FROM PERSEVERATION TO FLEXIBILITY:
REFLECTION AND THE DOWN-REGULATION OF CONFLICT DETECTION
UNDERLYING EXECUTIVE FUNCTION DEVELOPMENT

by

Stacey Dominique Espinet

A thesis submitted in conformity with the requirements
for the degree of Doctorate of Philosophy
Graduate Department of Psychology
University of Toronto

© Copyright by Stacey D. Espinet 2012
FROM PERSEVERATION TO FLEXIBILITY:
REFLECTION AND THE DOWN-REGULATION OF CONFLICT DETECTION
UNDERLYING EXECUTIVE FUNCTION DEVELOPMENT

Stacey D. Espinet
Doctorate of Philosophy
Department of Psychology
University of Toronto
2013

Abstract

Introduction. Executive function refers to the top-down neurocognitive processes involved in flexible, goal-oriented behavior. A number of studies have shown positive effects of EF training. The overall aim of these studies was to explore the neurocognitive processes that support the development of EF by understanding how EF training works and what the active ingredients are. Particular interest was in isolating the role of reflection in EF training to understand its top-down affect on ACC-mediated conflict detection. Method. In Exp. 1 the neural markers of EF were explored by comparing ERPs of preschoolers who passed the DCCS and preschoolers who failed. Exp. 2 represents an attempt to replicate the key findings of Kloo & Perner, (2003, Exp. 2) that reflection training improves preschoolers’ performance on the DCCS and demonstrates far transfer. A shortened version of the training protocol was also tested (Exp. 3). In Exp. 4, the neural correlates of reflection training in preschoolers were explored by examining changes in the neural marker of EF found in Exp. 1. Results. In Exp. 1, the N2 amplitude was smaller (less negative) for children who passed the DCCS and were able to
efficiently resolve the conflict in the stimuli than for children who failed and were unable to resolve the conflict. Exp. 2 replicated the findings of Kloo & Perner, (2003, Exp. 2) even using a brief (15 min) intervention targeting reflection (Exp. 3). In Exp. 4, one brief session of reflection training made children who initially failed the DCCS look like children who initially passed at both the behavioral and neural level (reduced N2 amplitude). **Conclusion.** Results suggest that reflective processing facilitates the development of EF in young children by teaching them to notice conflict, reflect on it, and formulate rules for resolving it, resulting in the down-regulation of ACC-mediated conflict detection.
Dedications

I dedicate this thesis to my husband, mother and sister, whose love, devotion and patience made completion of this dissertation possible. Thank you to my two boys who provided me with constant sources of inspiration throughout this process. I am grateful to everyone at Breaking the Cycle for their support, in particular, Dr. Mary Motz and Dr. Debra Pepler, for their unfailing understanding and encouragement. I would especially like to acknowledge and extend my sincerest gratitude to my advisor, Dr. Philip Zelazo, for his ongoing assistance, support, and guidance over the years.
Acknowledgments

Funding for this research was provided by NSERC of Canada, the Canadian Foundation for Innovation, and the Canada Research Chairs program. I thank Andrea Hankins-Palmer, Angela Scafe, Yuan Jiang, Cheryl Leung, Jasmine Cady, Leif Stennes, Misbah Mohammad, Nhi Thai, Kristin Walsh, Shintula Wijeyakulasingam, Xenia Zheng, Katie Karbo, and Lindsey Higgins for their assistance in data collection. Special thanks to Jake Anderson for deriving the ERP source data. Also, thank you to Jake Anderson and Dr. Phil Zelazo for their assistance in editing earlier drafts of the papers included herein. I’d also like to acknowledge the families who participated in these studies and made this research possible, as well as the daycares that generously provided space and time for testing purposes.
Table of Contents

Abstract...................................................................................................................................ii – iii
Dedications..................................................................................................................................iv
Acknowledgments..........................................................................................................................v
Table of Contents............................................................................................................................vi
List of Tables..................................................................................................................................vii
List of Figures...................................................................................................................................viii – ix
List of Appendices..........................................................................................................................x
Chapter 1: General Introduction........................................................................................................1
Chapter 2: N2 Amplitude as a Neural Marker of Executive Function in Young Children: An ERP
       Study of Children who Switch vs. Perseverate on the Dimensional Change Card
       Sort.................................................................................................................................................21
Chapter 3: Reflection Training Improves Executive Function in Preschool-age Children:
       Behavioral and Neural Effects.................................................................................................53
Chapter 4: General Discussion...........................................................................................................95
List of Tables

Chapter 2

Table 1: Mean N2 Amplitude (μV) and (SDs) at Frontal-Midline Sites for Children who Passed vs. Failed the Post-switch Phase of the DCCS .................................................................46

Table 2: Mean N2 Latency (ms) and (SDs) for Children who Passed vs. Failed the Post-Switch Phase of the DCCS...............................................................................................47

Chapter 3

Table 1: Mean Percentage Correct and (SDs) on Pre- and Post-Training DCCS, False Belief Task, and Relative Clause Task, and Post-Pre Differences (Δ), for the Reflection and Relative Clause Training Groups in Exps. 2 and 3.........................................................98 – 88

Table 2: Change in Mean N2 Amplitude (μV) from Pre-training Day 1 to Post-training Day 2 and (SDs) by Training Condition and at Four Frontal-Midline Sites.................................89
List of Figures

Chapter 2

Figure 1. Sequence of events for the computerized version of the Dimensional Change Card Sort .................................................................49
Figure 2. Sequence of events for the computerized version of the Training Dimensional Change Card Sort.................................................................50
Figure 3. 128-Channel Geodesic Sensor Net layout with medial-frontal sites of interest highlighted.................................................................51
Figure 4. Grand-averaged stimulus-locked waveforms at sites: 16, 11, 6, and 129, recorded during children’s performance on the post-switch phase of the pre-DCCS day 1 and the post-DCCS day 2. Waveforms depict the difference in N2 amplitude between pre-training and post-training waveforms for children who received reflection training, corrective feedback, and mere practice........................................................................................................52

Chapter 3

Figure 1. Sequence of events for the computerized version of the pre- and post-training Dimensional Change Card Sort .................................................................91
Figure 2. Sequence of events for the computerized version of the training Dimensional Change Card Sort.................................................................92
Figure 3. Grand-averaged stimulus-locked waveforms at sites 16, 11, 6, and 129, recorded during children’s performance on the post-switch phase of the pre-training DCCS (Day 1) and the post-training DCCS (Day 2). Waveforms depict the difference in
N2 amplitude between pre-training and post-training waveforms for children who received reflection training, corrective feedback, and mere practice……………93

Figure 4. Topographic plots of differences in peak N2 amplitude (post-switch trials) between Day 1 and Day 2 and corresponding source distributions for the reflection training group. (b). Topographic plots of differences in peak N2 amplitude (post-switch trials) between Day 1 and Day 2 and corresponding source distributions for the corrective feedback group. (c). Topographic plots of differences in peak N2 amplitude (post-switch trials) between Day 1 and Day 2 and corresponding source distributions for the mere practice group…………………………………………………………………………………………………………………………94
List of Appendices

Chapter 3

Appendix A: Computerized Reflection Training Protocol……………………………………….84

Appendix B: Relative Clause Control Training Protocol……………………………………….85 – 86
FROM PERSEVERATION TO FLEXIBILITY: REFLECTION AND THE DOWN-REGULATION OF CONFLICT DETECTION UNDERLYING EXECUTIVE FUNCTION DEVELOPMENT

Chapter 1: General Introduction

Flexible, goal-oriented behavior, particularly in the face of conflict (due to habit, predisposition or conflicting information in the environment), is the hallmark of executive function (EF) (Zelazo, Carlson, & Kesek, 2008). A number of processes contribute to the flexible control of behavior including shifting, inhibition, and working memory (Miyake, Friedman, Emerson, Witzki, Howerter, & Wager, 2000), which become increasingly integrated over time (Garon, Bryson & Smith, 2008). The development of EF follows a protracted course, first emerging in late infancy and peaking in young adulthood (Best, Miller & Jones, 2009; Zelazo & Muller, 2002; Zelazo, Anderson, Richler, Wallner-Allen, Beaumont & Weintraub, 2011), that underlies transformation of the impulsive, stimulus-driven tendencies of young children (also patients with prefrontal damage; e.g., Bechara, Damasio, Damasio, & Anderson, 1994; Milner, 1963; Koenigs & Tranel, 2007), into the flexible, goal-directed actions of adults.

A period of rapid development of EF occurs during the preschool years (Zelazo et al., 2008) just as children face sharp increases in the demand on EF as they transition to school. In the classroom context, children are required to follow multiple rules at once, switch between activities according to classroom demands, and attend selectively amidst distracters (e.g., peers, toys), all which rely on intact EF. Social skills that depend on the integrity of EF are also required such as emotion regulation (e.g., Carlson & Wang, 2007; Simonds, Kiera, Rueda, & Rothbart, 2007), perspective-taking (e.g., Frye, Zelazo, & Palfai, 1995; Sabbagh, Xu, Carlson, Moses, & Lee, 2006) and cooperative behavior (e.g., Ciairano, Visu-Petra, & Settanni, 2007).
Therefore, children with poor EF are at risk on multiple levels in an academic setting (e.g., McClelland, Morrison, & Holmes, 2000). Whereas children with poor EF are impulsive and driven by habit and salient aspects of their environment, children with good EF skills are able to learn in a top-down fashion allowing them to plan, act flexibly (Marcovitch, Jacques, Boseovski & Zelazo, 2008), and evaluate outcomes in an organized way (Anderson, 2002; Zelazo, Muller, Frye, & Marcovitch, 2003).

Indeed, a growing body of research demonstrates an important association between EF measured in early childhood and later academic performance (e.g., Blair & Razza, 2007; Brock, Rimm-Kaufman, Nathanson, Grimm, 2009; Bull & Scerif, 2001; McClelland, Acock, Morrison, 2006; McClelland, Cameron, Connor, Farris, Jewkes, & Morrison, 2007; Visu-Petra, Cheie, Benga, Micldea, 2011). Some evidence points to a domain-general contribution of EF to academic success (e.g., Best et al., 2011; Bull, Epsy, Wiebe, 2008), with individual differences in EF predicting social adjustment during the transition period to school (e.g., Hughes & Ensor, 2010), school readiness (e.g., Blair & Razza, 2007), and emergent academic skills in reading and math near the end of kindergarten, more strongly than measures of intelligence (McClelland, Cameron, Connor, Farris, Jewkes, & Morrison, 2007). The early development of EF seems to set a foundation for learning that influences later academic outcomes (e.g., Best Miller & Naglieri, 2011; Bull et al., 2008; Duncan, Dowsett, Claessens, Magnuson, Huston, Klebanov, et al., 2007), as far reaching as adolescent SAT performance (Shoda, Mischel, & Peake, 1990). Life-long success, as indicated by adult health, wealth, and criminality, is also predicted across a gradient of executive control in childhood, independently of measures of intelligence and socioeconomic status (Moffitt, Arseneault, Belsky, Dickson, Hancox, Harrington, et al., 2011).

**Influencing the Development of EF**

Clearly it is important to foster EF, particularly in early childhood (Altemeier, Jones,
Abbott, & Berninger, 2006). While individual differences in EF may remain relatively stable across development (e.g., Friedman, Miyake, Young, DeFries, Corley, & Hewitt, 2008), a number of studies have now shown that EF is amenable to environmental input. Factors found to influence the development of EF stem from the broad socio-cultural context (e.g., Lewis, Koyasu, Oh, S., Ogawa, Short, & Huang, 2009; Moriguchi, Evans, Hiraki, Itakura, & Lee, 2012; Moriguchi, Lee & Itakura, 2007; Sabbagh et al., 2006) to the more proximal familial environment (e.g., Bernier, Carlson, Whipple, 2010; Cole & Mitchel, 2000; Hughes, & Ensor, 2009). While certain environments characterized by neglect, trauma, and/or poverty tend to impede EF development (e.g., Bos, Fox, Zeanah & Nelson, 2009; DePrince, Weinzierl & Combs, 2009; Hackman & Farah, 2009, Mezzacappa, 2004; Noble, Norman, & Farah, 2005), positive and enriching experiences can facilitate EF (e.g., Bialystok, & Martin, 2004; see Diamond & Lee, 2011 for review).

Targeted interventions, for example, have demonstrated positive effects on EF including specific preschool curricula (e.g., Diamond, Barnett, Thomas, & Munro, 2007; Lillard & Else-Quest, 2006; Röthlisberger, Neuenschwander, Cimeli, Michel & Roebers, 2011), and certain extracurricular activities such as exercise, music, and martial arts (see Diamond & Lee, 2011 for review). Laboratory-based efforts targeting specific EF skills have also been effective in adults (see Dahlin, Bäckman, Stigsdotter Neely, & Nyberg, 2009 for a review), children (Klingberg, Forssberg, & Westerberg, 2002; Klingberg, Fernell, Olsen, Johnson, Gustafsson, Dahlstrom, et al., 2005; Holmes et al., 2009; Dowsett & Livesey, 2000; Jolles van Buchem, Crone, & Rombouts, 2011; Jolles, van Buchem, Rombouts, & Crone, 2012; Karbach & Kray, 2009; Kloo & Perner, 2003; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005; Rueda, Checa, & Cómbita, 2012; Tamm, Hughes, Ames, Pickering, Silver, Stavinoha, et al., 2010; Thorell, Lindqvist, Bergman Nutley, Bohlin & Klingberg, 2009; Mack, 2007), and even in infants (Wass,
Porayska-Pomsta & Johnson, 2011). A general finding of these studies is greater potential for transfer of trained skills in those with underdeveloped and/or compromised EF (e.g., Diamond and Lee, 2011).

Although there is clear potential for facilitating the early development of EF, most of these studies have been carried out with adults (e.g., Klingberg & McNab, 2009; Olesen P., Westerberg, H., & Klingberg, 2004) and older children (e.g., Jolles et al., 2011; 2012). Findings with twelve-year-old children are that six weeks of working memory practice led to increases in frontoparietal brain activation near adult levels (Jolles et al., 2012), but not connectivity (Jolles et al., 2011). Very few training studies have been carried out with young children, however (but see Kloo & Perner, 2003; Rueda et al., 2005; 2012; Thorell et al., 2009; Mack 2007). Researchers have also called for the combined use of neuroscientific measures with training to elucidate both the behavioral and neural mechanisms through which EF interventions yield positive results (e.g., Kloo & Perner, 2008), and to better inform academic practice (e.g., Posner & Rothbart, 2005). Rueda and colleagues have been amongst the first to examine the neurophysiological effects of EF training in young, preschool-age children. Five- and six-year-old children given several attention-training exercises over the course of a month demonstrated a more efficient and mature pattern of conflict processing (smaller N2 amplitudes and shorter latencies) relative to non-trained controls and younger children (Rueda et al., 2005; 2012).

