# A SYSTEMATIC ANALYSIS OF THE '*READING THE MIND IN THE EYES'* TASK IN CHILDREN AND WHAT IT MEANS FOR UNDERSTANDING SOCIAL PERSPECTIVE TAKING

by

TRACY GARDINER CASSELS

B.A., The University of California at Berkeley, 2001

M.A., The University of British Columbia, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2015

## Abstract

How well we understand social perspective taking is intricately linked to how well we assess this ability; however, there are factors that can influence its assessment, altering how we conceptualize social perspective taking and its development. The goal of the current dissertation was to systematically analyze one of the most popular measures of social perspective taking in children, the *Reading the Mind in the Eyes* task, by examining three specific measurement issues—what is being measured, response format, and coding scheme—to determine how these issues impact our understanding of social perspective taking more generally. **Methods:** Three studies were conducted with 249 children aged 4 to 13, including 54 children at-risk for affective perspective taking deficits. Two response formats (forced-choice vs. open-ended) and two coding schemes (term specific vs. valenced) were systematically compared on performance, relations to other abilities, and efficacy at measuring cognitive versus affective perspective taking. Comparison measures included dispositional empathy, cognitive perspective taking, and verbal ability. **Results:** There was a significant effect of response format, with the forced-choice format related to both cognitive and verbal abilities, suggesting that its performance is more apt to be influenced, unnecessarily, by the participants' vocabulary knowledge or use of alternate strategies. Furthermore, the forced-choice format was unrelated to dispositional empathy and cognitive perspective taking and failed to differentiate typically-developing from at-risk children. In contrast, the open-ended format was significantly related to dispositional empathy and differentiated at-risk from typically-developing children. Taken together these results a) raise concerns about the use of the *Reading the Mind in the Eyes* task as a measure of *cognitive* perspective taking and b) reveal that an open-ended format provides a better measure of *affective* perspective taking than the forced-choice format. The effect of coding scheme was

less clear, with evidence that term-specific coding was linked to vocabulary knowledge only in typically-developing children and only in a forced-choice response format. **Implications:** Findings are discussed in terms of implications on the type of information that can be gleaned from the *Reading the Mind in the Eyes* task and their relevance for studying social perspective taking more generally.

## Preface

A manuscript summarizing a subset of the data presented in Chapters 2 through 4 was published in the journal *PLoS One*. I am first author on the manuscript, with my supervisor, Dr. Susan Birch, as co-author. Reference: Cassels, T. G., & Birch, S. A. J. (2014). Comparisons of an open-ended vs. forced-choice 'mind reading' task: Implications for measuring perspective taking and emotion recognition. *PLoS One, 9,* e93653. DOI:10.1371/journal.pone.0093653.

A subset of the data from studies 2 and 3 were also included in a publication in the journal *Psychological Assessment.* The data were used to conduct specific statistical analyses on the measure of interest herein (the *Reading the Mind in the Eyes* task) using Item Response Theory and Confirmatory Factor Analysis. I am second author on the manuscript with Dr. Jasmine Carey as lead author. The reason for my second author status is that we agreed we each contributed equally so we chose to list authors alphabetically. The data were from my work, including the collection and design of the studies, Dr. Carey made a large contribution to the statistical analyses, and we wrote the manuscript together. Reference: Carey, J. & Cassels, T.G. (2013). Comparing two forms of a childhood perspective taking measure using CFA and IRT. *Psychological Assessment, 25,* 879-892.

Note that approximately half of the data collected in Study 2 (Chapter 3) were also part of my M.A. thesis. The analyses reported here are substantially different than those reported in the M.A. thesis, as the thesis focused on the ability to predict differences on various outcome measures in children higher and lower on psychopathic tendencies. The focus herein is on measurement issues and the comparisons of various modifications to the tasks to establish how certain factors may influence our understanding of social perspective taking. As such, the theoretical focus of this dissertation is substantially different from that of my M.A. thesis and

includes additional data, additional studies, and different analytical approaches. Reference: Cassels, T. G. (2007). Individual differences in psychopathic traits and identifying mental states and emotions in others. *Unpublished Master's thesis, University of British Columbia.* Similarly, the same subset of data from the participants in Study 2 (along with data from other subjects not included herein) examining the relationship with children's psychopathic traits were published in *SOJ Psychology*. Again I am first author on the manuscript with my supervisor, Dr. Susan Birch, as co-author. Reference: Cassels, T.G. & Birch, S.A.J. (2014). Outcomes associated with psychopathic traits in a non-clinical sample of children aged 6-13. *SOJ Psychology, 1,* 1-8.

