currently make predictions distinct enough to test against each other, it is still informative to consider how they account for the key processes that underpin CF development.

1.5.1. The Attentional Inertia Account

The Attentional Inertia account claims that perseveration on the DCCS arises from children's poor inhibitory control (Kirkham et al., 2003). This account proposes that because of the poor inhibitory control, 3-year-olds cannot inhibit their attention from the previous sorting dimension, and thus after a rule change they cannot shift their attention to the current sorting dimension. For example, while sorting based on a rule during the pre-switch phase (e.g., colour), children need to direct their attention to one dimension of the stimulus (e.g., the blue colour of a blue heart). After the rule changes (e.g., from colour to shape), children need to re-direct their attention to the other dimension (e.g., the heart shape of a blue heart). The Attentional Inertia Account argues that 3-year-olds cannot inhibit attention from the pre-switch phase; therefore during the post-switch phase, they keep attending to the dimension that of the pre-switch phase. This "attentional inertia" thus gives rise to perseveration (Diamond & Kirkham, 2005). Children with better inhibitory control are thought to be better able to suppress the no-longer relevant dimension of the stimulus, and therefore can better focus on the dimension that is currently relevant.

Support for this account is found in work where researchers sought to manipulate features of the DCCS to reduce inhibitory control demands, and which were found to improve performance. Inhibitory demands of the task can be reduced in different ways. One way is by verbally labelling the cards according to the currently relevant dimension. This would make the post-switch dimension more salient, and therefore allows children to better re-direct their attention to the post-switch dimension. For example, Kirkham and colleagues (2003) got children to label cards according to the relevant dimension (e.g., saying "red" while sorting by colour, *or* saying "a star" while sorting by shape), and found that children's performance improved. Another way is by reducing the selective attention demands of the task. For example, experimenters presented children with each sorting card and then covered it up, in order to reduce any interference from the irrelevant dimension depicted on the card. They found that children sorted cards better than when cards were left uncovered (Kirkham et al., 2003). These manipulations support that reducing inhibitory demands can improve children's CF performance.

1.5.2. The Graded Representations Account

The Graded Representations Account claims that perseveration on the DCCS arises from children's poor working memory (Morton & Munakata, 2002; Yerys & Munakata, 2006). This account suggests that poor working memory prevents children from appropriately updating the current sorting rule, leading to perseverative responding. In order to switch rules successfully, children need to update their mental representation of the current task rule from the pre-switch sorting dimension to the post-switch sorting dimension. To do that, the representation of the current rule needs to be sufficiently strong and active in mind to overwrite the representation of the previous rule. The Graded Representations Account suggests that during the pre-switch phase, children form a strong mental representation of that initial rule. After the rule changes, the current post-switch rule is not repeated as much as the pre-switch rule was, and thus is only weakly represented in memory. Thus, the *latent* trace of the previous rule is stronger than the *active* trace of the current rule – and this discrepancy leads to perseveration, when the weaker active rule cannot outcompete the stronger latent rule (Yerys & Munakata, 2006).

Evidence to support this account comes from studies in which researchers made manipulations to the DCCS task to target working memory (Morton & Munakata, 2002). For example, children do better on CF tasks when they are introduced a relatively unfamiliar values while sorting during the pre-switch phase (e.g., including greenish-blue or bluish-grey colours while sorting by colour), and are introduced a relatively familiar values while sorting during the post-switch phase (e.g., including circles or squares while sorting by shape) (Yerys & Munakata, 2006). This is likely to be because children's active representations arising from sorting with relatively familiar values during the post-switch phase (sorting by circles) could be stronger than their latent representations arising from sorting with relatively unfamiliar values during pre-switch phase (sorting by greenish-blue colour). Therefore, children's active representations while sorting during the post-switch phase in this task could compete with their latent representations during the pre-switch phase (Yerys & Munakata, 2006).

1.6. Recent Developments in Cognitive Flexibility

Recent CF research has focused on two key areas that have been helpful in advancing our understanding of how CF develops: the first is designing new CF tasks so that a wider range of switching behaviour can be examined, and the second is looking at CF across a wider age range. These two areas provide the motivation for the empirical work reported in this thesis. These approaches have advanced prior work which has focused very extensively on how children between the ages of three and four years are able to switch rules in the presence of response conflict. Although this approach has been convenient for many research questions – and has taught us a lot about some aspects of CF development – the full picture of how CF develops is likely to be more complex than this. Indeed, more recent work has shown that when we use tasks that allow for more diversity in how children can respond, and tasks that allow us to look at a wider age range, the picture that emerges is richer and more complex than previously thought.

In the following sections, I will focus on how recent research has advanced our understanding of CF. First, this research has used tasks designed to allow us to assess more varied forms of rule-switching behaviour. As we will see, results suggest that the extent to which children perseverate on CF tasks might have been overestimated. Second, this research has looked at CF in a wider and younger age range of children, by reducing incidental task demands. Results here suggest that CF might start to develop earlier than previously thought, with 2-year-olds being able to switch rules under some circumstances. These two areas are explained in more detail.

1.6.1. Designing CF tasks to assess wider range of behaviour

In most CF tasks, when the rule changes, children can make one of two possible responses: they can either switch successfully, or they can perseverate with the no longer relevant rule. This is because CF tasks typically present children with two response options: a correct response and an incorrect response. The correct response is the response that matches the target card according to the current rule, and the incorrect response is the response that matches the target card according to the previous rule (e.g.. Zelazo, 2006; Espy, 1997). An example is shown in Figure 1. After a rule change on the DCCS from colour to shape, the child must place the card depicting the blue star with the card depicting the red star, not with the card depicting the blue heart. The correct post-switch response is to sort by shape – so if the child places the blue star with the red star, then they make a correct response. If the child places the blue star with the blue heart, then they make an error, because sorting blue star with blue heart means matching the test card to the target card according to the previous rule. However, it is important to realise that because of the way these tasks are designed, the only error that it is possible to make on the task is to match the card by the previous rule – in other words, all errors will appear to be perseverative errors. Because most CF tasks have been designed in this way, it means that any errors that children make will necessarily appear to be perseverative in nature. Even if a child responded incorrectly because of inattention, or distractedness, or confusion, the error will still appear to be an error of perseveration.

However, evidence from less constrained switching tasks would seem to suggest that this is problematic – because when these tasks are used, children have been found to make errors for *other* reasons, such as inattention or distraction (Deák & Narasimham, 2003). For example, Chevalier and Blaye (2008) used the Preschool Attentional Switching Task-3 (PAST-3) where preschoolers needed to make an intradimensional switch (such as switching from selecting blue objects to selecting yellow objects). In the task, three objects with different colour and shape information were presented to children, and children needed to select one object based on a given rule. For example, when they were playing the *blue game*, children needed to select the object which is blue (e.g., among a blue cat, a yellow flower, and a green car). After playing blue game for several trials during the pre-switch phase, the rule changed from blue to yellow during the post-switch phase. When they were playing the *yellow game*, children needed to select the object which is yellow (e.g., among a yellow flower, a blue car, and a green cat). In the task, children needed to switch rules in the presence of both conflicting and distracting information. That is, during the post-switch phase (e.g., while playing the yellow game), children needed to switch rules in the presence of both conflicting information that was correct according to the previous rule (e.g., a blue car) and distraction information that was never relevant in the pre-switch phase or in the post-switch phase (e.g., a green cat). By using this kind of less constrained task design, the researchers were able to examine the incidence of not only perseverative errors, but also of distraction errors (Chevalier & Blaye, 2008). They found that 3-year-olds tended to make as many distraction errors as perseverative errors. This suggests that when switching tasks allow for more varied kinds of responses, children can and do make other types of switching errors (Chevalier & Blaye, 2008) – and, importantly, that tasks that measure only perseveration are likely to be missing important other kinds of behaviour. Consequently, there are strong grounds to think that the field may have overestimated the extent of perseveration in children's CF development.

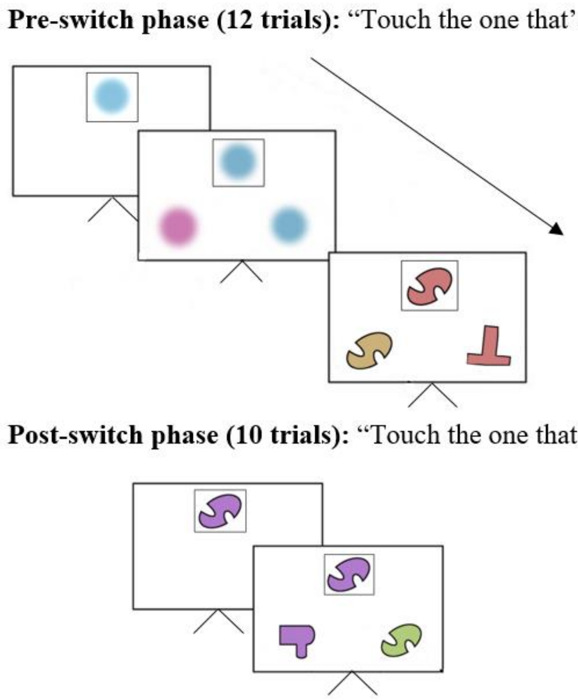
1.6.2. Designing CF tasks to assess wider age range of children

In recent years, tasks with reduced incidental demands have been developed, that enable us to examine CF in younger children. Most existing paradigms have task demands that mean they cannot be used with children younger than age three. For example, in addition to the core demand of switching rules, existing tasks typically require children to follow relatively lengthy instructions. They also require children to have sufficient selective attention to be able to attend to one dimension of a stimulus, while ignoring the other dimension (e.g., focusing on the shape of a blue heart, while ignoring its colour) (Carroll et al., 2016). These different task demands are *not* inherently part of CF, but they appear consistently in many CF tasks. This has had the effect of limiting CF research to children aged three years or older. This may mean, therefore, that early developments of CF in children younger than age three years have been missed.

The Switching, Inhibition and Flexibility Task (SwIFT) is a computerised rule-switching task that has been recently developed with reduced incidental demands; allowing us to compare CF in 2- to 4-year-olds (Blakey et al., 2016). In the SwIFT, children are asked to pick one of the two colourful shapes that matches a prompt stimulus according to the relevant dimension for a given trial (shape *or* colour; the rule changes mid-way through the task). The SwIFT requires children only to touch to an image on the screen; this response requirement is simple even for 2-year-olds. In addition, the language demands of the SwIFT are simple ("Touch the one that's the same [shape/colour]"), and thus are suitable for 2-year-olds. Furthermore, in the SwIFT, the stimuli used are initially very simple, and gradually increase in complexity, making it easy for 2-year-olds to understand the task (e.g., from a simple stimulus – a monochrome outline of a shape – to much more complex stimuli, combining shape and colour).

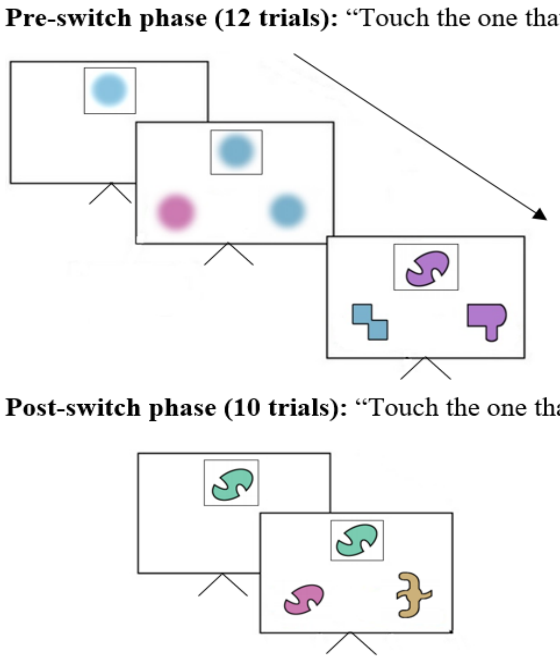
One important aspect of the SwIFT is that by using slightly different stimuli across different variations of the task, it is possible to separate out different task demands: for example, it is possible to separately assess switching when there is a response conflict, and switching when there is distracting information. This separation of previously conflated task demands allows researchers to study perseverative errors and distraction errors separately. On each trial, children are first presented with a prompt stimulus at the top; then, shortly after, two stimuli arrive under it (these are *correct response*, which matches the prompt on the current rule; and either a *conflicting response* or a *distractor response* which does not match the prompt on the current rule for different reasons). To illustrate this distinction, the figures below show an example of conflicting stimuli, and an example of distractor stimuli. In the Conflicting SwIFT, children have to switch rules while resolving response conflict (see Figure 2). That means after a rule change, children can either sort correctly, or they can continue to sort stimuli by using the previous rule (similar to the standard DCCS task). Conversely, in the Distracting SwIFT (see Figure 3), children have to switch rules while ignoring distracting, task-irrelevant information. That means after a rule change, children can either sort correctly, or they can select a stimulus unrelated to the task rules. They cannot continue to sort stimuli by the previous rule – in other words, it is not possible for them to make a perseverative error. They still need to update and maintain the currently relevant rule, but the to-be-sorted stimuli do not match the prompt stimulus on the previously relevant rule.

Figure 2. Conflicting SwIFT (from Blakey et al., 2016)



Note. Children are asked to touch to one of the stimuli at the bottom of the screen, following the prompt "Touch the one that's the same [colour/shape]". In the Conflicting SwIFT, the incorrect response in the post-switch phase matches the prompt stimulus on the previously relevant rule.

Figure 3. Distracting SwIFT (from Blakey et al., 2016)



Note. Children are asked to touch to one of the stimuli at the bottom of the screen, following the prompt "Touch the one that's the same [colour/shape]". In the Distracting SwIFT, the incorrect response in the post-switch phase does not match the prompt stimulus on the previously relevant rule.

When the performance of 2- to 4-year-old children was examined using these different versions of the SwIFT, an interesting developmental trajectory was found: switching successfully while resolving response conflict (as in the Conflicting SwIFT) and switching successfully while suppressing distracting information (as in the Distracting SwIFT) developed at different ages, and were supported by different executive functions (Blakey et al., 2016). Specifically, children become able to switch rules while resolving response conflict between the ages of three and four years, and that development was supported by children's working memory. Conversely, children become able to switch rules while ignoring distracting information between the ages of two and three years, and that development was supported by children's inhibitory control (Blakey et al., 2016). In addition, 2-year-olds tended to respond unsystematically after the rule change, by neither perseverating reliably nor switching reliably. This kind of performance, may reflect an absence of top-down control, indicating that this has yet to emerge in 2-year-olds.

One consequence of assessing CF with tasks that assess multiple kinds of error is that the data tends to be richer and more complex than on standard CF tasks. To analyse data of this kind, simply counting up the number of correct responses after a rule change would not be sufficient, since it would fail to account for valuable information about the different kinds of error children might make. Instead, techniques that allow us to identify diverse patterns of performance within the data set may offer a better way to explore the development of CF (Carroll et al., 2016; Deák, 2003). "Patterns of performance" describes switching behaviours that are *qualitatively* different from each other (van Bers, Visser, van Schijndel, Mandell, & Raijmakers, 2011). To identify distinct patterns of performance, advanced model-based analyses are required – analyses in which children’s performance is observed over the entire course of the post-switch phase. Based on the stability or variability of performance during the post-switch phase, distinct patterns of performance can be identified. For example, patterns of performance might include switching successfully across all post-switch trials; perseverating repeatedly across post-switch trials; alternating haphazardly between switching and perseverating; or perseverating initially, followed by switching successfully part way through the post-switch trials.

To the best of my knowledge, there are only three studies in the literature that use advanced statistical techniques to examine patterns of performance in CF. These studies raise the possibility that much more diverse patterns of performance can be observed, above and beyond simply switching or simply perseverating. One of those studies identified three distinct performance patterns in 3- to 5-year-olds, by using a model-based analysis to look at post-switch behaviour performance on the DCCS (van Bers et al., 2011). Researchers used Latent Markov models to examine trial-by-trial performance across the post-switch phase. Three patterns of performance were identified. One pattern reflected children who switched correctly by sorting by the current rule in the post-switch phase. A second pattern reflected children who perseverated with the previous rule. And a third pattern was also found, reflecting children who inconsistently alternated between perseverating and switching during the post-switch phase.

This view is broadly consistent with other research identifying *five* distinct CF performance patterns in 5- to 6-year-olds (Dauvier, Chevalier, & Blaye, 2012). Five and 6-year-olds were tested on the age-appropriate adaptation of the DCCS task (Dauvier et al., 2012). Finite-Mixture General Linear Models were used to examine the number of performance patterns children produced across the post-switch phase. Five distinct performance patterns were found. In addition to (i) children who switched rules successfully and (ii) children who perseverated with the previous rule, researchers also identified (iii) children who performed apparently randomly, by mostly perseverating and but by also sometimes switching rules correctly by chance (referred to as the “random switch” class); (iv) children who changed their behaviour *halfway* through the task, from switching to perseverating (referred to as the “strategy change” class); and (v) children who frequently alternated between successful switching and perseverating during the post-switch phase (referred to as the “mainstream” class) (Dauvier et al., 2012).

More diverse patterns of switching performance were also detected in a younger sample of 2- to 4-year-olds (Blakey et al., 2016). Researchers used two versions of the SwIFT with contrasting task demands. In the conflicting version, children had to switch rules in the presence of response conflict. In the distracting version, children had to switch rules while ignoring task-irrelevant, distracting information. Latent Markov Models were used to identify patterns of performance, by examining children’s trial-by-trial performance across the post-switch phase. Across different versions of the SwIFT, children demonstrated a total of three distinct performance patterns. In the conflicting SwIFT, three patterns of performance were found: (i) children who switched rules successfully, (ii) children who perseverated with the previous rule, and (iii) children who alternated between selecting the correct response and the conflicting response (referred to as “mixed responders”). This mixed responding performance was thought to reflect an absence of top-down control. In the distracting SwIFT, two patterns were reported: (i) children who switched rules successfully, and (ii) children who alternated between selecting the correct response and the distracting response (known as “mixed responders”). Children who showed the mixed responding performance were more likely to be young than children who showed successful rule switching.

In summary, the use of more advanced statistical techniques to look at CF development has shown that children's performance is more nuanced than that can be captured by simply counting the number of correct responses that children make. Children do not simply sort correctly or perseverate incorrectly when asked to flexibly follow a rule change. In fact, children may also show performance that is either reflective of transitioning between perseveration and switching, or reflective of an absence of any top-down control. Further research using this approach – and tracking how it aligns with the age of children – may help us to understand more comprehensively how CF emerges and develops.

1.7. Future Research Directions

Overall, this review has highlighted that children begin to show precursors to CF from infancy, after which CF develops rapidly during the preschool period. Much previous research has focused on studying the shift from perseveration to switching that occurs at around the three-year to four-year transition. This work has greatly improved our understanding of how CF develops. However, this period has arguably received so much attention more because of methodological convenience, than because this is when all CF development really takes place. More recent advances have shown that there are two particularly important gaps in how CF has been studied – gaps that might be sufficiently serious that our understanding of how CF develops needs to be revised.

First, many existing CF tasks have constrained the kinds of behaviour children can produce. That means that children can often only respond in one of two ways: either they switch rules successfully (by selecting the correct response), or they make a perseverative error (by selecting the incorrect response) (e.g., Espy, 1997; Zelazo, 2006). With these tasks, all errors children make will appear to be perseverative errors by definition, and any failure to switch rules will appear to be because children have problems overcoming perseveration. Performance on this kind of CF task can tell us nothing about errors that might happen due to inattention or distraction (Carroll et al., 2016; Deák, 2003). However, recent research has shown that errors of inattention or distraction do occur, and that overcoming these errors is likely to be an important aspect of CF development (Blakey et al., 2016; Chevalier & Blaye, 2008). We still know very little about how these two types of error – perseverative errors and distraction errors – co-occur during the development of CF. This is because tasks tend to involve either a response conflict *or* distracting information, but never both at the same time (though see Chevalier & Blaye, 2008). Therefore, it is important to develop new rule-switching tasks with both conflicting information and distracting information on the same sorting array that would allow children to show a greater variety of performance, and would allow researchers to better track the development of CF in all its facets.

Second, most existing CF tasks have incidental task demands that restrict the age of children that can be studied. Existing switching tasks are generally too complex to be used with children younger than three years of age. However as recent research has suggested, there appear to be important developments in CF happening prior to age three (Blakey et al., 2016). Therefore, early developments of CF in 2-year-olds may have been missed because of a lack of age-appropriate paradigms. It may be that further work with younger children may require us to revise the apparent consensus view that the first emergence of CF is when children become able to overcome perseveration between the ages of three and four years. By broadening the age span that we study, it may well be that we will better understand how CF emerges, and thus be better able to accurately characterise how it develops.

These two gaps will be examined in the following experimental studies in Chapter two and Chapter three of this thesis.

Chapter Two

How Does Cognitive Flexibility Develop in Early Childhood?

Cognitive flexibility has typically been assessed in preschool children using tasks where children sort complex stimuli first by one rule, and then by another, between the ages of three and four years. Over many studies it has been established that 3-year-olds tend to perseverate when the rule changes, whereas 4-year-olds are able to successfully switch rules. Consequently, theoretical accounts of CF development have solely explain how children overcome perseveration. Although this view has been very influential, recent research has suggested that this view may mischaracterise the development of CF. In this study, I aim to address this by examining two gaps in previous research: i) many preschool CF tasks have constrained the kinds of response children can produce so they can only perseverate or respond correctly and ii) many preschool CF tasks can only be used with 3-year-olds and older due to incidental task demands. By addressing these two gaps, the present study aimed to better characterise the development of CF. One hundred and eighty-three 2- to 4-year-olds performed on a novel rule-switching task that allowed children from a wider age range to produce more varied kinds of switching responses. The results showed that CF has a more protracted developmental trajectory, that starts earlier than previously thought. Specifically, CF comprises three distinct stages of increasing top-down control between the ages of two and four years. The results also suggest that perseveration, rather than being the main problem to be overcome in CF, is better considered as a developmental milestone on the way to flexible switching.

2.1. Introduction

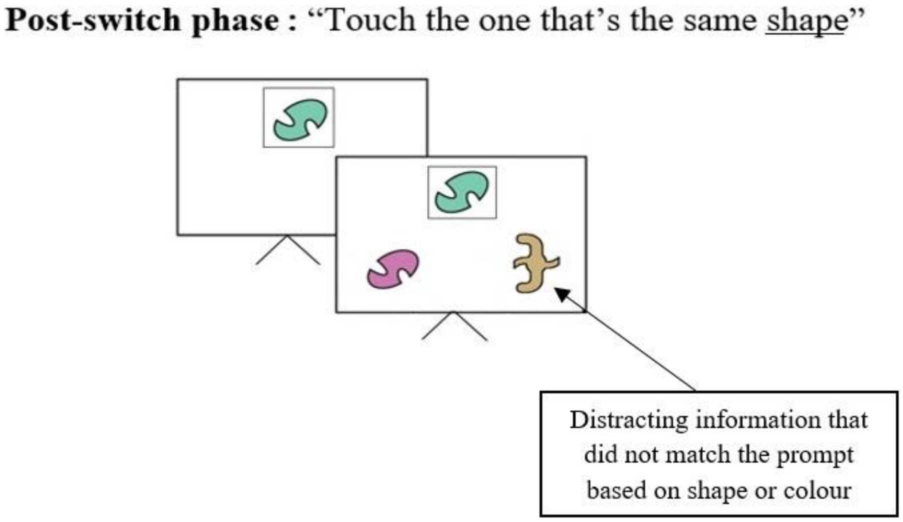
Cognitive flexibility has been typically thought as developing between the ages of three and four years, when children become able to sort complex stimuli first by one rule, and then by another (Diamond et al., 2005; Muller et al., 2006). For example, on the DCCS, preschoolers are asked to sort cards depicting colored shapes into two trays, first by one rule during the pre-switch phase (e.g., colour), then by another rule during the post-switch phase (e.g., shape) (Zelazo, 2006). Most preschoolers sort the cards successfully during the pre-switch phase. However, after the rule changes, most 3-year-olds continue to sort according to the previous rule – and therefore they perseverate incorrectly. In contrast, most 4-year-olds are able to switch rules successfully (Zelazo, 2006). This shift from perseverating incorrectly to switching successfully between the ages of three and four years has been presented as one of the key milestones in the development of CF (Munakata et al., 2012).

The transition from perseveration to successful switching during the preschool years has been influential in shaping theoretical accounts of CF, and in making testable predictions for how CF develops (Diamond & Kirkham, 2005; Munakata, 2001). However, some recent research indicates that the principal focus on perseveration may mischaracterise how CF emerges and develops (Carroll et al., 2016). There are two key gaps in the current literature that may explain why researchers have tended to focus on perseveration. The first gap is that most studies on the development of CF have used tasks that constrain the kinds of response children can produce. As a result, children can only respond in one of two ways: either switching rules correctly, or perseverating incorrectly. Consequently, accounts of CF have been based on explaining only these two patterns of response. The second gap is that most existing CF tasks, due to their design, can only be used with children between the ages of three and four. As a result of focusing only examining on this restricted age range, earlier competencies in CF may have been ignored. In the sections that follow, these gaps are explained in more detail.

The first gap is that most existing CF tasks have constrained the kinds of response children can produce. This has led children’s performance to be indexed in one of only two ways when the rule changes: they either switch correctly, or they perseverate incorrectly. No other kind of error is possible, which means that these tasks leave us unable to distinguish between cases of genuine rule perseveration and errors arising from other factors, such as inattention, impulsivity, or distractibility. For example, on many measures of CF (including the DCCS task and the Shape School task) when the rule changes, children can respond in one of two ways. They can either switch rules successfully by selecting the correct response, or they can make a perseverative error by selecting the response that matches according to the previous rule (Espy, 1997; Zelazo, 2006). This happens because on most CF measures, after the rule changes, children have to sort stimuli in the presence of response conflict. The response conflict occurs when the stimuli that children have to sort (e.g., a coloured shape) contain properties that prompt *both* the previous *and* the current rule on a given trial. In other words, the stimuli a child has to sort matches the response options on the task according to both colour *and* shape. Therefore, children can match the stimulus by the current rule (e.g., colour) or continue to match it according to the previous rule (e.g., shape) – both are potentially relevant according to the rules of the task, and no matter how the child responds, they will be following either the current rule or the previous rule. Consequently, any errors that children make will necessarily be perseverative errors, since the tasks do not allow children to make any other kind of error.