Although remarkable in their attempts to examine the neural mechanisms of EF as they emerge in early development, studies that have combined neuroscientific measures with training in young children have generally lacked active controls, thereby limiting the extent to which the effects of training can be disentangled from extraneous effects such as practice. Moreover, few studies have examined the relative efficacy of different kinds of EF training, making it difficult to determine what types of training work best and at what age. There is some evidence
highlighting the importance of working memory processes (Thorell et al., 2008), dosage, and
cognitive demand (e.g., Karbach & Kray, 2009) in EF training with young children. Despite
evidence for the latter, EF training to date has focused primarily on singular EF processes (e.g.,
working memory, inhibition). Training of higher-order EFs (e.g., flexible rule-use or shifting) is
needed to fill this gap in the literature, which will have important implications for educational
contexts that require these complex processes. The overall aim of the studies included herein was
to explore the neurocognitive processes that support the early development of EF by comparing
the neural correlates of EF in children with good versus poor EF and by exploring why EF
training is effective. Particular interest was in isolating the effect of reflection, or the reflective
reprocessing of information, in training effects. In these studies, children’s performance on the
Dimensional Change Card Sort (Zelazo, 2006, see Fig. 1), a complex measure of EF, was
assessed before and after reflection training, in combination with active controls.

**Dimensional Change Card Sort**

One widely studied measure of early EF development is the Dimensional Change Card
Sort (DCCS; Zelazo, 2006, see Fig. 1). Drastic improvements in children’s performance on this
task, which occur between the ages of three and six (e.g., Zelazo et al., 2003), capture the rapid
development of EF that characterizes the preschool years. The DCCS requires a shift in response
from sorting test cards according to one dimension, to sorting test cards according to another
dimension in the presence of two potential sources of conflict: bivalent test cards pulling for
competing response options and the need to switch on the post-switch phase from a prepotent
response to a previously inhibited response. In this task, children are shown two target cards
(e.g., a blue rabbit and a red boat) and asked to sort a series of bivalent test cards (e.g., red
rabbits and blue boats), first according to one dimension (e.g., color), and then according to the
other dimension (e.g., shape). Regardless of which dimension is presented first, most 3-year-olds
perseverate during the post-switch phase of the standard version, continuing to sort by the first dimension. Moreover, they do this despite being told the new rules on every trial, and despite correctly answering questions about the post-switch rules (e.g., “Where do the rabbits go in the shape game?”). In contrast, by 5 years of age, most children switch flexibly (e.g., Bialystok, 1999; Brace, Morton, & Munkata, 2006; Diamond, Carlson, & Beck, 2005; Dick, Overton, & Kovacs, 2005; Kirkham, Cruess, & Diamond, 2003; Kloo & Perner, 2005; Munakata & Yerys, 2001; see Garon et al., 2008 for review).

Several studies have shown that children’s performance on the DCCS cannot be attributed to an inability to sort by a particular category; a greater ability to sort by one dimension over the other; or a failure of working memory, per se (for a review see Zelazo et al., 2003). A number of accounts have been offered of the developing cognitive processes that underlie age-related changes on this task. Zelazo and colleagues (2003) have highlighted the importance of increases in the reflective reprocessing of information, allowing for the formulation, selection, and maintenance in working memory of higher-order rules. In the absence of reflection, behavior is determined by habit (e.g., responding according to the pre-switch rules on the post-switch phase). Reflection allows for deliberate response selection (e.g., responding according to the post-switch dimension on the post-switch phase) via top-down control (Zelazo et al., 2008). According to the Cognitive Complexity and Control theory-revised (CCC-r; Zelazo et al., 2003), children who perseverate have difficulty reflecting on their (conflicting) rule representations and formulating a hierarchical rule system that resolves the conflict inherent in the rules and the bivalent stimuli: “If I’m sorting by shape, and if it’s a red rabbit, then it goes here with the blue rabbit but, if I’m sorting by color, and if it’s a red rabbit, then it goes here with the red boat.”

Alternative accounts suggest that children’s difficulties on the DCCS are due to weak memory representations of the post-switch rules (e.g., Morton & Munakata, 2002), difficulty
inhibiting attention toward the previously relevant dimension (i.e., “attentional inertia”; Kirkham et al., 2003), and a failure of redescription (i.e., thinking about the blue boat as a “blue thing” in the colour game, and as a “boat” in the shape game). While both CCC-r and redescription accounts suggest that children must consider their rule representations to determine whether they should think of a red rabbit as a red thing or as a rabbit, redescription accounts suggest that children’s difficulties stem from their inability to think about the same stimulus in multiple ways), rather than a representational difficulty (the inability to formulate the structure of the task).

To date, only a handful of studies have used training on the DCCS to isolate the processes involved, rather than task modifications that tend to obscure underlying processes (e.g., Kloo & Perner, 2003; Towse, Redbond, Houston-Price, & Cook, 2000; Mack, 2007). Kloo and Perner (2003, Exp. 3) trained children over the course of several weeks to reflect on their rule representations during the DCCS (i.e., which dimension they were sorting by; see Appendix A in Chapter 3 for example). Children trained in this way showed more improvement on the DCCS than children trained to use relative clauses (an active control condition). These effects showed near transfer to a 3-boxes version of the DCCS and far transfer to a false belief task. Although training effects were attributed to children’s increased understanding of redescription (i.e., that it is possible to construe a stimulus in two different ways: as a red thing and as a rabbit), it is also possible to attribute the effects to children’s increased reflective capacity. Although the processes involved in the early development of EF remain a matter of debate, all accounts share a common emphasis on conflict, or the ability to resolve conflicting information.

**Neurophysiological Correlates of Conflict**

Support for the role of conflict in EF development (as measured by the DCCS) comes from recent neuroimaging studies indicating that what likely differentiates children with good EF
(passers) from those with poor EF (perseverators) is the ability to recruit ventrolateral prefrontal cortex in explicitly representing a higher-order rule for selecting among the conflicting response options and for resolving the conflict inherent in the stimuli (as indexed by lower amplitude N2s: Lamm, Zelazo, & Lewis, 2006 and an increase in oxygenated haemoglobin in ventrolateral prefrontal cortex: Moriguchi & Hiraki, 2009, in children who pass). Behavioral studies have also shown that performance on the DCCS improves if conflict is reduced (e.g., Diamond, Carlson, Beck, 2005; Jordan & Morton, 2008; Kloé & Perner, 2005).

Much like the development of EF, conflict processing shows a protracted developmental trajectory into adolescence, with children demonstrating a heightened response to conflict and an inability to use prior experience to resolve conflict efficiently compared to adults (Waxer & Morton, 2011). While a number of neuroimaging studies have associated conflict monitoring and detection processes with the Anterior Cingulate Cortex (ACC; e.g., Jonkman, 2006; Lamm et al., 2006; Forster, Carter, Cohen, & Cho, 2011, van Veen & Carter, 2002), there is evidence to suggest a functional link between ACC and lateral prefrontal cortex with the ACC signaling the need for top-down cognitive control mediated by lateral prefrontal cortex (e.g., Botvinik, Cohen, & Carter, 2004; Gehring & Knight, 2000; Walsh, Buonocore, Carter, & Mangun, 2011). In the neuroscience literature, the N2 has been consistently associated with the detection of conflict. The N2 is a negative deflection occurring approximately 200 ms in adults and between 300-500 ms in children (Lamm et al., 2006; Rueda et al., 2004, 2005; Waxer & Morton, 2011) following stimulus presentation on various measures of EF that has been source localized to ACC (e.g., Bokura, Yamaguchi, & Kobayashi, 2001; Lamm et al., 2006; Nieuwenhuis, Yeung, Van den Wildenberg, & Ridderinkhof, 2003; van Veen & Carter, 2002). Studies have shown that the N2 is sensitive to the presence of conflict and the need for behavioral control adjustments in response to conflict (Forster et al., 2011). Individual differences in N2 amplitude appear to be
related to EF in young children, independent of age (Lamm et al., 2006; Rueda et al., 2005; 2012). Despite some indication that neural systems underlying conflict detection play an important role in the development of EF, the neurocognitive processes supporting the early development of EF are not well understood.

The three studies included in this dissertation aimed at exploring the neurocognitive processes underlying early EF development. In Experiment 1 (Exp. 1; Chapter 2), neural markers of EF in young children were explored by comparing ERPs of children who passed the DCCS and resolved the conflict in the task efficiently, with children who failed and did not resolve the conflict. Exps. 2 to 4 are presented together in Chapter 3 as separate components of one study designed to isolate the role of reflection in EF development by examining whether reflection training improves children’s EF performance on the DCCS in comparison with active controls. In Experiment 2 (Exp. 2; Chapter 3) the key findings of Kloo & Perner, (2003, Exp. 2) were replicated, that reflection training improves young children’s performance on the DCCS and demonstrates far transfer. In Experiment 3 (Exp. 3, Chapter 3), a brief, single-session version of the training protocol was implemented to examine dosage effects (Exp. 3). In Experiment 4 (Exp. 4; Chapter 3), we explored the neural correlates of reflection training (used in Exp. 3) in young children by examining training induced changes in the neural marker of EF (N2 amplitude) found in Exp. 1.
References


Brace, J. J., Morton, J. B., & Munakata, Y. (2006). When actions speak louder than words:


Training, maturation, and genetic influences on the development of executive attention.

*Proceedings of the National Academy of Sciences of the United States of America, 102*, 14931–14936.


Chapter 2: N2 Amplitude as a Neural Marker of Executive Function in Young Children: An ERP Study of Children Who Switch Versus Persevere in the Dimensional Change Card Sort

Executive function (EF) refers to the deliberate, top-down neurocognitive processes involved in the regulation of thought, action, and emotion—processes such as cognitive flexibility, inhibitory control, and working memory (Miyake, Friedman, Emerton, Witzki, Howerter, & Wager, 2000). Individual differences in EF in childhood have been found to predict important developmental outcomes including math and reading skills in preschool and the early school grades (e.g., Blair & Razza, 2007), and SAT scores in adolescence (e.g., Shoda, Mischel, & Peake, 1990). Indeed, EF is often a better predictor of achievement than is IQ, and teachers often report that the most important determinant of classroom success in kindergarten and early school grades is the extent to which children can sit still, pay attention, and follow rules (e.g., McClelland, Cameron, Connor, Farris, Jewkes, & Morrison, 2007).

Children with poor EF may be at a disadvantage in educational contexts for a number of reasons, including poor attention, poor emotional control, an increased likelihood of causing behavioral disruptions, and teachers’ diminished expectations of children’s success. From a cognitive perspective, however, it has also been suggested that children with better EF may approach learning opportunities in a more reflective, self-directed way that allows them to be goal-directed and proactive in seeking out new information instead of learning in a more passive, incremental fashion (Marcovitch, Jacques, Boseovski, & Zelazo, 2008). For example, children who are more likely to reflect upon and monitor their own knowledge may display greater cognitive flexibility and be better able to override the influence of habits or predispositions that interfere with learning.
It is now well known that EF undergoes particularly marked changes between the ages of 3 and 6 years (Zelazo, Anderson, Richler, Wallner-Allen, Beaumont, & Weintraub, 2011), just as children face sharp increases in the demands placed on their self-regulation (e.g., as they transition to school). A widely used measure of EF during these years of rapid change is the Dimensional Change Card Sort (DCCS; Zelazo, 2006; see Figure 1). In the standard version of this task, children are shown two target cards (e.g., a blue rabbit and a red boat) and asked to sort a series of bivalent test cards (e.g., red rabbits and blue boats) first according to one dimension (e.g., color) and then according to the other (e.g., shape). Regardless of which dimension is presented first, most 3-year-olds sort correctly on the pre-switch trials but then perseverate during the post-switch trials, continuing to sort test cards by the first dimension despite being told the new rules on every trial, and despite correctly answering questions about the post-switch rules. By 5 years of age, most children switch flexibly (e.g., Bialystok, 1999; Bohlmann & Fenson, 2005; Brace, Morton, & Munakata, 2006; Diamond, Carlson, & Beck, 2005; Dick, Overton, & Kovacs, 2005; Kirkham, Cruess, & Diamond, 2003; Kloo & Perner, 2005; Munakata & Yerys, 2001; Zelazo, Muller, Frye, & Marcovitch, 2003). The DCCS shows excellent test-retest reliability during this age range ($ICC_s = .90-.94$; Beck et al., 2011).

How best to explain children’s performance on the DCCS is currently a matter of debate. According to the Cognitive Complexity and Control theory-revised (CCC-r; Zelazo et al., 2003), for example, children who perseverate on the DCCS have difficulty reflecting on the hierarchical structure of the task and formulating a hierarchical rule system that resolves the conflict inherent in the bivalent stimuli. The rapid development of self-reflection during the preschool years allows children to understand that they know two different ways of approaching the task: “If I’m sorting by color, then the red rabbits go here; but if I’m sorting by shape, then they go there.” This approach suggests that performance on the DCCS should provide a measure of self-
reflection, or monitoring one’s rule knowledge, that will predict the efficiency of later learning. In contrast to the CCC-r theory, other approaches emphasize other cognitive processes, such as the need to maintain rules in active working memory (e.g., Morton & Munakata, 2002) or to inhibit their attention to the pre-switch rules (e.g., Kirkham et al., 2003).

Regardless of how one characterizes the processes that make it possible for children to switch on the DCCS, however, there is general agreement that these processes depend on neural networks involving lateral prefrontal cortex, as shown in several recent neuroimaging studies using the DCCS (Moriguchi & Hiraki, 2009; Morton, Bosma, & Ansari, 2009; Waxer & Morton, 2011a). In the single study to date to examine preschoolers, Moriguchi and Hiraki (2009) used near-infrared spectroscopy (NIRS) to measure the concentration of oxygenated haemoglobin (oxy-Hb) in ventrolateral prefrontal cortex during performance on the task. Following presentation of the test cards, both 5-year-olds and those 3-year-olds who switched flexibly on the task showed an increase in oxy-Hb bilaterally, whereas 3-year-olds who failed the task did not.

The N2 component of the ERP is usually observed at medial-frontal sites between 200 and 400 ms following stimulus presentation on various measures of EF, including Go-Nogo tasks (e.g., Eimer, 1993; Falkenstein et al., 1999; Jodo & Kayama, 1992; Jonkman, Lansbergen & Stader, 2003; Lamm, Zelazo, & Lewis, 2006) and versions of the Eriksen flanker task (e.g., Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005; van Veen & Carter, 2002). The amplitude of the N2 has been found to vary as a function of conflict and the need for cognitive control. For example, amplitudes are larger (i.e., more negative) on Nogo trials than on Go trials in Go-Nogo tasks, they are larger when discriminability between Nogo and Go stimuli is difficult (see Folstein & van Petten, 2008, for review), and they increase as a function of target-distractor compatibility on a flanker task (Forster, Carter, Cohen, & Cho, 2011). For these and other
reasons, the N2 is often taken as an index of EF in general and conflict monitoring in particular (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Botvinik, 2007; Lahat, Todd, Mahy, & Zelazo, 2010; Lamm, Zelazo, & Lewis, 2006; Nieuwenhuis, Yeung, Van den Wildenberg, & Ridderinkhof, 2003; Rueda, Posner, Rothbart, & Davis-Stober, 2004; Waser & Morton, 2011a; Yeung & Nieuwenhuis, 2009). Conflict monitoring may initiate processes associated with lateral prefrontal cortex (e.g., Botvinick et al., 2001; Gehring & Knight, 2000; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004), such as high order rule use (e.g., Bunge and Zelazo, 2006) or the representation of attention-guiding rules (e.g., Waser and Morton, 2011a).