For all experiments reported in this dissertation, I had the primary role in the design of the experiments and data analyses, with assistance and feedback from Dr. Susan Birch. I collected a large portion of research myself, and five former research assistants (Winnie Chung, Sherilynn Chan, Samantha Bangayan, Janiene Chand, and Sophia Ongley) collected the rest of the data. I drafted all manuscripts and subsequently incorporated feedback from Dr. Susan Birch, Dr. Rachel Severson, Dr. Ara Norenzyan, and my dissertation committee (Drs. Mark Schaller and Kalina Christoff).

Ethics approval for all three experiments included in the present dissertation was obtained from the Behavioural Research Ethics Board of the University of British Columbia (Chapter 2: H09-01933; Chapters 3 and 4: H06-80630).

## Table of Contents

## List of Tables

## List of Figures

## Acknowledgements

This research would not have been possible without the support from many individuals and institutions. I would like to thank first and foremost my supervisor, Dr. Susan Birch, for all her assistance along this long and winding path I have found myself traveling, and for her understanding and care when I faced hardships I didn't expect on this journey. I would also like to express my gratitude to my dissertation committee, Dr. Mark Schaller and Dr. Kalina Christoff, for their patience and support as well as to Dr. Ara Norenzayan for his feedback early on. I also received invaluable help from other members of the K.I.D. Studies Centre in the running of these studies and I am immensely grateful to all of them, for without them, this research would not have been possible. Special thanks to Dr. Rachel Severson for feedback on this manuscript. This research could not have been conducted without the most important people: The participants. As such, I am forever grateful and indebted to the hundreds of children and parents that agreed to take part in my research, the after-school programs that welcomed us into their space, and the elementary schools (including Eaton Arrowsmith school for children with learning disabilities) who welcomed us into their classrooms. I have received financial support from the Social Sciences and Humanities Research Council of Canada (SSHRC) and the University of British Columbia (both the Faculty of Graduate Studies and the Psychology Department). I would also like to thank countless friends and family for their encouragement and belief in me, especially when I lacked it myself. Last, but certainly not least, I would like to thank my husband, Brian, for his never-wavering support and understanding during these many, *many* years as I could not have done any of this without having him in my corner.

## Dedication

I would like to dedicate this dissertation to the three amazing women I lost along the way and to the little girl I gained who has made my life forever richer.

Mom: I never expected to be here without you and it still doesn't seem real most days. From my days starting junior kindergarten at Brown to my years at Berkeley and onward you were a pillar of support for everything I chose to do, even when it may not have seemed like the most intelligent choice (and I know that was frequent). The hole that exists in my heart from losing you is something I grapple with every single day, but I hope that you are somewhere better and filled with the love that we all felt and continue to feel for you. I love you and miss you and I will for the rest of my life.

Betty and Grandma: You are two of the most amazing women I had the joy and luck of knowing, despite being two of the most different women I ever met. I am so blessed to have called you grandmothers and I thank you for all that you have taught me over the years. I hope I was able to make you both understand how much I appreciated and loved you. You may be gone, but you will never be forgotten.

Maddy: My wonderful sweetheart. When I think of what life was like before you, it seems like a distant memory of someone else's life, a far more boring life at that. From the day you were born, you have made my life immeasurably richer in every way and have given me the push to be the best person I can be, just so I can be that person for you. You never cease to amaze me with all that you do, and I imagine you will continue to amaze me in the years to come. I love you with all of my heart. A heart you help make bigger every day so that those holes left behind don't feel quite so big. Thank you for "filling my heart with love and happiness". Love, Mom.