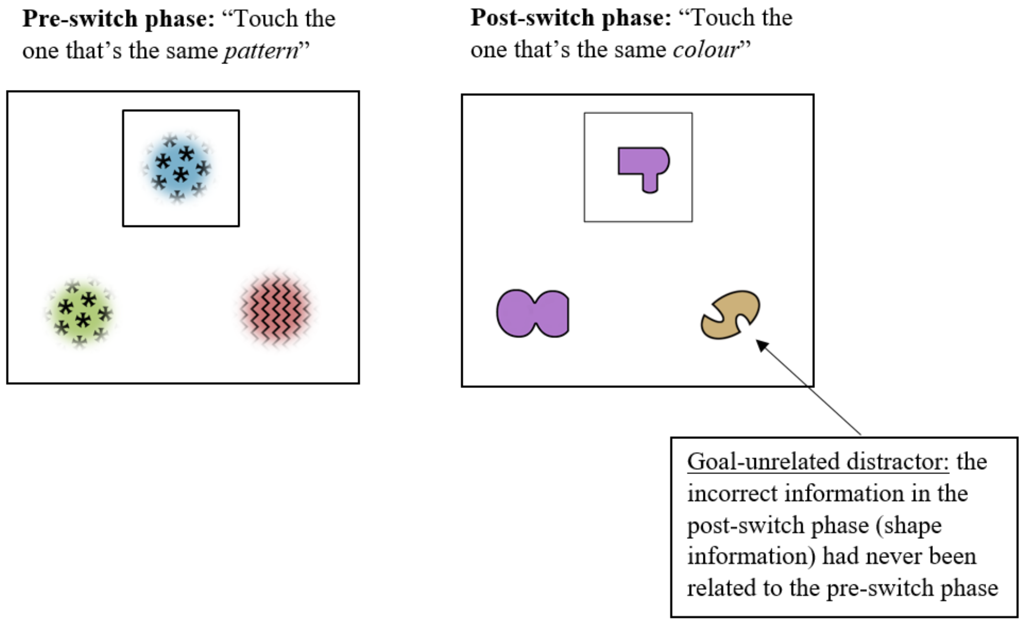
However, switching errors might occur for other reasons as well, such as distraction or inattention. Recent research has been able to examine the frequency of *other* kinds of error, through subtle experimental manipulations that change the type of information presented on the sorting array (Blakey et al., 2016; Chevalier & Blaye, 2008). For example, Blakey and colleagues (2016) used the SwIFT to look at different kinds of error. This rule-switching task involved children matching a coloured shape to one of two other coloured shapes according to a rule (e.g., colour) and then after several trials, switching to sorting them by a different rule (e.g., shape). In one task version (the Distracting SwIFT), children were presented with a sorting array where in the post-switch phase they could sort the stimuli correctly according to the current rule (e.g., during the shape game, the stimuli matched on shape), or they could sort incorrectly by selecting a stimulus that did not match on either the current rule *or* the previous rule (i.e., the incorrect response did not match on shape or colour). Instead, the stimuli had distraction shape or colour information that did not match any rules on the task (see Figure 4 for an example). A key feature of this condition was that children did not need to resolve response conflict after the rule changed. It was found that even when response conflict was entirely removed, while most children found it easier, many children still made rule-switching errors (so called “distraction errors”). Similarly, Chevalier and Blaye (2008) demonstrated that preschoolers can make rule-switching errors through distraction. In their study, researchers used the PAST-3 to explore whether children make switching errors for other reasons than only perseverating with the previous rule. This rule-switching task required children to make an intradimensional switch in the post-switch phase (e.g., switching from selecting blue objects to selecting yellow objects). In the task, children were required to select one of the three objects that matched the prompt according to the relevant colour. After the rule changed on the PAST-3 from sorting by blue to yellow, children could make a correct response by selecting the yellow object (e.g., yellow flower), or children could make a perseverative error by selecting the blue object whose colour was previously relevant (e.g., blue car), or they could make a distraction error by picking the green object whose colour was never relevant (e.g., green cat). The results showed that 3-year-olds made distraction errors with a similar proportion as they made perseverative errors. Together, these results suggest that when given the opportunity to, children can make errors that are not perseverative in nature, and instead arise through distraction.

Figure 4. Example of distracting information in the Distracting SwIFT (from Blakey et al., 2016)



Researchers have suggested that distraction errors are likely to occur because children fail to update the current sorting rule, perhaps due to constraints in working memory (Blakey & Carroll, 2018). This is evidenced by the recent research which assessed CF with tasks that allow children to make different kinds of distraction errors by varying the relevance of the distracting information presented in the stimuli. In their study, Blakey and Carroll (2018) assessed 2- to 3-year-olds' CF with a rule-switching task that involved four conditions that varied the level of distracting information but did not involve response conflict. These conditions differed in terms of the nature of information children need to ignore after the rule changed: baseline condition, goal-unrelated condition, goal-related condition, goal-related reactivation condition. In the *baseline condition*, children sorted by a single rule throughout the task (for example, they sorted by colour while ignoring shape information during the pre-switch phase *and* the post-switch phase). In other words, children were not required to switch rules in this condition. In the *goal-unrelated condition*, the incorrect information in the post-switch phase (shape information) had never been relevant in the pre-switch phase (e.g., children sorted by pattern while ignoring colour during the pre-switch phase; and then they sorted by colour while ignoring shape during the post-switch phase). Example stimuli for the goal-unrelated distractor can be found in Figure 5. However, in the *goal-related condition*, the incorrect information in the post-switch phase (shape information) was the information that had been related in the pre-switch phase (e.g., children sorted by shape while ignoring pattern during the pre-switch phase; and then they sorted by colour while ignoring shape during the post-switch phase). Example stimuli for the goal-related distractor can be found in Figure 6. Finally, in the *goal-related reactivation condition*, the incorrect information in the post-switch phase (shape information) was the information that children sorted by in the pre-switch phase (e.g., children sorted by shape while ignoring colour during the pre-switch phase; and then they sorted by colour while ignoring shape during the post-switch phase). The results showed that children made more distraction errors in the goal-related condition, where they were switching rules in the presence of goal-related distractors than in the goal-unrelated condition where they were switching rules in the presence of goal-unrelated distractors. Children made no additional errors in the goal-related reactivation condition, suggesting that re-activating a previously ignored stimulus was not the source of children's difficulties. In summary, the results suggest that children do not randomly make errors when switching in the presence of distraction (for example, because they do not understand the task). Furthermore, they do not make distraction error because they struggle to reactive previously ignored stimuli. Instead, children make distraction errors because the information related to the previous task goal continues to influence children's performance after the rule changes (Blakey & Carroll, 2018).

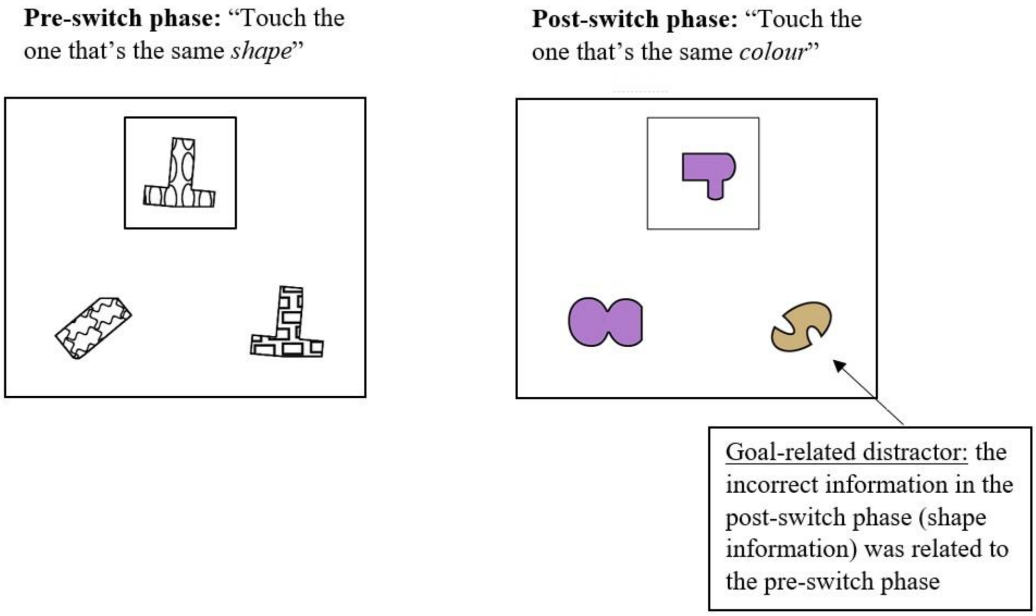
Figure 5. Example goal-unrelated distractor stimuli (from Blakey & Carroll, 2018)


Pre-switch phase: “Touch the one that’s the same *pattern*”

Post-switch phase: “Touch the one that’s the same *colour*”

Goal-unrelated distractor: the incorrect information in the post-switch phase (shape information) had never been related to the pre-switch phase

Figure 6. Example goal-related distractor stimuli (from Blakey & Carroll, 2018)


Pre-switch phase: “Touch the one that’s the same *shape*”

Post-switch phase: “Touch the one that’s the same *colour*”

Goal-related distractor: the incorrect information in the post-switch phase (shape information) was related to the pre-switch phase

Together, these few studies that have been conducted using tasks that allow for different kinds of error support the idea that overcoming distraction may also be a key milestone on the way to developing CF. If that were the case, then the field's heavy focus on explaining perseveration may have missed other important CF developments. However, to be able to judge how important a skill overcoming distraction is in the development of CF, we first need to know how common distraction errors are. At present, we know very little about the prevalence of different kinds of error when children are free to make multiple types of responses on the *same CF task*. This is because previous tasks tend to involve *either* requiring children to overcome response conflict, *or* requiring them to overcome distracting information when the rule changes – but crucially, usually never both at the same time (though see Chevalier & Blaye, 2008). Therefore, a new rule-switching task is needed that involves both conflicting information *and* distracting information within the same sorting array.

The second gap with existing research is that most existing CF tasks, due to their incidental task demands, can only be used with a narrow age range of children – typically between the ages of three and four. As a result, early competencies in CF may have been ignored. For decades, researchers have mainly focused on how CF develops in 3- to 4-year-olds. This narrow focus is understandable, given that the incidental demands of most rule-switching tasks make them very complex to use with younger than 3-year-olds. In particular, many rule-switching tasks have high language demands and high selective attention demands that make them challenging for young children. For example, the DCCS requires children to follow quite lengthy verbal instructions. In addition, it requires children to selectively focus on a single dimension of a bivalent stimulus (e.g., colour), while ignoring the other dimension (e.g., shape). However, these incidental demands are likely to be challenging in their own right, and may contribute to young children's difficulty on these tasks. If difficulty on CF tasks comes from incidental demands, rather than from the core requirement of flexibly updating one's own sorting behaviour, then early competencies in CF in children younger than age three years may be overlooked, and the development of CF itself may have been misrepresented (Carroll et al., 2016).

There is empirical evidence to support this assertion. Recent research has suggested that when incidental demands of CF tasks have been reduced, even 2-year-olds are capable of showing basic CF skills, by successfully sorting in the presence of distracting information (Blakey et al., 2016). This is an intriguing finding, because it suggests that CF may emerge earlier than previously thought. However, only two studies to date have examined this in 2-year-olds (Blakey et al., 2016; Blakey & Carroll, 2018), and therefore, we still know very little about how this early form of CF develops – or even if this finding itself is robust and generalisable. Therefore, if we are to gain a more comprehensive understanding of how CF emerges, we need more research using tasks with reduced incidental demands, in order to allow us to study CF from an earlier age.

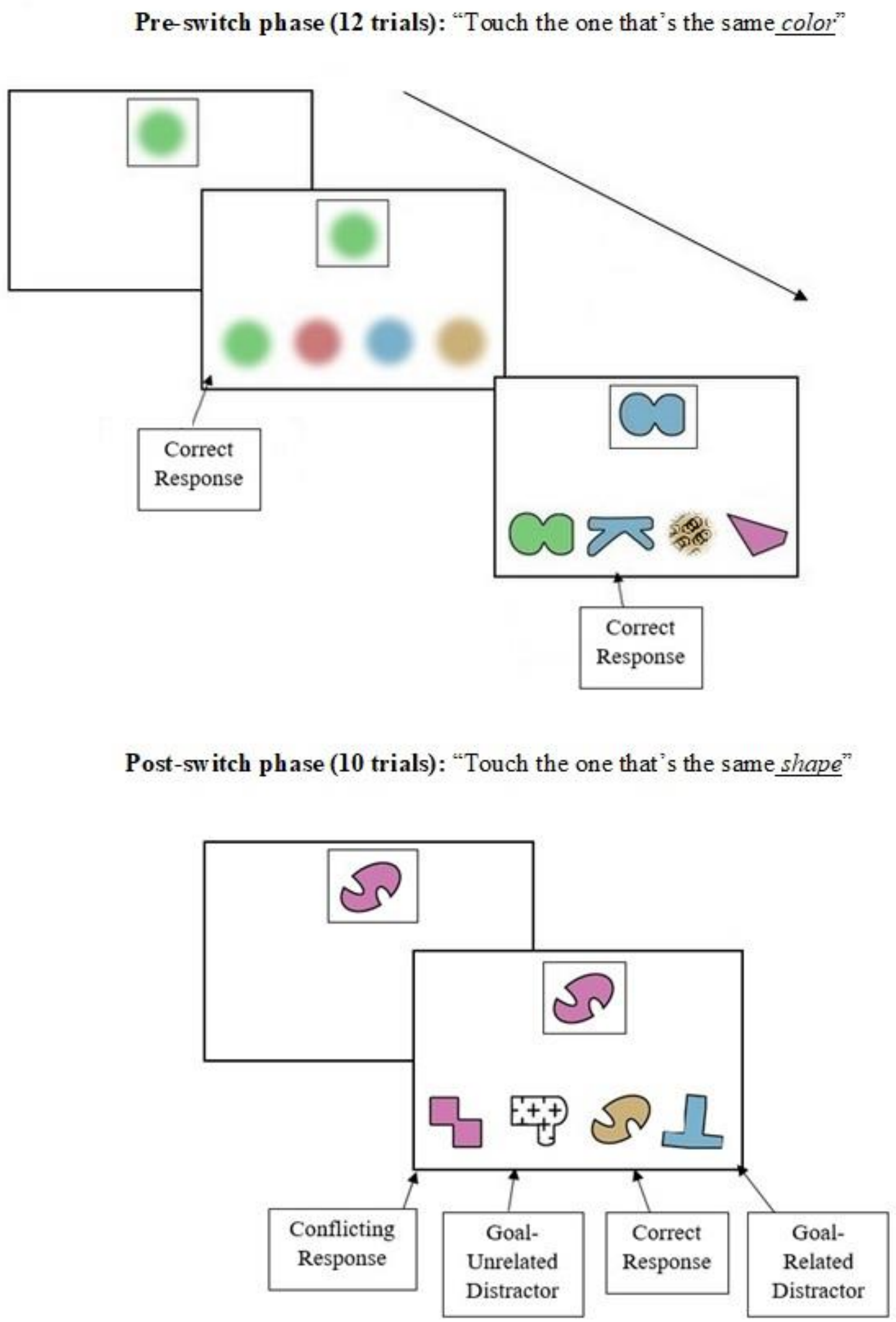
The combination of these two gaps means that previous ways of assessing CF may well have mischaracterised how CF develops. In particular, the idea that CF development represents a clear-cut transition between the ages of three and four years from perseveration to switching may be misleading. If we were able to broaden the kinds of response that children can produce on CF tasks, and were able also to widen the age range of children that can complete the tasks, then it may be that we will find that young children are capable of more diverse forms of switching performance – and that CF itself has a far more nuanced developmental trajectory than previously thought.

The current study aimed to address these gaps, with the goal of better characterising how CF develops. For the current study, we developed a new rule-switching task that i) allows us to look at multiple kinds of error on a single task, and ii) can be used with children between the ages of two and four years. The Switching, Inhibition, and Flexibility Task-4 (SwIFT-4) is an adapted version of the SwIFT (Blakey et al., 2016), renamed to reflect the fact that there are four possible response options on every trial. The task was presented on a touchscreen computer and required children to sort colourful stimuli according to different rules. In the SwIFT-4, the child had to decide which of four stimuli on the sorting array matched a prompt stimulus, first by the pre-switch rule (e.g., colour), and then by the post-switch rule (e.g., shape) (as shown in Figure 7). Cognitive flexibility was assessed by examining children's responses during the post-switch phase. The SwIFT-4 task array features both conflicting information and two different kinds of distracting information. Thus, on any post-switch trial, children could make four different kinds of response: they could select the correct response; they could make a perseverative error, by continuing to sort according to the previous rule; or they could make one of two types of distraction error, by selecting a stimulus that does not match the prompt stimulus by either of the task rules. The SwIFT-4 has minimal motor and language demands (the child follows the prompt instruction "Touch the one that's the same colour/shape"), and the task requires the child to make a straightforward response of touching the picture that they want to select.

In this study, I aimed to use this new task to examine how CF develops in 2- to 4-year-olds. However, given the greater variety of responses that the SwIFT-4 can assess, to analyse task performance it would not be sufficient to simply adopt the analyses used in earlier research. Most previous studies have indexed children's CF performance simply by counting the number of correct sorting responses children made after the rule changed. This is a reasonable thing to do on those tasks, because the most common patterns of performance are for preschoolers either to sort most cards correctly (scoring around 100%) or to perseverate incorrectly (scoring around

0%). In other words, when children scored numerically low on switching tasks, researchers interpreted this as perseveration, and when children scored high on switching tasks, researchers interpreted this as successful switching. (There are some criticisms one could make of this approach, though they are beyond the focus of the present discussion). However, the present study uses a task where there are *four* possible responses the child could make, in order that we can identify different kinds of switching error. Therefore, it is likely that children will produce a broader range of responses, such that simply counting the number of correct switches will miss out potentially informative information. So to analyse these data, it will be necessary to use an analytic approach that can look at the proportions of different kinds of response children make, rather than simply noting how many correct sorts they make after a rule change. Importantly, it is perfectly conceivable that children may make multiple kinds of error across a single iteration of the task. For example, rather than consistently perseverating *or* showing distraction errors, children could make *both* kinds of error, in different proportions. By looking at the different patterns of performance that may arise on the task, I aim to be able to offer a more finer-grained picture of how CF develops in early childhood.

Figure 7. Response options in the pre- and the post-switch phase of the SwIFT-4



Note. Figure 7 shows the pre- and post-switch phase of the SwIFT-4. Children were first presented with the prompt stimulus (a colourful shape) and asked to find the colourful shapes that matched on either colour or shape. Children completed 12 trials on the pre-switch phase and then they completed 10 trials on the post-switch phase. In each trial, children had to make a single response. In the post-switch phase, children were presented with a correct matching option, an option that matched according to the pre-switch rule (conflicting response), and two options that contained different amount of distraction information and that did not match on the current rule nor the previous rule (goal-related distractor and goal-unrelated distractor). Rule order was counterbalanced such that half of children started with colour rule and half started with shape rule.

One benefit of this new task is that by including multiple kinds of response option, it was possible to identify qualitatively different patterns of CF performance across different children. Therefore, comparing these patterns across age groups may offer a new perspective on how CF develops. It is plausible that distinct patterns of CF performance may be observed, as previous research suggests that when CF tasks are less constrained and have more response options, more varied performance patterns are seen, beyond simply perseverating or switching. For example, recent research – using either novel CF tasks or more advanced statistical techniques to examine performance across the entire post-switch block – have identified multiple performance patterns in children's CF performance, including unsystematic mixed responding by selecting haphazardly from all available response options (Blakey et al., 2016, Dauvier et al., 2012), and children fluctuating back and forth between sorting correctly and perseverating on the post-switch block (van Bers et al., 2011). Based on this previous research, it is possible that different children may demonstrate distinct performance patterns on the SwIFT-4. This could involve children sorting correctly by selecting the correct response; or children consistently perseverating with the previous rule; or children alternating between sorting correctly and sorting perseveratively; or children unsystematically selecting from all possible response options.

In order to accurately characterise the different performance patterns that children might show, advanced analysis techniques are required to identify how many performance patterns are observed that can maximally explain the largest amount of variance on the task. One promising technique in this regard is Latent Class Analysis (LCA; Williams & Kibowski, 2016). The main principle of LCA is to identify patterns – or *classes* – that best explain a cognitive process in relation to multiple variables in a data set. It operates by using a number of different models to try to explain variance in a data set, and by comparing these models with each other. The models vary in terms of the number of classes they use to explain variance. Latent Class Analysis, therefore, is able to identify the optimal model – that is, the one that explains the process with the least unexplained variance. In the current study, I will use LCA to determine the optimal number of performance patterns in order to explain CF performance on the SwIFT-4. After finding the optimal number of performance patterns to account for variance in the data, I will also examine how these performance patterns vary by age. By comparing the performance patterns across ages (in 2-year-olds, 3-year-olds and 4-year-olds), this will offer a new way to characterise how CF develops in the early years.

An additional question that this study will address is to what extent individual differences in executive functions – specifically, working memory and inhibitory control – predict children's CF performance. A growing number of studies have suggested that working memory and inhibitory control may underpin successful rule switching (Cragg & Chevalier, 2012; Kirkham et al., 2003; Morton & Munakata, 2002). That is, working memory and inhibitory control are thought to help children to better maintain and update the current sorting rule, and also to better suppress the previously relevant rule. The fact that previous research has focused on examining the role of executive functions in overcoming perseveration is understandable – because in many paradigms, the only alternative to switching is perseverating with the previous rule. However, in the current study, the SwIFT-4 allowed children to produce more varied kinds of responses. Therefore, this provides a good opportunity for us to examine how executive functions might contribute to distinct patterns of performance on a new and more complex CF task, as well as seeing how executive functions might contribute to changes in patterns of performance across age groups. For example, emerging research has suggested that executive functions may make specific contributions to CF, depending on the kinds of information children are presented with when they are switching (Blakey et al., 2016). In this study, children's working memory was related to switching in the presence of *conflicting* information; whereas inhibitory control was related to switching in the presence of *distracting* information. This is a promising study in terms of differentiating the roles of executive functions in more diverse forms of CF. However, in the study by Blakey et al. (2016), children were only asked to switch rules *either* in the presence of conflicting information *or* in the presence of distracting information. Crucially, these distinct demands were never assessed together. Furthermore, previous studies have tended to just use one measure for working memory and inhibitory control. In order to account for the issue of task impurity – specifically, executive function tasks that on their own may not accurately capture the intended measure of interest – it is better to take multiple measures of a given construct. Therefore, the current study uses multiple measures to assess working memory and inhibitory control, to explore how executive functions contribute to CF when children are asked to switch rules in the presence of both conflicting information and distracting information on the same sorting array.

In summary, the present study aimed to better characterise the development of CF, by using a novel task switching measure that allowed children from a wider age range to produce a wider variety of switching responses. To shed further light on how CF develops, the study also tested how individual differences in executive functions contribute to CF. To assess children's CF, I used a new task – the SwIFT-4 – to examine CF in a sample of 2- to 4-year-olds. To assess working memory, the Spatial Span task (Muller, Liebermann-Finestone, Carpendale, Hammond, & Bibok, 2012) and the Spin the Pots task (Hughes & Ensor, 2007) were used. To assess inhibitory control, separate tasks were used for 2.5- and 3-year-olds, and for 3.5- and 4-year-olds, because there was no single task to assess inhibitory control across 2- to 4-year-olds. Specifically, the Reverse Categorization task (Carlson et al., 2004) and the Fruit Stroop task (Kochanska, Murray, & Harlan, 2000) were used for 2.5- and 3-year-olds, and the Black and White task (Simpson & Riggs, 2005) and the Peg Tapping task (Diamond & Taylor, 1996) were used for 3.5- and 4-year-olds. For the purposes of age-related analyses, children were divided into four age groups using six-month age bands (see also Blakey et al., 2016): 2.5-, 3-, 3.5-, and 4-year-olds.

2.2. Method

2.2.1. Participants

One hundred and eighty-three typically developing 2- to 4-year-old children (96 males, 87 females) were recruited from volunteer local families and nurseries. Data from a further 18 children were excluded: Eleven 3.5-year-olds and three 4-year-olds failed to understand the task instructions, and four 2.5-year-olds did not complete a full set of test trials. In terms of sample size, in line with recommendations for carrying out LCA, it is considered necessary to recruit at least 100 participants (Nylund, Asparouhov, & Muthén, 2007). To address the follow-up question of whether there are differences in age and executive functions across performance, a two-way Analysis of Variance (ANOVA) was required, with age and performance pattern as independent variables and working memory and inhibitory control scores as dependent variables. Previous work indicated that it would be reasonable to expect three performance patterns (Blakey et al., 2016). Based on this estimate, the power calculation found that at least 158 participants would be required for 80% power to detect a medium effect in a two-way ANOVA (performance pattern with three levels and age group with four levels) (with an alpha = .05 and df = 2 for performance pattern and df = 3 for age group) (G*Power 3.1.9.2; Faul, Erdfelder, Lang, & Buchner, 2007).Therefore, the sample size of the current study (*N* = 183) was considered adequate to meet the objectives of the study.

The sample was divided into four age groups: 2.5-year-olds (*N* = 47, M_{age} = 2;8 years, range = 2;5 years - 2;11 years), 3-year-olds (*N* = 42, M_{age} = 3;2 years, range = 3;0 years - 3;5 years), 3.5-year-olds (*N* = 50, M_{age} = 3;8 years, range = 3;6 years - 3;11 years), and 4-year-olds (*N* = 44, M_{age} = 4;2 years, range = 4;0 years - 4;5 years). Children were predominantly White British and were monolingual. The study received ethical approval from the department's ethics committee, and informed consent and verbal assent were obtained from all participating parents and children respectively. Participating children received a small gift for taking part in the study. Testing took place between June 2018 and January 2019.

2.2.2. Procedure

First, children played a warm-up game in which their understanding of the words used in the task rules was checked; such as "shape", "colour", "different", or "same". Once their understanding of the words was ensured, then children completed five tasks, in a single testing session that lasted around 20 minutes. Specifically, 2.5-year-olds and 3-year-olds completed the following tasks (in this order): the SwIFT-4, the Reverse Categorization task, the Spatial Span task, the Fruit Stroop task, and the Spin the Pots task. Three-and-a-half-year-olds and 4-year-olds completed the following tasks (in this order): the SwIFT-4, the Black and White task, the Spatial Span task, the Peg Tapping task, and the Spin the Pots task.

2.2.3. Measures

Warm-up Game. Because 2-year-olds have limited verbal ability, it was possible that 2-year-olds could not understand the task instructions and that they might perform poorly on the tasks due to their limited vocabulary. Therefore, it was important to check whether 2-year-olds understand the words like “shape”, “colour”, “different”, or “same”. To check children’s (particularly 2-year-olds’) understanding of the words used in the tasks, a short warm-up game was used. In this game, the experimenter placed cards with different colourful shapes in front of the child. Then the experimenter asked the child to find the cards with the same colour and with the same shape. If children matched the cards in terms of their colour and their shape, then they continued to play with the rule-switching task.

2.2.3.1. Assessing Cognitive Flexibility

Switching, Inhibition, and Flexibility Task-4 (SwIFT-4). To assess CF, the SwIFT-4 was used. This task was a new adaptation of a task previously used by Blakey et al., (2016), though with four response options on each trial, rather than two. The task was created in E-Prime Psychology Software version 2 (PST, Pittsburgh, PA) and was administered on a touchscreen computer. It required the child to sort colourful stimuli either according to their colour or their shape; firstly during a pre-switch phase, and then, following a rule change, during a post-switch phase. On every trial, a prompt stimulus appeared, showing a colourful shape. Then four possible response options appeared, each showing a different colourful shape, only one of which matched the prompt stimulus according to the current rule. Midway through the task, the sorting rule changed. Children had to decide which of four response options on the sorting array matched the prompt stimulus according to the current rule.