The present experiment used EEG to examine conflict-related processing during a standard version of the DCCS in preschool age children. In particular, the N2 component of the ERP was examined in relation to children’s performance (passing vs. failing) on the post-switch phase of the task. In contrast to Go-Nogo tasks, in which greater conflict may be expected on Nogo versus Go trials (e.g., Crone, Donohue, Honomichl, Wendelken & Bunge, 2006), the standard version of the DCCS has the potential to elicit conflict on all trials (i.e., both pre- and post-switch trials) because all trials involve bivalent stimuli, which pull for competing response options. Source analysis of the N2 in adults and children has suggested cortical generators in both cingulate cortex and right orbitofrontal cortex (Bokura, Yamaguchi, & Kobayahsi, 2001; Jonkman, Sniedt, & Kemner, 2007; Lewis, Lamm, Segalowitz, Stieben, & Zelazo, 2006, Lewis, Granic, lamm, Zelazo, Stieben, Todd, Lewis, Meusel, & Zelazo, 2008; Nieuwenhuis et al., 2003; for review, see van Veen and Carter, 2002), although in children the cingulate sources may be relatively posterior compared to older adolescents and adults (Lamm et al., 2006).

Developmental research on the N2 has generally shown a decrease with age in both amplitude and latency (Davis, Bruce, Snyder, & Nelson, 2003; Johnstone, Pleffer, Barry, Clarke, & Smith, 2005, Johnstone, Dimoska, Smith, Barry, Pleffer, Chiswick, et al., 2007; Jonkman, 2006;
Jonkman et al., 2003; Lewis, Lamm, Segalowitz, Stieben, & Zelazo, 2006; Rueda et al., 2004), and individual differences in amplitude, in particular, appear to be related to EF in school age children and adolescents (Lamm et al., 2006).

A prediction derived from CCC-r theory is that children who understand the hierarchical structure of the task and effectively resolve the conflict inherent in the bivalent stimuli should not only switch flexibly but also show smaller N2 amplitudes on the pre- and post-switch trials, consistent with the down-regulation of ACC-mediated conflict detection via top-down control (e.g., Forster et al., 2011). Children who fail to understand the hierarchical structure of the task should not only perseverate on the task but also display evidence of unresolved conflict between the competing affordances of the bivalent stimuli, manifested as larger N2 amplitudes. A similar prediction could be made on the basis of accounts emphasizing working memory (e.g., Morton & Munakata, 2002). According to inhibitory account (e.g., Kirkham et al., 2003), however, children who perseverate on the DCCS understand the hierarchical structure of the task but simply cannot inhibit attention to the pre-switch rules on the post-switch phase (i.e., once the pre-switch rules have become prepotent through use). Therefore, inhibitory control accounts would presumably predict that compared to children who switch, children who perseverate should only experience higher levels of response conflict (and higher N2 amplitudes) on the post-switch phase, but not on the pre-switch phase. Do children who switch versus perseverate on the DCCS differ in N2 amplitude, and do these differences appear on the pre-switch phase, the post-switch, or both?

**Method**

**Participants**

The final sample included 99 children (36 males; aged 35 to 54 months: $M$ age = 41.79; $SD = 4.29$). These children were predominantly Caucasian (82%), from middle to upper-middle class backgrounds. Parents and their children were recruited through a database of parents who
had expressed interest in participating in studies and through advertisements posted in the community. An additional 42 children (20 males; $M_{age} = 40.24$, $SD = 4.37$) were tested but excluded from the final sample for the following reasons: (a) they failed to perform better than chance on the pre-switch phase of the DCCS and/or did not unambiguously pass or fail the post-switch phase (see pass/fail criteria below; $n = 24$); (b) they were inattentive and moved or vocalized excessively ($n = 10$); (c) their ERP data did not contain 15 or more usable post-switch trials ($n = 8$). Exclusion of all participants was carried out by an experimenter blind to children’s pass/fail status.

**Procedure**

Children were tested individually (with their parent present) in a room decorated according to a space theme. They were shown an electrode sensor net, referred to as a “space hat” and informed that they were going to drive a rocket ship once the experimenter put on the hat. When children were comfortable, the net was applied and children were seated 45 cm in front of a 17” computer monitor, next to an experimenter. Overhead lights were turned off, and the room was illuminated by a 100-watt lamp located 200 cm away from the monitor. Children were asked to sit as silently and as still as possible throughout the session. Children were administered a computerized-version of the DCCS.

*The Dimensional Change Card Sort (DCCS).* The DCCS was adapted from the standard version (Zelazo, 2006). Shape was always the pre-switch dimension and color was the post-switch dimension. On each trial, a test stimulus (blue rabbit or red boat) was presented at a central location above and between two target stimuli (blue boat on the left and red rabbit on the right). Test stimuli remained onscreen until children responded by pressing (using the index finger of their dominant hand) one of two laminated replicas of the target stimuli that were affixed to buttons on a response pad (see Figure 1). After a response, the test stimulus
disappeared while a fixation-cross appeared until the next trial. The experimenter initiated the next trial by pressing the space bar on a keyboard after a minimum of 1000 ms, with an average inter-trial interval of 2668.19 ms (SD = 418.86). Stimuli were presented in the same pseudo-random order for all children, such that a particular stimulus was never presented more than twice in a row and each of the two stimulus types (blue rabbit and red boat) were presented equally often on the post-switch phase.

To begin, children were given six practice trials. The pre-switch rules were stated before each of these trials. On the first practice trial, the experimenter demonstrated a correct response, emphasizing the importance of pressing gently and quietly. Children were then asked to respond without assistance on the five remaining practice trials and were rewarded with a sticker on each trial that they sat still, remained silent, and pressed the correct button without extraneous hand movements. Anytime children hesitated or were unsure of the correct response, the experimenter pointed to the correct button. The practice trials were designed to ensure that children understood the task and practiced responding in an efficient fashion. Children then received 15 pre-switch and 30 post-switch trials that were administered in the same way as the last 5 training trials except that (a) children did not receive stickers or feedback of any kind, (b) children were only told the relevant rules on the initial trial of each phase and every 5th trial thereafter, and (c) if children hesitated, the experimenter provided only general encouragement to respond (and never pointed to the correct button). Between the pre-switch and the post-switch phases, children were explicitly told to switch: “Now we’re going to play a new game—the color game—and the color game is different. In this game if you see a blue one, press this one [the experimenter pointed to the correct button], but if you see a red one, press that one [the experimenter pointed to the correct button]. Okay?”
**EEG Testing.** EEG data were recorded using a 128-channel Geodesic Sensor Net (Tucker, 1993). Data were sampled at 250 Hz and recorded using NetStation 4.1.2 (EGI Software: EGI, Eugene, OR). Impedances were kept below 80 kΩ at the beginning of the session. All channels were referenced to Cz during recording and later rereferenced to an average reference. An FIR 1-30 Hz bandpass filter was applied. Data were stimulus-locked to the onset of the test stimulus. Segments were based on 400 ms pre-stimulus to 700 ms post-stimulus. For each segment, channels were marked as bad and replaced through interpolation of neighboring electrodes if the fast average amplitude exceeded 200 µV; the differential average amplitude exceeded 100 µV; and/or the channel had zero variance. Ocular artifacts were corrected using the ocular artifact removal tool in Net Station. The EOG sensitivity, which uses the slope of the eye blink to differentiate eye blinks from eye movements so that separate correction factors can be applied to every EEG channel (EGI; Net Station Waveforms Technical Manual, 2006), was set at 1.35 µV. This sensitivity level produced a good balance between rejection of unusable segments and preservation of segments relatively unaffected by artifacts. Data for each child were also individually inspected to ensure that the algorithms used were appropriate, with good trials included and bad trials excluded. Finally, segments were excluded if they contained more than 25 bad channels and/or eye blinks or eye movements (using a100 µV threshold). As a result of the artifact detection process, 27% of pre-switch segments and 30% of post-switch segments were rejected (similar to rejection levels seen in previous research with young children, e.g., Rueda et al., 2004). Averaged ERPs from individual children were grand-averaged and adjusted using the 200 ms prior to stimulus as baseline.

As is common in research with young children (e.g., Lewis et al., 2006; Lewis, Todd, & Honsberger, 2007; Rueda et al., 2004; 2005; Todd et al., 2008), amplitude of the N2 was calculated based on the peak within the coded region (300 to 500 ms post-stimulus) that
occurred after the N1 and (b) had a fronto-central topography. Latency was calculated as the time in milliseconds from stimulus onset to peak amplitude. An experimenter blind to participants’ pass/fail status carried out the coding of the ERPs.

Results

Preliminary Analyses

By design, all children included in the final sample passed the pre-switch phase ($p < .05$, based on the binomial theorem), sorting correctly on 11 or more trials out of 15 ($M = 13.92, SD = 1.17$). Preschoolers’ post-switch performance on the DCCS is typically bimodal, with children either clearly passing or clearly failing, although the inclusion in the current version of many more trials than is usual, as well as the presence of the sensor net, was expected to reveal intermediate patterns of responding and/or increase the likelihood of occasional random responding due to distraction. Children were classified as passing the DCCS as a whole (i.e., passing the post-switch phase) if they were correct on 20 or more trials out of 30 on the post-switch phase ($p < .05$), and children were classified as failing the DCCS if they made 20 or more errors on the post-switch phase ($p < .05$). Altogether, 45 children passed, with an average of 26.38 ($SD = 2.53$) post-switch trials correct, whereas 54 children failed, making an average of 26.83 ($SD = 2.55$) post-switch errors.

Preliminary analyses did not reveal any significant differences between children who passed and children who failed the DCCS in terms of sex, ethnicity, mean number of pre-switch errors, median pre- or post-switch reaction times, or mean number of trials contributing to the ERP waveforms for the pre- or post-switch phases. Subsequent analyses collapsed across these variables. Children who passed were significantly older than children who failed, however, by approximately 3 months (passers: $M = 43.40$ months, $SD = 4.11$; failers: $M = 40.44$ months, $SD = 3.99$, $F(1, 97) = 3.62, p < .01$), so age was used as a covariate in the following analyses.
ERP Analyses

Analyses of ERPs during the post-switch trials were based on correct trials for passers and incorrect trials for failers. The mean number of trials contributing to the post-switch N2s was 20.27 (SD = 4.22), comparable to trial counts used in previous neurodevelopmental research (e.g., Rueda et al., 2005; Todd et al., 2008). In order to increase the ratio of ERP signal to noise, however, we employed a much larger sample of participants than is typical. N2 amplitudes were examined at four frontal midline sites including sites 129, 6, 11, and 16, corresponding roughly to Cz, Fcz, Fz, and Afz respectively (Luu & Ferree, 2000), across which amplitudes were maximal. See Figure 2 for grand-averaged waveforms at these sites, and Table 1 for mean N2 amplitudes for passers and failers.

N2 Amplitude

Post-switch N2 amplitudes (µV) at each of the four fronto-central sites were examined via a repeated measures analysis with performance status (passers vs. failers) as a between-subjects variable and age (in months) as a covariate. Results revealed a significant effect of performance status, \( F(1, 96) = 4.92, p = .03, \eta^2_p = .05 \), shown in Figure 3. Univariate results were significant at three sites: site 129, \( F(1, 96) = 6.22, p = .01, \eta^2_p = .06 \), site 6, \( F(1, 96) = 3.82, p = .05, \eta^2_p = .04 \), and site 16, \( F(1, 96) = 3.93, p = .05, \eta^2_p = .04 \), but not at site 11, \( F(1, 96) = 2.38, p = .13, \eta^2_p = .04 \), with smaller N2 amplitudes for passers. Age was not a significant covariate, \( F(1, 96) < 1.00 \), and the same pattern of results was obtained when trial count or N2 latency was included as a covariate.

The mean number of pre-switch trials available for analysis was only 10.79 (SD = 2.48), so analysis of ERPs on pre-switch trials should be considered exploratory. Despite the low trial count, however, a repeated measures analysis (sites) with performance status (passers vs. failers) as a between-subjects variable and age (in months) as a covariate revealed a significant effect of
performance status $F(1, 96) = 10.71, p < .01$. Univariate results were significant at all four sites: site 129, $F(1, 96) = 7.95, p < .01, \eta^2_p = .08$; site 6, $F(1, 96) = 8.67, p < .01, \eta^2_p = .08$; site 11, $F(1, 96) = 8.64, p < .01, \eta^2_p = .08$; and site 16, $F(1, 96) = 9.12, p < .01, \eta^2_p = .09$, with smaller pre-switch amplitudes for passers. Age was not a significant covariate, $F(1, 96) < 1.00$, and the same pattern of results was obtained when trial count or N2 latency was included as a covariate.

**N2 Latencies**

For all children, the peak of the N2 considered across both the pre-switch and post-switch phases and across all four sites occurred at 438.29 (SD = 48.40) ms post-stimulus. To determine whether N2 latencies differed between children who passed and children who failed the DCCS, a set of repeated measures analyses analogous to those used to examine amplitude were carried out on N2 latencies (repeated measures analysis with sites as a within-subjects variable, performance status as a between-subjects variable, and age as a covariate). Results revealed no significant effects of performance status at any site on either the pre- or the post-switch phase, all $F$s(1, 96) < 2.55, $p > .12$ (see Table 2 for means and standard deviations). Age was not a significant covariate, $F(1, 96) < 1.00$, and the same pattern of results was obtained when trial count was included as a covariate.

**N2 Topography and Source Analysis**

The topography of the N2 is shown separately for passers and failers in Figure 4 (top row). Also shown is the topography of a difference wave created by subtracting the peak N2 amplitudes for passers from the peak N2 amplitudes for failers at each electrode site. This difference wave thus captures potential differences between failers and passers in N2 topography. As can be seen, there was a broad negativity across a broad fronto-central region, and differences between failers and passers were located within this region. EEG source imaging was performed using the standardized low-resolution brain electromagnetic tomography
(sLORETA) method (Pascual-Marqui, 2002), with a Sun-Stok 4-shell Sphere head model and Tikhonov regularization of $1 \times 10^{-4}$) as implemented in the NetStation GeoSource 2.0 software package. Results are shown in Figure 4 (bottom), separately for failers and passers, and for the difference wave, which allowed us to focus on the sources of scalp electrical activity that differentiated children who passed versus failed the DCCS. As can be seen in the bottom right panel, there were several frontal and cingulate cortical source activations that likely contributed to the N2 differences. These sources included: ventral anterior cingulate cortex (BA 25, 24, & 32), central cingulate cortex (BA 23 & 31), posterior cingulate cortex (BA 29 & 30), medial orbitofrontal cortex (BA 11), and lateral orbitofrontal cortex (BA 47).