# 1 Introduction

## 1.1 Overview and Objectives

The ability to understand the mental states of other individuals, otherwise known as social perspective taking, is critical to human socialization. Perspective taking lies at the heart of social cognition, providing the foundation for empathic concern, reduced prejudice, and moral regard for others. From very early in development children are motivated to engage in perspective taking and readily attempt to infer the contents of others' minds (e.g., their knowledge, intentions, thoughts, emotions, and beliefs) to facilitate learning, communication, and decision-making (e.g., Harris, 2007). Vast individual differences in perspective taking are present early and continue into adulthood, with better perspective taking predicting a myriad of positive outcomes (e.g., fewer relationship problems, higher academic achievement, more prosocial behavior, and better quality of life indices to name just a few; see Chandler & Birch, 2010 for review).

Despite its early emergence and pervasiveness, perspective taking is remarkably error prone. Limitations and biases in social perspective taking are prevalent and give rise to a range of problems across various domains including (but not limited to) education, medicine, law, politics, business, and literature (see Birch & Bernstein, 2007 and Hawkins & Hastie, 1990 for reviews). More pronounced deficits in social perspective taking are linked to various psychological disorders such as depression, Autism Spectrum Disorders (Autism Spectrum Disorders), conduct disorder, psychopathy, and schizophrenia (for reviews, see Baron-Cohen, 2000; Dawel, O'Kearney, McKone, & Palermo, 2012; Kohler, Walker, Martin, Healey, & Moberg, 2009). Not surprisingly, individual differences in social perspective taking in typically-

developing populations are similarly related to social competency and prosocial behaviour (e.g., Blair & Coles, 2000; Lalonde & Chandler, 1995; Walker, 2005).

Clearly then, there are widespread applications for accurately measuring one's social perspective taking aptitude. As with all cognitive or social abilities, the understanding gained from research is limited by how well we are able to assess the construct(s) of interest. Unfortunately, quality measurement tools for studying individual differences in social perspective taking aptitude are surprisingly scarce—especially scarce are measures that can be used to examine individual differences in perspective taking across development, from the preschool period into adulthood. Instead, the field is largely comprised of myriad kinds of measures designed to either tap social perspective taking at a particular point in development (e.g., Wellman, Cross, & Watson, 2001) or to differentiate those with clinical levels of perspective taking problems from those without (e.g., those with Autism Spectrum Disorders; Baron-Cohen, Wheelwright, Scahill, Lawson, & Spong, 2001). Although these approaches have great utility in their own right (e.g., in clinical settings) there is considerable merit in developing measures that can capture individual variance in perspective taking, and its components, in a format that is appropriate for use in both children and adults.

The primary goals of this dissertation are to: a) highlight important methodological considerations for researching social perspective taking across development from preschool into adulthood, and b) systematically compare and contrast the efficacy of using different response formats (open-ended vs. forced choice) and coding schemes in one popular measure of social perspective taking. These goals serve the ultimate objectives of enhancing our knowledge of how best to measure social perspective taking across development and to improve upon the existing measurement tools that are so vital to the quality of both basic and applied research initiatives. I

will briefly address the three specific methodological considerations of interest herein that are critical to develop a full and coherent picture of the development of social perspective taking skills, and their associated outcomes.

The first measurement issue is the degree to which a measure assesses the two components of social perspective taking – cognitive and affective – as one construct instead of as separate constructs. Research that has compared performance on these social perspective taking components has found that the two components are quite distinct, with no significant overlap in performance (see Kurdek & Rodgon, 1975) and with distinct associated outcomes (a review is forthcoming in Sections 1.3.1 and 1.3.2). Yet measures, especially in childhood, often conflate the two, claiming to measure social perspective taking broadly speaking, the results of which are unknown.

The second issue is how the response format of the measure can allow participants to effectively bypass the need to employ social perspective taking abilities by using compensatory strategies such as 'process of elimination'. Though not all individuals will use compensatory strategies, this issue is of particular importance when measuring social perspective taking in populations that are apt to use such strategies to mask any deficits. For instance, individuals high on psychopathic traits have been found to have specific affective perspective taking deficits (e.g., Blair, Budhani, Colledge, & Scott, 2005; Blair & Coles, 2000; Blair, Colledge, Murray, & Mitchell, 2001; Kosson, Suchy, Mayer, & Libby, 2002; Marsh & Blair, 2008); however, when given a measure that allows for compensatory strategies to be used, these deficits disappear, though the areas of the brain activated differ from controls (Gordon, Baird, & End, 2004).