The crucial novel feature of the task was that on each trial, the child was presented with *four* possible response options (Figure 7 shows an example trial). On every trial, one response option was the correct response – that is, the response that matched the prompt stimulus according to the current rule. The remaining three response options were all incorrect – they did not match the prompt stimulus according to the current rule – though they were importantly different from each other. On each trial, there was one conflicting response, and two types of distracting response. The conflicting response matched the prompt stimulus according to the *previously* relevant rule – that is, the rule that was correct during the pre-switch phase. For example, as shown in Figure 7, while sorting stimuli by *shape* rule during the post-switch phase, the conflicting response continues to match the prompt stimulus according to the previous *colour* rule, therefore it creates response conflict. The child would be likely to select a conflicting response if they were unable to update or to maintain the current task rule.

The two distracting responses, conversely, did not match the prompt stimulus according to either the current rule or the previous rule. Therefore, they did not create any response conflict. However, they did contain various degrees of distracting information that could potentially influence switching behaviour (Blakey & Carroll, 2018). For example, the *goal-related distractor* did not match the prompt stimulus either by the current shape rule or by the previous colour rule. However, it still had information from a dimension *(colour)* that was related to the previous rule. In contrast, the *goal-unrelated distractor* also did not match the prompt on either the current or the previous rule. However, it had information from a dimension (*pattern*) that was not related to the previous rule. Neither of the two distracting responses matched the prompt stimulus by either rule, so the child would only be likely to select one of these responses if they failed to maintain either the current rule or the previous rule.

The task consisted of three phases: a practice phase (of four trials), a pre-switch phase (of 12 trials), and a post-switch phase (of 10 trials). In the practice phase, the child sorted stimuli that had *only* colour information *or* only shape information for four trials (a blurry patch of colour, or a monochrome outline of a shape) (see Figure 8 for example stimuli). In the pre-switch phase, the child continued to sort stimuli with only colour or shape information for the first six trials. Then, for the next six pre-switch trials, stimuli with both colour and shape information were used (see Figure 9 for example stimuli). Finally, in the 10 post-switch trials, stimuli with both colour and shape information were used (see Figure 7).

Figure 8. Example stimuli in the practice phase on the SwIFT-4 that have only colour information

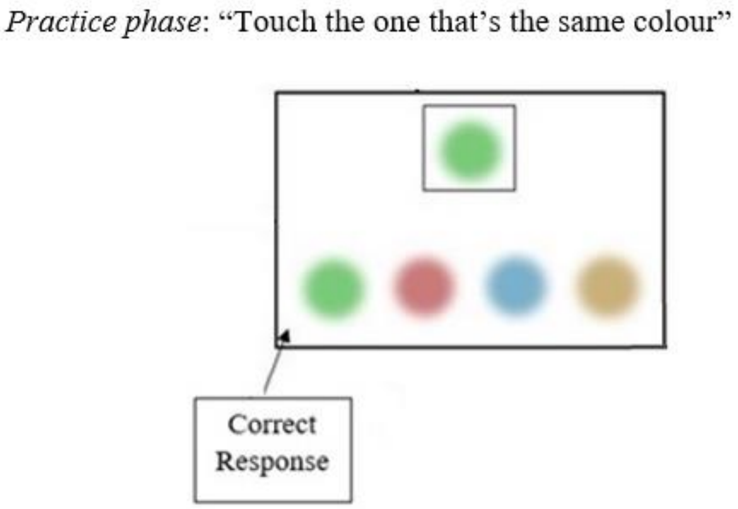
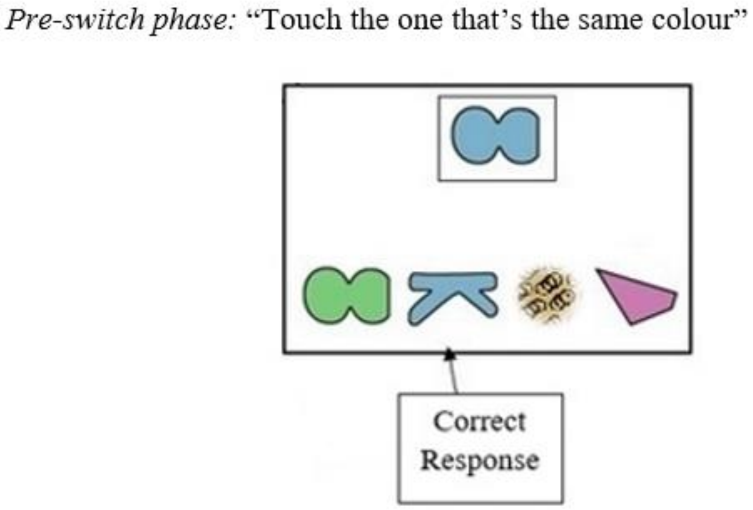


Figure 9. Example stimuli in the pre-switch phase on the SwIFT-4 that have both colour and shape information



Each trial began with a prompt stimulus showing at the top of the screen. After a 1000-ms delay, four response stimuli showed in the lower part of the screen, along with an auditory instruction of "Touch the one that's the same colour/shape". If the child selected the correct response option, positive feedback appear in the place of the selected stimulus that features a musical cartoon animation. If the child selected an incorrect response option, the selected image disappeared without any animation, and the next trial began. If the child did not respond, the experimenter repeated the prompt wording. The dependent variable was the frequency of making each kind of response during the post-switch phase (correct response, conflicting response, distracting responses).

2.2.3.2. Assessing Working Memory

Spatial Span Task. The task required the child to watch and then reproduce a sequence of tapping actions on an array of six everyday objects (Muller et al., 2012). The task used two toy wands (for tapping the objects), and six objects that were put in front of the child in a fixed order (a book, a peg, a toy phone, a sock, a hat, and a cup). The task consisted of a practice phase and a testing phase. The practice phase had four trials where the experimenter tapped one object per trial, and asked the child to tap the same objects. The testing phase had up to 18 trials, where the child copied tapping sequences of increasing length. The number of objects the child had to tap was gradually increased from one object to six objects, and the child received three trials for each span length. If the child correctly copied at least two trials out of three in any object span, the span length was increased by one, and the child was given a further three trials at the longer span length. The task ended if the child incorrectly copied at least two trials out of three in any object span. The dependent variable was the number of correctly copied trials in total (out of 18).

Spin the Pots Task. This was a simple search task that required the child to find stickers hidden in an array of boxes (Hughes & Ensor, 2007). The task used eight visually distinct pots attached to a rotating tray, and a black cloth to cover the boxes. The experimenter hid stickers in six out of the eight pots, in full view of the child, and pointed out the empty boxes. Then the experimenter covered the tray with the cloth and rotated it for a few seconds. The cloth was then removed, and the child was asked to look for a sticker. If the child found a sticker, he or she kept it. If the child did not find a sticker, the tray was covered again and rotated, and a new trial began. The task ended when the child found all the stickers, or after 16 trials. The dependent variable was the total number of searching trials (out of 16).

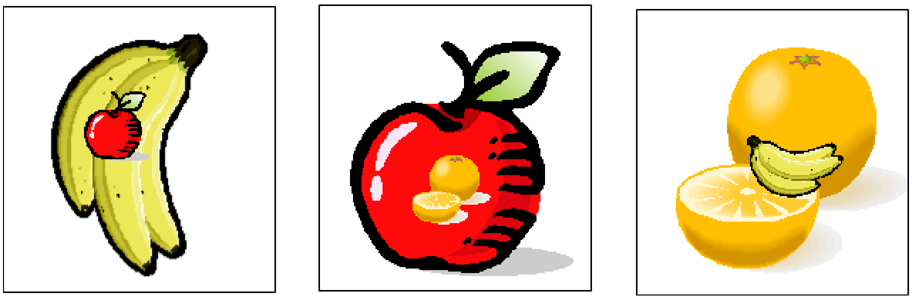
2.2.3.3. Assessing Inhibitory Control

Reverse Categorization Task. The task required the child to sort colourful blocks into different-coloured boxes (Carlson et al., 2004). It used 12 blocks (six yellow and six blue), and two coloured boxes (one yellow and one blue). The task consisted of a practice phase and a testing phase. In the practice phase, the experimenter demonstrated sorting by placing three yellow cubes in the yellow box, and three blue cubes in the blue box. Then the child was instructed to sort the remaining three yellow cubes and three blue cubes into the same-coloured boxes. Then, in the testing phase, the experimenter emptied the boxes and asked the child to sort the cubes into *different*-coloured boxes: yellow cubes into the blue box, and blue cubes into the yellow box. The dependent variable was the number of cubes that were correctly sorted into different-coloured boxes (out of 12).

Fruit Stroop Task. In this task, the child was presented with pictures of small pieces of fruit embedded within larger pieces of fruit. When the experimenter said the name of a fruit, the child had to point to the *small* fruit, and not to the larger, more salient fruits (Kochanska et al., 2000). The task had a practice phase and a testing phase. In the practice phase, the experimenter showed both small and larger pictures of the three fruits – an orange, a banana, and an apple – and named them in turn. Then small pictures of the fruits were removed, and the child was asked to point to each of the larger fruits (to check that the child knew the names of the fruits).

Then in the testing phase, the experimenter presented three pictures showing small fruits presented within larger different fruits (see Figure 10 for example pictures). The child was then instructed to point to each of the small fruits. On each trial, performance was scored out of 2: the child received a score of 0 if he or she pointed to the larger version of the fruit being named; a score of 1 if he or she pointed to the larger fruit, and then self-corrected to point to the smaller fruit; and a score of 2 if he or she pointed to the correct small fruit on their first response. The dependent variable was the summed score from three testing trials (giving a total score out of six).

Figure 10. Stimuli from the testing phase of the Fruit Stroop Task



Black and White Task. In this task, the child was shown one of two cards (a black card and a white card). They were told that when the experimenter said “white”, they should point to the black card; and when the experimenter said “black”, they should point to the white card (Simpson & Riggs, 2005). The task consisted of a demonstration phase, a practice phase, and a testing phase. In the demonstration phase, the experimenter introduced a monkey puppet to help the child grasp the task rules. When the experimenter said “white”, she made the puppet point to the black card, and when the experimenter said “black”, she made the puppet point to the white card. The child was then asked to complete four check trials. In the practice phase, the child completed six trials with feedback, in a pre-set order of BWBWBW. In the testing phase, the child completed 12 trials without feedback, in a pre-set order of BWBWWBBWBWWB. The dependent variable was the total number of correct responses in the testing phase (out of 12).

Peg Tapping Task. In this task, the child was asked to tap on the table twice with a wooden peg when the experimenter tapped once, and to tap once when the experimenter tapped twice (Diamond & Taylor, 1996). The task consisted of a demonstration phase, a practice phase, and a testing phase. In the demonstration phase, the experimenter introduced the monkey puppet to help the child grasp the task rules. When the experimenter tapped once, she made the puppet tap on the table twice, and when she tapped twice, she made the puppet tap on the table once. The child was then asked to complete four check trials. In the practice phase, the child completed six trials with feedback, in a pre-set order of 121212. In the testing phase, the child completed 12 trials without feedback, in a pre-set order of 112122122121. The dependent variable was the total number of correct responses in the testing phase (out of 12).

2.3. Results

2.3.1. Preliminary Analyses

First, preliminary analyses were performed to test if performance varied by gender or rule order. To test whether gender had an effect on CF, on working memory, or on inhibitory control, independent-sample t-tests were conducted separately for 2.5-year-olds, 3-year-olds, 3.5-year-olds, and 4-year-olds. For most age groups, there was no gender effect on CF; the exception was for 3.5-year-olds, where girls (N = 23, M = 5.17, SD = 4.31) performed better than boys (N = 27, M = 2.89, SD = 3.01) (p = .03, medium effect size, Cohen's d = .61). There were no gender effects on working memory or on inhibitory control in any age group (ps ranged between .07 and .98). Second, to test whether there was an effect of rule order on switching performance on the SwIFT-4, an independent-sample t-test was conducted. There was no effect of rule order on SwIFT-4 performance (*p* = .96), indicating that children's switching accuracy was no different if they sorted by colour first, or by shape first. Third, since the SwIFT-4 was a new task and was conducted with very young children, it was important to check whether children could engage successfully with the task. To assess this, children's pre-switch performance was compared against chance performance (chance level performance would be performing around 25% correct responses because children could make four different kinds of responses on the SwIFT-4). All children comfortably exceeded chance-level performance in the pre-switch phase (2.5-year-olds performed 79% correct responses; 3-year-olds performed 92% correct responses; 3.5-year-olds performed 94% correct responses; and 4-year-olds performed 97% correct responses). This suggests that all children were able to engage well with the task.

2.3.2. How Many Patterns of Performance Best Explain Switching on the SwIFT-4?

Unlike previous measures of CF, the SwIFT-4 allowed children to make any one of four different kinds of post-switch response. Therefore, a simple comparison of the total number of correct responses would be poorly suited to identify any switching patterns within the present sample. Instead, the Latent Class Analysis (LCA; Williams & Kibowski, 2016) was used to identify the number of different performance patterns children produced on the post-switch phase of the SwIFT-4. A series of LCAs were performed in Mplus 8.3 demo version (Muthén & Muthén, 2019). In line with recommendations, seven fit indices were used – three likelihood ratio tests and four information criterion indices (Fryer, 2017; Kam, Morin, Meyer, Topolnytsky, 2016).

The three likelihood ratio tests used were: i) the Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (Vuong, 1989); ii) the Lo-Mendell-Rubin Likelihood Ratio Test (Lo, Mendell, & Rubin, 2001); and iii) the Bootstrap Likelihood Ratio Test (McLachlan & Peel, 2000). These tests were used to examine whether adding one additional performance pattern to the model led to a statistically significant improvement (Fryer, 2017). Each performance pattern was identified according to two criteria: maintaining as much similarity *within* a pattern, while also maintaining as much difference *between* different patterns, as possible (Lanza & Cooper, 2016). The present study compared models with one- to five-performance patterns, because that number reflected our maximum hypothesized plausible range of switching patterns that children might produce on the SwIFT-4. The model fit indices for one- to five-performance-pattern solutions are reported in Table 1. As Table 1 shows, the one-performance-pattern model was not the best solution, and adding extra performance patterns significantly improved the model fit. However, these likelihood ratio tests did not offer clear grounds for choosing between the two- to five-performance-pattern models, and so other indices were required. To further determine which model fits the data best, four information criterion indices were used.

Table 1. Model fit evaluation information for the SwIFT-4

	VLMRT	**LMRT**	**BLRT**	**Entropy**	**AIC**	**BIC**	**SABIC**
1-performance pattern	*NA*	*NA*	*NA*	*NA*	858.112	877.303	858.3
2-performance patterns	$p < .001$	$p < .001$	$p < .001$	1	667.41	708.99	667.819
3-performance patterns	$p < .001$	$p < .001$	$p < .001$	1	579.555	643.525	580.183
4-performance patterns	$p < .05$	$p < .05$	$p < .001$	1	581.008	667.367	581.857
5-performance patterns	$p < .05$	$p < .05$	$p < .05$	1	587.958	696.707	589.027

Note. VLMRT = Voung-Lo-Mendell-Rubin Likelihood Ratio Test. LMRT = Lo-Mendell-Rubin Likelihood Ratio Test. BLRT = Bootstrapped Likelihood Ratio Test. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. SABIC = Sample-size Adjusted Bayesian Information Criterion. NA = Not applicable.

The four information criterion indices used were: i) Entropy (Celeux & Soromenho, 1996); ii) Akaike's Information Criterion (AIC; Akaike, 1987); iii) Bayesian Information Criterion (BIC; Schwartz, 1978); and iv) Sample-size Adjusted BIC (SABIC; Sclove, 1987). For Entropy, values closer to 1 (the maximum value) indicate better fit, because this criterion reflects the overall quality of the model classification (Geiser, 2012). Conversely, for the other three information criterion indices, the lower the value the better the model, because lower values indicate less unexplained variance in the model (Fryer, 2017; Nylund-Gibson, Grimm, Quirk, & Furlong, 2014). As Table 1 shows, the Entropy value was "1" for all models, so Entropy did not indicate that any model was better than the others. However, the other three information criterion indices indicated that the three-performance-pattern model was best, as it had the lowest AIC (579.555), BIC (643.525), and SABIC (580.183) values. In other words, the most parsimonious way to think about children's switching ability on the SwIFT-4 was to consider it involving three distinct performance patterns.

2.3.3. What Are These Three Performance Patterns?

The Latent Class Analysis suggested that three distinct performance patterns best explain switching on the SwIFT-4. The next question then, is how can we best describe these three performance patterns? These performance patterns differed across three dimensions: probability of making a *correct response*, probability of making a *conflicting response*, and probability of making a *distracting response* (the goal-related distractor and goal-unrelated distractor were combined as a single type of distracting response, because these different types of distractor did not influence the performance patterns of children). Table 2 shows how the three performance patterns differed across these three dimensions. Children showing the first performance pattern performed very well at switching: they made correct responses frequently (91%), and made conflicting responses or distracting responses rarely (8% and 1% respectively). Therefore, those children are referred to as "Switchers" (*N* = 69, 38%). Children showing the second performance pattern performed poorly at switching: they made conflicting responses frequently (92%), and made correct responses or distracting responses rarely (6% and 2% respectively). Therefore, those children are referred to as “Perseverators” (*N* = 54, 30%). Finally, children showing the third performance pattern performed moderately at switching: they made *both* correct responses and conflicting responses reasonably frequently (43% and 45% respectively), and rarely made distracting responses (12%). Therefore, those children are referred to as “Mixed Responders” (*N* = 58, 32%).

Table 2. Probabilities of making different kinds of response across three performance patterns on the SwIFT-4

	Correct Response	**Conflicting Response**	**Distracting Response**
Switchers (38%)	High probability	Low probability	Low probability
Perseverators (30%)	Low probability	High probability	Low probability
Mixed Responders (32%)	Moderate probability	Moderate probability	Low probability

Note. High probability = 8-10 trials out of 10.

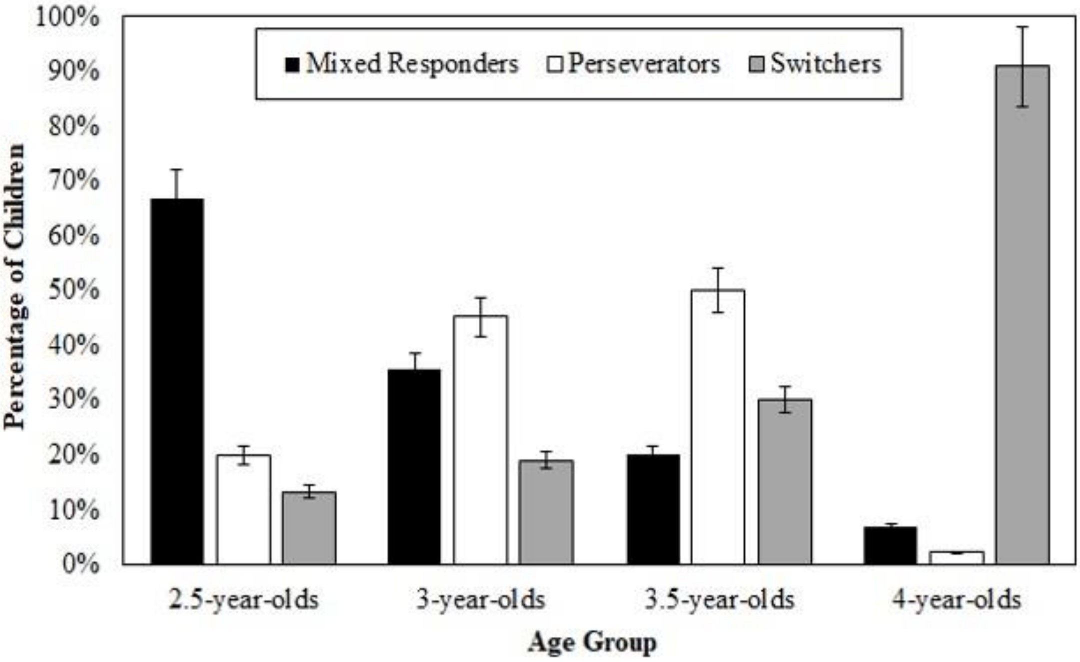
Moderate probability = 3-7 trials out of 10.

Low probability = 0-2 trials out of 10.

In order to determine whether these three performance patterns varied by age, age-related changes in those performance patterns were examined. This would give us some indication of whether CF development followed a particular trajectory. Figure 11 shows the proportion of children showing different performance patterns presented for each of the four 6-month age spans. Most 2.5-year-olds were Mixed Responders (around 67%); around half of 3-year-olds (48%) and most 3.5-year-olds (around 54%) were Perseverators; and most 4-year-olds were

Switchers (around 91%). Figure 11 illustrates an interesting developmental trend: the number of Mixed Responders decreased with age, while the number of Switchers increased with age. Interestingly, the number of Perseverators increased between the ages of two-and-a-half and three years, but by age four years, few children showed this pattern.

Figure 11. Percentage of children showing different performance patterns as a function of age



2.3.4. How Do Executive Functions Contribute to CF Performance on the SwIFT-4?

Working Memory: To look at how individual differences in working memory related to switching ability, a composite working memory score was computed. This was an average of the standardized z-scores for the Spatial Span task and the Spin the Pots task. A Pearson's correlation found a significant correlation between working memory tasks, $r(181) = .48$, $p < .001$; suggesting that it is therefore reasonable to aggregate the scores on these two tasks.

To examine how working memory related to performance on the SwIFT-4, a two-way ANOVA was run with age group and performance pattern (mixed responding, perseveration, and switching) on working memory scores. It found significant main effects of age group, *F*(3, 167) = 7.74, $p < .001$, $\eta^2 = .12$, and of performance patterns, *F*(2, 167) = 18.60, $p < .001$, $\eta^2 = .18$ on working memory. There was no significant interaction between age group and performance patterns, *F*(6, 167) = 1.05, $p = .40$, $\eta^2 = .04$.

To further test the main effect of age group on working memory, a Bonferroni post-hoc test was conducted. It showed that the 4-year-olds group had a higher working memory score (*N* = 44, *M* = .91, *SD* = .56) than the 3-year-olds group (*N* = 42, *M* = -.30, *SD* = .51) and the 2.5-year-olds group (*N* = 43, *M* = -.61, *SD* = .60) ($p = .01$ and $p < .001$). In addition, the 3.5-year-olds group (*N* = 50, *M* = .04, *SD* = .69) had a higher working memory score than the 3-year-olds and 2.5-year-olds groups ($p < .001$). The difference between the 4-year-olds and 3.5-year-olds groups was not significant ($p = .06$).

To further test the main effect of performance patterns on working memory, a Bonferroni post-hoc test was conducted. Figure 12 shows working memory performance of children with each of the three different performance patterns. Switchers (*N* = 69, *M* = .71, *SD* = .63) had a higher working memory score than both Perseverators (*N* = 54, *M* = -.29, *SD* = .51) and Mixed Responders (*N* = 58, *M* = -.55, *SD* = .62) ($p < .001$). There was no significant working memory difference between Perseverators and Mixed Responders ($p = .74$).

Inhibitory Control: To look at how individual differences in inhibitory control related to switching ability, composite scores of inhibitory control were computed (one score for 2.5- and 3-year-olds, and one score for 3.5- and 4-year-olds). Specifically, an average of the standardized z-scores for the Reverse Categorization task and the Fruit Stroop task was calculated for 2.5- to 3-year-olds. A Pearson's correlation found significant correlations between the tasks, $r(81) = .46$, $p < .001$; suggesting that it is therefore reasonable to aggregate the scores on these two tasks. An average of the standardized z-scores for the Black and White task and the Peg Tapping task was calculated for 3.5- to 4-year-olds. A Pearson's correlation found significant correlations between the tasks, $r(90) = .66$, $p < .001$; suggesting that it is therefore reasonable to aggregate the scores on these two tasks.

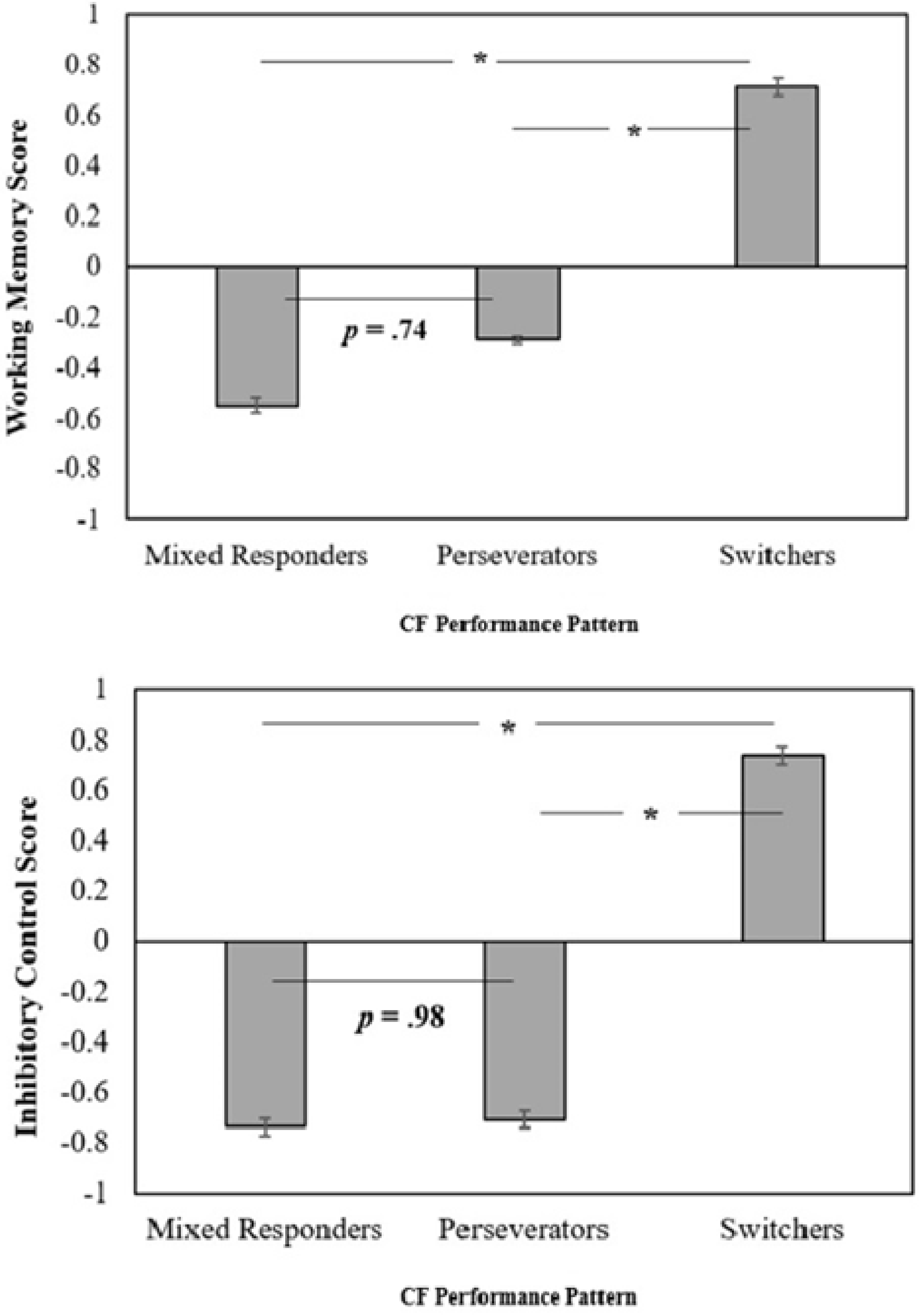
To examine how inhibitory control related to performance patterns on the SwIFT-4, two separate two-way ANOVAs were run with age group and performance pattern (mixed responding, perseveration, and switching) on inhibitory control scores for younger children and older children (as they completed separate inhibitory control tasks).