The results of source analyses should always be interpreted with caution, and this is especially true when the analyses are based on EEG data from preschool age children because these data tend to be more variable than those of older participants and because there in no age-appropriate head model (i.e., model of how neural activation propagates from neural sources to the scalp). See Luck (2005) and Michel, Murray, Lantz, Gonzalez, Spinelli, & Grave de Peralta (2004) for thoughtful reviews of EEG source analysis methodology.

**Discussion**

Compared to children who perseverated on the DCCS, children who switched flexibly on the DCCS showed smaller N2 amplitudes during the post-switch phase of the task. Exploratory analyses revealed that children who passed also showed smaller N2 amplitudes on pre-switch trials. Source analysis of the N2 suggested sources in cingulate and orbitofrontal regions. Consistent with previous research (e.g., Lamm et al., 2006), the cingulate sources were more posterior than is typically found with adults (Lamm et al., 2006). The source analysis findings need to be replicated using pediatric head models when these models become available.
Previous research has found that the amplitude of the N2 generally decreases with age and is associated with EF (e.g., Lamm et al., 2006). In the present experiment, this more mature pattern of N2 amplitudes was observed in children who switched versus perseverated on the DCCS even when age, the number of trials contributing to the ERPs, and latency were controlled for statistically. No differences between children who switched and children who perseverated were observed in N2 latencies, although the N2 occurred later than is normally observed in adults, consistent with previous early pediatric research on ERP latencies in general (Thierry, 2005).

Children who switched were older than children who perseverated by about 3 months, but age was used as a covariate in the analyses, so it is difficult to explain the differences in N2 amplitude in terms of age differences or incidental changes such as changes in skull thickness. In addition, the number of usable trials did not differ between children who switched and children who perseverated (i.e., artifacts were not more problematic for one group than the other), so it is unlikely that systematic differences in artifact contamination played a role. Instead, the smaller N2 amplitudes seen in children who passed versus children who failed appear to provide a reliable neural marker of early individual differences in performance on the DCCS, complementing previous research (Lamm et al., 2006) and complementing previous work establishing a role for ventrolateral prefrontal cortex (Moriguchi & Hiraki, 2009). It should be noted, however, that recent research shows intra-subject variability in N2 amplitude depending on the specific processing context (Forster et al., 2011; Waxer & Morton, 2011a,b).

In adults, the N2 has been related to conflict monitoring (Carter, Braver, Barch, Botvinick, Noll, & Cohen, 1998) and detection of the need for top-down cognitive control, which in turn is mediated by lateral prefrontal cortex (e.g., Gehring & Knight, 2000; Ridderinkhof et al., 2004). Children who pass the DCCS may resolve the conflict inherent in the bivalent stimuli
more efficiently than children who fail, resulting in smaller N2 amplitudes. This interpretation is consistent with the CCC-r theory (Zelazo et al., 2003), according to which children who switch flexibly are children who understand the hierarchical structure of the task. That is, according to this theory, children who pass the DCCS immediately resolve the conflict inherent in the bivalent stimuli (i.e., even during the pre-switch) by recognizing that they know two ways of sorting the stimuli.

Although the findings of this experiment are consistent with CCC-r theory, other interpretations are also possible. From a working memory perspective (e.g., Morton & Munakata, 2002), children who switch flexibly may resolve the conflict inherent in the stimuli by keeping the relevant rules firmly in mind. It is less clear, however, how the observed differences in N2 amplitude could be explained by accounts emphasizing inhibitory control (e.g., Kirkham et al., 2003). Although poor inhibitory control might lead to both perseveration and larger N2 amplitudes on the post-switch phase (i.e., in the presence of response conflict), there should be no differences in N2 amplitude during the pre-switch phase, in the absence of both response conflict and inhibitory demands.

The most obvious source of conflict in the DCCS is the conflict inherent in the bivalent stimuli, but for children who perseverate, there is additional conflict on the post-switch phase between the explicit post-switch instructions (i.e., sort by color) and children’s perseverative behavior (i.e., sorting by shape). The finding that differences in N2 amplitudes were evident even on the pre-switch phase, while preliminary, suggests that the most relevant source of conflict may have been that inherent in the stimuli, however. Moriguchi and Hiraki (2009) also found that children who switch and children who perseverate differed on pre-switch as well as post-switch trials. In that study, children who perseverated on the task showed much less lateral prefrontal activation on both pre- and post-switch trials. In other words, it has now been found in
two independent studies with preschoolers that what differentiates children who pass from children who fail the DCCS is not something that occurs only during the post-switch phase. Taken together with Moriguchi and Hiraki’s findings, the current findings suggest the following account. Children who pass the DCCS detect the conflict inherent in the bivalent stimuli and then reflect upon their conflicting rule representation, recruiting lateral prefrontal cortex as they resolve the conflict by formulating (and keeping in mind) a representation of the hierarchical structure of the task (Bunge & Zelazo, 2006; cf. Badre & D’Esposito, 2007; Botvinick, 2008; Christoff & Gabrieli, 2000; Koechlin, Ody, & Frederique, 2003; Goldberg & Bilder, 1987) and/or by keeping the appropriate post-switch rules in mind (e.g., Waxer & Morton, 2011a,b). Children who perseverate on the DCCS also detect the conflict inherent in the stimuli but they fail to reflect upon their rule representations, fail to recruit lateral prefrontal cortex, and fail to understand the hierarchical structure of the task. For these children, unresolved conflict continues to be processed throughout the task, resulting in larger N2 amplitudes (present experiment) and reduced activation in lateral prefrontal cortex (Moriguchi and Hiraki, 2009). To confirm that N2 amplitudes as measured in the present experiment are indeed associated with conflict processing, it will be necessary to compare ERPs on trials involving bivalent stimuli to ERPs on trials involving univalent stimuli, as in the study by Waxer & Morton (2011b). The current interpretation generates the strong prediction that both children who switch and children who perseverate would show larger N2 amplitudes on bivalent compared to univalent trials.

Future research should also address more directly the hypothesized link between indices of conflict monitoring, such as N2 amplitude, and indices of prefrontally mediated hierarchical rule use. It is possible that what develops during the preschool period is a functional network linking ACC to lateral prefrontal cortex. It is also possible that understanding the ways in which ACC interacts with lateral prefrontal cortex will shed light on the co-occurrence during early
childhood of changes in EF, error monitoring, metacognition, and uncertainty monitoring (reviewed in Lyons & Zelazo, 2012). Although these constructs are often studied independently of one another, similar patterns of development are observed across all four constructs, with substantial improvement observed in early childhood and more gradual improvements evident well into adolescence.

The current findings complement previous research with young children pointing to the N2 as a neural marker of individual differences in EF (e.g., Lamm et al., 2006), and extend previous work with adolescents and adults (Morton et al., 2009; Waxer & Morton, 2011a) by examining ERPs in relation to performance on a standard version of the DCCS in preschool-age children. The findings support a characterization of performance on this task that highlights the role of conflict detection in a cascade of processes that eventuate in cognitive and behavioral flexibility.

Conflict detection and top-down EF are such fundamental neurocognitive skills that it is perhaps not surprising that individual and developmental differences in these skills are associated with a wide range of developmental outcomes. Moffitt, Arseneault, Belsky, Dickson, Hancox, Harrington, et al., (2011), for example, found that EF in childhood predicts (as a gradient) physical health, substance dependence, socioeconomic status, and the likelihood of a criminal conviction at age 32 years, even after controlling for social class of origin and IQ. Although much remains to be learned about the nature of these far-reaching longitudinal correlations, the proposed account of individual differences in EF as measured by the DCCS may help shed light at least on the relation between EF and educational achievement (e.g., Blair, 2002; McClelland, Acock, & Morrison, 2006; McClelland et al., 2007). Reflecting on conflict, including gaps in one’s understanding, may go hand in hand with the adoption of a more active, goal-directed, top-down approach to learning. Children with better EF may spontaneously monitor their own
understanding and seek actively to improve it. Further research might usefully investigate the conditions under which reflection is facilitated, in an effort not only to exercise children’s EF, but also potentially to transform the way in which children learn.

**Conclusion**

Children who switch flexibly on the DCCS, a key measure of EF during the preschool years, show smaller N2 amplitudes, consistent with the down-regulation of ACC-mediated conflict detection in children who reflect upon their conflicting rule representations and apprehend the hierarchical structure of the task. These findings implicate conflict detection in EF, and they point to the possible emergence during this period of a functional neural network linking ACC and lateral prefrontal cortex.
References


**Table 1**

*Mean N2 Amplitude (µV) and (SDs) at Frontal-Midline Sites for Children who Passed vs. Failed the Post-switch Phase of the DCCS*

<table>
<thead>
<tr>
<th>Phase</th>
<th>Site</th>
<th>Passers</th>
<th>Failers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-switch</td>
<td>129</td>
<td>-5.51(4.94)</td>
<td>-9.04 (6.10)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-6.46 (5.41)</td>
<td>-10.68 (7.55)</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>-7.08 (6.16)</td>
<td>-10.35 (6.71)</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>-6.52 (6.28)</td>
<td>-9.62 (6.35)</td>
</tr>
<tr>
<td>Post-switch</td>
<td>129</td>
<td>-5.02 (4.25)</td>
<td>-7.73 (4.17)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-5.66 (4.84)</td>
<td>-8.13 (4.82)</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>-5.75 (5.52)</td>
<td>-7.70 (6.07)</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>-5.19 (5.74)</td>
<td>-7.56 (6.33)</td>
</tr>
</tbody>
</table>
Table 2

*Mean N2 Latency (ms) and (SD) for Children who Passed vs. Failed the Post-Switch Phase of the DCCS*

<table>
<thead>
<tr>
<th>Phase</th>
<th>Site</th>
<th>Passers</th>
<th>Failers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-switch</td>
<td>129</td>
<td>427.11 (67.59)</td>
<td>426.67 (68.42)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>452.27 (65.04)</td>
<td>444.96 (71.16)</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>457.60 (70.86)</td>
<td>447.78 (77.93)</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>444.53 (74.80)</td>
<td>454.59 (79.85)</td>
</tr>
<tr>
<td>Post-switch</td>
<td>129</td>
<td>406.76 (59.04)</td>
<td>427.11 (54.87)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>418.31 (61.91)</td>
<td>435.56 (59.19)</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>442.84 (74.41)</td>
<td>444.59 (66.31)</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>433.69 (75.53)</td>
<td>444.37 (62.95)</td>
</tr>
</tbody>
</table>
Figure Captions

Figure 1. Sequence of events for the computerized version of the Dimensional Change Card Sort.

Figure 2. Grand-averaged pre-switch and post-switch ERPs at sites: 129, 6, 11, and 16, for children who passed vs. failed the post-switch phase of the Dimensional Change Card Sort.

Figure 3. Mean pre-switch and post-switch N2 amplitudes across sites 129, 6, 11, and 16 for passers and failers (* indicates significant at $p < .05$; error bars represent 95% confidence intervals).

Figure 4. Topographic maps of the N2 at the latency of peak amplitude are presented for Failers, Passers, and the difference wave of Failers minus Passers. The scalp voltages were submitted to sLORETA, and the resulting source activations for each group are plotted below their respective topographic map.
Experimenter states rules.
Fixation remains until experimenter presses space bar to advance test stimulus.

Test card appears.
Participant makes response on response pad.

Experimenter imposes a 2 sec delay on average before advancing next test stimulus.

Participant responds on response pad.
Topographic Plots of N2 Peak

Source Distribution at N2 Peak (sLORETA)
Chapter 3: Reflection Training Improves Executive Function in Preschool-age Children:

Behavioral and Neural Effects

Executive function (EF) refers to the top-down neurocognitive processes involved in flexible, goal-directed problem solving (Zelazo, Carlson, & Kesek, 2008). There is currently considerable interest in the early development of EF, in part because EF measured in early childhood predicts important developmental outcomes, including math and reading skills in preschool and the early school grades (e.g., Blair & Razza, 2007), cognitive control (Go/Nogo performance), SAT scores in adolescence (e.g., Eigsti, Zayas, Mischel, Shoda, Ayduk, Dadlani, et al., 2006), and socioeconomic status in adulthood (Moffit, Arseneault, Belsky, Dickson, Hancox, Harrington, et al., 2011). While this research suggests that individual differences in EF remain relatively stable from early childhood to adulthood, a growing number of studies have now shown that EF can be trained (see Diamond & Lee, 2011, for review). Much of this research has focused on the preschool years, a period of rapid development of EF that occurs just prior to a sharp increase in the demands placed on children’s developing EF (i.e., as they transition to school).

Interventions shown to improve EF include specific preschool curricula (e.g., Diamond, Barnett, Thomas, & Munro, 2007; Lillard & Else-Quest, 2006), certain extra-curricular activities (e.g., music, exercise, martial arts), and laboratory-based efforts targeting specific EF skills (e.g., Autin & Croizet, 2012; Dowsett & Livesey, 2000; Jolles, van Buchem, Crone, & Rombouts, 2011; Jolles, van Buchem, Rombouts, & Crone, 2012; Karbach & Kray, 2009; Klingberg, Fernell, Olesen, Johnson, Gustafsson, Dahlstrom, et al., 2005; Kloo & Perner, 2003; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005, Rueda, Checa, & Cómbita, 2012; Tamm, Hughes, Ames, Pickering, Silver, Stavinoha, et al., 2010; Thorell, Lindqvist, Bergman, Bohlin, & Klingberg, 2009). While these studies clearly reveal the considerable plasticity of EF in early
childhood, key questions remain concerning why EF interventions are effective: What are the active ingredients and how do they act? Three experiments (Exps. 2 to 4) are presented in this chapter, all with a common interest in isolating the role of reflection in EF training in preschool age children.

A well studied developmental transition in EF occurs in children’s performance on the Dimensional Change Card Sort (DCCS; Zelazo, 2006; see Fig. 1), in which children are required to sort bivalent test cards (e.g., red rabbits and blue boats) first according to one dimension (e.g., color) and then according to the other (e.g., shape). Most 3-year-olds perseverate on the post-switch phase of the DCCS, continuing to sort by the pre-switch dimension, whereas by 5 years of age, most children switch flexibly. Performance on more difficult versions of the task (e.g., speeded versions with more frequent, unpredictable switches between dimensions) continues to improve during adolescence, reaches a peak in young adulthood, and then declines in senescence (e.g., Diamond & Kirkham, 2005; Zelazo, Muller, Frye, & Marcovitch, 2003; Zelazo, Anderson, Richler, Wallner-Allen, Beaumont, & Weintraub, 2011).