Third, and finally, the manner of coding chosen by researchers may influence our measurement, and thus understanding, of social perspective taking. Specifically, coding can be

done on a term-specific basis (i.e., the individual must provide a very exact answer to have it deemed correct) or a more general or valence-based basis (i.e., the individual can get the "gist" of the answer). The type of coding used in the research may either limit or enhance our knowledge of the abilities of the populations of interest. For example, Widen and Russell (2008) found children's emotion understand begins in a valence-based manner and becomes term-specific with age. The use of specific coding with young children may therefore mask the level of emotion understanding they have whereas valence-based coding is arguably more appropriate. In sum, the first of the three aforementioned measurement issues pertains to our understanding of what construct is being measured, the second and third issues address how certain types of methodological procedures (i.e. response format and coding schemes) can influence the assessment of social perspective taking.

In light of these measurement issues, I devised three studies to systematically compare a widely-used measure of social perspective taking to modified versions reflecting changes to response format and coding. This dissertation compares and contrasts the efficacy of the original measure with my modified measures with a particular focus on: 1) The ability of each measure to assess cognitive and affective perspective taking specifically, 2) The effect of measurement type (forced-choice versus open-ended) and the potential use of compensatory strategies on performance, and 3) The effect of coding scheme (specific versus general) on performance.

To this end, I have taken one of the most popular measures of perspective taking—the *Reading the Mind in the Eyes Task* (Baron-Cohen, Joliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Baron-Cohen, Wheelwright, Scahill, et al., 2001)—which is a forced-choice measure that utilizes term-specific coding and claims to assess perspective taking more generally (i.e., a combination of affective and cognitive

perspective taking)—and modified it to systematically examine the aforementioned methodological issues. The *Reading the Mind in the Eyes* task is one of the few social perspective taking measures that includes both a child and adult version, allowing comparisons across development. Prior to describing the studies in more detail, the remainder of this introduction will provide definitions of social perspective taking and the associated affective and cognitive components, along with an overview of the literature on the components of social perspective taking, their development, and their measurement before moving to a review of the literature on the *Reading the Mind in the Eyes Task* as well as a more detailed overview of the methodological issues, including their relevance to the *Reading the Mind in the Eyes Task* specifically. I close this chapter with the main questions of interest, a review of the studies conducted, and the specific hypotheses associated with the questions of interest.

## 1.2 What is Social Perspective Taking?

*Scenario 1: At a park, a child is playing in a sandbox, attempting to build a sand castle. Each time the child tries to pack the sand in his pail and turn it over, the sand falls out. A slightly older girl comes along and watches the boy as he attempts to build his castle. After a moment, she goes over, adds some water to the sand, and helps him pack it down and flip it over without the boy saying a word.*

*Scenario 2: Riding the bus home from work one day a man sees a woman get on and sit down across from him. She's not crying and she's not saying a word to anyone, but he looks to her face and leans over and asks, "Are you okay?" The woman shakes her head and starts to cry.*

These are both examples of the types of social perspective taking that individuals engage in that assist us in making sense of and being a part of the social world around us. Although

researchers and lay people often speak of social perspective taking as a unitary construct (e.g., Flavell, 2004), it is actually comprised of two different types of perspective taking abilities that are conceptually related, though practically distinct from one another: Cognitive and affective (or emotional) (Hinnant & O'Brien, 2007; Kurdek, 1977; Kurdek & Rodgon, 1975).

Social perspective taking as its own individual construct involves taking the mindset of another individual; however, two further divisions of cognitive and affective are needed to fully understand the construct. Cognitive perspective taking generally refers to one's ability to infer cognitive mental states, including intentions, knowledge, thoughts, beliefs, and desires (for a review, Astington, Harris, & Olson, 1988). In the first scenario above, the older girl is able to infer the little boy's intent even though he has failed to produce anything; she sees what he is *trying* to accomplish by watching his behavior and because she is able to take other people's perspective and knows that humans have goals (e.g., Meltzoff, 1995; Tomasello, Carpenter, Call, Behne, & Moll, 2005). Combining her own general knowledge of human behavior with what she sees the boy attempt, she can act to help him without even having to utter a word. Cognitive perspective taking is itself a multifaceted concept, including a variety of abilities that emerge at different points in the developmental trajectory. An abbreviated overview of the literature on the developmental trajectory of, and outcomes associated with, cognitive perspective taking is provided in Section 1.3.1.