For 2.5- to 3-year-olds, the two-way ANOVA found no significant main effect of age group, $F(1, 73) = 1.83$, $p = .18$, $\eta^2 = .02$, or of performance patterns, $F(2, 73) = 2.38$, $p = .10$, $\eta^2 = .06$ on inhibitory control. That is, there was no significant inhibitory control difference between the 2.5-year-olds group ($N = 38$, $M = -.15$, $SD = .91$) and the 3-year-olds group ($N = 41$, $M = .19$, $SD = .75$). In addition, there was no significant inhibitory control difference between Switchers ($N = 14$, $M = .49$, $SD = .63$), Perseverators ($N = 26$, $M = .04$, $SD = .88$), and Mixed Responders ($N = 39$, $M = -.15$, $SD = .84$) (*p*s > .10). The interaction between age group and performance patterns was also not significant, $F(2, 73) = 1.18$, $p = .31$, $\eta^2 = .03$.

For 3.5- to 4-year-olds, the two-way ANOVA found significant main effects of age group, $F(1, 75) = 4.44$, $p = .04$, η2 = .06, and of performance patterns on inhibitory control, $F(2, 75) = 8.62$, *p* < .001, η2 = .19. However, there was no significant interaction between age group and performance patterns, *F*(2, 75) = .37, *p* = .70, η2 = .01.

The significant main effect of age group showed that the 4-year-olds group (*N* = 41, *M* = .60, *SD* = .50) performed better on inhibitory control tasks than the 3.5-year-olds group (*N* = 40, *M* = -.54, *SD* = .88) (*p* = .04). To further test the main effect of performance patterns on inhibitory control, a Bonferroni post-hoc test was conducted. Figure 12 shows inhibitory control performance of children with each of the three different performance patterns. Switchers (*N* = 52, *M* = .46, *SD* = .62) had a higher inhibitory control score than both Perseverators (*N* = 17, *M* = -.71, *SD* = .92) and Mixed Responders (*N* = 12, *M* = -.74, *SD* = .83) (*p* = .02 and *p* = .01). However, the difference between Perseverators and Mixed Responders was not significant (*p* = .98).

Figure 12. Working memory and inhibitory control score by performance pattern on the SwIFT-4



In summary, Switchers had better working memory and inhibitory control performance than both Perseverators and Mixed Responders. However, there were no significant differences in working memory or inhibitory control performance between Perseverators and Mixed Responders.

2.4. Discussion

The present study aimed to better characterise how CF develops, through the use of a novel measure that assessed multiple kinds of sorting response across a wider preschool age range. It also tested how individual differences in executive functions contribute to CF. Two novel findings in particular emerged from these results: First, it was found that CF had a stage-like and more protracted developmental trajectory than previously thought. Second, executive functions underpinned some of the stage-like development of CF – but not all. I now explore these findings in more detail.

The first major finding is that CF had a stage-like and more protracted developmental trajectory than previously thought. Rather than development progressing in a linear fashion, the results suggest that there are three distinct qualitative stages in the development of CF between the ages of two and four years. This finding supports previous recent work suggesting that CF development may proceed through three distinct stages (Blakey et al., 2016), but importantly, it extends this work by showing these three stages are indeed apparent on a less constrained task with a wider variety of response options. I will refer to these stages, in chronological order, as the mixed responding stage, the perseveration stage, and the switching stage given that the analyses comparing the patterns with children's mean age suggested that these stages progress with children's age. The difference between these stages is best seen in the types of error children make at each stage, and *not* simply from the total number of correct responses they make. The starting point of CF development is the mixed responding stage. In this stage, children are well able to systematically sort stimuli according to a given rule. They have sufficient top-down control to reliably produce coherent, goal-oriented behaviour. However, when they are required to switch from the previous rule to a current rule, this top-down control seems to almost disappear: after a rule change, children neither flexibly switch to the current rule, nor do they systematically continue to sort with the previous rule. Rather, their post-switch sorting behaviour appears to fluctuate between the two possible rules. The perseveration stage appears to be the next developmental stage, in which children are still unable to flexibly switch from one rule to another, but where their post-switch performance is stable and systematic, if incorrect: they perseverate by continuing to sort with the previous rule. The switching stage appears to be the final developmental stage of the preschool period, in which children can switch from one rule to a another rule.

When asked to switch rules on the SwIFT-4, most 2-year-olds in the sample engaged in mixed responding, by unsystematically selecting from all response options. The LCA in the current study identified mixed responding as a distinct pattern of performance in the development of CF. This is noteworthy because previous studies have usually found two patterns of performance, including switching and perseveration (Diamond et al., 2005; Muller et al., 2006; Zelazo, 2006). Mixed responding was found to be associated with poorer working memory and poorer inhibitory control, consistent with the idea that Mixed Responders had difficulty in maintaining the current sorting rule in mind, and in suppressing task-irrelevant information. This stage, then, might best be characterised as one in which children have such weak top-down control that the requirement to switch rules overwhelms their capacity for cognitive control – so that not only are they unable to switch rules, but they are also unable to maintain the rule they had just been using. However, although Mixed Responders appeared quite haphazard in their post-switch responding, it is important to note that these responses were not entirely lacking in top-down control. Mixed Responders did not select all four possible responses with equal likelihood. Instead, they were relatively well able to ignore rule-irrelevant distraction responses, to instead select either the correct response or the conflicting response. As they selected rule-relevant responses much more often than rule-irrelevant responses, Mixed Responders showed rudimentary forms of top-down control.

The difference between mixed responding and perseveration seen in the present study indicates that an important part of CF development is the gradual increase in the kind of top-down control that children are able to produce. Children who perseverated with the previous rule by definition produced a very low number of successful switches – fewer than children who showed mixed responding. So if one were to index CF performance by the number of correct switches, there would appear to be a U-shaped developmental curve, where Perseverators did worse than both Mixed Responders (who tended to be younger) and Switchers (who tended to be older) (Carroll et al., 2016). However, an alternative view would instead be to consider *qualitative* differences in development, and to characterise the difference between mixed responding and perseverative responding as an important increase in top-down control. Children who perseverate on the SwIFT-4 are producing stable but outdated behaviour (in the sense that it reliably follows the previous rule). This is nevertheless something of an improvement on the unstable behaviour seen in the mixed responding stage. In this sense, one might well argue that perseveration is a developmental milestone in the development of CF – and not simply as a problem that children have to overcome. In a quantitative sense, one might argue that Perseverators performed worse than Mixed Responders, because they made fewer correct responses than Mixed Responders – by definition, Perseverators made errors on almost every post-switch trial, whereas Mixed Responders did sometimes select the correct response. However, I would suggest that a simple numerical tally of the number of correct switches misses the more important qualitative point: even though they made errors on almost every trial, Perseverators responded systematically after a rule change. While they did not successfully switch to the current rule, they nevertheless managed to continue to produce systematic behaviour. In contrast, Mixed Responders after a rule change were unable to maintain and follow a single rule. Therefore, I would suggest that the perseveration stage is best considered as a qualitative improvement than as a quantitative step backwards in the development of CF.

The second major finding is that executive functions did underpin some of the stage-like development of CF, but not all of it. Specifically, executive functions could account for the later development in CF - the difference between perseveration and switching between the ages of three-and-a-half and four years. However, they could not account for the early development in CF – the difference between mixed responding and perseveration between the ages of two-and-a-half and three years. The finding that executive functions explained the later development in CF is consistent with the previous research. For example, recent studies have suggested that working memory and inhibitory control are associated with better switching ability (Blakey et al., 2016; Chevalier et al., 2012; Marcovitch et al., 2010), likely by allowing children to maintain the current task rule and suppress rule-irrelevant information. Performance in the present study indicated that for older children, this indeed appeared to be the case. However, it is surprising that executive functions did not appear to explain the early development in CF – the difference between mixed responding and perseveration. This is interesting, because almost by definition, perseveration involves the continuation of a mentally represented task set, both across time and across changes in task demands. One might expect working memory to be the crucial enabling development between Mixed Responders (whose post-switch behaviour is characterised by an apparent absence of sorting rule) and Perseverators (whose post-switch behaviour seems to reflect a strongly represented previous sorting rule). Indeed, this seems to be the case when we consider flexible behaviour in infancy. For example, in their study, Clearfield et al. (2009) found that infants' perseverative reaches in the A-not-B task came about in line with improvements in their working memory. That is, when the toy was hidden into Location B after placing it into the Location A for several trials, while 8-months-olds continued to searched for the toy in the previous Location A, 5-months-olds unsystematically searched for the toy in both locations. The authors proposed that developments in working memory drove the shift from unsystematic behaviour to perseveration, as improvements in memory enabled 8-months-olds to successfully maintain at least the first hiding location. However, this appears not to be the case for children in the present study, as working memory did not explain the difference between Mixed Responders and Perseverators. Thus, there must be other factors that emerge during this period that support children's goal-oriented flexible behaviour.

It is somewhat surprising that executive functions did not predict the early development in CF, since CF is characteristically thought to be underpinned by core cognitive skills, through which children are able to update the rules governing their goal-oriented behaviour, while also suppressing task-irrelevant information. However, the fact that neither working memory nor inhibitory control were able to explain the early development in CF – through explaining the difference between mixed responding to perseverating raises an obvious question: if this early emerging CF is not explained by executive functions, then what other factors could explain it? This is likely to be a focus of future study. Two potentially informative directions would be to look at children's representational abilities, and their attentional control. There is some evidence indicating that representational abilities can play an important part in young children's goal-oriented behaviours (Hendry et al., 2016; Wiebe et al., 2010). For example, Miller and Markovitch (2015) examined how language and attention abilities at 14 months of age contributed to rudimentary executive functions at 18 months of age. They found that as the child's early vocabulary increased, and the frequency of directing attention to the behaviours of others increased, toddlers' performance on a range of executive function tasks also improved (Miller & Markovitch, 2015). In addition, Devine and colleagues (2019) showed that early attention abilities predicted performance on very basic measures of executive functions at 14 months of age (Devine, Ribner, & Hughes, 2019). Language and attention may well play important roles in the early emergence of CF, because the ability to verbally label a stimulus (or an aspect of a stimulus) as well as the ability to focus on the relevant part of that stimulus can strength the representation of that key task-related information in mind (Marcovitch & Zelazo, 2009) and thus can help children guide their behaviour accordingly (Kuhn et al., 2016). Given this prior work showing that early representational abilities and control of attention are related to better control of behaviour, these abilities could potentially explain the early development in CF – the difference between mixed responding and perseveration between the ages of two-and-a-half and three years.

The present study has identified that early CF development consists of three distinct stages, reflecting qualitative improvements in top-down control of goal-oriented behaviour. This is a potentially important finding, reshaping the prevailing view of CF development in the preschool years. However, it raises two important questions to be addressed: given the novelty of these findings, how generalisable are these three stages of CF development across other CF tasks? And within the development of an individual, do all children pass through all stages? There has been criticism raised of previous CF research for being overly focused on results from a single paradigm. It would therefore strengthen our confidence in the present findings if they replicated, and – no less importantly – if they were extended to different CF tasks. In addition, it would offer valuable context for the present results if we were able to determine whether the three-stage developmental pattern was something seen in all children, or whether instead it were possible to reach the stage of being a successful switcher without first passing through the two stages identified by the present study. These are important questions to address. This will be taken up in Chapter Three.

By using a novel experimental approach, the present study offers a new characterisation of how early CF develops. These results suggest that CF emerges across three distinct stages, marked by increasing levels of top-down control. Furthermore, while the later development in CF appear to be underpinned by executive functions, the earlier development in CF are not. An important question for future research is to examine what factors underpin the early development in CF, possibly through looking at early representational abilities (e.g., language) or early attention during toddlerhood.

Chapter Three

Does the Three-Stage Pattern of Cognitive Flexibility Development Generalise Across Tasks?

The results of Chapter two suggest that the development of CF during the preschool period involves three stages: CF development begins at around two years of age with mixed responding; then with perseveration; and finally with successful switching, emerging by around four years of age. The three-stage pattern of CF development is important because it offers a new view of how CF develops, in particular with regard to CF having a longer trajectory than previously thought. However, it is not yet clear whether this three-stage pattern reflects CF in general, or whether it simply reflects performance on a single task. Chapter three addresses this question by comparing performance on the SwIFT to an entirely new CF task. One hundred and thirty-five 2- to 4-year-olds performed on two different CF tasks, and the developmental patterns on each task were examined. The results showed that the three-stage pattern of CF development generalised across the two tasks. That is, both tasks yielded performance showing three distinct stages, characterised as mixed responding, perseveration, and switching. However, surprisingly the within-child performance across the two tasks moderately but not highly consistent across the two tasks. That is, while mixed responding and switching remained relatively consistent, perseveration varied across the two tasks. This suggests that the three-stage pattern of CF development does generalise across tasks – but also that the degree of perseveration seen when children are asked to switch rules is likely to be substantially determined by the task used to measure CF.

3.1. Introduction

In chapter two, I examined CF development in children between the ages of two and four, on a task that allowed children to make multiple kinds of response. The results revealed a three-stage pattern of CF development, with some children perseverating with the previous rule, some children switching rules correctly, and some children responding in an unsystematic way (referred to as mixed responding). These three stages were related to the age of children, with 2-year-olds being more likely to be Mixed Responders, 3-year-olds more likely to be Perseverators, and 4-year-olds more likely to be Switchers. The finding that CF has three stages is important because it potentially reshapes our understanding of how CF develops – in particular, by demonstrating that CF has a longer developmental trajectory than previously thought. This finding improves our understanding of CF development and suggests that improvements in CF are more than just overcoming perseveration. Indeed, these results offer that perhaps the first milestone children overcome is shifting from responding unsystematically to recruiting their top-down control to respond more systematically (even if that means systematically incorrectly).

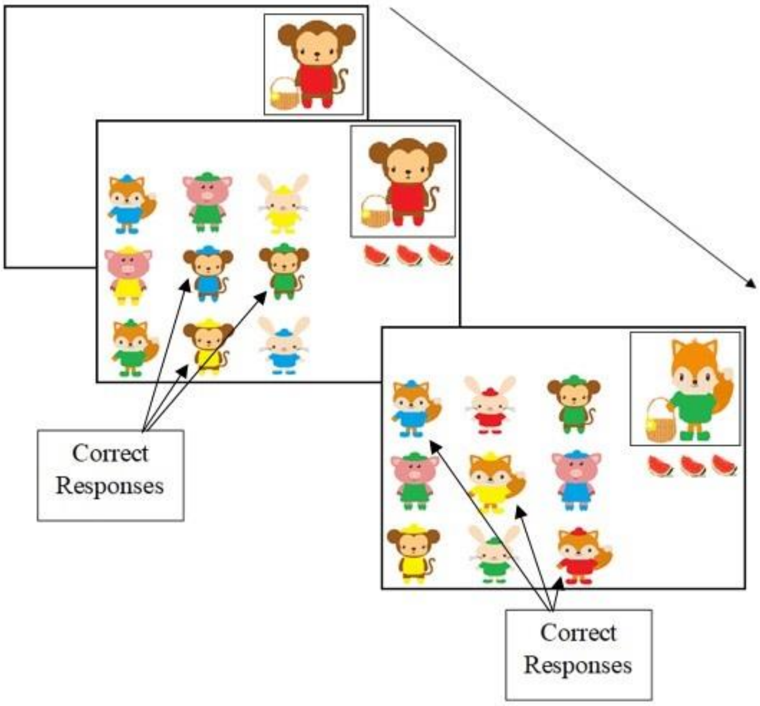
While the three-stage pattern of CF development is a potentially important finding, a necessary next step is to see whether this pattern generalises beyond a single task. The three-stage pattern of CF development has so far been reported on only a single task – the SwIFT (Blakey et al., 2016; see also Chapter two). Since the empirical basis for this three-stage pattern of CF development is new and relatively narrow, it would seem prudent both to i) see whether this finding replicates; and ii) see also whether it generalises beyond a single task, to test whether we can be confident this is a feature of CF development more generally. Broadly speaking, there are two possibilities we should consider. It could be that the three-stage pattern is a genuine, robust and widespread phenomenon that accurately captures how CF develops. In contrast, it could be that the three-stage pattern reported in Chapter two is simply reflecting performance on one specific task. If that were the case, it would clearly be premature to entirely reject the earlier two-stage view of how CF develops (though it would still be informative to learn that the developmental trajectory of CF may vary according to the task used to assess it).

The main aim of the present study was to test whether or not the three-stage pattern of CF development generalises across tasks. This question was addressed by comparing children's CF on two intentionally different rule-switching tasks – the SwIFT, and a novel task designed specifically for this chapter. Two further questions were addressed in the study. First, using this design, we were able to examine whether the three-stage pattern of CF development we observed in Chapter two (using the SwIFT-4) replicates in the present study. In doing so, we would be able to test how robust the three-stage pattern is. Second, we aimed to examine the contributions of executive functions to performance on the novel CF task. This would enable us to examine how working memory and inhibitory control contribute to switching behaviour across distinct CF tasks.

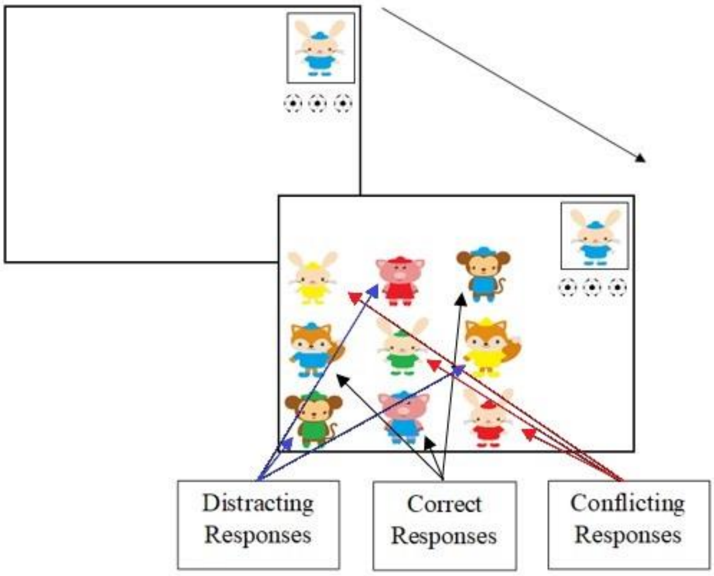
In order to see whether the three-stage pattern generalises beyond a single task, I designed a novel CF task called the Multiple Item Switching Task (the MIST). The MIST was a computerised rule-switching task in which children were instructed to sort animals according to one rule during the pre-switch phase (e.g., by the same species), and by a different rule in the post-switch phase (e.g., by the same colour clothing). Figure 13 shows example trials illustrating the response options in the pre-switch phase and in the post-switch phase. The MIST had simple instructions (e.g., to instruct children to sort by same species rule, children were told "Can you help the mummy animal find her children?") and had simple response requirements (children simply needed to touch images on a touchscreen). Therefore, the MIST was suitable to be used with children as young as two years of age. There are two noteworthy features of the MIST: first, it shared a crucial core demand with SwIFT in that it required children to switch from using one rule to using another rule in the presence of both conflicting information and distracting information. Second, the MIST was also designed to be very different from the SwIFT at a surface level, in terms of its array and its responses. These two features will now be explained in more detail.

Figure 13. Response options in the pre- and the post-switch phase of the MIST

Pre-switch phase (4 trials): "Can you help the *mummy* find her *children*?"



Post-switch phase (4 trials): "Can you help the child find their friends in the *same color clothes?*"



Note. Figure 13 shows the pre- and post-switch phase of the MIST. Children were first presented with the prompt stimulus (an animal) and asked to find the animals that matched on either animal species or clothing colour. In each phase, children completed four trials and had to select three responses per trial (giving a total score of 12 for each phase of the task). In the post-switch phase, children were presented with three correct matching options, three options that matched according to the pre-switch rule (conflicting responses), and three options that contained task irrelevant information and did not match on the current rule nor the previous rule (distractors). Rule order was counterbalanced such that half of children started with animal species rule and half started with colour rule.

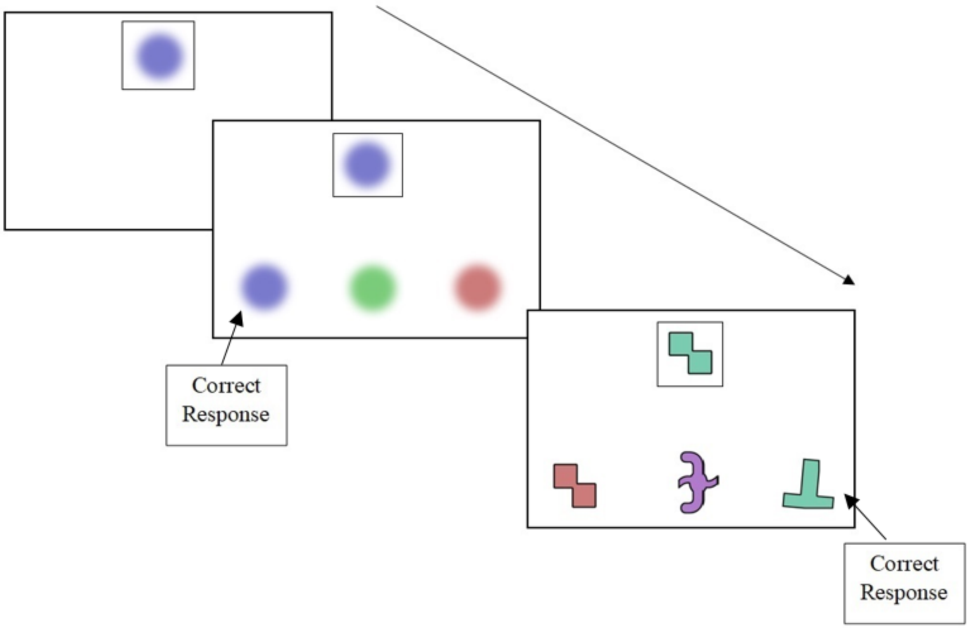
Firstly, when designing the MIST, it was important that the task matched the SwIFT on the crucial core demand: that of switching from one rule to another in the presence of both conflicting and distracting information. That is, both in the new task and in the SwIFT, children were required to switch rules in the presence of a correct answer(s), conflicting information and distracting information. By using this task design, it enabled us to examine different kinds of sorting response on each trial (such as correct responses, perseverative errors, and distraction errors). As we have seen from chapter two, by using a task that allows for more varied responses in this way, we are able to identify more diverse patterns of response in children's behaviour.

Second, the MIST was also designed to be different from the SwIFT at a surface level. The biggest differences were that the MIST had a larger sorting array, and that children were required to make multiple responses on a single trial. Other surface-level differences included using different stimuli, different instructions, and different feedback. Existing CF tasks have usually used relatively small task arrays with a limited number of response options. They also tend to require children to make a single response on each trial (though see Deák & Wiseheart, 2015). However, in order to test whether the three-stage pattern generalises across different CF tasks, it was important to ensure that there were significant differences between the MIST and the SwIFT – while also ensuring that the core switching demand remained the same. Specifically, on the MIST, children were presented with a sorting array of nine items, meaning there were nine possible response options on every trial. Among these nine response options, children could make three different kinds of response: on each trial, three response options were *correct responses* that matched the prompt stimulus according to the current rule; three response options were *conflicting responses* that matched the prompt stimulus according to the previous rule; and three response options were *distracting responses* that did not match the prompt stimulus according to either rule. Among these nine response options, children had to make three responses in total in order to proceed to the next trial. There were four trials, so the total score for the task was out of 12.

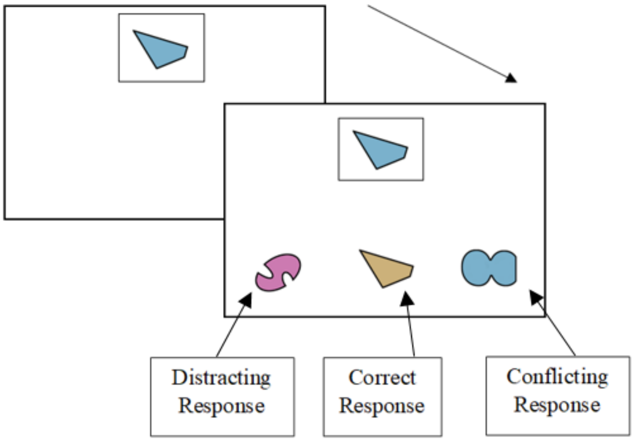
For the purposes of the present study, children's performance on the MIST would be compared to their performance on the Switching, Inhibition, and Flexibility Task-3 (SwIFT-3). The SwIFT-3 was similar to the SwIFT-4 used in Chapter two, with the key difference being that there were three possible response options, instead of four, on each trial. Figure 14 illustrates example response options in the pre-switch phase and in the post-switch phase of the SwIFT-3. In the present study, only one type of distractor was used – goal-related distractors – on both the SwIFT-3 and the MIST. This was because Chapter two found that different types of distractor had little influence on children's performance; goal-related distractors were retained (in preference to goal-unrelated distractors) as these were slightly more common. In addition to being a comparison task for the MIST, the SwIFT-3 would also allow us to test whether the three-stage pattern of CF development found in Chapter two replicates (albeit on a task with this minor change). In the SwIFT-3, children were presented with three response options on the sorting array in total. Among three response options, children could make three different kinds of response: one response option was *correct response* that matched the prompt stimulus according to the current rule; one response option was *conflicting response* that matched the prompt stimulus according to the previous rule; and one response option was *distracting response* that did not match the prompt stimulus according to either rules. Among these response options, children had to make a single response in order to proceed to the next trial. The SwIFT-3 had 12 post-switch trials, and thus the total score for the task was out of 12.

Figure 14. Response options in the pre- and the post-switch phase of the SwIFT-3

Pre-switch phase (12 trials): "Touch the one that's the same *color*"



Post-switch phase (12 trials): "Touch the one that's the same *shape*"



Note. Figure 14 shows the pre- and post-switch phase of the SwIFT-3. Children were first presented with the prompt stimulus (a colourful shape) and asked to find the colourful shapes that matched on either colour or shape. In each phase, children completed 12 trials and had to make a single response per trial. In the post-switch phase, children were presented with a correct matching option, an option that matched according to the pre-switch rule (conflicting response), and an option that contained task irrelevant information and did not match on the current rule nor the previous rule (distractor). Rule order was counterbalanced such that half of children started with colour rule and half started with shape rule.