Several different accounts have been offered of the developing cognitive processes that underlie age-related behavioral changes on this task (e.g., Bunge & Zelazo, 2006; Kirkham, Cruess, & Diamond, 2003; Morton & Munakata, 2002; Zelazo et al., 2003). For example, Zelazo and colleagues (2003) have highlighted the importance of increases in the reflective reprocessing of information, allowing for the formulation, selection, and maintenance in working memory of higher-order rules. According to the Cognitive Complexity and Control theory-revised (CCC-r; Zelazo et al., 2003), children who persevere on the DCCS have difficulty reflecting on their (conflicting) rule representations and formulating a hierarchical rule system that resolves the conflict inherent in the rules and the bivalent stimuli. The rapid development of self-reflection during the preschool years allows children to recognize that they know two different ways of
approaching the task: “If I’m sorting by color, then the red rabbits go here; but if I’m sorting by shape, then they go there.”

Other approaches emphasize different cognitive processes such as active (working) memory (Morton & Munakata; 2002), inhibition of attention (Kirkham et al., 2003), and redescription (the understanding that a stimulus can be redescribed from a different perspective; Kloo & Perner, 2003). Despite their differences, however, all of these accounts acknowledge that a key challenge for younger children is resolving, by one means or another, how to respond flexibly to the conflicting information in the task (i.e., the conflicting rules and/or the conflict inherent in the bivalent stimuli).

As an experimental test of their hypotheses, Kloo and Perner (2003, Exp. 3) trained children for approximately 30 min (2 x 15 min training sessions over the course of several weeks) to reflect on their rule representations during the DCCS (i.e., which dimension they were sorting by; see Kloo & Perner, 2003, Exp. 3). After sorting incorrectly (perseverating) on the DCCS, children were given corrective feedback and taught to (a) reflect on the relevant dimension (or higher-order setting condition): “That’s wrong. We are not playing the color game, the game with yellow and green, anymore. Now, we are playing the shape game—the game with apple and house”, then to (b) consider the antecedent conditions: “In the shape game, when you see an apple...”; and finally to (c) specify the consequents associated with each antecedent: “…you have to press the button with apple on it.” Children trained in this way showed more improvement on the DCCS than children trained to use relative clauses (an active control condition). These effects showed near transfer to a 3-boxes version of the DCCS and far transfer to a false belief task. Although training effects were attributed to children’s increased understanding of redescription (i.e., that it is possible to think of the same stimulus from two different perspectives: as a red thing and as a rabbit), the training itself can be also construed as
reflection training.

Three experiments are presented in this chapter with the overall goal of isolating the role of reflection in EF training in preschool age children. In Exp. 2, we replicated the key findings of Kloo and Perner (2003, Exp. 2). In Exp. 3, we confirmed that similar results were obtained using a brief (15 min), single session version of the reflection training protocol. In Exp. 4, we turned our attention to the neural correlates of reflection training and training-induced changes in children’s DCCS performance. In particular, we examined the N2 component of the event-related potential (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Botvinick, 2007; Lahat, Todd, Mahy, Lau, & Zelazo, 2010; Lamm, Zelazo, & Lewis, 2006; Nieuwenhuis, Yeung, Van den Wildenberg, & Ridderinkhof, 2003; Rueda, Posner, Rothbart, & Davis-Stober, 2004; Waxer & Morton, 2011; Yeung & Nieuwenhuis, 2009), which has consistently been associated with anterior cingulate cortex (ACC)-mediated detection of conflict in a variety of EF tasks, including Go-Nogo tasks, flanker tasks, and the DCCS. In Exp. 1, 99 preschoolers (35 to 54 m) were tested on a computerized version of the DCCS and found that N2 amplitude was smaller (less negative) for children who switched flexibly (passed) than it was for children who perseverated (failed), regardless of age. Children who pass the DCCS evidently resolve the conflict inherent in the task more efficiently than children who fail, resulting in smaller N2 amplitudes. One possibility is that for these children, conflict detection initiated reflection and higher-order rule use (mediated by lateral prefrontal cortical networks) that effectively resolved the conflict inherent in the stimuli and down-regulated ACC activation (Bunge & Zelazo, 2006; cf. Badre & D’Esposito, 2007; Botvinick, 2008; Christoff & Gabrieli, 2000; Koechlin, Ody, & Frederique, 2003). In fact, young children who pass the DCCS do show an increase in oxygenated haemoglobin in ventrolateral prefrontal cortex in response to the presentation of the stimuli (Moriguchi & Hiraki, 2009).
Rueda and colleagues have examined ERP effects of EF training in early development (Rueda et al., 2005; 2012). Children given 5 to 10, 45-min attention training sessions (over the course of a month) showed smaller (less negative) and earlier (shorter latency) N2 amplitudes that looked more adult-like. In Exp. 4, we predicted that reflection training would produce not only behavioral improvements, but also reductions in N2 amplitudes, compared to children who simply received corrective feedback: “That’s right/that’s wrong. You (are supposed to) press the button with the apple on it (experimenter points to correct button).” An immediate effect of corrective feedback was expected on the training DCCS, but the effects were not expected to transfer to the post-training DCCS (Bohlmann & Fenson, 2005). A mere practice condition (without corrective feedback) was also included to control for simple practice effects in the absence of training.

Method

Participants

A total of 113, 2- to 4-year old children (52 males; $M$ age = 41 months; $SD = 4.5$; range = 31 to 54 months) participated in this study. All participants were tested in a university laboratory or at a daycare center in a large metropolitan area. Although limited demographic information was collected, children were from diverse ethnic backgrounds and the daycare centers were located in economically diverse regions. Parents gave written consent for their children to participate in the experiments. Children received stickers or a small toy for their participation, and families in Exp. 4 received $15 in compensation for their participation. All experiments were conducted with IRB approval.

In Exp. 2, twenty-one additional children were tested and excluded for the following reasons: (1) they passed the post-switch phase of the DCCS, sorting all 5 trials correctly ($n = 12$); (2) they passed the control task, answering all 4 questions correctly ($n = 1$); (3) they failed to
return or otherwise refused to complete two or more post-training assays \((n = 7)\), or (4) they were persistently inattentive \((n = 1)\). Children who were included versus excluded did not differ significantly in terms of age, \(F(1, 46), p = .87\), verbal intelligence (PPVT standard score), \(F(1, 39), p = .28\), or gender, \(\chi^2(1, 50) = .02, p = .89\).

In Exp. 3, nineteen additional children were tested and excluded for the following reasons: (1) they failed to demonstrate understanding of the DCCS on the pre-switch \((n = 2)\); (2) they passed the post-switch phase of the DCCS \((n = 9)\); (3) they performed randomly (i.e., did not pass or fail) \((n = 2)\); (4) they were persistently inattentive or refused to complete one or more post-training tasks \((n = 5)\); or (4) they failed to return \((n = 1)\). Those who were included did not differ from those who were excluded in terms of age, \(F(1, 45), p = .68\), or gender, \(\chi^2(1, N = 47) = .01, p = .59\). There was, however, a marginally significant difference in terms of verbal intelligence, \(F(1, 36) = 3.93, p = .06\), with those excluded having higher standardized PPVT scores \((110, SD = 20.90)\) than those who were included \((99, SD = 12.13)\). This difference, which was expected, was driven by those who passed the pre-training DCCS on Day 1 and were excluded for this reason.

In Exp. 4, an additional 54 preschoolers were tested who passed the pre-training DCCS \((n = 41)\) or did not unambiguously fail the pre-training DCCS (i.e., 20 or more errors out of 30 post-switch trials to fail; \(p < .05\); see below; \(n = 13)\). These children were excluded from further testing. An additional 40 children 18 males; \(M\) age = 40.18, \(SD = 4.58\) were tested but excluded from the final sample for the following reasons: (a) they did not return for post-testing \((n = 4)\); (b) equipment failure \((n = 4)\); (c) error in testing/training administration \((n = 4)\); or (d) children moved or vocalized excessively, or otherwise refused to participate \((n = 15)\).

EEG data from five children (two in the reflection training condition and three in the mere practice condition) were unusable for the following reasons: (a) the sensor net was not
fitted well resulting in high impedance and very few usable ERP segments \((n = 1)\) or excessive artifacts (extremely high amplitude ERPs \((> 4 \text{ SDs} \text{ away from the mean}; \ n = 1)\); (b) stimulus events were not recorded during the EEG session \((n = 2)\); (c) random performance on Day 2 \((15/30 \text{ trials correct})\) and not enough usable ERP segments \((n = 1)\). Day 1 (pre-training) data from 35 children were used as part of Exp. 1.

**Design**

Twenty-nine children participated in Exp. 2, 28 in Exp. 3, and 56 in Exp. 4. For each experiment, children were randomly assigned to a reflection training condition (Exp. 2: \(n = 15\); Exp. 3: \(n = 14\); Exp. 4: \(n = 20\)) or a control condition (Exps. 2 & 3: relative clause training, \(n = 14\); Exp. 4: corrective feedback, \(n = 16\); mere practice, \(n = 20\)). On the first day, all participants received a pre-training DCCS. Children in Exps. 2 and 3 received additional pre-training assays: a relative clause test (control task) and a far-transfer false belief task (see Kloo & Perner, 2003, Exp. 2). On the last day, all participants received the same pre-training assays as on the first day. Children in Exps. 2 and 3 received an additional near transfer task (i.e., a 3-boxes DCCS with different test and target cards) on the last day.

**Procedures.** In general, children first received pre-training assays followed by reflection training or control training (either one session on the same day: Exps. 3 and 4, or two sessions separated by one week intervals: Exp. 2), and then they received post-training assays (either two days later: Exps. 3 and 4, or one week later: Exp. 2). The procedures used in Exps. 2 and 3 were based on Kloo and Perner (2003, Exp. 2) with the following exceptions: 1) the Piagetian number-conservation training and the false belief training were dropped because our theoretical interest was in reflection; 2) the Peabody Picture Vocabulary Test-III (PPVT-III; Dunn, & Dunn, 1997) was used instead of the K-ABC as a test of verbal intelligence, and it was always administered after all post-training assays on the last day; 3) children in Exps. 3 and 4 were given half the
training dosage given to children in Exp. 2. That is, they received only one training session (reflection training: 20 trials; control training: 4 scenarios) that occurred immediately following pre-testing on Day 1, and there was a 1-day interval between training and post-training tests.

**Pre- and Post-Training DCCS**

The pre- and post-training DCCS used in Exps. 2 and 3 is the same as that used in Klo and Perner (2003, Exp. 2). The pre-and post-training DCCS used in Exp. 4 is the same as that used in Exp. 1 (see Fig. 1). An experimenter blind to children’s experimental assignment administered all pre- and post-training assays, and a different experimenter administered training.

**Reflection Training**

**Training DCCS.** The training procedure used in Klo and Perner (2003, Exp. 2) was used in all three experiments, and each 15 minute session involved 20 trials with three switches between dimensions (4 blocks of 5 trials each). Any time children made a mistake they were trained. In general, reflection training consisted of asking the child to name the relevant dimension, providing an example of a correct sort, and then asking the child to sort with assistance (see Fig. 2). At the end of each session, four rapid switch trials were administered in which each test card had to be sorted by one dimension and then by the other.

Reflection training was the same in all three experiments with the following exceptions: In Exp. 4, training was adapted for use with a computerized version of the DCCS and the training script was adjusted accordingly (e.g., if you see a yellow one, press this button; see Appendix A). In Exps. 2 and 3, the dimensions used in the Training DCCS were number and color, as in Klo and Perner (2003), whereas in Exp. 4, the dimensions were shape and color. In Exps. 2 and 3, two sets of training DCCS cards with different attributes were used (Set 1 with two yellow apples and one green apple vs. Set 2 with one red house and two blue houses). In
Exp. 2, a different set was used for each training session, and in Exp. 3, the sets were counterbalanced such that half of the participants received one set, half the other. In Exp. 4, only attributes from Set 1 were used.

**Control Training**

In Exps. 2 and 3, control training focused on the use of relative clauses (Appendix B). The procedure carried out by Kloo and Perner (2003, Exp. 2) was replicated except that a baby doll was used instead of Ernie. In Exp. 4, two different control conditions were used: corrective feedback and mere practice. Children in the corrective feedback condition received the training DCCS and were simply given corrective feedback, without any further explanation or reflection training, on all trials: “That’s right/that’s wrong you are supposed to press this button” [the experimenter pointed to the correct button]. Children in the mere practice condition received the Training DCCS and were not given feedback of any kind.

**Electroencephalographic Recording**

In Exp. 4, EEG data were collected during the computerized version of the pre- and post-training DCCS, and were recorded and processed identically to Exp. 1. The mean inter-trial interval between children’s response to the test stimulus and when the experimenter initiated the following trial was 3164.38 ms ($SD = 1038.63$ ms) for the pre-training DCCS and 3106.67 ms ($SD = 1215.87$ ms) for the post-training DCCS.

In keeping with Exp. 1, the N2 was coded as the largest negative deflection after the N1 (100 to 300 ms) with a frontal-central topography and a latency of 300 to 500 ms. ERP amplitudes were calculated based on the peak within the coded region, and latencies were calculated as the time in milliseconds from stimulus onset to peak amplitude. ERP amplitudes were examined at four frontal, midline sites including sites 16, 11, 6, and 129, corresponding
roughly to Afz, Fz, Fcz, and Cz respectively (Luu & Ferree, 2000). N2 amplitudes were maximal over these fronto-central sites.

**Data Analysis**

Following intent-to-treat principles, designed to minimize the introduction of a sampling bias through attrition, all participants were included in all analyses whenever it was possible and appropriate to do so (e.g., children were only included if they failed the DCCS but were retained regardless of their performance during training). Exceptions are discussed further in the Results section. We were unable to include data from children who refused to cooperate or who did not return for all sessions, and this may limit the generalizability of the findings to children who are relatively attentive or cooperative. In all three experiments, the same logic was used to test our hypothesis that reflection training would produce improvements in EF performance compared to control training. Children who failed the DCCS were randomly assigned to the experimental or control condition(s), performance was assessed by individuals who were uninformed about the participants’ group status, and we used repeated-measures analyses of variance (ANOVAs) to test for Group (reflection training vs. control(s)) X Time (pre- vs. post-training) interactions in children’s EF performance. In all three experiments, we also tested for between-groups effects using ANOVAs on percentage change in performance. The efficacy of reflection training to improve EF performance would be supported by the presence of a significant omnibus interaction, along with paired-samples *t*-tests (Exps. 2 and 3) or multiple post-hoc comparisons with Bonferroni adjustment (Exp. 4) indicating an effect of time (i.e., improvement) for the reflection training group but not for the control condition(s). Mann-Whitney U and Wilcoxon Matched Pairs tests were also carried out to obtain analogous non-parametric results. Effect sizes (Cohen’s *d*; Cohen, 1988) were calculated correcting for the dependence between means for
within-subjects effects (Cepeda, 2008). Following Cohen (1988), effect sizes were considered small ($d < .20$), moderate ($0.20 \leq d \leq 0.80$), or large ($d > 0.80$).