The second component of social perspective taking is affective perspective taking, which refers to the ability to identify the emotional mental states of another individual. For instance, in the second scenario above, the man is using his affective perspective taking abilities when he assesses how the woman on the bus is feeling without her having to say anything or express blatant cues like crying. Non-verbal cues to emotion, such as facial expressions, are important

means for humans to decipher underlying affective mental states and effectively tailor our behavior toward others. As with cognitive perspective taking, affective perspective taking is also complex and the various facets of affective perspective taking develop at different times. An abridged overview of the literature on the developmental trajectory of, and outcomes associated with, affective perspective taking is provided in Section 1.3.2.

The remainder of this literature review will focus on why we ought to care about perspective taking (Section 1.3) as well as the individual trajectories and outcomes for cognitive (Section 1.3.1) and affective (Section 1.3.2) perspective taking. This will be followed by a discussion of the measurement of perspective taking and the specific measure chosen for comparison herein, the *Reading the Mind in the Eyes Task* (Baron-Cohen, Wheelwright, Hill, et al., 2001; Baron-Cohen, Wheelwright, Scahill, et al., 2001) (Section 1.4). The three measurement issues raised above—1) the degree to which a measure assesses cognitive or affective perspective taking (Section 1.5.1); 2) the effect of response format (Section 1.5.2); and 3) the effect of coding (Section 1.5.3)—will then be reviewed before returning to a more detailed description of the current research goals and specific hypotheses (Section 1.6).

## 1.3 Why is Social Perspective Taking Important?

*"All else being equal, they [good social perspective takers] are likely to be the most tactful advisors, the most diplomatic officials, the most effective negotiators, the most electable politicians, the most productive salespersons, the most successful teachers, and the most insightful therapists." ~ William Ickes (1997, p.2)*

Recognizing, reasoning about and responding to the mental states of others are fundamental aspects of social interaction and development. The capacity to understand the minds of others allows humans to engage in the social world around them and is apparent from very

early in development. The interest in how children come to understand others and the effects of social perspective taking on other realms of life is not limited to psychologists. The study of Economics has a large focus on how individuals develop strategies for dealing with other people and situations—either financially or practically—which critically involves understanding and utilizing social perspective taking (e.g., Sally & Hill, 2006). Cooperation and fairness rely, at least partially, upon social perspective taking. Marketing and consumerism require one to understand how other people will perceive what is being advertised or presented in order to be successful. Education requires social perspective taking on the part of both the student and the teacher, for without it, teachers cannot effectively pass on the knowledge they aim to share and students can fail to understand the knowledge that is being shared. Notably, much of what children learn about the world prior to formal education comes from taking the perspective of another individual (see Section 1.3.1 for more information), and also by being able to assume and build upon shared knowledge during these interactions (Keysar, Barr, Balin, & Brauner, 2000). There are thousands of papers written on the topic and dozens of books, including Paul Bloom's (2002) *How Children Learn the Meaning of Words*, William Ickes' (1997) *Empathic Accuracy*, Ian Apperly's (2012) *Mindreaders: The Cognitive Basis of "Theory of Mind"*, and Simon Baron-Cohen's (1997) *Mindblindness*.