In order to test whether the three-stage pattern of CF development generalised beyond a single task, we needed to compare performance patterns on the two tasks. Most existing CF tasks measure performance by counting the total number of correct post-switch responses. However, because the MIST and the SwIFT-3 allow children to make multiple kinds of response (that is, correct responses, conflicting responses, and distracting responses), it was necessary to use the same Latent Class Analysis (LCA)-led analyses that were used in Chapter two. That is, first, we would need to calculate how many times children made each kind of response during the post-switch phase. Second, we would need to identify different patterns of performance derived from the frequency of making different kinds of response on the sorting array. Third, we would need to use LCA to help us find the most parsimonious model to best explain the switching behaviour on the MIST and on the SwIFT-3. If the LCA suggested similar models to explain switching behaviour both on the MIST and SwIFT-3, then this would be consistent with the view that the three-stage pattern of CF development generalised beyond a single task. However, if the LCA suggested different models as the optimal way to explain switching behaviour on the MIST and on the SwIFT-3, then this would mean that the three-stage pattern of CF development did not generalise across tasks. In addition, we also examined *within-child performance* across the MIST and the SwIFT-3 in order to see to what extent children showed the same performance pattern on the MIST and the SwIFT-3.

A secondary goal of this study was to examine how executive functions contribute to CF performance. As the MIST was a novel task with different structural features to previous tasks, it would be informative to test how working memory and inhibitory control support performance on this task. In the MIST, children are presented with multiple response options, and multiple kinds of response. That is, there are three correct responses, three conflicting responses, and three distracting responses. In addition, children must make three responses in total among these nine response options. Because of these surface level differences on the MIST, there could, for example, be a need for stronger inhibition and stronger working memory. When the rule changes, for example, children might need to inhibit an increased number of response options in order to select the correct responses on the MIST. Likewise, children might need to use their working memory to strongly maintain the relevant rule in mind on the MIST in the presence of multiple competing stimuli. In sum, the different task demands on the novel CF task could draw on executive functions differentially compared to the SwIFT-3. Therefore, by using different measures for testing working memory and inhibitory control, we can examine how basic executive functions support CF in different contexts on different tasks that have varied surface features.

In summary, the present study aimed to test whether or not the three-stage pattern of CF development generalises beyond a single task. To do this, we compared CF performance on two tasks: the MIST, a newly developed rule-switching task, and the SwIFT-3 based on the task used in Chapter two. This study design also allowed us to test whether the three-stage pattern of CF reported in Chapter two replicates. And it enabled us to ask how individual differences in executive functions contribute to CF on both CF tasks. In this study, children completed four tasks in total: the SwIFT-3, the MIST, a working memory task (the Spatial Span task), and an inhibitory control task (the Reverse Categorization task for younger children and the Black and White task for older children). For the purposes of age-related analyses, children were divided into four age groups using six-month age bands (see also Blakey et al., 2016): 2.5-, 3-, 3.5-, and 4-year-olds.

3.2. Method

3.2.1. Participants

One hundred and thirty-five typically developing 2- to 4-year-olds took part in the study (63 males, 72 females). Data from a further 16 children were excluded: eight children were excluded due to interference from other children in the nursery; two children were excluded due to parental interference; and six children were excluded due to being inattentive during the games. To carry out LCA, a sample size of at least 100 participants is recommended (Nylund et al., 2007). In addition, a power analysis was run to determine the required sample size for comparing children's CF performance across the two CF tasks using Chi-Square analysis. As this research on the three-stage pattern is novel and no prior studies have examined how robust it is (but it has replicated across studies, e.g., Blakey et al., 2016), we opted to power for a medium effect size. The power calculation found that 108 participants would be required for 80% power to detect a medium effect on the Goodness-of-fit Chi-Square tests (with an alpha = .05 and df = 2) (G*Power 3.1.9.2; Faul et al., 2007). Thus, the sample size in the present study ($N = 135$) was sufficient to run the required analyses.

The sample was divided into four age groups, each covering a six-month age span: 2.5-year-olds ($N = 36$, $M_{\text{age}} = 2;8$ years, range = 2;5 years - 2;11 years), 3-year-olds ($N = 31$, $M_{\text{age}} = 3;3$ years, range = 3;0 years - 3;5 years), 3.5-year-olds ($N = 37$, $M_{\text{age}} = 3;8$ years, range = 3;6 years - 3;11 years), and 4-year-olds ($N = 31$, $M_{\text{age}} = 4;1$ years, range = 4;0 years - 4;5 years). Children were recruited from local nurseries and from volunteer local families who had expressed an interest in participating in research. All children were monolingual, and the sample was predominantly White British. The study received ethical approval from the department's ethics committee, and informed consent and verbal assent were obtained from participating parents and children respectively. A small gift was given to children as a thank you for taking part in the study. Testing took place between July 2019 and December 2019.

3.2.2. Procedure

First, children played a warm-up game in which their understanding of the words used in the task rules was checked; such as "shape", "colour", "different", or "same". Once their understanding of the words was ensured, then children completed four tasks in a single testing session that lasted around 25 minutes. Specifically, 2.5-year-olds and 3-year-olds completed four tasks in the following order: the SwIFT-3, the Reverse Categorization task, the Spatial Span task, and the MIST. Three-and-a-half-year-olds and 4-year-olds completed four tasks in the following order: the SwIFT-3, the Black and White task, the Spatial Span task, and the MIST.

3.2.3. Measures

Warm-up Game. This game was identical to the one used in Chapter two, with a child being asked to find the cards with the same colour and with the same shape, after the experimenter placed cards with different colourful shapes in front of the child. If children matched the cards in terms of their colour and their shape, then they continued to play with the rule-switching task.

3.2.3.1. Assessing Cognitive Flexibility

Multiple Item Switching Task (MIST). The MIST was a rule-switching task that was created using E-Prime Psychology Software Version 2 (PST, Pittsburgh, PA). The MIST required children to sort different animals according to a rule first during the pre-switch phase (e.g., by the same species) and then after a rule change, to sort them according to a different rule (e.g., by the same colour clothing). The task was administered on a touchscreen computer, and children made their response by touching the animals on the screen. On every trial, first a prompt stimulus appeared, showing an animal wearing colourful clothing. Then nine response options appeared, each one showing an animal wearing colourful clothing. Among the nine response options, only three matched the prompt stimulus according to the relevant rule (there were thus three correct responses on each trial). Halfway through the task, the rule that the child was asked to sort by changed. Children had to decide which three of nine response options on the sorting array matched the prompt stimulus according to the current rule.

On every trial of the post-switch phase, the child was presented with nine response options on the sorting array. Crucially, each response option was designed so that selecting it would indicate that the child had made a particular *type* of response – a correct response, a perseverative error, or a distraction error. Among nine response options, *three* of them were the correct responses that matched the prompt stimulus according to the current rule. The remaining six response options – three conflicting responses and three distracting responses – were incorrect responses that did not match the prompt stimulus according to the current rule (though for different reasons). The conflicting responses matched the prompt stimulus according to the previously relevant rule. If the child was unable to update or to maintain the current task rule, they would be likely to select the conflicting responses. Conversely, the distracting responses did not match the prompt stimulus *either* according to the current rule *or* according to the previously relevant rule. Instead, the distracting responses contained rule-irrelevant information. If the child failed to maintain either the current rule or the previous rule, it is very possible that they would select a distracting response.

The MIST included three phases: a practice phase (of 2 trials), the pre-switch phase (of 4 trials), and the post-switch phase (of 4 trials). On each trial, children made three responses. Therefore, children made 12 responses in the pre-switch phase, and 12 responses in the post-switch phase. In the practice phase, the child was required to sort stimuli that matched the prompt stimulus only according to one of the rules for two trials. For example, as Figure 15 shows, the stimuli on the sorting array could only match the prompt stimulus (e.g., Mummy monkey) according to the same species rule – matching little monkeys with the Mummy monkey. The stimuli did not match the prompt stimulus by the same colour clothing rule. Similarly, in the pre-switch phase, the child was required to continue to sort stimuli that matched the prompt stimulus only by one of the rules for the first two trials. Then the child was required to sort the stimuli that match the prompt stimulus according to both of the rules for the next two trials. For example, as Figure 16 shows, the stimuli on the sorting array could match the prompt stimulus (e.g., Mummy Fox) both in terms of the same species rule (matching little foxes with the Mummy Fox) and in terms of the same colour clothing rule (matching little animals with green clothing with the Mummy Fox wearing green clothing). Finally, in the post-switch phase, the child was required to sort the stimuli that matched the prompt stimulus according to both rules for four trials.

Figure 15. Example stimuli in the practice phase on the MIST that match the prompt stimulus only by one of the rules

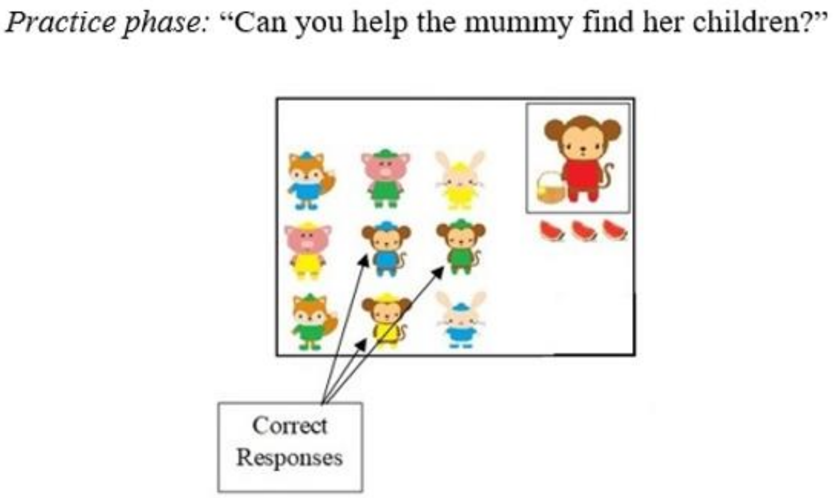
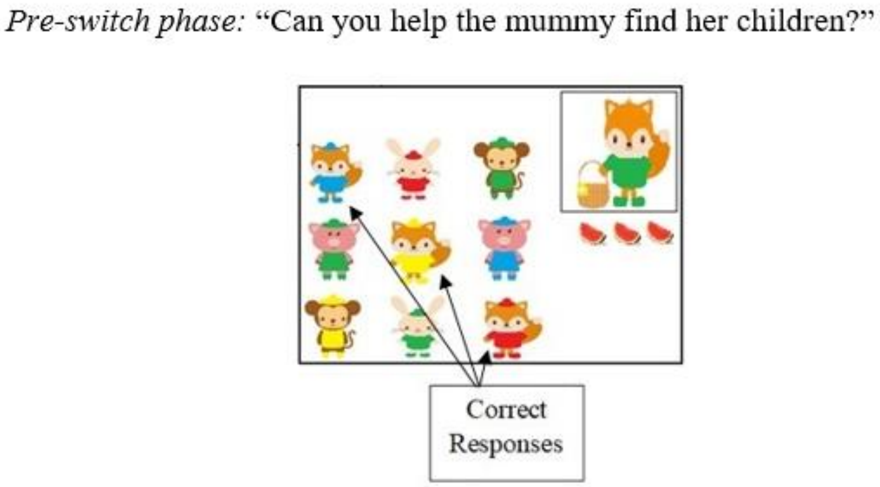


Figure 16. Example stimuli in the pre-switch phase on the MIST that match the prompt stimulus by both rules



Each trial started with a prompt stimulus that showed at the top right corner of the screen. After a 1000-ms delay, nine response options showed at the left part of the screen with an auditory instruction of "Can you help the mummy find her children?" or "Can you help the child find their friends in the same colour clothes?". If the child made a correct response, positive feedback was given that featured a musical animation of the animal selected. If the child made an incorrect response, the selected response option disappeared without any animation, and the task continued to the next trial. If the child did not give any answer, the instructions were repeated by the experimenter. The dependent variable was the frequency of making each kind of response during the post-switch phase (correct response, conflicting response, distracting response).

Switching, Inhibition, and Flexibility Task-3 (SwIFT-3). This rule-switching task was an adaptation of the task used in Chapter two, with the key difference being that the SwIFT-3 used three response options on the sorting array, instead of four. The task was designed in E-Prime Psychology Software Version 2 (PST, Pittsburgh, PA) and was administered on a touchscreen computer. The SwIFT-3 required children to sort colourful images according to one rule first during the pre-switch phase (e.g., colour), and then after a rule change, to sort them according to a different rule (e.g., shape). Children gave their answers by touching an image on the screen. The SwIFT-3 measured children's ability to switch rules, on an array that allows them to make multiple kinds of sorting response. On every trial, first a prompt stimulus appeared, showing a colourful shape. Then three response options appeared, each of which showed a different colourful shape, only one of which matched the prompt stimulus according to the relevant rule. Children must decide which of three response options on the sorting array matched the prompt stimulus according to the current rule.

On every trial of the post-switch phase, the child was presented with three response options. Crucially, each response option was designed so that selecting it would indicate that the child had made a particular *type* of response – a correct response, a perseverative error, or a distraction error. Among three response options, one of them was the correct response that matched the prompt stimulus according to the current rule. The remaining two response options

– a conflicting response and a distracting response – were incorrect responses, as they did not match the prompt stimulus according to the current rule.

The SwIFT-3 included three phases: a practice phase (of 4 trials), the pre-switch phase (of 12 trials), and the post-switch phase (of 12 trials). In the practice phase, the child was asked to sort stimuli that have only shape information or colour information for four trials (e.g., a monochrome outline of a shape or a blurry patch of colour). In the pre-switch phase, the child was asked to sort stimuli that have only shape or colour information for the first six trials. Then, the child was asked to sort stimuli with both shape and colour information for the next six trials. Finally, in the post-switch phase, the child was asked to sort stimuli with both shape and colour information for 12 trials.

Each trial began with a prompt stimulus that appeared at the top of the screen. After a 1000-ms delay, three response options appeared on the lower half of the screen, accompanied by an auditory instruction telling the child to respond according to the relevant rule for that phase of the task (either "Touch the one that's the same *colour*" or "Touch the one that's the same *shape*"). If the child made the correct response, positive feedback was given which featured a musical cartoon animation appearing in place of the selected response option. If the child made an incorrect response, the selected response option disappeared without any animation, and the task continued to the next trial. If the child did not make any response, the instructions were repeated by the experimenter. The dependent variable was the frequency of making each kind of response during the post-switch phase.

3.2.3.2. Assessing Working Memory

Spatial Span Task. This task was identical to the one used in Chapter two, with a child watching and then reproducing a sequence of tapping actions on an array of six everyday objects (Muller et al., 2012). The dependent variable was the number of correctly copied trials in total (out of 18).

3.2.3.3. Assessing Inhibitory Control

Reverse Categorization Task. This task was identical to the one used in Chapter two of this thesis, with a child sorting colourful blocks into different coloured boxes (e.g., blue blocks being placed in a yellow box, and yellow blocks being placed in a blue box) (Carlson et al., 2004). This task was used with 2.5-year-olds and 3-year-olds. The dependent variable was the number of cubes that were correctly sorted into different-coloured boxes (out of 12).

Black and White Task. This task was identical to the one used in Chapter two, with a child pointing to the black card when the experimenter said “white”, and pointing to the white card when the experimenter said “black” (Simpson & Riggs, 2005). This task was used with 3.5-year-olds and 4-year-olds. The dependent variable was the total number of correct responses (out of 12).

3.3. Results

This section is divided into four parts: First, preliminary analyses were performed to check if there were any effects of rule-order on the MIST or SwIFT-3, and any effect of gender on any of the tasks. In addition, because the MIST and the SwIFT-3 were new tasks, we wanted to check how children (particularly younger children) engaged with the tasks, by comparing children’s pre-switch performance with chance. Second, a LCA was performed to test how many performance patterns best explained CF on the MIST. In addition, to assess the contributions of executive function to CF, a series of two-way ANOVAs was performed to test whether there were executive function differences between these performance patterns on the MIST. Third, a further LCA was performed to test how many performance patterns best explained CF on the SwIFT-3, as well as two-way ANOVAs to look at the contributions of executive functions to performance patterns on the SwIFT-3. Fourth, a Chi-Square analysis was performed to directly compare how children performed across the MIST and the SwIFT-3.

3.3.1. Preliminary Analyses

Firstly, to test whether there was an effect of gender on CF, working memory, or inhibitory control, independent-samples t-tests were performed for each age group. There was no effect of gender on CF (*p*s ranged between .10 and .96 for the MIST, and between .14 and .85 for the SwIFT-3); no effect on working memory (*p*s ranged between .18 and .68); and no effect on inhibitory control (*p*s ranged between .06 and .37) in any age group. Second, to test whether there was an effect of rule order on CF performance on the MIST or the SwIFT-3, independent-samples t-tests were conducted on post-switch performance in each age group. For three of the four age groups, there was no effect of rule order on MIST performance (*p*s ranged between .14 and .91); the exception was for 4-year-olds, who performed better when they started with same species rule (N = 12, M = 9.92, SD = 4.42) than with the same colour clothing rule (N = 19, M = 5.79, SD = 4.87) (p = .02, d = .89). For three of the four age groups, there was no effect of rule order on SwIFT-3 performance (*p*s ranged between .34 and .56); the sole exception was for 3-year-olds, who performed better when they started with the shape rule (N = 11, M = 5.27, SD = 4.38) than with the colour rule (N = 20, M = 1.85, SD = 2.32) (p = .01, d = .98). This indicated that rule-switching accuracy of most of the children was not affected by which rule they sorted by first. Third, to check whether children could engage with these tasks, children's pre-switch performance on the MIST and on the SwIFT-3 were compared against chance performance (chance level performance would be performing around 33% correct responses because children could make three different kinds of responses on the sorting array on both tasks). Children clearly exceeded chance performance in both the pre-switch phase of the MIST (2.5-year-olds performed 76% correct responses; 3-year-olds performed 86% correct responses; 3.5-year-olds performed 91% correct responses; and 4-year-olds performed 98%

correct responses) and the pre-switch phase of the SwIFT-3 (2.5-year-olds performed 75% correct responses; 3-year-olds performed 87% correct responses; 3.5-year-olds performed 94% correct responses; and 4-year-olds performed 98% correct responses). Thus, all children appeared to engage well with both tasks.

3.3.2. Analyses for the MIST

3.3.2.1. How many patterns of performance best explain switching on the MIST?

To determine the number of performance patterns that best explained children's performance on the post-switch phase of the MIST, a Latent Class Analysis (LCA; Williams & Kibowski, 2016) was conducted in Mplus 8.3 demo version (Muthén & Muthén, 2019). As part of the LCA, seven fit indices were used – three likelihood ratio tests and four information criterion indices (Fryer, 2017; Kam et al., 2016).

The likelihood ratio tests were used to examine whether adding one additional performance pattern to the model would lead to a statistically better fit (Fryer, 2017). Those tests were i) the Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (Vuong, 1989), ii) the Lo-Mendell-Rubin Likelihood Ratio Test (Lo et al., 2001), and iii) the Bootstrap Likelihood Ratio Test (McLachlan & Peel, 2000). Table 3 indicates the model fit evaluation information for models with increasing number of performance patterns. As shown in Table 3, the model with one performance pattern was not the best solution, and adding extra performance patterns significantly improved the model fit. One of these likelihood ratio tests suggested that the significance value for differentiating between performance patterns started to decrease from four performance patterns and up. But still they did not offer clear grounds for choosing between the models with two to five performance patterns. Thus, to determine which model fits the data best, four other information criterion indices were used.

These four information criterion indices were i) Entropy (Celeux & Soromenho, 1996), ii) Akaike's Information Criterion (AIC; Akaike, 1987), iii) Bayesian Information Criterion (BIC; Schwartz, 1978), and iv) Sample-size Adjusted BIC (SABIC; Sclove, 1987) (see Chapter 2 for a description of these indices). As Table 3 shows, the Entropy values for the models with four and five performance patterns (.80 and .73) were lower than the Entropy values for the models with two and three performance patterns (.97 and .97); therefore models with four and five performance patterns were eliminated. However, Entropy value did not suggest which model was the best between the models with two and three performance patterns. But AIC, BIC, and SABIC values indicated that the model with three performance patterns fitted the data best, as it reflected the lowest AIC (480.60), BIC (529.22), and SABIC (475.45) values. In other words, children's performance on the MIST was best explained as involving three different performance patterns.

Table 3. Model fit evaluation information for the MIST

	VLMRT	**LMRT**	**BLRT**	**Entropy**	**AIC**	**BIC**	**SABIC**
1 performance pattern	*NA*	*NA*	*NA*	*NA*	678.73	693.03	677.22
2 performance patterns	$p < .001$	$p < .001$	$p < .001$	0.97	550.55	582.01	547.22
3 performance patterns	$p < .001$	$p < .001$	$p < .001$	0.97	480.60	529.22	475.45
4 performance patterns	$p < .05$	$p < .05$	$p = .67$	0.80	490.97	556.74	483.10
5 performance patterns	$p < .05$	$p < .05$	$p = 1$	0.73	502.97	585.90	494.18

Note. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. SABIC = Sample-size Adjusted Bayesian Information Criterion. VLMRT = Voung-Lo-Mendell-Rubin likelihood ratio test. LMRT = Lo-Mendell-Rubin likelihood ratio test. BLRT = Bootstrapped likelihood ratio test. *NA* = Not applicable.

3.3.2.2. What Are These Three Performance Patterns on the MIST?

Table 4 indicates how children in the three performance patterns on the MIST performed in terms of making correct responses, conflicting responses, and distracting responses. Children who showed the first performance pattern performed very well at switching. They made the correct response very often (95%), but made the conflicting response and the distracting response rarely (3% and 2% respectively). Therefore, those children are referred to as "Switchers" (*N* = 37, 29%). Conversely, children who showed the second performance pattern performed poorly at switching. They made the conflicting response very often (90%), but made the correct response and the distraction response rarely (4% and 6% respectively). Therefore, those children are referred to as "Perseverators" (*N* = 24; 18%). Finally, children who showed the third performance pattern performed unsystematically at switching. They made both the correct response and the conflicting response relatively often (41% and 43% respectively), but made the distracting response rarely (16%). Therefore, those children are referred to as "Mixed Responders" (*N* = 68; 53%).

Table 4. Probabilities of making different kinds of response across three performance patterns on the MIST

	Correct Response	**Conflicting Response**	**Distracting Response**
Switchers (29%)	High probability	Low probability	Low probability
Perseverators (18%)	Low probability	High probability	Low probability
Mixed Responders (53%)	Moderate probability	Moderate probability	Low probability

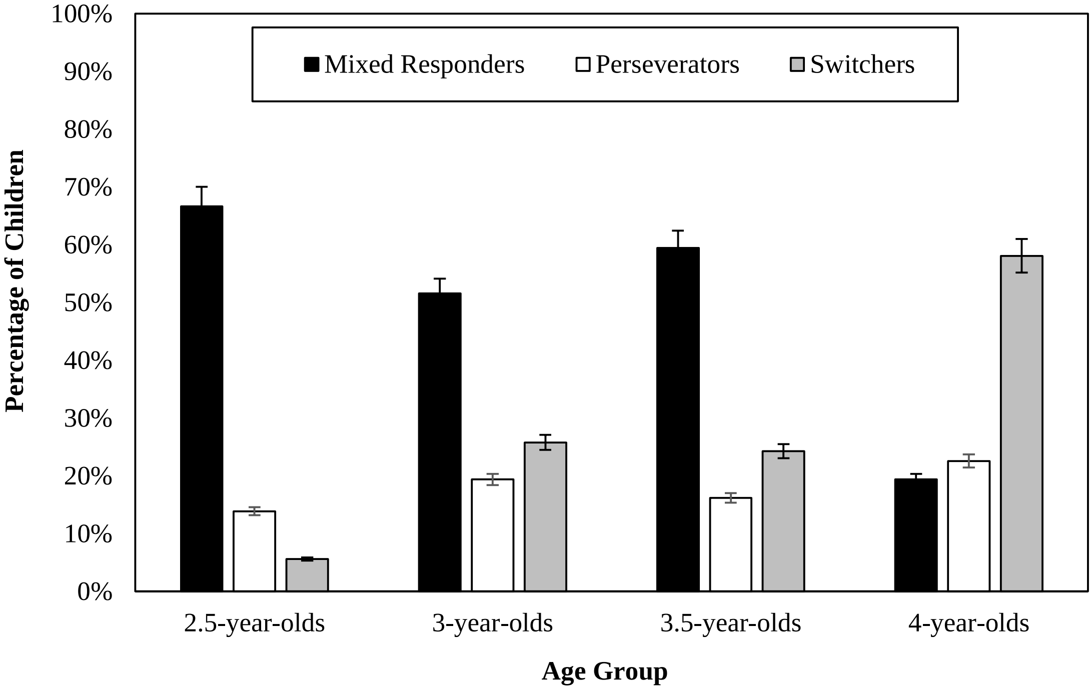
Note. High probability = 10-12 trials out of 12.

Moderate probability = 3-9 trials out of 12.

Low probability = 0-2 trials out of 12.

To explore the relationship between age and CF, we examined age-related changes in performance patterns. Figure 17 shows age-related changes in the proportion of children in different performance patterns, presented across the four six-month age spans. The developmental pattern in Figure 17 suggests three notable findings: First, most 2.5-year-olds (68%), 3-year-olds (53%), and 3.5-year-olds (61%) were Mixed Responders, with the proportion of Mixed Responders decreasing by four years of age. Second, the number of Perseverators slightly increased between the ages of two-and-a-half and three years, but the overall number of Perseverators in the sample as a whole was lower compared to Mixed Responders and Switchers. Third, the number of Switchers appeared to show a marked increase between the ages of two-and-a-half and three years, and then again between three-and-a-half and four years.

Figure 17. Percentage of children showing different performance patterns as a function of age on the MIST



3.3.2.3. How Do Executive Functions Contribute to CF Performance on the MIST?

Working Memory: To examine how working memory related to performance on the MIST, a two-way ANOVA was run on working memory scores, with age group and performance pattern (mixed responding, perseveration, and switching) as factors. There was a significant main effect of age group, $F(3, 111) = 16.25$, $p < .001$, $\eta^2 = .31$, and of performance patterns on working memory, $F(2, 111) = 6.50$, $p = .01$, $\eta^2 = .11$. There was no significant interaction between age group and performance patterns, $F(6, 111) = 1.16$, $p = .34$, $\eta^2 = .06$.

To further examine the main effect of age group on working memory, a Bonferroni post-hoc test was performed. It showed that 4-year-olds had a higher working memory score ($N = 31$, $M = 7.16$, $SD = 1.27$) than both the 3-year-olds ($N = 29$, $M = 5.14$, $SD = 1.30$) ($p < .001$) and the 2.5-year-olds ($N = 26$, $M = 4.31$, $SD = .97$) ($p < .001$). In addition, the 3.5-year-olds ($N = 37$, $M = 6.27$, $SD = 1.17$) had a higher working memory score than both the 3-year-olds and the 2.5-year-olds ($p < .001$). There were no differences between the 3-year-olds and the 2.5-year-olds ($p = .06$), or between the 4-year-olds and the 3.5-year-olds ($p = .37$).