In Exp. 4, to test our hypothesis that reflection training would produce reductions in N2 amplitude and source activation compared to control training, we used ANOVA to examine the effect of group on changes in N2 amplitude and source activation, and independent samples $t$-tests were used to deconstruct this effect. The efficacy of reflection training in helping children resolve the conflict inherent in the DCCS would be supported by a significant effect of group, with children who received reflection training showing smaller N2 amplitudes and reductions in medial frontal source activation. Improvements in EF performance were expected to correlate with electrophysiological changes, based on Pearson correlations.

**Results: Exps. 2 & 3**

**Scoring**

Children were given a score out of 5 for each phase of the pre- and post-training DCCS (i.e., pre-switch and post-switch), a score out of 4 for the relative clause test, a score out of 2 for false belief performance, and a score out of 9 for each phase of the 3-boxes DCCS. Scores were transformed into a percentage out of the total possible score for the purpose of analysis. In contrast to Kloo and Perner (2003, Exp. 1), whose sample included children who correctly sorted as many as 3 out of 5 trials on the post-switch phase of the DCCS, and potentially included children who responded correctly to 4 out of 4 control task scenarios, children were included in this experiment only if they clearly passed the pre-switch phase of the pre-training DCCS (i.e., less than 2 pre-switch errors out of 5, $p < .06$) and clearly failed the post-switch phase (4 or more post-switch errors out of 5, $p < .06$), committed at least one error on the relative clause test, and failed at least one of two false-belief questions. The pre-switch criterion (i.e., clearly passed the pre-switch phase) was included in order to ensure that children understood the basic sorting
requirements and the post-switch criteria were designed to minimize the likelihood of Type 1 error (i.e., false positive responses to intervention).

**Preliminary Analyses**

Six preliminary analyses were carried out to test for pre-training differences between the groups. Results based on ANOVA/Chi-square did not reveal any significant differences between the training conditions in children’s age, PPVT standard scores, gender, or pre-training performance, with the exception of pre-training false belief performance in Exp. 3, $F(1, 26) = 4.45, p = .05$; Mann-Whitney U test $Z = -1.99, p = .05$), where children in the relative clause training condition answered more false belief questions correctly than children in the reflection training condition.

**Post-Training Performance**

Results of the pre- and post-training DCCS, relative clause test, and false belief test for Exps. 2 and 3 are shown in Table 1. For comparative purposes, data were analyzed using the parametric tests carried out by Kloo and Perner (2003, Exp. 2). Given that DCCS data are typically bimodal and that the distribution for all dependent measures departed from normality (Shapiro-Wilk, all $p$s < .01), however, non-parametric results for each analysis are also presented. Results were nearly identical regardless of statistical approach. To be conservative, all reported tests of significance are two-tailed, although one-tailed test results are also appropriate because analyses were based on a priori uni-directional predictions.

**Post-training DCCS.** Children who sorted 80% of post-switch trials correctly (i.e. at least 4 out of 5 correct; $p = .06$) on the post-training DCCS were scored as passing. Eight of fifteen (53%) children in Exp. 2 who received reflection training passed, as did 7 of 14 children (50%) in Exp. 3. In contrast, only 2 children (14%) who received relative clause training passed the post-training DCCS in each experiment.
A 2 (day: Day 1, pre-training vs. Day 4, post-training) X 2 (training condition: reflection vs. relative clause) repeated measures analysis of variance (ANOVA) was carried out on the percentage of correct post-switch trials. Results of both experiments revealed a significant main effect of day (Exp. 1: $F(1, 27) = 21.48, p < .001, \eta^2_p = .44$; Exp. 3: $F(1, 26) = 15.54, p < .001, \eta^2_p = .37$), as well as a significant Day X Training Condition interaction (Exp. 2: $F(1, 27) = 5.66, p = .03, \eta^2_p = .17$; Exp. 3: $F(1, 27) = 5.76, p = .02, \eta^2_p = .18$). Children who received reflection training showed a significant improvement in DCCS performance (Exp. 2: $t(14) = -4.39, p = .001$, Wilcoxon Matched Pairs test $Z = -2.87, p < .004$; Exp. 3: $t(13) = -3.96, p = .002$, Wilcoxon Signed Ranks test $Z = -2.71, p < .007$), whereas children in the control condition did not (Exp. 2: $t(13) = -1.94, p = .08, Z = -1.83, p = .07$; Exp. 3: $t(13) = -1.29, p = .22, Z = -1.09, p = .28$). Also, a one-way ANOVA on change scores (i.e. percentage correct post-training DCCS – percentage correct pre-training DCCS) showed that improvement on the DCCS was greater for the reflection training group (Exp. 2: 53%, SD = 47%; Exp. 3: 53%, SD = 50%) than for the relative clause training group (Exp. 2: 17%, SD = 33%, $F(1, 27) = 5.66, p = .03$, Cohen’s $d$: 0.90; Mann-Whitney test $Z = -2.23, p = .03$; Exp. 3: 13%, SD = 37%, $F(1, 26) = 5.76, p = .02$, Cohen’s $d$: 0.90; $Z = -2.34, p = .02$).

**False belief (far transfer).** Group differences in false belief test scores were examined using a 2 (day) X 2 (training condition) repeated measures ANOVA. In Exp. 2, but not in Exp. 3, there was a main effect of day, $F(1, 27) = 7.69, p = .01, \eta^2_p = .22$, indicating that both groups performed better on Day 2 (29% correct, SD = 5%) than on Day 1 (14% correct, SD = 4%). No Day X Training Condition interaction was found in either Exp. 2 or Exp. 3, consistent with the findings of Kloo and Perner (2003, Exp. 2). Given that a group difference in improvement between Day 1 and Day 2 was predicted a priori, paired-samples t-tests were carried out on change scores. In Exp. 2, children in the reflection training condition showed a significant
improvement in performance, $t(14) = -2.82, p = .01$; Wilcoxon Matched Pairs test $Z = -2.33, p = .02$, whereas those in the relative clause condition did not, $t(13) = -1.00, p = .34; Z = -1.00, p = .32$. In Exp. 3, the paired samples $t$-tests did not show a significant facilitation effect for either the relative clause training group, $t(13) = .37, p = .72; Z = -.38, p = .71$, or the reflection training group, $t(13) = -1.47, p = .17; Z = -1.41, p = .16$. A one-way ANOVA on change scores confirmed that improvement on the false belief test was greater for the reflection training group (Exp. 2: 23.3%, $SD = 32%$; Exp. 3: 14.3%, $SD = 36%$) than for the relative clause training group in Exp. 2 (7%, $SD = 27%$, $F(1, 27) = 5.66, p = .03$, Cohen’s $d = .54; $Mann-Whitney test $Z = -1.33, p > .1$) but not Exp. 3 (0%, $SD = 37%$, $F(1, 26) = 1.68, p > .1$, Cohen’s $d = .38; Z = -1.18, p > .1$).

**Relative clause test (control task).** Performance on the relative clause test was analyzed using a 2 (day) X 2 (training condition) repeated measures ANOVA. This analysis revealed a significant main effect of day (Exp. 2: $F(1, 27) = 13.50, p = .001, \eta^2_p = .33$; Exp. 3: $F(1, 26) = 13.50, p = .003, \eta^2_p = .30$), as well as a significant Day X Training Condition interaction in Exp. 2: $F(1, 27) = 6.66, p = .02, \eta^2_p = .20$, but not Exp. 3. There was a significant improvement in performance for the relative clause training group, $t(13) = -3.17, p < .01$; Wilcoxon Matched Pairs test $Z = -2.41, p = .02$, but not the reflection training group, $t(14) = -1.87, p = .08; Z = -1.73, p = .08$. A one-way ANOVA on change scores also revealed that the pre- to post-training improvement on the relative clause test was greater for the relative clause training group (29%, $SD = 34%$) than for the reflection training group (5%, $SD = 10%$, $F(1, 27) = 5.66, p = .03$, Cohen’s $d$: 1.09; Mann-Whitney test $Z = -2.23, p = .03$.

In contrast to Exp. 2, in Exp. 3, there were significant improvements from pre- to post-training in both the relative clause training group (20% improvement, $t(13) = -2.47, p = .03$; Wilcoxon Matched Pairs test $Z = -2.04, p = .04$) and the reflection training group (18% improvement; $t(13) = -2.22, p = .05; Z = -2.04, p = .04$).
3-Boxes DCCS (near transfer). As with the standard DCCS, children who correctly sorted 80% of trials (i.e., at least 7 out of 9) on the 3-boxes DCCS were considered to have passed. Based on this criterion, 67% of children who received reflection training passed the 3-boxes DCCS in Exp. 2, and 71% passed in Exp. 3. In comparison, only 14% of children who received relative clause training passed the 3-boxes DCCS in Exp. 2, and only 7% did so in Exp. 3. A one-way ANOVA was carried out on the mean percentage of post-switch trials correct. Results revealed higher percentages for reflection training (Exp. 2: 73.3%, \(SD = 38.2\%\); Exp. 3: 77.0%, \(SD = 35.3\%\)) than for relative clause training (Exp. 2: 28.6%, \(SD = 33.4\%\); Exp. 3: 13.5%, \(SD = 29.3\%\)) in both experiments, Exp. 2: \(F(1, 27) = 11.21, p = .002\), Cohen’s \(d\): 1.25; Mann-Whitney test \(Z = -2.89, p = .004\); Exp. 3: \(F(1, 26) = 26.86, p = .001\), Cohen’s \(d\): 1.97; \(Z = -3.41, p = .001\).

Dosage Effects

In order to assess the effects of reflection training dosage/duration on DCCS performance, data from Exp. 2 were compared to data from Exp. 3 using Wilcoxon-Mann-Whitney tests. No significant differences were observed in post-training performance or the percentage change in performance (all \(Zs < 1\), all \(ps > .30\)).

Results: Exp. 4

Preliminary Analyses

All children passed the pre-switch phase of the DCCS on Day 1 and Day 2, making 6 or fewer errors out of a possible 15 (\(M\) pre-training = 1.41 errors, \(SD = 1.56\); \(M\) post-training = .80; \(SD = 1.42\)). On the post-switch phase, children were classified as passing if they made 10 or fewer errors out of 30 (\(p = .05\)) and were classified as failing if they made 20 or more errors (\(p \sim .05\)).
Eight preliminary analyses were carried out to test for pre-training differences between the groups. Results based on ANOVA/Chi-square were that children in the three training conditions did not differ in age, sex, ethnicity, median post-switch RT on Day 1, mean number of post-switch errors on Day 1, mean number of trials contributing to the post-switch ERP waveforms on Day 1 or Day 2, or verbal ability (as measured by standard scores on the PPVT-III; Dunn & Dunn, 1997). Subsequent analyses collapsed across these variables.

**Behavioral Analyses**

**Training DCCS performance.** Children who received reflection training made 1.80 (SD = 1.58) errors on the training DCCS, compared to 3.25 (SD = 3.05) for corrective feedback, and 7.10 (SD = 4.22) for mere practice. A one-way ANOVA revealed a significant effect of condition on errors made, \( F(2, 53) = 15.04, p < .001, \eta_p^2 = .36 \). Multiple comparisons with Bonferonni adjustment indicated that both children who received reflection training and children who received corrective feedback made fewer errors (Cohen’s \( d = .060 \) and 1.66 respectively) than children who received mere practice (all \( ps < .01 \)).

**Post-training DCCS performance.** Eleven children (55%) in the reflection training condition passed the post-training DCCS (one of whom made 11 errors out of 30 instead of 10), in contrast to two children (13%) in the corrective feedback condition and two children (10%) in the mere practice condition. Children who received reflection training made 14.90 post-switch errors on the post-training DCCS (SD = 11.39), compared to 25.00 (SD = 8.20) for corrective feedback, and 25.50 (SD = 7.32) for mere practice. A repeated measures ANOVA revealed a main effect of day, \( F(1, 53) = 1.44, p < .001, \eta_p^2 = .21 \), and a Day X Training Condition interaction, \( F(2, 53) = 8.66, p < .01, \eta_p^2 = .25 \). Multiple comparisons with Bonferonni adjustment showed that the interaction was driven by a difference on Day 2 between reflection training and corrective feedback, and also between reflection training and mere practice, with children who
received reflection training making the least errors (both \( ps \leq .01 \)). There was a significant reduction in post-switch errors from Day 1 to Day 2 only for children who received reflection training (\( p < .01 \)).

A one-way ANOVA on change scores (i.e., percentage correct post-training DCCS – percentage correct pre-training DCCS) showed that improvement on the DCCS differed significantly between groups, \( F(2, 53) = 8.66, p < .01 \). Multiple comparisons with Bonferonni adjustment revealed that this difference was driven by greater improvement on the DCCS for reflection training (39%, \( SD = 36\% \)) than for corrective feedback (5%, \( SD = 29\% \), Cohen’s \( d = 1.04 \)) and mere practice (3%, \( SD=24\% \), Cohen’s \( d: 1.18 \)), both \( ps < .01 \).

A repeated measures ANOVA on post-switch RTs revealed a main effect of day, \( F(1, 49) = 1.68, p < .001, \eta_p^2 = .26 \), and a Day X Training Condition interaction, \( F(2, 49) = 6.22, p < .01, \eta_p^2 = .20 \) (RT data are missing from 4 participants due to equipment failure). Multiple comparisons with Bonferonni adjustment showed that children who received reflection training had significantly longer RTs on the post-training DCCS than children who received mere practice (\( p = .02 \), Cohen’s \( d = .86 \)). There was a significant increase in RTs between Day 1 and Day 2 for children who received reflection training and for children who received corrective feedback (both \( ps < .01 \)), but not for children who received mere practice. Median post-switch RTs (and \( SDs \)) on Day 1 and Day 2 were as follows: reflection training, 1463.58 ms (607.81) and 1796.56 ms (688.95); corrective feedback, 1317.60 ms (396.62) and 1611.83 (642.08); mere practice, 1371.68 ms (361.62) and 1336.82 ms (300.98).

**ERP analyses**

ERP analyses focused on the N2 on post-switch trials during the pre-training and post-training DCCS (see Fig. 3 for the pre- and post-DCCS waveforms represented separately for each condition). For the pre-training DCCS, N2 amplitudes were always based on incorrect
trials, whereas for the post-training DCCS, they were based on correct trials for children who passed, and incorrect trials for children who failed. Out of the original sample of 56 children, EEG data were available for 51 children: 18 children in the reflection training condition, 16 in the corrective feedback condition, and 17 in mere practice.