Interest in social perspective taking is not limited to academics. Parents strive to understand what their children are thinking and there is a plethora of popular books on the topic, including the more academic, such as Alison Gopnik's (1999) *The Scientist in the Crib* (and 2009's *The Philosophical Baby*), and the more humourous whereby parents attempt to take the mind of their child, such as Bunmi Laditan's (2013) *The Honest Toddler*, a book written through the eyes of a very precocious toddler. Educators have special curricula to help teach students

with social perspective taking deficits the skills needed to navigate a social world (Ordetx, 2012). Clinical psychologists have perspective taking-based therapies to help individuals learn to differentiate their own mental processes from others in order to build successful relationships (Bateman & Fonagy, 2012). Finally, there are the hundreds of books written on specific forms of social perspective taking, such as empathy and the effects on one's social life, such as Bruce Perry's (2011) *Born for Love*, or the related concept of emotional intelligence (EQ), which has now been argued to be even more important to life outcomes than cognitive intelligence (IQ) (Goleman, 2005). The developmental trajectories of cognitive and affective perspective taking highlight their distinctiveness, but also how the two can work together (for a review, see Saxe, Carey, & Kanwisher, 2004).

### 1.3.1 Cognitive Perspective Taking

There are multiple trajectories of cognitive perspective taking representing the various types of mental states that children must infer from others. Research into cognitive perspective taking started with the initial examination of false belief understanding, or the concept of mental representation and *mis*representation, in preschoolers and has since worked both forwards and backwards in development to understand the full trajectory (for an overview, see Flavell, 2004). Despite the desire to delineate the full trajectory for the emergence of various components of cognitive perspective taking, no one clear trajectory has been found; this is at least in part because different components overlap in their development and because there is individual variability in which components come online first and at which age. As such, it is perhaps more informative to examine the subcomponents of cognitive perspective taking individually.

#### 1.3.1.1 Goal and Intention Understanding

In the first year and a half of life, children develop the ability to understand that other individuals have specific goals and intentions that can be reflected in their actions. For example, infants as young as six months will habituate differently to actions that seem intentional versus random (Woodward, 1998). In line with this development of intention understanding in the second year of life, numerous studies have shown that toddlers will not imitate a failed or accidental action (Carpenter, Akhtar, & Tomasello, 1998; Meltzoff, 1995), will only imitate direct action when the goal for the action is clear (Carpenter, Call, & Tomasello, 2005), and children will selectively learn based on an understanding of the other's goals or intentions (Carpenter, Call, & Tomasello, 2002). This ability to understand and to be critical of other's actions and intentions is thought to be necessary to collaborate together via shared intentions (Tomasello et al., 2005).

#### 1.3.1.2 Desire and Preference Understanding

A second type of understanding that develops in the first two years of life is the understanding that others may have desires and preferences different from our own. Understanding that others have different desires than us is critical to social interaction as it allows us to cater our behaviours in response to another person. This ability is most famously demonstrated in studies looking at toddlers' understanding of food preferences. In these studies, 18-month-olds are able to realize that another person may prefer different foods to them and offer them the food *the other person* wants; prior to 18 months, they offer the food *they* prefer, even after seeing the other individual display disgust for that food (Repacholi & Gopnik, 1997). Although this may seem trivial to an adult, the act shows the child is able to not only recognize

the cues of the other individual as to what a preference is, but also that the child can override his or her preference to provide the person with something more in line with that person's desires.

This ability to understand others' desires and preferences continues to develop and by 2 ½ years of age children understand that not only do people have different desires and preferences, but that they act upon these desires and preferences (Wellman & Woolley, 1990). Perhaps more impressively, these children are able to make emotion attributions that are dependent upon the failure or success of the act aimed to fulfill a desire. For example, they know that a child who wanted a firetruck for Christmas and received one would be happy; however, a child who wanted a dog for Christmas but received a firetruck would be sad. Furthermore, children understand the distinction between desiring something and acting to achieve it (Astington & Gopnik, 1991; Hadwin & Perner, 1991; Yuill, 1984). By preschool age, children are quite adept at understanding and talking about desires and preferences as well as utilizing them to make predictions about people's actions or to explain already-occurred actions.

#### 1.3.1.3 Knowledge Attribution

Understanding that others have *knowledge* they can share starts early in the first year of life. This is most likely due to the importance knowledge transmission has in how children are socialized and learn about the world around them. Early forms of this understanding occur in the form of gaze following and referential pointing, which develop in the first year of life (Baldwin, 1993; D'Entremont, 2000; D'Entremont, Hains, & Muir, 1997; Johnson, Slaughter, & Carey, 1998; Povinelli, Reaux, Bierschwale, Allain, & Simon, 1996). Gaze following and referential pointing both involve cognitive perspective taking as children must be aware that others have information to share with them *and* that they are taking measures to share this information (Liebal, Behne, Carpenter, & Tomasello, 2009; Tomasello, Carpenter, & Liszkowski, 2007).