To further examine the main effect of performance pattern on working memory, a Bonferroni post-hoc test was performed. Figure 18 shows the working memory performance of children in each of the three performance patterns. Both Switchers ($N = 37$, $M = 6.78$, $SD = 1.48$) and Perseverators ($N = 23$, $M = 6.22$, $SD = 1.38$) had higher working memory scores than Mixed Responders ($N = 63$, $M = 5.10$, $SD = 1.35$) ($p = .01$ *and* $p = .01$). The difference between Switchers and Perseverators was not significant ($p = .93$).

Inhibitory Control: To examine how inhibitory control related to performance patterns on the MIST, two separate two-way ANOVAs were run with age group and performance pattern (mixed responding, perseveration, and switching) on inhibitory control scores for younger children and older children (as they completed separate inhibitory control tasks).

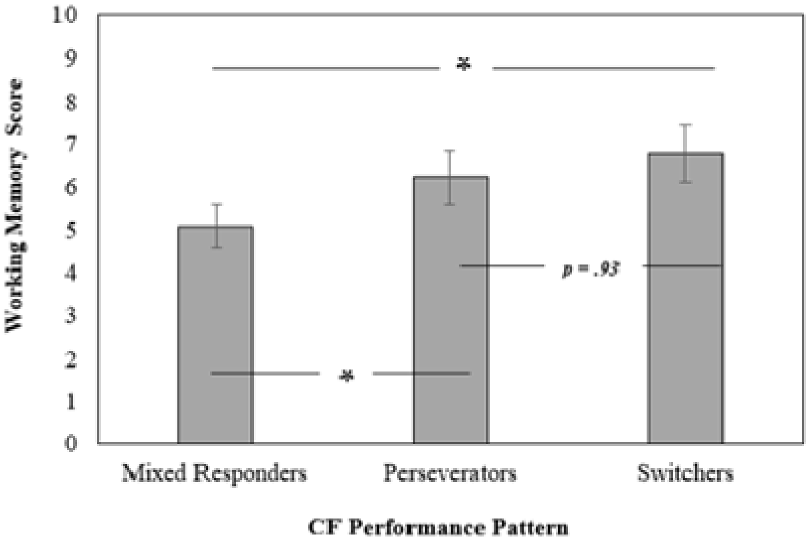
For 2.5- to 3-year-olds, the two-way ANOVA found no significant effect of age group on inhibitory control as measured by the Reverse Categorisation task, $F(1, 55) = .79$, $p = .38$, $\eta^2 = .01$. That is, 2.5-year-olds ($N = 31$, $M = 4.81$, $SD = 5.50$) did not perform significantly differently from 3-year-olds ($N = 30$, $M = 8.43$, $SD = 5.08$) ($p = .38$). However, there was a significant main effect of performance patterns on inhibitory control, $F(2, 55) = 7.89$, $p = .01$, $\eta^2 = .22$. The interaction between age group and performance patterns was not significant, $F(2, 55) = .22$, $p = .80$, $\eta^2 = .01$.

To further examine the main effect of performance patterns on inhibitory control for the younger children, a Bonferroni post-hoc test was performed. Figure 18 shows the inhibitory control performance of children in the three performance patterns. Both Switchers ($N = 10$, $M = 11.90$, $SD = .32$) and Perseverators ($N = 11$, $M = 9.09$, $SD = 5.09$) had significantly higher inhibitory control scores than Mixed Responders ($N = 40$, $M = 4.58$, $SD = 5.27$) ($p = .01$ *and* $p = .02$). The difference between Switchers and Perseverators was not significant ($p = .23$).

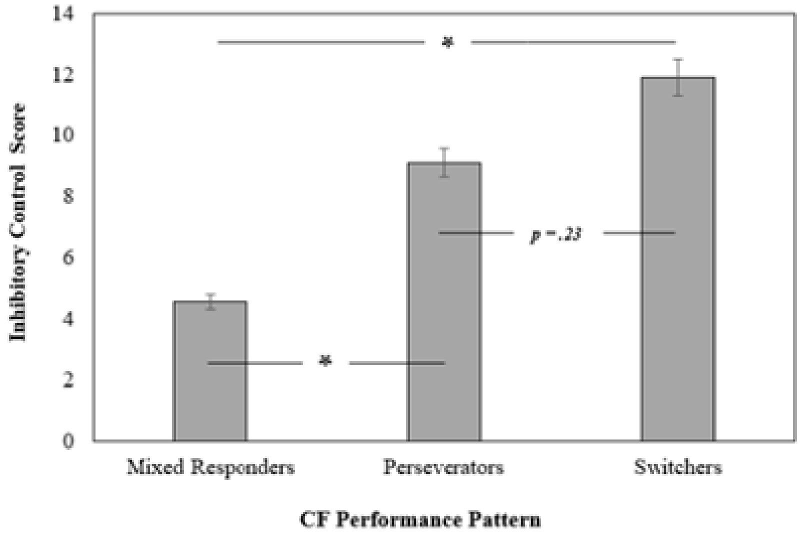
For 3.5- to 4-year-olds, the two-way ANOVA found significant main effects of age group, $F(1, 55) = 5.54$, $p = .02$, $\eta^2 = .09$ and of performance patterns, $F(2, 55) = 4.00$, $p = .02$, $\eta^2 = .13$ on inhibitory control as measured by the Black White Stroop task. However, there was no significant interaction between age group and performance patterns, $F(2, 55) = 1.00$, $p = .37$, $\eta^2 = .04$. The significant main effect of age group showed 4-year-olds ($N = 30$, $M = 9.63$, $SD = 2.31$) had a higher inhibitory control score than 3.5-year-olds ($N = 31$, $M = 7.23$, $SD = 2.69$) ($p = .02$).

To further test the main effect of performance patterns on inhibitory control, a Bonferroni post-hoc test was conducted. Figure 18 shows the inhibitory control performance of children according to the three performance patterns. Switchers (*N* = 26, *M* = 9.88, *SD* = 2.36) had a significantly higher inhibitory control score than both Perseverators (*N* = 12, *M* = 7.17, *SD* = 2.25) and Mixed Responders (*N* = 23, *M* = 7.39, *SD* = 2.76) (*p* = .01 and *p* = .04). However, the difference between Perseverators and Mixed Responders was not significant (*p* = .47).

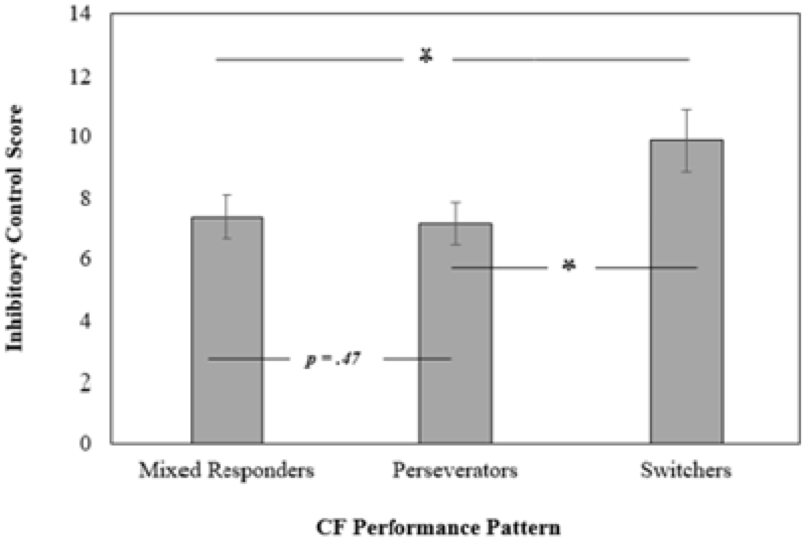
Figure 18. Working memory and inhibitory control score by performance pattern on the MIST



Inhibitory control score by performance pattern for 2.5- to 3-year-olds



Inhibitory control score by performance pattern for 3.5- to 4-year-olds



In summary, both working memory and inhibitory control significantly helped to explain the three-stage pattern of CF development on the MIST. However, while working memory only contributed to the earlier development in CF through explaining the difference between mixed responding and perseveration, inhibitory control explained both the earlier development in CF (through explaining the difference between mixed responding and perseveration) and the later development in CF (through explaining the difference between perseveration and switching).

3.3.3. Analyses for the SwIFT-3

3.3.3.1. How many patterns of performance best explain switching on the SwIFT-3?

To determine the number of performance patterns that best explained children's performance on the post-switch phase of the SwIFT-3, a further LCA (Williams & Kibowski, 2016) was conducted in Mplus 8.3 demo version (Muthén & Muthén, 2019). The same method was used here as in the LCA for the MIST. Table 5 indicates the model fit evaluation information for models with increasing number of performance patterns. As shown in Table 5, the model with one performance pattern was not the best solution; and adding extra performance patterns significantly improved the model fit, up to and including the model with four performance patterns. Starting from the model with four performance patterns, the performances started to be not significantly differentiated (*p* = .06 and *p* = .12); therefore, the models with four and five performance patterns were eliminated. However, these likelihood ratio tests did not offer clear grounds for choosing between models with two and three performance patterns. Thus, to determine which model fits the data best, four other information criterion indices were used, as in the LCA for the MIST above.

As Table 5 shows, the Entropy values for the models with two and three performance patterns were all "1", so Entropy did not indicate that any model was better than any other. However, AIC, BIC, and SABIC values indicated that the model with three performance patterns was best, as it reflected the lowest AIC (438.19), BIC (487.20), and SABIC (433.43) values. In other words, children's performance on the SwIFT-3 was best explained as involving three different performance patterns - as was the case for the MIST, and for the SwIFT-4 in Chapter two.

Table 5. Model fit evaluation information for the SwIFT-3

	VLMRT	LMRT	BLRT	Entropy	AIC	BIC	SABIC
1 performance pattern	*NA*	*NA*	*NA*	*NA*	673.87	688.29	672.47
2 performance patterns	$p < .001$	$p < .001$	$p < .001$	1	524.27	555.98	521.18
3 performance patterns	$p < .05$	$p < .05$	$p < .001$	1	438.19	487.20	433.43
4 performance patterns	$p = .06$	$p = .07$	$p = .43$	1	446.43	512.74	439.99
5 performance patterns	$p = .12$	$p = .12$	$p = .67$	0.84	458.01	541.60	449.88

Note. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. SABIC = Sample-size Adjusted Bayesian Information Criterion. VLMRT = Young-Lo-Mendell-Rubin likelihood ratio test. LMRT = Lo-Mendell-Rubin likelihood ratio test. BLRT = Bootstrapped likelihood ratio test. *NA* = Not applicable.

3.3.3.2. What Are These Three Performance Patterns on the SwIFT-3?

Table 6 indicates how children in the three performance patterns on the SwIFT-3 performed, in terms of making correct responses, conflicting responses, and distracting responses. Children who showed the first performance pattern performed very well at switching. They made the correct response very often (93%), and made the conflicting response and the distracting response rarely (6% and 1% respectively). Therefore, those children are referred to as "Switchers" ($N = 33$; 25%). Conversely, children who showed the second performance pattern performed poorly at switching. They made the conflicting response very often (95%), but made the correct response and the distraction response rarely (4% and 1% respectively). Therefore, those children are referred to as “Perseverators” (*N* = 38; 30%), as their most common response was to continue sorting by the previous task rule. Children who showed the third performance pattern performed unsystematically at switching. They made *both* the correct response and the conflicting response relatively often (41% and 48% respectively), but made the distracting response rarely (11%). Therefore, those children are referred to as “Mixed Responders” (*N* = 61; 45%).

Table 6. Probabilities of making different kinds of response across three performance patterns on the SwIFT-3

	Correct Response	**Conflicting Response**	**Distracting Response**
Switchers (25%)	High probability	Low probability	Low probability
Perseverators (30%)	Low probability	High probability	Low probability
Mixed Responders (45%)	Moderate probability	Moderate probability	Low probability

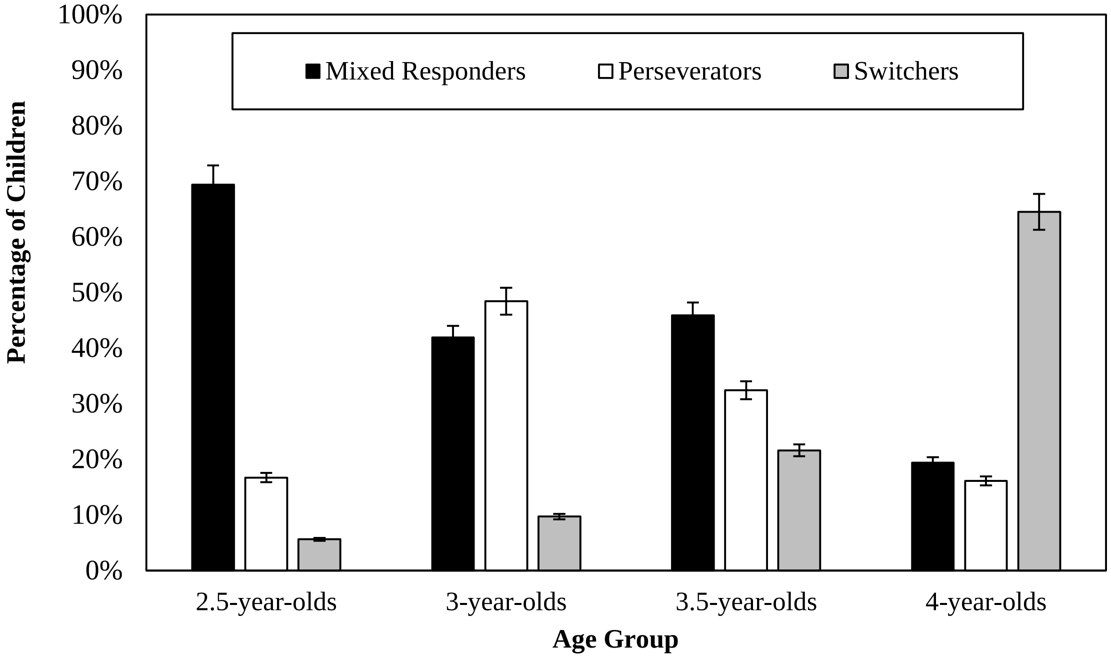
Note. High probability = 10-12 trials out of 12.

Moderate probability = 3-9 trials out of 12.

Low probability = 0-2 trials out of 12.

To explore how the three performance patterns on the SwIFT-3 varied by age, we examined what proportion of children in each age group fell into each of the three patterns. Figure 19 shows age-related changes in the proportion of children in different performance patterns, presented across the four six-month age spans. The developmental pattern in Figure 19 suggests three notable findings: First, in the youngest group, most 2.5-year-olds (70%) were Mixed Responders. Second, the number of children categorised as Perseverators increased between the ages of two-and-a-half and three years, then decreased thereafter. Third, the number of children categorised as Switchers increased between the ages of three-and-a-half and four years.

Figure 19. Percentage of children showing different performance patterns as a function of age on the SwIFT-3



3.3.3.3. How Do Executive Functions Contribute to CF Performance on the SwIFT-3?

Working Memory: To examine how working memory related to performance on the SwIFT-3, a two-way ANOVA was run on working memory scores, with age group and performance pattern (mixed responding, perseveration, and switching) as factors. The two-way ANOVA found significant main effects of age group, $F(3, 113) = 12.04$, $p < .001$, $\eta^2 = .24$, and of performance patterns on working memory, $F(2, 113) = 11.27$, $p < .001$, $\eta^2 = .17$. There was no significant interaction between age group and performance patterns, $F(6, 113) = 2.13$, $p = .06$, $\eta^2 = .10$.

To further examine the main effect of age group on working memory, a Bonferroni post-hoc test was performed. It showed that the 4-year-olds had a higher working memory score (*N* = 31, *M* = 7.16, *SD* = 1.27) than both the 3-year-olds (*N* = 30, *M* = 5.20, *SD* = 1.32) and the 2.5-year-olds (*N* = 27, *M* = 4.33, *SD* = .96) (*p* = .01 and *p* < .001). In addition, the 3.5-year-olds (*N* = 37, *M* = 6.27, *SD* = 1.17) had a higher working memory score than both the 3-year-olds and 2.5-year-olds (*p* = .02 and *p* < .001). Furthermore, the 3-year-olds group had a higher working memory score than 2.5-year-olds group (*p* = .02). The difference between the 4-year-olds and 3.5-year-olds groups was not significant (*p* = .19).

To further examine the main effect of performance patterns on working memory, a Bonferroni post-hoc test was performed. Figure 20 shows the working memory performance of children in the three performance patterns. Switchers (*N* = 33, *M* = 7.12, *SD* = 1.34) had a higher working memory score than both Perseverators (*N* = 38, *M* = 5.87, *SD* = 1.40) and Mixed Responders (*N* = 54, *M* = 4.98, *SD* = 1.25) (*p* = .03 and *p* < .001). In addition, Perseverators had a higher working memory score than Mixed Responders (*p* = .01).

Inhibitory Control: To examine how inhibitory control related to performance patterns on the SwIFT-3, two separate two-way ANOVAs were run on inhibitory control scores, with age group and performance pattern (mixed responding, perseveration, and switching) as factors. There were separate ANOVAs for younger children and older children, as they completed different inhibitory control tasks.

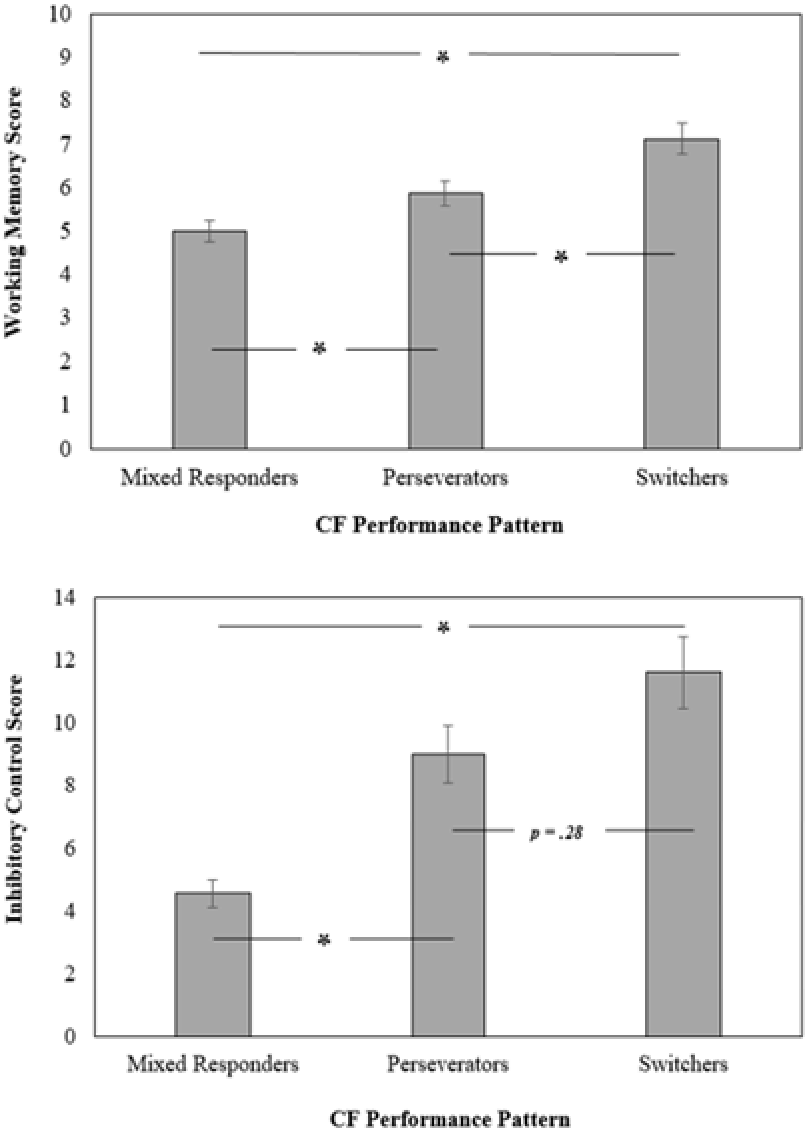
For 2.5- to 3-year-olds, the two-way ANOVA found no significant main effect of age group, *F*(1, 58) = .35, *p* = .56, η^2 = .01. That is, there was no difference in inhibitory control between the 2.5-year-olds (*N* = 33, *M* = 4.88, *SD* = 5.54) and the 3-year-olds (*N* = 31, *M* = 8.35, *SD* =

5.01). However, there was a significant main effect of performance patterns on inhibitory control, *F*(2,58) = 5.90, *p* = .01, η^2 = .17. The interaction between age group and performance patterns was not significant, *F*(2, 58) = .70, *p* = .50, η^2 = .02.

To further examine the main effect of performance patterns on inhibitory control for 2.5- and 3-year-olds, a Bonferroni post-hoc test was performed. Figure 20 shows the inhibitory control performance of children in the three performance patterns. Both Switchers (*N* = 5, *M* = 11.60, *SD* = .55) and Perseverators (*N* = 21, *M* = 9.00, *SD* = 4.70) had higher inhibitory control scores than Mixed Responders (*N* = 38, *M* = 4.55, *SD* = 5.41) (*p* = .01 and *p* = .01). However, the difference between Switchers and Perseverators was not significant (*p* = .28).

For 3.5- to 4-year-olds, the two-way ANOVA found a significant main effect of age group, *F*(1, 55) = 5.95, *p* = .02, η^2 = .10. That is, 4-year-olds (*N* = 30, *M* = 9.63, *SD* = 2.31) had a higher inhibitory control score than 3.5-year-olds (*N* = 31, *M* = 7.23, *SD* = 2.69) (*p* = .02). However, there was no significant main effect of performance patterns on inhibitory control, *F*(2, 55) = 1.67, *p* = .20, η^2 = .06, and no significant interaction between age group and performance patterns, *F*(2, 55) = .34, *p* = .71, η^2 = .01.

Figure 20. Working memory and inhibitory control score by performance pattern on the SwIFT-3



In summary, both working memory and inhibitory control significantly helped to explain the three-stage pattern of CF development on the SwIFT-3. However, while inhibitory control only contributed to the earlier development in CF through explaining the difference between mixed responding and perseveration, working memory contributed to both the earlier development in CF (through explaining the difference between mixed responding and perseveration) and the later development in CF (through explaining the difference between perseveration and switching).

3.3.4. Did Children Perform Similarly Across the MIST and the SwIFT-3?

Both on the MIST and on the SwIFT-3, children showed one of three patterns of performance: mixed responding, perseveration, or switching. However, although all three performance patterns were found on both tasks, they occurred in different proportions. That is, there were differences between the MIST and the SwIFT-3 in terms of how frequently children engaged in the same performance patterns. Notably, on the MIST, most 2-year-olds, 3-year-olds, and 3.5-year-olds were classed as Mixed Responders, while most 4-year-olds were Switchers. Therefore, perseveration emerged as a less frequent performance pattern on the MIST. However, on the SwIFT-3, a relatively clearer age-related pattern emerged which mirrors the findings seen in Chapter two: most 2-year-olds were Mixed Responders, most 3- and 3.5-year-olds were Perseverators, and most 4-year-olds were Switchers. This difference led us to question how these three patterns of performance were similar across the two tasks.

To test this, we used Chi-Square analysis. The Crosstabs allowed us to test how similarly an individual child performed across the SwIFT-3 and the MIST by comparing how frequently children showed the same performance pattern across the SwIFT-3 and the MIST. The results of the Chi-Square analysis showed consistencies in performance across the tasks in some performance types, but not in others; X^2 (4, N = 129) = 51.90, p < .01. Table 7 shows how similarly children performed across the SwIFT-3 and the MIST, presented as a percentage. There was some within-child consistency in performance patterns: specifically, of the children categorised as Mixed Responders on the SwIFT-3, most (81.4%) were also categorised as Mixed Responders on the MIST. In addition, of all the children categorised as Switchers on the SwIFT-3, most (63.6%) were also categorised as Switchers on the MIST. This suggests mixed responding and switching are performance patterns that are consistent in children across two different CF tasks. A different picture emerged for perseveration: of all the children categorised as Perseverators on the SwIFT-3, fewer than half (40.5%) were also categorised as Perseverators on the MIST. Of the remaining children categorised as Perseverators on the SwIFT-3, some were categorised as Mixed Responders on the MIST (35%) and some were categorised as Switchers on the MIST (24%). This suggest that perseveration may not be a stable performance pattern for children: most children who perseverated on one task showed another kind of performance pattern on the other task.

Table 7. Comparing within-child performance patterns across the SwIFT-3 and the MIST

		MIST		
		Mixed Responders	**Perseverators**	**Switchers**
	Mixed Responders (*N* = 59)	81.4% *N* = 48	6.8% *N* = 4	11.9% *N* = 7
SwIFT-3	**Perseverators** (*N* = 37)	35.1% *N* = 13	40.5% *N* = 15	24.3% *N* = 9
	Switchers (*N* = 33)	21.2% *N* = 7	15.2% *N* = 5	63.6% *N* = 21

3.4. Discussion

The present study aimed to test whether or not the three-stage pattern of CF development found in Chapter two generalises beyond a single task. The two tasks used in the present study shared the same core demand – to switch rules in the presence of conflicting and distracting information – but differed on various surface features. Two important findings emerged from these results: First, the three-stage pattern of CF development was found across two different tasks, with analyses showing (i) that three performance patterns best captured variability on both of the CF tasks, and (ii) that these performance categories appeared to be the same for both tasks. When considering within-child consistency in performance patterns, it was found that most children who were classed as Mixed Responders and most children who were classed as Switchers on the SwIFT-3 tended to be classed in the same way on the MIST. However, it was not the case that children classed as Perseverators on the SwIFT-3 were most likely to also be classed as Perseverators on the MIST. Instead, there was much less perseveration reported on the MIST, relative to the SwIFT-3. Around 60% of children who perseverated on the SwIFT were classed as either Mixed Responders or Switchers on the MIST. This suggests that the degree of perseveration seen when switching rules may be substantially determined by the task used to assess CF, as well as the cognitive abilities of the child. Second, working memory and inhibitory control explained important aspects of performance on both tasks. However, successful switching was underpinned by different executive functions on different tasks. Successful switching was explained by *working memory* in the SwIFT-3, but by *inhibitory control* in the MIST. I now consider these results in more detail.