**Pre- to Post-training Change in N2 Amplitude**

A series of ANOVAs compared the effect of training on changes in N2 amplitude (change scores = post-training DCCS N2 amplitude – pre-training DCCS N2 amplitude). A significant effect of training condition across all sites, $F(2, 48) = 3.21, p = .05, \eta_p^2 = .12$, was mainly driven by a significant difference between conditions at site 16, $F(2, 48) = 3.17, p = .05, \eta_p^2 = .12$ (site 6, $F(2, 48) = 3.00, p = .06, \eta_p^2 = .11$; site 11, $F(2, 48) = 2.55, p = .09, \eta_p^2 = .10$; site 129, $F(2, 48) = 1.87, p = .16, \eta_p^2 = .07$). Independent samples $t$-tests to assess the effect of training condition showed that reductions in N2 amplitude were greater for reflection training than mere practice at site 16, $t(33) = 2.74, p = .01$, Cohen’s $d$: .92; site 11, $t(33) = 2.40, p = .02$, Cohen’s $d$: .82; site 6, $t(33) = 2.07, p = .02$, Cohen’s $d$: .83; and across all sites $t(33) = 2.64, p = .01$, Cohen’s $d$: .90; but not at site 129, $t(33) = 1.49, p = .15$ (see Table 2). Degree of amplitude change did not differ significantly between reflection training and corrective feedback at any site (all $ts < 1.30$, all $ps > .10$). Amplitude change differed significantly between corrective feedback and mere practice only at site 129, $t(31) = 2.06, p = .05$, Cohen’s $d$: .72, with the corrective feedback condition showing a greater reduction.

**Pre- to Post-training Change in N2 Latencies**

An analogous set of ANOVAs revealed no significant effect of training condition on change in post-switch N2 latency from Day 1 to Day 2, at any particular site, or across all sites, all $Fs(2, 48) < 2.11, p > .13$.

**Source Analyses**
The topography and source distribution of the N2 are shown separately for each condition in Figure 4. Also shown is the topography and source distribution of a difference wave created by subtracting the peak post-switch N2 amplitudes on Day 1 from the peak post-switch amplitudes on Day 2 separately for each of the groups at each electrode site. Visual inspection of the topographies reveals a broad negativity across fronto-central regions and a pattern of increasing reduction in negativity with each level of training (practice to corrective feedback to reflection).

EEG source imaging was performed as per Exp. 1 using the standardized low-resolution brain electromagnetic tomography (sLORETA) method (Pascual-Marqui, 2002), as implemented in the NetStation GeoSource 2.0 software package. This method relies on a Sun-Stok 4-shell Sphere head model with a Tikhonov regularization of 110^{-4}. There were several cingulate and mediofrontal cortical source activations that likely contributed to the N2, as well as more posterior sources in occipital and precuneus areas. Source activation levels on both days were greatest in mediofrontal regions. A series of one-way ANOVAs were carried out to compare the changes in activation levels (Day 2 – Day 1) across training conditions. There was a lateralized effect of training condition on left central cingulate activation, $F(2, 48) = 3.11$, $p = .05$, with children in the reflection training and corrective feedback conditions showing a greater reduction in activation ($M = -.02$, $SD = .08$; $M = -.03$, $SD = .11$, respectively) than children in the practice group (Cohen’s $d$: .82 and .80 respectively), who showed an increase in activation ($M = .05$, $SD = .09$). A similar modulatory effect of training condition on right central cingulate source activation was not seen.

Correlations

Number correct on the training DCCS was significantly associated with the magnitude of pre- to post-training improvement on the DCCS (post-switch errors on the pre-training DCCS –
post-switch errors on the post-training DCCS; \( r = .35, p = .01 \) and marginally associated with a reduction in N2 amplitude \( (r = .26, p = .07) \) and increase in latency \( (r = .34, p = .08) \). Reduction in N2 amplitude was associated with an increase in post-switch RT \( (r = .33, p = .02, N = 47) \). Passing (versus failing) the DCCS after training was associated with an increase in post-switch RT \( (r = .34, p = .02, N = 47) \), and a chi-square analysis further revealed that post-training DCCS performance status (pass versus fail) was associated with the direction of N2 amplitude change (increase versus decrease), \( \chi^2(1, 51) = 4.60, p = .03 \) such that children who passed tended to show a decrease in amplitude.

**N1 Amplitude**

A one-way ANOVA was used to also examine differences in the post-switch ERP (Day 2 – Day 1) within an N1 timeframe typical for preschool age children (100–300 ms; Sanders, Stevens, Coch, & Neville, 2006; Stevens, Lauinger, & Neville, 2009). Both children who received reflection training and children who received corrective feedback showed a significantly greater reduction in peak N1 amplitude (reflection training: \( M = 1.80, SD = 4.79 \); corrective feedback: \( M = 1.88, SD = 4.79 \)) than children who received mere practice (\( M = -1.28, SD = 2.65 \), \( F(2,48) = 3.36, p = .04, \eta^2_p = .12 \). These differences were pronounced at more frontal sites (11 and 16; \( p < .05 \)).

**Discussion**

The results of Exp. 2 closely mirrored the findings of Kloo and Perner (2003, Exp. 2). Whereas Kloo and Perner found improvements on the DCCS of 57% and 17% for reflection training and relative clause training, respectively, we found improvements of 53% and 17%. Children in Exps. 2 and 3 also showed near transfer to a more complicated 3-boxes DCCS task version, and children in Exp. 2 showed far transfer to a false belief task. Effect sizes associated with the effect of reflection training on DCCS performance and near transfer were large, whereas
those associated with far transfer task were moderate. These results are consistent with the suggestion that reflection plays a key role in EF and its early development. Reflection training may teach children to notice the conflict inherent in the bivalent stimuli, reflect on it, and formulate rules appropriate to the hierarchical structure of the task.

The results of Exp. 3 showed that even a single brief, 15 min session of reflection training led to substantial improvements in DCCS performance, although there was less evidence for far transfer to false belief performance. These results were associated with large and moderate effect sizes, respectively. In Exp. 4, children who received a brief session of reflection training also showed substantial improvement in performance (55% passed). Importantly, these dramatic improvements in performance co-occurred with corresponding changes at the neurophysiological level. Children who received reflection training showed a substantial reduction in post-switch N2 amplitude during DCCS performance, as well as an increase in RT. All associated effect sizes were large. Slowing down may afford the time needed to reflect on the hierarchical nature of the task. There was also a pattern of reduced source activation with increasing levels of training, consistent with the neurodevelopmental literature noting a shift from diffuse to more localized activation with development (e.g., Casey, Galavan, Hare, 2005; Durston, Davidson, Tottenham, Galvan, Spicer, Fossella, et al., 2006; Lamm et al., 2006). These results add to a growing body of literature demonstrating that neural networks involved in processing conflict can be modified with EF training in early childhood (see Rueda et al., 2005; 2012) and they further support the suggestion from Exp. 1 that children who pass the DCCS reflect on their rule representations and efficiently resolve the conflict inherent in the task. The neurophysiological effects of reflection training (reduction in N2 amplitudes) resemble those that occur naturally with increasing EF (see Exp. 1; Lamm et al., 2006; Jonkman, 2006).
Despite demonstrating fewer errors on the training-DCCS, children in the corrective feedback condition in Exp. 4 showed little improvement on the post-training DCCS. These findings replicate Bohlmann and Fenson (2005), who also found an immediate benefit of corrective feedback on DCCS performance but no transfer to a subsequent DCCS task. It is possible, however, that corrective feedback may have initiated some spontaneous reflection on children’s rule representations. These children did show an increase in RT on the post-training DCCS, at a level intermediate between children in the reflection training and mere practice conditions, and they did show a reduction in N2 amplitudes. It is possible that they slowed down in anticipation of feedback, and that in the absence of feedback on the post-training DCCS, they assumed that they were sorting correctly (and hence discounted the conflict in the task). The N1 findings are consistent with this suggestion indicating that both children who received reflection training and corrective feedback attended selectively (indicated by reduced N1 amplitudes) but children who received reflection training attended to the relevant dimension and switched flexibly, whereas children who received corrective feedback attended to the irrelevant dimension and perseverated.

Similar to children who received corrective feedback, children who received mere practice also showed little improvement on the post-training DCCS, however, they did not slow down and they showed increases in both N1 and N2 amplitudes. The extent of these ERP effects was greater than that of reflection training. These results suggest that children who received mere practice did not attend selectively to the relevant dimension and experience sustained, unresolved conflict. Although a number of studies have demonstrated practice effects on singular, lower-order measures of EF (e.g., Shalev, Yehoshua, Mevorach, 2007; Rueda et al., 2005; Klingberg et al., 2005; Dowsett et al., 2000), bottom-up teaching strategies (e.g., mere practice) may increase
conflict on more complex, higher-order measures of EF in the absence of reflective processing and a hierarchical understanding of the task.

Taken together, the findings of this set of experiments suggest that performance on measures of EF, such as the DCCS, responds readily to top-down teaching strategies (e.g., reflection training; see also Autin & Croizet, 2012; Brace, Morton, & Munakata, 2006; Mack, 2007; Iseman & Naglieri, 2011). These findings have particular relevance for education, particularly in the early school years. Effective strategies for inducing flexible, adaptive behavior, as well as transfer, include encouraging children to reflect on their rule representations, rather than simply telling them when a behavior is incorrect or inappropriate.

Previous efforts to train EF in young children include broad curricular interventions (e.g., through academic and extra-curricular programming such as Montessori education, piano training, martial arts and general physical activity; see Diamond & Lee, 2012 for a review). However, the current findings also point to the value of more focused interventions that target key processes underlying EF (see also Dowsett & Livesey, 2000; Karbach & Kray, 2009; Klingberg et al., 2005; Jolles et al., 2012; Klo & Perner, 2003; Rueda et al., 2005; 2012; Tamm et al., 2010; Thorell et al., 2009).

Individual differences in EF predict important social, academic, and lifelong outcomes (Best, Miller, & Naglieri, 2011; Moffit et al., 2011). The results of this set of experiments suggest that the development of EF can be facilitated in the preschool years, just prior to school entry when individual differences in early EF abilities begin to significantly influence academic trajectories (Best et al., 2011). Although previous studies have demonstrated neurophysiological changes in older children in response to working memory training (Söderqvist, Bergman, Nutley, Peyrard-Janvid, Matsson, Humphreys, Kere, et al., 2012; Jolles et al., 2011; 2012) and switch training (Karbach et al., 2009), and in response to attention training in younger children (Rueda
et al., 2005; 2012), Exp. 4 represents the first study to isolate an effect of reflection training on neural responses to conflict, indexed by the amplitude of the N2 component of the ERP.

Interventions designed to promote the development of EF may work by challenging children to reflect on their mental representations in relation to broader contextual considerations. For example, children who participated in an academic intervention found to facilitate EF (Tools of the Mind; Bodrova & Leong, 1996) learned to restrain from talking when another child was talking, by considering contextual constraints (e.g., Is it my turn to talk or is it my partner’s turn?), which were symbolized by a picture they held of either an ear (not my turn) or a mouth (my turn) (Diamond et al., 2007). Teaching children to reflect on their rule representations may help them to pursue goals in a more top-down, evaluative fashion, allowing conflicts and inconsistencies in their representations to be more readily detected and resolved.

**Conclusion**

Children who were taught how to reflect on their rule representations in the DCCS responded more slowly, were more likely to pass the task, and showed reductions in N2 amplitude. These findings point to conflict monitoring and reflective processing of information as key mechanisms in the development of flexible rule use, and they are consistent with a model of EF in which conflict detection initiates reflective reprocessing (mediated by lateral prefrontal cortical networks) that has the potential to resolve the conflict and down-regulate ACC activation (Bunge & Zelazo, 2006; cf. Badre & D’Esposito, 2007; Botvinick, 2008; Christoff & Gabrieli, 2000; Koechlin et al., 2003).
References


Appendix A

Computerized Reflection Training Protocol

Statement: “When you saw the yellow one, you pressed the button with yellow on it and that means you looked at the color. But, we are not playing the color game, the game with yellow and green, anymore. Now, we are playing the shape game - the game with apple and house [experimenter corresponding buttons].”

Question: “What game are we playing now?”

Feedback: “Right/No. We are playing the shape game—the game with apple and house. So, you have to look at what shape is on the button,”

Question: “Look, what shape is this [experimenter points to test stimulus]?”

Feedback: “Right/No. It is an apple.”

Question: “Look, what shape is on this button [experimenter points to target stimulus on response pad]?”

Feedback: “Right/No. It is an apple. In the shape game, when you see an apple right here [experimenter points to test stimulus], you have to press the button with apple on it [experimenter points to target stimulus on response pad].”

Question: “So, which button do you press when you see an apple in the shape game?”

Feedback: “Right/No. When you see an apple right here [experimenter points to test stimulus], you press the button with an apple on it [experimenter points to target stimulus on response pad].”

Question: “Look, here’s an apple. Which button do you have to press in the shape game [experimenter ensures that child responds correctly]?”
Appendix B

Relative Clause Control Training Protocol

Training Set 1

Scene 1

The following scene was enacted: “This girl is jumping up and down. This girl is shaking her head. The baby comes along and kisses the girl who was shaking her head.”

The twin dolls were hidden under the table, and the children were asked: “Which girl did the baby kiss?” In the case of no response, the two answer alternatives were provided: “Did the baby kiss the girl who had been jumping up and down or did the baby kiss the girl who had been shaking her head?”

The children received positive or negative feedback: “Right/No, the baby kissed the girl who had been shaking her head.” Finally, the relevant actions were enacted once again: “This girl is shaking her head. The baby comes and kisses the girl who was shaking her head.” That is, the correct relative clause was modeled once again.

Scene 2

The following scene was enacted: “This girl is wearing a green hat. That girl is wearing a blue hat. The baby comes along and baby hugs the girl who is wearing a green hat.”

The twin dolls were hidden under the table, and the children were asked: “Which girl did the baby hug?” In the case of no response, the two answer alternatives were provided: “Did the baby hug the girl who was wearing a green hat or did the baby hug the girl who was wearing a blue hat?”

The children received positive or negative feedback: “Right/No, the baby hugged the girl who was wearing a green hat.” Finally, the relevant actions were enacted once again: “This girl is wearing a green hat. The baby comes and hugs the girl who is wearing a green hat.”

Scene 3

The following scene was enacted: “This girl is sitting on the floor. That girl is lying on the floor. The baby comes along and trips over the girl who is lying on the floor.”

The twin dolls were hidden under the table, and the children were asked: “Which girl did the baby trip over?” In the case of no response, the two answer alternatives were provided: “Did the baby trip over the girl who was sitting on the floor or did the baby trip over the girl who was lying on the floor?”
The children received positive or negative feedback: “Right/No, the baby tripped over the girl who was lying on the floor.” Finally, the relevant actions were enacted once again: “This girl is lying on the floor. The baby comes and trips over the girl who is lying on the floor.”

**Scene 4**

The following scene was enacted: “This girl is turning a somersault. That girl is stretching her arms over her head. The baby comes along and taps the girl on the shoulder who was turning a somersault.”

The twin dolls were hidden under the table, and the children were asked: “Which girl did the baby tap on the shoulder?” In the case of no response, the two answer alternatives were provided: “Did the baby tap the girl on the shoulder who had been turning a somersault or did the baby tap the girl on the shoulder who had been stretching her arms over her head?”