Understanding and using referential pointing have been linked to later language production and comprehension (Baldwin, 1993; Butterworth & Morrissette; 1996), suggesting that these early forms of communication are grounds for later language development, one of our more prominent forms of communication as humans. Supporting this causal view, the relationship holds longitudinally with individual differences in infant referential pointing predicting later language abilities in toddlerhood (Butterworth & Morrisette, 1996). Tomasello and colleagues (2007) also argue that infant pointing is an attempt by the infant to influence the mental state of the other individual (often a parent) which inherently involves the understanding of another's mind (at least at the basic level).

This understanding of the knowledge states of others continues into toddlerhood and the preschool years where children learn to distinguish not only between those who are knowledgeable and those who are not (Brosseau-Liard & Birch, 2010; Gillis & Nilsen, 2013; Jaswal & Neely, 2006; Koenig & Harris, 2005; Taylor, Cartwright, & Bowden, 1991), but also who is knowledgeable about different *types* of information (Lucas & Lewis, 2011; Lutz & Keil, 2002; Sobel & Corriveau, 2010). Over the preschool years children also learn to differentially value various cues to knowledge, such as a person's prior accuracy or level of confidence (Brosseau-Liard, Cassels, & Birch, 2014). Children also demonstrate their awareness of others' knowledge states by altering their own communication to said partners (Akhtar, Carpenter, & Tomasello, 1996; Bahtiyar, & Küntay, 2009; Nadig & Sedivy, 2002; Nilsen & Graham, 2009; Tomasello & Akhtar, 1995). Young children are therefore making important inferences about the contents of other people's minds in order to learn about the world around them. Indeed, the ability to infer more nuanced aspects of knowledge attribution follows the development of children's ability to understand ambiguity (Nilsen & Graham, 2012), a skill that is highly

relevant to navigating social situations, which are often rife with ambiguity. Thus the continued development of this skill is not surprising and serves to highlight its importance to human socialization.

#### 1.3.1.4 Beliefs and False Beliefs

Perhaps the most well-known component of cognitive perspective taking, inferring and understanding others' beliefs, has received the lion's share of both academic focus and discussions of cognitive perspective taking more generally. Understanding false beliefs begins to develop in late infancy or early toddlerhood, but was long-believed to be the sole domain of preschoolers, though recent research has shown otherwise. Between 13 and 18 months of age, when given a task in which the child must predict the behavior of an individual who holds a false belief (e.g., Sally puts her ball away in Box A and leaves the room then Ann comes in and moves the ball from Box A to Box B and leaves; Sally returns and the child has to predict where Sally will look for her ball), the child is unable to make an accurate prediction, but consistently gazes to the correct location of where such an individual would look (e.g., if asked about Sally, the toddler would gaze at Box A, but point to Box B) (Scott & Baillargeon, 2009; Song & Baillargeon, 2008; Song, Onishi, Baillargeon, & Fisher, 2008; Southgate, Senju, & Csibra, 2007; Surian, Caldi, & Sperber, 2007; Yott & Poulin-Dubois, 2012). This provides evidence that these younger children are able at least partially reason that others have false beliefs. However, children seem unable to explicitly indicate that others will act upon a false belief until later in development (between 3 and 5 years of age), even when the task has been simplified to reduce cognitive load (for a review, see Wellman, Cross, & Watson, 2001).

As mentioned, the ability to predict actions based on false belief explicitly (e.g., verbally) does not emerge until the preschool years (Wellman et al., 2001) and evidence is mixed about

how younger preschoolers use false belief information to explain behavior. In one study, three-year-olds were able to use false beliefs to reason about other people's behaviours in retrospect (though still not able to use that information to make *predictions* about behaviours) (Dunn, Brown, Slomkowski, Tesla, & Youngblade, 1991) whereas in another, three-year-olds failed to explain their own actions in terms of false-belief understanding (Atance & O'Neill, 2004). The aforementioned findings using eye gaze becomes even more complicated when we consider research by Ruffman and colleagues that suggests children are *unaware* of their eye gaze towards the correct location (Ruffman, Garnham, Import, & Connolly, 2001). In this study children were asked to place bets on where they thought a character would look (based on a false belief) and despite their eye gaze looking towards the "correct" location (i.e., the one consistent with a false belief), children bet highly on the location they provided explicitly, suggestive of high confidence in their explicit answer.