Finding that the three-stage pattern of CF development generalises across the two tasks allows us to be more confident that CF has a broader developmental trajectory than previously thought. For the past few decades, CF has typically been characterised as developing from perseverating with the previous rule (at around age three) to switching rules successfully (at around age four; Muller et al., 2006; Munakata et al., 2012). However, the results of this thesis suggest instead that CF starts to develop earlier than this. Specifically, CF has a three-stage developmental pattern beginning in 2- to 4-year-olds: when asked to switch from one rule to another, young children tend to engage in unsystematic mixed responding around age two years. The present study shows that this mixed responding is not merely an artefact from a single task, but is seen across distinct paradigms. It therefore seems appropriate to regard mixed responding as the earliest stage of CF development. It is notable that mixed responding – both in the MIST and the SwIFT-3 – was associated with poorer executive functions. This may well help to explain why Mixed Responders performed unsystematically after the rule changed: their relatively weak executive function meant that these children were unable to maintain *either* the current sorting rule *or* the previous sorting rule. As a result, their post-switch phase sorting behaviour was unsystematic, and lacking in strong top-down control. Together, these results suggest that earlier views of CF development as transitioning from perseveration to switching between the ages of three and four years mischaracterise how CF develops, missing out an important initial part of this development. (It is also important to note that the three-stage pattern of CF development reported in Chapter two was replicated in the present study. This is important, as it increases our confidence that this is a genuine and robust finding.)

One intriguing aspect of the results of the present chapter relates to the surprisingly low levels of perseveration on the MIST. Previous research strongly indicates that perseverative responding is a common hallmark of CF during development. However, the low level of perseveration seen on the MIST – a task that requires children to update their sorting rule, and which should therefore be prone to perseverative responding – suggests that the incidence of perseveration in CF development might have been overstated, or at least mischaracterised. Performance on both the SwIFT-3 and the MIST were both best explained by a three-stage developmental pattern, and for both tasks, one of those three patterns was best described as representing perseveration. However, the proportion of children who perseverated on the MIST (around 18%) was far smaller than those who perseverated on the SwIFT-3 (around 30%). This is also in contrast to previous research reporting high levels of perseveration on CF tasks, especially in 3-year-olds (Muller et al., 2006; Zelazo, 2006). How should this difference in levels of perseveration be explained? One possible explanation relates to the different number of response options presented on the sorting array in each task. Specifically, it may be that the more competing responses there are on a task, the less salient any individual response option will be – and responses that are less salient are less likely to be selected. The SwIFT-3 presents children with three response options, whereas the MIST presents them with nine options. While searching for the correct response among three response options on the SwIFT-3, the conflicting information in the array is likely to be highly salient, and a strong competitor for selection, and could induce children to incorrectly select the conflicting stimulus. In contrast, while searching for the correct response among nine response options on the MIST, the conflicting responses may be less salient, and thus less likely to capture children's attention. If this view is correct, then it would seem to suggest that the larger the task array, the less likely children would be to make consistent perseverative responses. If that were the case, then it would appear that the degree of perseveration in CF development may be as much a function of the task used to measure CF as it is of the cognitive ability of the child.

The present results showed that executive functions play an important role in CF development, and more specifically that switching on the two tasks was underpinned by different executive functions. Successful switching on the MIST was underpinned by inhibitory control, whereas successful switching on the SwIFT-3 was underpinned by working memory. Again, this difference may be related to the different task arrays, and potentially to the role of attentional processes. The SwIFT-3 has a small task array, with three response options. As the number of response options was low, attention given to each individual response option on the SwIFT-3 was comparatively high. When children were asked to switch rules on an array with small number of response options, the competition between competing responses might be resolved at the top-down cognitive level, by actively maintaining the current rule in mind and finding the correct response among small number of alternatives. Conversely, the MIST has a larger task array, with nine response options. As the number of response options was relatively high, children's attention given to each individual response option was lower than in the SwIFT-3. This could mean that when children are asked to find multiple correct responses within a large task array, remembering the current rule may not be sufficient to select correctly. Instead, inhibitory control may be required to help them suppress the increased amount of distracting information and competing responses in the array. This view is consistent with previous research on the role of attentional processes on working memory (Cowan, Morey, AuBuchon, Zwilling, & Gilchrist, 2010). In their study, Cowan and colleagues examined working memory capacity across two conditions that differed in terms of their attentional demands (e.g., remembering relevant objects on a 4-object array condition versus remembering relevant objects on 6-object array condition). They found that for the smaller array size (that is, the 4-object condition), children with better working memory capacity could better remember the relevant objects on the array. However, for larger arrays (the 6-object condition), children who were better able to suppress distractor objects better remember the correct objects on the sorting array. Overall, this suggests that when switching rules on larger arrays, inhibitory control is likely to play an important role in explaining successful switching behaviour.

The present study shows that the three-stage pattern of CF development does not merely replicate, but also extends across different tasks. While this advances our understanding of how CF develops, it also raises an important question: Do all children go through these three stages of CF development in the same fixed order, or would it be possible (say) for a child to move from the mixed responding stage to the switching stage? The present results showing that 2- to 4-year-olds engaged in three distinct performance patterns – both in the MIST and in the SwIFT-3 – were based on cross-sectional studies. As such, these data cannot tell us whether children go through each of these stages in order – though they do allow us to speculate on this point. Based on the lower than expected levels of perseveration reported on the MIST, it may well be that perseveration is not a phase that all children have to pass through on all tasks. The data from the MIST suggest that in some situations, children can switch rules without necessarily showing perseverative responding, meaning that it is quite possible that some children may not show perseverative responding at all. This suggestion is consistent with findings from previous studies. For example, Deák and colleagues assessed children's CF using three different tasks in order to see whether children performed similarly on each of the three tasks (Deák & Wiseheart, 2015). They found that while some perseveration was reported on two of the tasks, children did not perseverate on the third task, consistent with the idea that perseveration is not an inherent part of CF development. So it may be that on some measures of CF, we will see reduced (or potentially even no) perseveration. However, for the CF tasks that do show perseverative responding, we cannot yet be sure whether the three performance patterns seen on these tasks also reflect three developmental stages that all children pass through. It could be that children progress with age from showing mixed responding, to perseveration, to successful switching. Or it could be that before children achieve reliable successful switching, their errors vary freely between unsystematic responding and perseverative responding. An ideal way to address this question would be to conduct a longitudinal study in which children's CF is assessed at regular intervals from toddlerhood to the end of preschool age. This would allow us to better understand whether the different patterns of responding reflect stable developmental stages, or whether the developmental trajectory of CF is more fluctuating.

To sum up, the present study showed that the three-stage pattern of CF development reported in Chapter two was not merely a task artefact, but rather generalised across two different CF tasks. On both the MIST and the SwIFT-3, the development of CF was best explained by three distinct performance patterns: mixed responding, perseveration, and switching. While this pattern was broadly consistent across the two tasks, there were nevertheless some important differences between them. Notably, there was much less perseveration on the MIST than one would expect (either from the SwIFT-3, or from the broader CF literature). It may be that perseveration is in part a function of the CF task chosen, rather than being an inherent and inevitable part of CF development itself. In addition, although both working memory and inhibitory control supported key aspects of CF development, successful switching was underpinned by different executive functions across different tasks. On the SwIFT-3, successful switching was underpinned by working memory, while on the MIST, successful switching was underpinned by inhibitory control. These results are broadly consistent with the idea that as the number of possible responses on the sorting array increased, the role of inhibitory control becomes increasingly important. However, although we have increased confidence in the view that CF develops as a three-stage pattern, it remains to be determined whether all children go through all these stages in the same order. Longitudinal studies are likely to be particularly helpful in addressing this question in future.

Chapter Four

General Discussion

The aim of this research was to better characterise how CF develops in early childhood, by taking two main approaches: First, children's CF was assessed with tasks that would allow children to make a greater variety of responses, by testing rule-switching on multi-item arrays, in the presence of both conflicting information and distracting information. Second, children's CF was assessed in a wider and younger age range, from two to four years of age. Previous studies have typically characterised CF as emerging between the ages of three and four years, when children gradually learn to overcome perseveration (Munakata et al., 2012). However, the research in this thesis instead suggests that CF in early childhood has a more protracted developmental trajectory than previously thought, and one that involves more than merely overcoming perseveration.

This chapter will first give an overview of the main findings from the two empirical studies, and will discuss how the research has helped us improve our understanding of how CF develops in early childhood. The implications of these findings for the broader literature are then discussed, and the chapter will finish with suggestions for future research.

4.1. Summary of the main findings

Chapter two aimed to better characterise the development of CF, by using a novel measure of CF that allowed children to produce a greater variety of potential responses. Two- to 4-year-old children were tested on the SwIFT-4, a task that required children to switch rules in the presence of *both* conflicting information and distracting information (including two types of distracting stimuli: a goal-unrelated distractor and a goal-related distractor). When asked to switch rules, therefore, children could make one of three different kinds of response: they could switch rules successfully; they could make a perseverative error, by selecting the conflicting stimulus and thus continuing to sort by the previous rule; or they could make a distraction error, by selecting a stimulus unrelated to either task rule. The study also tested how individual differences in executive functions contribute to CF. I was able to use LCA, which allowed me to explore the extent to which coherent performance patterns emerged in children's switching behaviour. The results from the LCA showed that the development of CF was best explained by three distinct performance patterns, characterised as follows: *mixed responding*, in which children sorted unsystematically between the previous rule and the current rule; *perseveration*, in which children continued to sort based on the previous (but no-longer relevant) rule; and *switching*, in which children successfully sorted by the current rule. The next step was to see if these performance patterns were related to children's age, to see whether these patterns comprised distinct stages of development. When I examined how these performance patterns varied by age, it was found that most 2-year-olds engaged in mixed responding, most 3-year-olds engaged in perseveration, and most 4-year-olds engaged in switching. In addition, executive functions contributed to later development in CF, but not to earlier development in CF. That is, Switchers had significantly better working memory and inhibitory control than Perseverators, but there was no difference in executive functions between Mixed Responders and Perseverators. This suggests that in order to switch rules successfully, children need both to maintain the current sorting rule in mind and also to suppress attending to information related to the previous sorting rule. The results from Chapter two suggest that the development of CF is stage-like, and more protracted than previously thought; and that while executive functions underpinned some of the key developments in CF, there appear to be other important aspects of CF development that cannot be accounted for by executive functions.

Chapter three aimed to determine whether the three-stage pattern of CF development generalised beyond a single task. This was an important question, since the prevailing view of how CF develops – shifting from perseveration to switching – has largely been based on findings from a single paradigm (the DCCS; Zelazo, 2006). To increase confidence that the findings of Chapter two are reliable, it was important to see whether this finding generalised beyond a single paradigm. To address this, 2- to 4-year-olds were tested on two different CF tasks – one that had been used before (the SwIFT-3) and one entirely novel task designed specifically for this study (the MIST). These tasks shared the same core demand, of having to switch rules in the presence of conflicting and distracting information. In this sense, they were both conventional CF tasks. However, they had intentionally very different surface features: the MIST had a larger sorting array, of nine items, from which children had to select three responses. The tasks also varied in terms of the stimuli, instructions and feedback. I also explored how individual differences in executive functions contributed to CF performance. The results showed that performance on both tasks was best explained by three distinct performance patterns, and crucially, that these patterns were similar on both tasks. Thus, the three-stage pattern of CF development generalised beyond a single task, suggesting that the findings of Chapter two really can inform us about CF itself. However, there was an important difference between the two tasks. Surprisingly, when within-child performance was examined across the two tasks, there was some divergence in terms of what performance patterns children produced on the different tasks. That is, while *mixed responding* and *switching* broadly generalised across tasks – in other words, a child who was classed as a Mixed Responder or a Switcher on the SwIFT-3 was very likely to be classed in the same way on the MIST – perseveration did not appear to generalise in the same way. In other words, most children who perseverated on the

SwIFT-3 did not perseverate on the MIST – and instead there was a much lower incidence of perseveration on the MIST which was only 18%.

A further noteworthy finding from Chapter three was that executive functions made significant contributions to CF, but these contributions differed by age and by task. Working memory and inhibitory control both contributed to the early development in CF, with a significant difference in working memory and inhibitory control between Mixed Responders and Perseverators. That is, working memory and inhibitory control were poorer in children classed as Mixed Responders compared to Perseverators and Switchers. This suggests that mixed responding may in part be driven by the inability children have to maintain the current rule in working memory, and the inability to suppress attention to information related to the previous rule. However, the contribution of executive functions to the later development of CF was slightly more complex. The difference between Perseverators and Switchers was explained by working memory on the SwIFT-3, but was explained by inhibitory control on the MIST. These results suggest that although we do see similarities between the two tasks – notably that both yield three similar performance patterns – there are also important differences between them. These differences lead to differential contributions of individual executive functions. This raises the possibility that our best explanations for how CF develops may well vary as a function of the task you choose to measure CF.

4.2. How do these findings contribute to our understanding of the development of cognitive flexibility?

In the opening chapter of this thesis, I identified two gaps that needed to be addressed in order to accurately characterise the development of CF in early childhood. First, CF needed to be assessed by less constrained tasks that allowed children to make a greater range of responses. Second, CF needed to be investigated in a wider and younger age range in order to better understand how CF first emerges. In this section, I explain how the research in this thesis has improved our understanding of the development of CF, by addressing these two gaps. Overall, the thesis suggests that CF has a more protracted developmental trajectory that starts to emerge at around two years of age, with 2-year-olds showing unsystematic mixed responding when they try to change sorting rules. In addition, there are grounds for thinking that perseveration may not always be as common a behaviour as previous research might suggest, and instead that it may arise as a function of the task used to measure CF. These ideas are now explained in more detail.

Findings from Chapter two and Chapter three improved our understanding of the development of CF by indicating that CF has a stage-like developmental trajectory that begins from two years of age. By using more sophisticated tasks and more advanced statistical methods, it was shown that children's CF development involves three qualitatively distinct patterns of performance: mixed responding, perseveration, and switching. The age-related analyses further suggest that these stages may progress in this order, as 2-year-olds tended to be Mixed Responders, 3-year-olds tended to be Perseverators, and 4-year-olds tended to be Switchers. I would suggest that the key difference between these performance patterns is the level of top-down control involved. Specifically, in mixed responding, after a rule change, children neither flexibly switch to the current rule, nor do they systematically continue to sort with the previous rule. Rather, their post-switch sorting behaviour appears to vary unsystematically between both available rules. From here, CF develops to the perseveration stage. While children in this stage are still not able to flexibly switch to a current sorting rule, their post-switch performance is systematic, as they continue to sort by the previous rule. Thus, while they are not able to switch rules, Perseverators can still produce stable, systematic, rule-driven behaviour. In that sense, children who perseverate appear to demonstrate better top-down control than Mixed

Responders. Finally, in the switching stage, children become able to switch from the previous rule to the current rule.

Perhaps the most important finding from the research in this thesis is that the earliest developmental stage in CF development is mixed responding – in other words, when young children are asked to switch from one rule to another, they are initially not only unable to switch to the current rule, but they also seem to lose the ability to maintain the previous rule. Previous research has typically considered perseveration as the starting point of CF development (Bunge, Kahn, Wallis, Miller, & Wagner, 2003; Muller et al., 2006; Munakata et al., 2012). However, the present research showed that CF development starts before that, with 2-year-olds' post-switch behaviour involving unsystematic fluctuating between the two available rules. Mixed responding was usually experienced by the youngest children in the sample. This was true for both Chapter two and Chapter three, and was consistent across different tasks (the SwIFT-4, the SwIFT-3, and the MIST). There was also notable within-individual consistency, with children classed as Mixed Responders on the SwIFT-3 also likely to be Mixed Responders on the MIST. Therefore, there are clear grounds for suggesting that mixed responding is the earliest stage in the development of CF.

These results support recent previous research that has used the SwIFT with a wider and younger age range and have identified this three-stage pattern of CF development which starts with mixed responding in 2-year-olds (Blakey et al., 2016). However, importantly the present results go beyond that work. Specifically, the results presented in this thesis show that these three performance patterns also emerge when children are presented with a less constrained task with both conflicting *and* distracting information. This approach was particularly helpful in telling us more about the first stage of CF development, mixed responding. In particular, by using this less constrained paradigm we could better understand whether mixed responding reflects a *complete absence* of top-down control (as children would select all response options randomly regardless of whether they were conflicting or distracting) or whether children who show mixed responding are pulled more towards the conflicting response option and thus are evidencing some memory of the previous no longer relevant rule.

The first proposed stage of CF development, mixed responding, was related to poorer working memory and inhibitory control across all CF tasks used in this research. Specifically, in all tasks, Mixed Responders were found to have poorer working memory and inhibitory control than Switchers. In addition, in the SwIFT-3 and in the MIST, Mixed Responders were also found to have poorer working memory and poorer inhibitory control than Perseverators. Poor executive functions may well help to explain why Mixed Responders had difficulty in reliably switching to a current rule. Mixed Responding by definition might appear to reflect weak top-down control, since it is characterised by children being not only unable to reliably establish the current rule, but also being unable to maintain the previous rule that they had just sorted by. Executive functions in general – and working memory in particular – are likely to be integral in allowing children to maintain and update sorting rules (as also found in Blakey et al., 2016). Nevertheless, it is not the case that Mixed Responders show a *complete* lack of top-down control. In Chapter two, children were presented with four response options on the SwIFT-4, of which the two distracting responses did not match the prompt stimulus by the previous rule or by the current rule. Mixed Responders did not select all four possible responses equally often, as one might expect if they were entirely lacking top-down control. Instead, they were much more likely to select response options that matched either the previous rule (the conflicting response) or the current rule (the correct response), than to select response options unrelated to either rule. This suggests that Mixed Responders are capable of rudimentary top-down control – though this is very limited.

There was an unexpected finding relating to between-task differences in the levels of perseveration observed. Far more children produced a perseverative response pattern on the SwIFT-3 than they did on the MIST. This is something of a surprise, given that both tasks require children to ignore conflicting responses in order to sort according to a current rule. This has consistently been shown to be difficult for preschool children (Diamond et al., 2005; Espy, 1997; Zelazo, 2006), so it is interesting to note the lesser levels of perseveration seen on the MIST. This between-task divergence also raises the question of how one should think about perseveration. According to the results of Chapter two, most 2-year-olds were Mixed Responders, most 3-year-olds were Perseverators, and most 4-year-olds were Switchers. Thus, perseveration would appear to be an intermediate stage in CF development, reflecting a partial increase in children's top-down control. However, some results from Chapter three indicate that caution may be needed in characterising perseveration as a robust developmental stage. Chapter three reported that contrary to what one would expect from prior research, there were only 18% of children engaged in perseveration on the MIST. Moreover, while mixed responding and switching generalised across tasks and within-individual, perseveration did not: around 60% of children who perseverated on the SwIFT-3 did not perseverate on the MIST. If it were the case that perseveration was a robust developmental stage, then it would be surprising if children who clearly perseverated on the SwIFT were able to produce a quite different response pattern on a separate task intending to measure the same construct. And yet Chapter three reported this very pattern of results. What should we conclude from these findings?

The results of Chapters two and three clearly show that perseveration *is* a common pattern of responding on all three tasks used in this thesis. In both chapters, one of the three performance patterns identified by LCA was characterised by children sorting almost entirely by the previous (and no longer correct) task rule. So it does appear that perseveration is a common response pattern – one that appears to arise after mixed responding, and before successful switching. However, there was a striking reduction in the number of children who perseverated on the MIST, compared to the same children's performance on the SwIFT-3. Perseveration was present on this task, but it was far less common than on other CF measures (for a meta-analysis, see Doebel & Zelazo, 2015). This is an important finding, since it indicates that perseveration is not simply a function of the child's own stage of development. Rather, this suggests that the degree to which a child perseverates is determined, in part, by the features of the situation in which they must switch rules. When asked to switch rules, the same child may perseverate with the previous rule on the SwIFT-3, but show mixed responding, or successful switching, on the MIST. So while perseveration does appear to be an intermediate stage in the development of CF, it is not the case that all tasks index this equally – or that a child will perform in the same way across different tasks. It seems entirely possible that a child may appear to be at a different stage of CF development depending on the task. This also raises the possibility that, on some tasks, children may show a developmental progression that bypasses perseveration altogether. It may be that on tasks like the MIST, children may progress from mixed responding directly to successful switching. The present data cannot definitively show whether this is the case. To do that, longitudinal research is needed, to examine transitions in switching performance over time. This future research suggestion will be discussed in more detail later on in the chapter.

Another key finding of this research is that executive functions are essential in explaining several distinct aspects of CF development. Both working memory and inhibitory control could explain differences between Perseverators and Switchers on the SwIFT-4 in Chapter two. These results are consistent with previous research showing that children who successfully switch rules have better working memory and inhibitory control than children who perseverate (Kirkham et al., 2003; Yerys & Munakata, 2006). However, the results in Chapter three go beyond this. In addition to contributing to the later development in CF through explaining the difference between perseveration and switching, both working memory and inhibitory control explained the *earlier* development in CF development, as they accounted for differences between Mixed Responders and Perseverators – in both the SwIFT-3 and the MIST. That is, children who perseverated with the previous rule both on the SwIFT-3 and on the MIST had better working memory and inhibitory control than children who showed mixed responding. These results go beyond previous research by suggesting that executive functions contribute to developments in CF from the outset, and that even before a child is able to successfully switch rules, their executive functions are nevertheless driving qualitative improvements in their top-down control.

The present findings have suggested that the existing theories of CF (e.g., attentional inertia account, graded representations account) might be insufficient to explain the three-stage pattern of CF development. Those theories generally seek to explain why children fail to switch rules and instead perseverate by the previous rule between the ages of three and four years. They propose that either poor inhibitory control or poor working memory is responsible for children's perseverating with the previous rule after the rule changed (Kirkham et al., 2003; Morton & Munakata, 2002; Yerys & Munakata, 2006). However, the results from the present research has offered that CF develops as a three-stage pattern in which children are likely to transfer from mixed responding at age two to perseveration at age three and to switching at age four. In light of the present findings, the existing theories seem to miss an important initial part of the development of CF, which is the transfer from mixed responding to perseveration between the ages of two and three years. Therefore, it has been suggested that the existing theories need to be revised as to include the initial part of CF development from mixed responding to perseveration in explaining how CF develops. While explaining the early development in CF from mixed responding to perseveration, in addition to working memory and inhibitory control, theories also need to consider other factors as potential mechanisms; such as representational abilities (e.g., vocabulary), and/or attentional control. This is because solely working memory and inhibitory control did not consistently explain the early development of CF in the present study. Instead, as recent research has shown, early vocabulary and attentional abilities are associated with better control of behaviour in many executive function paradigms, including CF (Devine & Hughes, 2019; Miller & Marcovitch, 2015). Therefore, the early development in CF from unsystematic mixed responding to systematic (but incorrect) perseveration can be explained by early vocabulary and early attentional control.

One limitation of the individual differences analyses in the present study is that they were underpowered for looking at how inhibitory control contributes to CF, particularly in Chapter three. Because there is not a single task that can be used to assess inhibitory control across 2-, 3- and 4-year-olds, I had to use separate measures of inhibitory control for the younger and older halves of my sample, which resulted in some rule-switching performance patterns having very few children in them. For example, when looking at how inhibitory control relates to CF in younger children, there were only five children who were categorised as Switchers. Therefore, there is a need for caution in interpreting how inhibitory control relates to CF in this thesis. Nevertheless, the reported associations between inhibitory control and the different performance patterns suggest that inhibitory control does play a role in helping children with CF – likely by enabling them to better suppress the previous rule – because Switchers had higher inhibitory control compared to Perseverators both on the SwIFT-4 and on the MIST. Inhibitory control also seems to help children better suppress rule-irrelevant distractors, since Perseverators had higher inhibitory control compared to Mixed Responders on the SwIFT-3.

However, although inhibitory control explained important aspects of CF development in Chapter two and Chapter three, it explained different parts of CF development across different CF tasks. Specifically, in Chapter two, inhibitory control contributed to the later development in CF through explaining the difference between perseveration and switching on the SwIFT-4 (see also Kirkham et al., 2003). In Chapter three, inhibitory control contributed to *earlier* development on the SwIFT-3, through explaining the difference between mixed responding and perseveration, though it contributed to *later* development on the MIST, through explaining the difference between perseveration and switching. It is not obvious why these differences would arise, since one would expect that inhibitory control would both help children to suppress attention to distracting information, and also help them to suppress attention to information related to the no longer relevant rule. It is possible that the lack of power for these analyses may explain the discrepancy in findings across the tasks. Another possible explanation for the difference in the role of inhibitory control in CF across the SwIFT-3 and the MIST in Chapter three could be explained by different number of response options on the sorting array across the two tasks. As the number of response options on the sorting array increased on the MIST, it is likely that children would need to use more inhibition to suppress the increased number of competing responses. That could explain why inhibitory control explained successful switching on the MIST, but not on the SwIFT-3.

Perhaps the most unexpected finding regarding executive functions was that working memory appeared to explain different aspects of CF development on the SwIFT-4 and the SwIFT-3. It was found that on the SwIFT-4, working memory explained the difference between Perseverators and Switchers, and on the SwIFT-3 it explained the difference between Mixed Responders and Perseverators. The present associations between working memory and different performance patterns suggest that working memory does help children with CF – through allowing them to better maintain the current rule – because Switchers had better working memory compared to Perseverators on the SwIFT-4 (see also Blakey et al., 2016; Chevalier et al., 2012; Marcovitch et al., 2007; 2010). Working memory also seems to allow children to better maintain systematic goal-oriented behaviour, because Perseverators (who systematically sorted based on a previous rule) had better working memory compared to Mixed Responders (who unsystematically sorted between both available rules) on the SwIFT-3. However, although working memory explained important aspects of CF development both in the SwIFT-4 and in the SwIFT-3, it explained different parts of CF development across the two tasks. This is surprising because these tasks are very similar, and there are no *a priori* reasons why one would expect working memory to contribute differently between tasks. One possible explanation might be the methodological difference in how working memory was assessed. In Chapter two, working memory was assessed using aggregated scores from multiple tasks, whereas in Chapter three, it was assessed with a single task. However, this does not seem sufficient to explain the difference. In Chapter two, I combined scores from the Spatial Span task and the Spin the Pots task to give a single working memory score – whereas in Chapter three, I only used scores from the Spatial Span task. However, when the analyses in Chapter two are re-run with only Spatial Span scores – as in Chapter three –, the difference between Mixed Responders and Perseverators in the SwIFT-4 was still not significant. Thus, even when this methodological difference is removed, the difference in findings remains. Therefore, it can be suggested that the role of working memory in CF might differ depending on the CF measure chosen to assess CF. Still, further investigation is needed to better elucidate the differential role of working memory in different CF tasks, probably using more diverse measures of working memory that would allow us to examine potentially diverse dynamics of this association.

4.3. Future research directions

The research in this thesis has advanced our understanding of how CF develops. However, a number of important questions remain to be addressed. First, while there is increasing evidence suggesting that CF has a three-stage developmental pattern, we do not know whether children pass through all three performance patterns in the same order. Specifically, I found that the three-stage pattern of CF development seems robust across tasks, and that it relates to a child's age, with children showing increasing top-down control between the ages of two and four years. However, to be sure that this is truly a developmental progression, we would need longitudinal data tracking children over time. Indeed, we do not yet know whether performance patterns are stable within child at a given point in development, or if the same child might show different patterns of performance from day to day. Second, there is strong evidence that working memory and inhibitory control make important contributions to CF, though the nature of these contributions remains to be fully elucidated. Third, the early development of CF between the ages of two and three years – when children move from mixed responding to perseveration – is a relatively new area of investigation, and so further exploring the factors that contribute to CF development at this age would be very informative. These points are discussed further below.