The children received positive or negative feedback: “Right/No, the baby tapped the girl on the shoulder who had been turning a somersault.” Finally, the relevant actions were enacted once again: “This girl is turning a somersault. The baby comes and taps the girl on the shoulder who was turning a somersault.”

**Training Set 2**

Scenes 5-8 followed the same general structure. (The particular actions of the twins remained the same. But now, the baby performed a different action, and furthermore the baby interacted with the alternative twin.) Therefore, only the “enacted scene” is described.

**Scene 5**

The following scene was enacted: “This girl is jumping up and down. That girl is shaking her head. The baby comes along and dances with the girl who was jumping up and down.”

**Scene 6**

The following scene was enacted: “This girl is wearing a green hat. That girl is wearing a blue hat. The baby comes along and tickles the girl who is wearing a blue hat.”

**Scene 7**

The following scene was enacted: “This girl is sitting on the floor. That girl is lying on the floor. The baby comes along and takes the hand of the girl who is sitting on the floor.”

**Scene 8**

The following scene was enacted: “This girl is turning a somersault. That girl is stretching her arms over her head. The baby comes along and whispers in the ear of the girl who was stretching her arms over her head.”
Table 1

Mean Percentage Correct (and Standard Deviations) on Pre- and Post-Training DCCS, False Belief Task, and Relative Clause Task, and Post-Pre Differences (\(\Delta\)), for the Reflection and Relative Clause Training Groups in Exps. 2 and 3

<table>
<thead>
<tr>
<th>Training Condition</th>
<th>DCCS</th>
<th>False Belief</th>
<th>Relative Clause</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>(\Delta)</td>
</tr>
<tr>
<td>Reflection</td>
<td>1.3</td>
<td>54.7</td>
<td>53.3**</td>
</tr>
<tr>
<td>Exp. 1</td>
<td>(5.2)</td>
<td>(46.3)</td>
<td>(47.0)</td>
</tr>
<tr>
<td>Relative Clause</td>
<td>1.4</td>
<td>18.6</td>
<td>17.1††</td>
</tr>
<tr>
<td></td>
<td>(5.3)</td>
<td>(36.3)</td>
<td>(33.2)</td>
</tr>
<tr>
<td>Reflection</td>
<td>0.0</td>
<td>52.9</td>
<td>52.9*</td>
</tr>
<tr>
<td>Exp. 2</td>
<td>(0.0)</td>
<td>(50.0)</td>
<td>(50.0)</td>
</tr>
<tr>
<td>Relative Clause</td>
<td>1.4</td>
<td>14.3</td>
<td>12.9</td>
</tr>
<tr>
<td></td>
<td>(5.3)</td>
<td>(36.3)</td>
<td>(37.3)</td>
</tr>
</tbody>
</table>

Note. \(\Delta\) = % correct post-training – % correct pre-training; DCCS = Dimensional Change Card Sort.

* Two-tailed t-test results: \(p < .05\).
** Two-tailed t-test results: $p \leq .005$.

† One-tailed t-test results: $p = .04$.

‡‡ One-tailed t-test results: $p = .08$
Table 2

*Change in Mean N2 Amplitude (µV) from Pre-training Day 1 to Post-training Day 2 and (SDs) by Training Condition and at Four Frontal-Midline Sites*

<table>
<thead>
<tr>
<th>Phase</th>
<th>Site</th>
<th>Reflection</th>
<th>Corrective</th>
<th>Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-switch</td>
<td>16</td>
<td>-1.95 (17.57)</td>
<td>2.41 (11.25)</td>
<td>-2.40 (9.17)</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>-0.29 (16.90)</td>
<td>0.44 (10.20)</td>
<td>2.67 (9.93)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.66 (14.05)</td>
<td>3.00 (9.63)</td>
<td>-1.02 (9.05)</td>
</tr>
<tr>
<td></td>
<td>129</td>
<td>0.14 (10.44)</td>
<td>1.60 (4.53)</td>
<td>-0.63 (8.28)</td>
</tr>
<tr>
<td></td>
<td>All sites</td>
<td>-0.36 (13.68)</td>
<td>1.86 (7.67)</td>
<td>-1.68 (8.63)</td>
</tr>
<tr>
<td>Post-switch</td>
<td>16</td>
<td>1.40 (4.94)</td>
<td>0.88 (8.56)</td>
<td>-3.83 (6.30)</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>2.32 (7.71)</td>
<td>0.74 (8.21)</td>
<td>-2.98 (4.98)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2.34 (6.11)</td>
<td>-0.29 (5.65)</td>
<td>-2.12 (4.47)</td>
</tr>
<tr>
<td></td>
<td>129</td>
<td>1.45 (6.67)</td>
<td>2.21 (5.71)</td>
<td>-1.35 (4.04)</td>
</tr>
<tr>
<td></td>
<td>All sites</td>
<td>1.88 (5.61)</td>
<td>0.88 (6.25)</td>
<td>-2.57 (4.21)</td>
</tr>
</tbody>
</table>
Figure Captions

Figure 1. Sequence of events for the computerized version of the pre- and post-training Dimensional Change Card Sort.

Figure 2. Sequence of events for the computerized version of the training Dimensional Change Card Sort.

Figure 3. Grand-averaged stimulus-locked waveforms at sites 16, 11, 6, and 129, recorded during children’s performance on the post-switch phase of the pre-training DCCS (Day 1) and the post-training DCCS (Day 2). Waveforms depict the difference in N2 amplitude between pre-training and post-training waveforms for children who received reflection training, corrective feedback, and mere practice.

Figure 4 (a). Topographic plots of differences in peak N2 amplitude (post-switch trials) between Day 1 and Day 2 and corresponding source distributions for the reflection training group. (b). Topographic plots of differences in peak N2 amplitude (post-switch trials) between Day 1 and Day 2 and corresponding source distributions for the corrective feedback group. (c). Topographic plots of differences in peak N2 amplitude (post-switch trials) between Day 1 and Day 2 and corresponding source distributions for the mere practice group.
Experiment states rules.
Fixation remains until experimenter presses space bar to advance test stimulus.

Test card appears.
Participant makes response on response pad.

Experimenter imposes a 2 sec delay on average before advancing next test stimulus.

Participant responds on response pad.
**Examining N2 of Day 1 vs. Day 2: Reflection Training**

Topographic Plots of N2 Peak @ 398 ms

Source Distribution at N2 Peak (sLORETA)

**Examining N2 of Day 1 vs. Day 2: Corrective Feedback**

Topographic Plots of N2 Peak @ 398 ms

Source Distribution at N2 Peak (sLORETA)

**Examining N2 of Day 1 vs. Day 2: Mere Practice**

Topographic Plots of N2 Peak @ 398 ms

Source Distribution at N2 Peak (sLORETA)
Chapter 4: General Discussion

Early individual differences in executive function (EF), the ability to flexibly shift behavior according to contextual demands, are associated with children’s socioemotional, academic, and life-long outcomes (e.g., Moffit, Arsenault, Belsky, Dickson, Hancox, Harrington et al., 2011). A number of studies have now demonstrated promising effects of EF training and curricula designed to facilitate EF on children’s academic performance (e.g., Diamond, Barnett, Thomas, & Munro, 2007; Holmes, Gathercole, & Dunning, 2009), working memory (Klingberg, Forssberg, Westerberg, 2002; Klingberg, Fernell, Olesen, Johnson, Gustafsson, Dahlstrom, et al., 2005; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009); attention skills (Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005; Rueda, Checa, & Cómbita, 2012; Tamm, Hughes, Ames, Pickering, Silver, Stavinoha, 2010), and fluid intelligence (Rueda et al., 2005; 2012). The neurocognitive processes supporting the early development of EF are not well understood, however. The overall aim of the studies included herein was to help clarify these processes by examining how training interventions work. Specific interest was in isolating the role of reflection in EF training and its effect on conflict detection mediated by the ACC.

The present set of experiments make a number of important contributions to the literature on the processes contributing to the early development of EF. In three separate experiments using reflection training (Exps. 2, 3 and 4), strong support was provided for the idea that reflective processing facilitates preschool children’s performance on the DCCS (e.g., Zelazo et al., 2003). Exps. 2 and 3 replicated the results of Kloo & Perner, (2003, Exp. 2), including the finding of transfer to a distinct cognitive domain (ToM) in Exp. 2. Taken together with the findings of Kloo and Perner, the findings presented in Exps. 2 to 4 show that it is possible to accelerate EF development by approximately two years (the DCCS performance of children who received reflection training resembled that of most 5-year-olds), by training children to reflect on
their rule representations and to formulate a hierarchical understanding of the task. The replication of near (Exps. 2 and 3) and far transfer effects (Exps. 2) as found in Kloo et al., (2003, Exp. 2) supports the possibility of a domain-general mechanism underlying training effects. Moreover, Exp. 4 shows that the effects of reflection training are associated not only with performance improvements but also with a more efficient response to conflict at the neural level.

Despite the suggestion that training may need to be extensive to assist children in overcoming the strength of the perseverative effect on the DCCS (e.g., Kloo et al., 2003; Zelazo & Frye, 1998), Exps. 3 and 4 demonstrate that 3-year-olds can be taught to pass the task after just one brief, single-session of training targeting reflection. Transfer effects are weaker after just one day of training, however, compared to two days of training over several weeks. This dosage effect is consistent with previous findings in the adult literature indicating greater training effects (on working memory) with more extensive practice (Dahlin, Bäckman, Stigsdotter Neely, & Nyberg, 2009). Reflection training may need to be extensive for transfer to occur and for long-lasting improvement in EF.

In Exp. 1, the neural correlates of children’s performance on the DCCS were explored. N2 amplitude (but not latency) was found to be a reliable neural marker of early individual differences in this measure of executive function. Children with good EF (those who switched) had lower N2 amplitudes than children with poor EF (those who perseverated), consistent with previous research (Lamm, Zelazo, & Lewis, 2006). This finding suggests that the capacity for top-down regulation of ACC-driven conflict detection differentiates children with good EF from children with poor EF, complementing previous work establishing a role for ventrolateral prefrontal cortex in children’s successful performance on the DCCS (Moriguchi & Hiraki, 2009). By incorporating a training manipulation targeting reflection in Exp. 4, strong support was
achieved for the role of reflective processing in the down-regulation of ACC-driven conflict
detection. A particular strength of Exp. 4 is the incorporation of active controls, the lack of
which has been a limitation of previous studies (e.g., Rueda et al., 2012). These controls made it
possible to rule out extraneous influences unrelated to reflection including experimenter effects,
test-retest effects, the effect of corrective feedback, and the effect of practice in sorting and
switching.

A second important strength of Exp. 4 was that improvements in behavioral performance
on the DCCS were significantly associated with changes in the neural systems that respond to
conflict. Such specific effects of training on EF performance and on the neural networks
supporting flexible behavior have been difficult to demonstrate in young children (e.g., Jolles,
van Buchem, Rombouts, & Crone, 2012; Rueda et al., 2005; 2012). Exp. 4 represents the first
attempt in the literature to examine training-related effects on preschoolers’ flexible rule use on
the DCCS at the neural level. More generally, it is also amongst the first studies to look at the
neurophysiological effects of EF training on the development of EF in the first few years of life
(e.g., Rueda et al., 2005; 2012; Wass, Porayska-Pomsta, & Johnson, 2011).

Compared to a vast literature demonstrating robust practice effects on tasks tapping
lower-order, singular EF skills (e.g., Holmes, Gathercole, & Dunning, 2009; Klingberg et al.,
2002; Klingberg et al., 2005; Rueda et al., 2005; 2012; Thorell et al., 2009), the findings of Exp.
4 suggest that practice may not be sufficient to improve more complex EF skills in young
children (see also Bohlmann & Fenson, 2005; cf. Karbach & Kray, 2009 with older children).
Top-down training techniques (e.g., reflection training) may be necessary to improve
performance on tasks involving these higher-order EF skills. It seems necessary that children
learn to take a hierarchical, reflective stance for them to take the broader context into account
and to flexibly choose an appropriate response. These findings add to a small set of studies
identifying factors influencing EF training efficacy (e.g., Karbach & Kray, 2009; Morrison & Chein, 2011; Thorell et al., 2008), highlighting a key role for higher-order reasoning. Future training studies should target this kind of reasoning in children, possibly by promoting their use of abstract representations (e.g., Synder & Munakata, 2010) or metacognitive reframing (e.g., Autin & Croizet, 2012).

Although the findings of all four experiments presented herein are consistent with CCC-r theory, other interpretations are also possible. From a working memory perspective (e.g., Morton & Munakata, 2002), children who received reflection training in Exps. 2 to 4 may have resolved the conflict and passed the task by learning to keep the rules firmly in mind. Children who naturally resolved the conflict and passed the DCCS in Exp. 1 may have been also better able to keep the rules in mind than children who perseverated. It is less clear, however, how the observed differences in N2 amplitudes between children who switched and children who perseverated in Exp. 1 can be explained by accounts emphasizing inhibitory control (e.g., Kirkham, Cruess, & Diamond, 2003). The observed differences in N2 amplitudes should not be expected during the pre-switch phase, in the absence of both response conflict and inhibitory demands. Inhibitory accounts should presumably also predict improvements in performance after learning to attend to the relevant dimension through corrective feedback. In the absence of a hierarchical understanding of the task and feedback, however, children who received corrective feedback attended more readily to the irrelevant dimension (increased N1 amplitudes) and continued to sort according to the pre-switch rules, while believing that they were sorting correctly and discounting the conflict (reduced N2 amplitudes).

Although the experiments discussed herein make important contributions to the literature on the nature of EF by helping to elucidate the neural networks and cognitive processes that support its development, certain limitations must be considered. First, transfer effects were not
assessed in Exp. 4 precluding the possibility of exploring the neural systems that support transfer in early development. The adult literature points to certain neurocognitive constraints on transfer such that the transfer task must engage a similar cognitive process and brain region as the skill targeted (Dahlin et al., 2009). Future work should explore the neurocognitive processes that support and constrain transfer of EF training effects in the young, developing brain. Second, it is difficult to differentiate reflective from redescription processes based on the manipulations used in the present set of experiments. Future work should consider these processes more closely by comparing training targeting redescription (e.g., Mack, 2007) with training targeting reflective processes (e.g., Exps. 2, 3, and 4). A final consideration is that the source analyses used in Exps. 1 and 4 derive from head models based on adult anatomy due to the current lack of available head models based on relevant child data. Therefore, the interpretation of these findings needs to be treated cautiously, and replication is needed with head models optimized for pediatric use when they become available.

The wide-ranging potential for EF training to support children’s academic outcomes and long-term development is becoming increasingly apparent with studies demonstrating facilitative effects on children’s working memory, attention, and self-regulation (Holmes et al., 2009; Rueda et al., 2005; Rueda et al., 2012; Tamm et al., 2010; Wass et al., 2011). EF training paradigms have the potential to also elucidate the cognitive processes and neural mechanisms underlying EF across development. The findings of the studies presented herein indicate a key role for reflective processing and the top-down regulation of ACC-driven conflict detection in the emergence of flexible thought in early childhood.
References