First, it is important to determine whether these three performance patterns in CF development reflect three stable and successive stages of development. I have speculated that children go through each of these patterns in order (mixed responding, perseveration, successful switching), since the mean age of children in each performance category is consistent with a chronological improvement. However, it may instead be that there are individual differences in CF development, whereby some children may transfer from mixed responding to switching without an intermediate perseveration stage. Alternatively, it may be that rule switching is a highly variable skill, and a single child might be capable of producing different performance patterns at a single time-point. The ideal way to assess this would be to conduct a longitudinal study in which children's CF is assessed at regular intervals, from toddlerhood to the end of the preschool period. I designed a longitudinal study on these lines, though unfortunately I was not able to run the study due to COVID-19. The study would assess children's CF four times in total at four-month intervals, using both the MIST and the SwIFT-3 at each time point. Children would be tested at 30 months (time 1), 34 months (time 2), 38 months (time 3), and 42 months (time 4). This design would allow us to see in more detail whether children's CF shows a period of consistency (e.g., from Mixed Responding at time 1 to Perseveration at time 2 to Perseveration at time 3 to Switching at time 4), or whether it was more variable (e.g., from Mixed Responding at time 1 to Mixed Responding at time 2 to Switching at time 3 and to Switching at time 4). In addition, I aimed to test how growth rate (or stability) of executive functions across the four time points would be related to CF. I would hypothesise that, based on the results from Chapter three, we would see a broadly consistent pattern of mixed responding, followed by perseverating, followed by successful switching; however, I would also speculate that some children would skip perseveration entirely, transitioning directly from mixed responding to switching (a speculation broadly consistent with the low rate of perseveration on the MIST). In addition, given that working memory and inhibitory control helped children better switch rules (see also Blakey et al., 2016; Chevalier et al., 2012), I would speculate that children who reach successful switching quicker would have higher growth rates in executive functions across testing points. I hope to be able to conduct this study in the future.

Second, the role of executive functions remains to be fully elucidated. One reason why the present results were not straightforward to interpret is due to the use of different inhibitory control measures for the younger and older halves of the sample. Both in Chapter two and Chapter three of this thesis, separate tasks had to be used to assess inhibitory control in younger children (2.5- to 3-year-olds) and older children (3.5- to 4-year-olds). This is because there was no single task that could be used to assess inhibitory control across 2- to 4-year-olds. As a consequence, some performance patterns for inhibitory control analysis contained a relatively small number of data points (e.g., only five children were categorised as Switchers for 2.5- to 3-year-olds). It also meant that inhibitory control was being assessed in slightly different ways for the two groups, which could introduce small but important incidental differences. This makes it hard to offer strong conclusions about how inhibitory control contributes to CF. Therefore, it would be an important step forward to develop measures of inhibitory control that are effective across a wider age span. Recently, one possible such task was developed, called the Early Childhood Inhibitory Touchscreen Task (ECITT; Holmboe, 2021). This task can be used from infancy to adulthood. It is a simple response inhibition task in which participants are trained to touch the smiley face at the top of the screen. Once this response tendency is established, they then need to overcome this prepotent response in order to touch the smiley face at the bottom of the screen. According to initial results, the task was effective at measuring inhibitory control for 24- to 30-month-olds (Holmboe, 2021). That suggests that using the ECITT in future research may be a fruitful way to assess the role of inhibitory control in CF.

Third, it is important to further investigate the early development in CF, particularly the difference between mixed responding and perseveration. While the later development in CF – the difference between perseveration and switching – has received a lot of attention in prior research (Buss & Spencer, 2014; Cragg & Chevalier, 2012; Kirkham et al., 2003, Yerys &

Munakta, 2006), there has obviously been less attention paid to early development in CF. The present work makes a start on that, though as noted previously, there was a small sample size in some performance patterns to be able to look at how executive functions contributed to CF. Thus, it will be important for future research to further investigate the factors that underpin the early development in CF. These factors will likely include working memory and inhibitory control, though as present results indicate, there are also likely to be important contributing skills other than executive function.

In addition to working memory and inhibitory control, what else could explain the early development in CF from mixed responding to perseveration? To address this, I would suggest two potentially informative directions for future research: representational abilities (including language), and control of attention. Language and attention are plausible mechanisms to support the early emergence of CF because being able to verbally label a stimulus (or an aspect of a stimulus), as well as being able to selectively focus on the relevant part of a stimulus are likely to lead to stronger mental representations of key task information (Marcovitch & Zelazo, 2009). These stronger mental representations can help children to better control and direct their behaviour (Kuhn et al., 2016). There is some evidence to suggest that representational and attentional abilities during the first years of life can support young children's goal-oriented behaviours (Hendry et al., 2016; Wiebe et al., 2010). For example, Miller and Markovitch (2015) investigated the role of language and attention abilities at 14 months of age in predicting executive function scores at 18 months of age. The results showed that children's performance on a range of executive function tasks was positively associated with their language ability (as indexed by their early vocabulary) as well as with their attentional ability (as indexed by the frequency of directing attention to others' behaviours). Furthermore, associations between executive functions and attention have been reported in even younger samples. For example, Devine and colleagues (2019) found that attention abilities at four months of age predicted executive function abilities at 14 months of age (Devine et al., 2019). Given their demonstrated role in the emergence of executive functions, early representational abilities and control of attention are likely to potentially contribute to the early development in CF. Indeed, they may be crucial factors in explaining the transition from mixed responding to perseveration.

Finally, it is important to pre-register experimental studies and to acknowledge the value of open science practices. Pre-registering is important for many reasons, including enhancing the quality of research by receiving earlier feedback from the outset, increasing the chance of publishing research, and improving the chance of publishing null or negative results (Willroth, Graham, & Mroczek, 2021). However, the research in this thesis were not pre-registered because there was some challenges for the present studies that prevented me to pre-register my studies. For example, for pre-registering, we had to be entirely clear about the whole research process, including the analysis we aimed to make (Yamada, 2018). However, it was not certainly clear which modelling analysis would better suit to my research questions before I started my research because of the novelty of my CF paradigms. Therefore, I could not pre-register my experimental studies in my thesis. Nonetheless, I support the open science practices across many domains, including open data set, open tasks sharing, and open research publishing. I believe research should be open because it improves accessibility and visibility of research, which is likely to improve knowledge across many areas. Therefore, I aim to pre-register my studies and to make my entire research open in the future.

Overall, the research in this thesis helps us improve our understanding of CF development in two ways: *First*, CF has a stage-like and more protracted developmental trajectory than previously thought that starts at two years of age, with 2-year-olds showing unsystematic mixed responding when they are asked to switch rules. My research has found that CF is best characterised as involving three distinct performance patterns in 2- to 4-year-olds where children progress from mixed responding, to perseveration, and then to switching. In contrast to previous work in the field, rather than perseveration being the starting point of CF development instead it appears to be an intermediate pattern that some (but not all) children demonstrate. In addition, executive functions contribute to the important aspects of this broader developmental trajectory of CF, but their contribution varies according to tasks used to measure CF. *Second*, perseveration may not be always as frequently experienced as previous research might suggest; instead perseveration might emerge as a function of the task used to assess CF rather than being a stable or inherent part of CF development. Therefore, the focus of previous research on explaining the shift from perseveration to switching may be misguided and too narrowly focused. However, although this research has shown that CF develops through three performance patterns, it is still unclear whether all children experience these performance patterns in the same order or whether there are individual differences in development where some children might skip some performance patterns. To address this, longitudinal research is needed to examine children's CF from toddlerhood to the end of preschool in order to see whether these performance patterns are stable developmental stages that are experienced by all children.

References

Ackerman, D. J., & Friedman-Krauss, A. H. (2017). Preschoolers' executive function: Importance, contributors, research needs and assessment options. *ETS Research Report Series*, *2017*(1), 1-24. doi:10.1002/ets2.12148

Atkinson, J., Hood, B., Wattam-Bell, J., & Braddick, O. (1992). Changes in infants' ability to switch visual attention in the first three months of life. *Perception*, *21*(5), 643-653. doi:10.1068/p210643

Akaike, H. (1987). Factor analysis and AIC. In *Selected Papers of Hirotugu Akaike* (pp. 371-386). Springer, New York, NY.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, *29*(3), 180-200. doi:10.1016/j.dr.2009.05.002

Birgisdóttir, F., Gestsdóttir, S., & Thorsdóttir, F. (2015). The role of behavioral self-regulation in learning to read: A 2-year longitudinal study of Icelandic preschool children. *Early Education and Development*, *26*(5-6), 807-828. doi:10.1080/10409289.2015.1003505

Blackwell, K. A., Cepeda, N. J., & Munakata, Y. (2009). When simple things are meaningful: Working memory strength predicts children’s cognitive flexibility. *Journal of Experimental Child Psychology*, *103*(2), 241-249. doi:10.1016/j.jecp.2009.01.002

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false-belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, 78, 647–63. doi:10.1111/j.1467-8624.2007.01019.x

Blakey, E., & Carroll, D. J. (2018). Not all distractions are the same: Investigating why preschoolers make distraction errors when switching. *Child Development*, *89*(2), 609 - 619. doi:10.1111/cdev.12721

Blakey, E., Visser, I., & Carroll, D. J. (2016). Different executive functions support different kinds of cognitive flexibility: Evidence from 2-, 3-, and 4-year-olds. *Child Development*, *87*(2), 513 - 526. doi:10.1111/cdev.12468

Blaye, A., & Chevalier, N. (2011). The role of goal representation in preschoolers' flexibility and inhibition. *Journal of Experimental Child Psychology, 108*(3), 469-483. doi:10.1016/j.jecp.2010.09.006

Bunge, S. A., Kahn, I., Wallis, J. D., Miller, E. K., & Wagner, A. D. (2003). Neural circuits subserving the retrieval and maintenance of abstract rules. *Journal of Neurophysiology*, *90*(5), 3419-3428. doi:10.1152/jn.00910.2002

Bull, R., Espy, K. A., Wiebe, S. A., Sheffield, T. D., & Nelson, J. M. (2011). Using confirmatory factor analysis to understand executive control in preschool children: Sources of variation in emergent mathematic achievement. *Developmental Science, 14*(4), 679-692. doi:10.1111/j.1467-7687.2010.01012.x

Buss, A. T., & Spencer, J. P. (2014). The emergent executive: A dynamic field theory of the development of executive function. *Monographs of the Society for Research in Child Development*, *79*(2), vii. doi:10.1002/mono.12096

Carlson, S., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology*, *40*(6), 1105 – 1122. doi:10.1037/0012-1649.40.6.1105

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595-616. doi:10.1207/s15326942dn2802_3

Carroll, D. J., Blakey, E., & FitzGibbon, L. (2016). Cognitive flexibility in young children: Beyond perseveration. *Child Development Perspectives*, *10*(4), 211-215. doi:10.1111/cdep.12192

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, *13*, 195-212. doi:10.1007/BF01246098

Cepeda, N. J., & Munakata, Y. (2007). Why do children perseverate when they seem to know better: Graded working memory, or directed inhibition? *Psychonomic Bulletin & Review*, *14*(6), 1058-1065. doi:10.3758/BF03193091

Chevalier, N., & Blaye, A. (2008). Cognitive flexibility in preschoolers: The role of representation activation and maintenance. *Developmental Science*, *11*(3), 339–353. doi:10.1111/j.1467-7687.2008.00679.x

Chevalier, N., & Blaye, A. (2009). Setting goals to switch between tasks: effect of cue transparency on children's cognitive flexibility. *Developmental Psychology*, *45*(3), 782. doi:10.1037/a0015409

Chevalier, N., Blaye, A., Dufau, S., & Lucenet, J. (2010). What visual information do children and adults consider while switching between tasks? Eye-tracking investigation of cognitive flexibility development. *Developmental Psychology*, *46*(4), 955. doi:10.1037/a0019674

Chevalier, N., Sheffield, T. D., Nelson, J. M., Clark, C. A.C., Wiebe, S. A., & Espy, K. A. (2012). Underpinnings of the costs of flexibility in preschool children: The roles of working memory and inhibition. *Developmental Neuropsychology*, *37*(2), 99–118. doi:10.1080/87565641.2011.632458

Clark, C. A., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Development*, *84*(2), 662-677. doi:10.1111/j.1467-8624.2012.01854.x

Clearfield, M. W., Diedrich, F. J., Smith, L. B., & Thelen, E. (2006). Young infants reach correctly in A-not-B tasks: On the development of stability and perseveration. *Infant Behavior and Development*, *29*(3), 435-444. doi:10.1111/j.1467-8624.2006.00955.x

Clearfield, M. W., Dineva, E., Smith, L. B., Diedrich, F. J., & Thelen, E. (2009). Cue salience and infant perseverative reaching: Tests of the dynamic field theory. *Developmental Science*, *12*(1), 26-40. doi:10.1111/j.1467-7687.2008.00769.x

Clearfield, M. W., & Niman, L. C. (2012). SES affects infant cognitive flexibility. *Infant Behavior and Development*, *35*(1), 29-35. doi:10.1016/j.infbeh.2011.09.007

Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E., & Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science, 13*(1), 120-133. doi:10.1111/j.1467-7687.2009.00864.x

Cragg, L., & Chevalier, N. (2012). The processes underlying flexibility in childhood. *Quarterly Journal of Experimental Psychology, 65*(2), 209-232. doi:10.1080/17470210903204618

Cragg, L., & Nation, K. (2009). Shifting development in mid-childhood: The influence of between-task interference. *Developmental Psychology, 45*(5), 1465. doi:10.1037/a0015360

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078. doi:10.1016/j.neuropsychologia.2006.02.006

Dauvier, B., Chevalier, N., & Blaye, A. (2012). Using finite mixture of GLMs to explore variability in children's flexibility in a task-switching paradigm. *Cognitive Development*, *27*(4), 440-454. doi:10.1016/j.cogdev.2012.07.004

Deák, G. O. (2003). The development of cognitive flexibility and language abilities. In R. V. Kail (Ed.), Advances in child development and behavior (Vol. 31, pp. 271–327). San Diego, CA: Academic Press.

Deák, G. O., & Narasimham, G. (2003). Is perseveration caused by inhibition failure? Evidence from preschool children's inferences about word meanings. *Journal of Experimental Child Psychology, 86*(3), 194-222. doi:10.1016/j.jecp.2003.08.001

Deák, G. O., & Wiseheart, M. (2015). Cognitive flexibility in young children: General or task-specific capacity? *Journal of Experimental Child Psychology*, 138, 31-53. doi:10.1016/j.jecp.2015.04.003

Devine, R. T., Ribner, A., & Hughes, C. (2019). Measuring and predicting individual differences in executive functions at 14 months: a longitudinal study. *Child Development, 90*(5), e618-e636. doi:10.1111/cdev.13217

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development*, 868-883. doi:10.2307/1130099

Diamond, A., & Kirkham, N. (2005). Not quite as grown-up as we like to think: Parallels between cognition in childhood and adulthood. *Psychological Science, 16*(4), 291-297. doi:10.1111/j.0956-7976.2005.01530.x

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do". *Developmental Psychobiology, 29*(4), 315-334. doi:10.1002/(SICI)1098-2302(199605)29:4<315::AID-DEV2>3.0.CO;2-T

Diamond, A. (2013). Executive functions. *Annual Review of Psychology,* 64, 135–168. doi:10.1146/annurev-psych-113011-143750

Diamond, A., Carlson, S., & Beck, D. (2005). Preschool children's performance in task switching on the dimensional change card sort task: Separating dimensions aids the ability to switch. *Developmental Neuropsychology*, 28(2), 689-729. doi:10.1207/s15326942dn2802_7

Doebel, S., & Zelazo, P. D. (2015). A meta-analysis of the Dimensional Change Card Sort: Implications for developmental theories and the measurement of executive function in children. *Developmental Review, 38,* 241-268. doi:10.1016/j.dr.2015.09.001

Ellis, A. E., & Oakes, L. M. (2006). Infants flexibly use different dimensions to categorize objects. *Developmental Psychology*, *42*(6), 1000. doi:10.1037/0012-1649.42.6.1000

Espy, K. A. (1997). The Shape School: Assessing executive function in preschool children. *Developmental Neuropsychology*, *13*(4), 495-499. doi:10.1080/87565649709540690

Espy, K. A., Bull, R., Martin, J., & Stroup, W. (2006). Measuring the development of executive control with the shape school. *Psychological Assessment*, *18*(4), 373–381. doi:10.1037/1040-3590.18.4.373

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223-241. doi:10.1177/1745691612460685

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. doi:10.3758/BF03193146

Fryer, L. K. (2017). (Latent) transitions to learning at university: A latent profile transition analysis of first-year Japanese students. *Higher Education*, *73*(3), 519–537. doi:10.1007/s10734-016-0094-9

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60. doi:10.1037/0033-2909.134.1.31

Garon, N., Smith, I. M., & Bryson, S. E. (2014). A novel executive function battery for preschoolers: Sensitivity to age differences. *Child Neuropsychology*, *20*(6), 713-736. doi:10.1080/09297049.2013.857650

Geiser, C. (2012). Data analysis with Mplus. Guilford press.

Hendry, A., Jones, E. J., & Charman, T. (2016). Executive function in the first three years of life: Precursors, predictors and patterns. *Developmental Review*, *42,* 1-33. doi:10.1016/j.dr.2016.06.005

Holmboe, K., Fearon, R. P., Csibra, G., Tucker, L. A., & Johnson, M. H. (2008). Freeze-Frame: A new infant inhibition task and its relation to frontal cortex tasks during infancy and early childhood. *Journal of Experimental Child Psychology*, *100*(2), 89-114. doi:10.1016/j.jecp.2007.09.004

Holmboe, K. (2021). *Investigating the Development of Inhibitory Control with the Early Childhood Inhibitory Touchscreen Task*. Society for Research in Child Development (SRCD), 7- 9 April, Minnesota(Virtual).

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*(6), 1447–1459. doi:10.1037/0012-1649.43.6.1447

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions?. *Child and Adolescent mental Health*, *7*(3), 131-142. doi:10.1111/1475-3588.00024

Hunnius, S., Geuze, R. H., & Van Geert, P. (2006). Associations between the developmental trajectories of visual scanning and disengagement of attention in infants. *Infant Behavior and Development*, *29*(1), 108-125. doi:10.1016/j.infbeh.2005.08.007

Jennings, K. D. (2004). Development of goal-directed behaviour and related self-processes in toddlers. *International Journal of Behavioral Development*, *28*(4), 319-327. doi: 10.1080/01650250444000036

Johansson, M., Marciszko, C., Brocki, K., & Bohlin, G. (2016). Individual differences in early executive functions: A longitudinal study from 12 to 36 months. *Infant and Child Development, 25(6)*, 533-549. doi:10.1002/icd.1952

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience*, 3(4), 335-344. doi:10.1162/jocn.1991.3.4.335

Kam, C., Morin, A. J., Meyer, J. P., & Topolnytsky, L. (2016). Are commitment profiles stable and predictable? A latent transition analysis. *Journal of Management*, *42*(6), 1462-1490. doi:10.1177/0149206313503010

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, *6*(5), 449-467. doi:10.1111/1467-7687.00300

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*(2), 220–232. doi:10.1037//0012-1649.36.2.220

Kuhn, L. J., Willoughby, M. T., Vernon-Feagans, L., Blair, C. B., & Family Life Project Key Investigators. (2016). The contribution of children’s time-specific and longitudinal expressive language skills on developmental trajectories of executive function. *Journal of Experimental Child Psychology, 148,* 20-34. doi:10.1016/j.jecp.2016.03.008

Kulke, L., Atkinson, J., & Braddick, O. (2015). Automatic detection of attention shifts in infancy: Eye tracking in the fixation shift paradigm. *PloS One*, *10*(12), e0142505. doi:10.1371/journal.pone.0142505

Lanza, S. T., & Cooper, B. R. (2016). Latent class analysis for developmental research. *Child Development Perspectives*, *10*(1), 59-64. doi:10.1111/cdep.12163

Lerner, M. D., & Lonigan, C. J. (2014). Executive function among preschool children: Unitary versus distinct abilities. *Journal of Psychopathology and Behavioral Assessment, 36*(4), 626-639. doi:10.1007/s10862-014-9424-3

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767-778. doi:10.1093/biomet/88.3.767

Lonigan, C. J., Allan, D. M., Goodrich, J. M., Farrington, A. L., & Phillips, B. M. (2017). Inhibitory control of Spanish-speaking language-minority preschool children: Measurement and association with language, literacy, and math skills. *Journal of Learning Disabilities, 50*(4), 373-385. doi:10.1177/0022219415618498

Marcovitch, S., Boseovski, J. J., & Knapp, R. J. (2007). Use it or lose it: Examining preschoolers' difficulty in maintaining and executing a goal. *Developmental Science, 10*(5), 559-564. doi:10.1111/j.1467-7687.2007.00611.x

Marcovitch, S., Boseovski, J. J., Knapp, R. J., & Kane, M. J. (2010). Goal neglect and working memory capacity in 4-to 6-year-old children. *Child Development, 81*(6), 1687-1695. doi:10.1111/j.1467-8624.2010.01503.x

Marcovitch, S., O'Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cognitive Development, 33*, 40-55. doi:10.1016/j.cogdev.2014.07.001

Marcovitch, S., & Zelazo, P. D. (2009). A hierarchical competing systems model of the emergence and early development of executive function. *Developmental Science, 12*(1), 1-18. doi:10.1111/j.1467-7687.2008.00754.x

McLachlan, G. J., & Peel, D. (2000). Multivariate normal mixtures. Finite Mixture Models, 81-116. John Wiley & Sons.

Memisevic, H., & Biscevic, I. (2018). Exploring the link between inhibitory control and cognitive flexibility in preschool children. *Cognition, Brain, Behavior*, *22*(1), 1-11. doi:10.24193/cbb.2018.22.01

Miller, S. E., & Marcovitch, S. (2015). Examining executive function in the second year of life: Coherence, stability, and relations to joint attention and language. *Developmental Psychology, 51*(1), 101–114. doi:10.1037/a0038359

Miller, M. R., Muller, U., Giesbrecht, G. F., Carpendale, J. I., & Kerns, K. A. (2013). The contribution of executive function and social understanding to preschoolers' letter and math skills. *Cognitive Development*, *28*(4), 331-349. doi:10.1016/j.cogdev.2012.10.005

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-10. doi:10.1006/cogp.1999.0734

Monette, S., Bigras, M., & Lafrenière, M. A. (2015). Structure of executive functions in typically developing kindergarteners. *Journal of Experimental Child Psychology*, *140*, 120-139. doi:10.1016/j.jecp.2015.07.005

Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration, dissociation, and decalage. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *40*(3), 255-265. doi:10.1002/dev.10033

Muller, U., Liebermann-Finestone, D. P., Carpendale, J. I., Hammond, S. I., & Bibok, M. B. (2012). Knowing minds, controlling actions: The developmental relations between theory of mind and executive function from 2 to 4 years of age. *Journal of Experimental Child Psychology, 111*(2), 331-348. doi:10.1016/j.jecp.2011.08.014

Muller, U., Dick, A. S., Gela, K., Overton, W. F., & Zelazo, P. D. (2006). The role of negative priming in preschoolers' flexible rule use on the dimensional change card sort task. *Child Development*, *77*(2), 395-412. doi:10.1111/j.1467-8624.2006.00878.x

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences, 5*(7), 309-315. doi:10.1016/S1364-6613(00)01682-X

Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing cognitive control: Three key transitions. *Current Directions in Psychological Science*, *21*(2), 71–77. doi:10.1177/0963721412436807

Muthén, L.K.; Muthén, B.O. Mplus: Statistical Analysis with Latent Variables: User's Guide; Muthén & Muthén: Los Angeles, CA, USA, 2019.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535-569. doi:10.1080/10705510701575396

Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modelling: A Multidisciplinary Journal*, *21*(3), 439-454. doi:10.1080/10705511.2014.915375

Phipps, D. J., Hagger, M. S., & Hamilton, K. (2019, April 24). A Meta-Analysis of Implicit and Explicit Attitudes in Children and Adolescents. doi:10.31234/osf.io/52jrs

Piaget, J. (1954). *The construction of reality in the child.* (M. Cook, Trans.). Basic Books. doi:10.1037/11168-000

Simpson, A., & Riggs, K. (2005). Factors responsible for performance on the day–night task: Response set or semantics? *Developmental Science*, *8*(4), 360–371. doi:10.1111/j.1467-7687.2005.00424.x

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6(2),* 461-464.

Sclove S.L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333–343. doi:10.1007/BF02294360

van Bers, B. M. C. W., Visser, I., van Schijndel, T. J. P., Mandell, D. J., & Raijmakers, M. E. J. (2011). The dynamics of development on the dimensional change card sorting task. *Developmental Science, 14(5),* 960-971. doi:10.1111/j.1467-687.2011.01045.x

van Dijk, R., Deković, M., Bunte, T. L., Schoemaker, K., Zondervan-Zwijnenburg, M., Espy, K. A., & Matthys, W. (2017). Mother-child interactions and externalizing behavior problems in preschoolers over time: Inhibitory control as a mediator. *Journal of Abnormal Child Psychology, 45*(8), 1503-1517. doi:10.1007/s10802-016-0258-1

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society, 57*(2), 307-333. doi:10.2307/1912557

Wang, X., Liu, X., & Feng, T. (2021). The continuous impact of cognitive flexibility on the development of emotion understanding in children aged 4 and 5 years: A longitudinal study. *Journal of Experimental Child Psychology, 203,* 105018. doi:10.1016/j.jecp.2020.105018

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102*(1), 43. doi:10.1037/a0016738

Wiebe, S. A., & Bauer, P. J. (2005). Interference from additional props in an elicited imitation task: When in sight, firmly in mind. *Journal of Cognition and Development, 6*(3), 325-363. doi:10.1207/s15327647jcd0603_2

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*(2), 575. doi:10.1037/0012-1649.44.2.575

Wiebe, S. A., Lukowski, A. F., & Bauer, P. J. (2010). Sequence imitation and reaching measures of executive control: A longitudinal examination in the second year of life. *Developmental Neuropsychology*, *35*(5),522–538. doi:10.1080/87565641.2010.494751

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, 108, 436–452. doi:10.1016/j.jecp.2010.08.008

Williams, G. A., & Kibowski, F. (2016). Latent class analysis and latent profile analysis. *Handbook of methodological approaches to community-based research: Qualitative, Quantitative, and mixed methods* (pp. 143-151). Oxford University Press.

Willroth, E. C., Graham, E. K., & Mroczek, D. K. (2021). Challenges and opportunities in preregistration of coordinated data analysis: A tutorial and template. *Psychology and Aging*. Advance online publication. doi:10.1037/pag0000611

Yamada, Y. (2018). How to crack pre-registration: Toward transparent and open science. *Frontiers in Psychology*, *9*, 1831. doi:10.3389/fpsyg.2018.01831

Yerys, B. E., & Munakata, Y. (2006). When labels hurt but novelty helps: Children's perseveration and flexibility in a card-sorting task. *Child Development*, *77*(6), 1589-1607. doi:10.1111/j.1467-8624.2006.00961.x

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297–301. doi:10.1038/nprot.2006.46

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37-63. doi:10.1016/S0885-2014(96)90027-1