One possible reconciliation of the mixed findings by age is rooted in a perspective taking bias observed even in adults: Adults show a difficulty inhibiting knowledge they have (that others do not) when predicting other people's knowledge and beliefs (Birch & Bernstein, 2007; Birch & Bloom, 2004). Specifically, adults will internally correct for the fact that they have privileged information, but their correction is rarely as sufficient as it should be; they cannot completely ignore what they know when making predictions about the knowledge of other individuals. Birch and colleagues have argued that this curse of knowledge exists across development, albeit to a lesser degree later in development (Birch, 2005; Birch & Bloom, 2003, 2004). The <u>curse of knowledge</u> refers to an innate bias that all individuals have to privilege their own information over another's. This negatively influences their ability to take a naïve perspective and, notably, it is linked to inhibitory control, among other factors. The ability to

inhibit one's own perspective is argued to be necessary to take another's, and importantly, inhibitory control continues to improve with age and the development of the prefrontal cortex. Indeed, the development of inhibitory control has been linked to preschoolers' ability to communicate to another based on the others' perspective (Nilsen & Graham, 2009). Interestingly, the decline in cognitive perspective taking that occurs in older-age adults (Bernstein, Thorton, & Sommerville, 2011) may also reflect a greater failure to inhibit our own knowledge and age-related degradation of the prefrontal cortex

#### 1.3.1.5 Outcomes Associated with Cognitive Perspective Taking

The failure to develop cognitive perspective taking abilities, or what is sometimes termed a 'theory of mind', results most generally in deficits in social competence. Much of the research on cognitive perspective taking deficits focuses on the development of Autism Spectrum Disorders, with deficits in cognitive perspective taking strongly associated with these disorders (Baron-Cohen et al., 1997; Baron-Cohen, Wheelwright, Scahill, et al., 2001; Craig, Hatton, Craig, & Bentall, 2004; Happé, 1993; Frith & Happe, 1994; for a review, see Baron-Cohen, 2000). However, there are also outcomes associated with individual differences in the varying components of cognitive perspective taking even in typically developing populations. For example, greater cognitive perspective taking is associated with socio-emotional maturity in 3-year-olds (Lalonde & Chandler, 1995) and prosocial behaviour in 3- to 5-year-old girls (Walker, 2005). In older children, cognitive perspective taking has been associated with peer-rated social competence (Bosacki & Astington, 1999) and parent-rated social competence (Charman, Carroll, & Sturge, 2001; although for results suggesting no relationship between social competence and theory of mind, see Repacholi, Slaughter, Pritchard, & Gibbs, 2003). Interestingly, cognitive perspective taking has predicted *greater* bullying behaviour in young boys (Walker, 2005), in

line with ideas that some forms of bullying require advanced understanding of others' minds (for a discussion, see Sutton, Smith, & Swettenham, 1999). Finally, there is a link between the ability to understand other people's intentions and moral reasoning and moral behavior in children (Baird & Astington, 2004), suggesting that cognitive perspective taking plays a role in our moral development as well. Generally speaking, cognitive perspective taking is associated with positive social behavior, but can also be exploited and may explain the link to negative behaviours that require a degree of understanding or manipulation of other people's minds (e.g., bullying, conning; for a review, see Repacholi & Slaughter, 2003).

### 1.3.2 Affective Perspective Taking

As with cognitive perspective taking, affective perspective taking is not a unitary construct with one clear trajectory for all people. It is influenced by experience and there are individual differences in its development and trajectories. However, taken as a whole, it is thought to come online in a fuller form earlier than cognitive perspective taking, though there are elements that certainly come online later (for a review, see Saxe et al., 2004). The focus on the affective component of social perspective taking means the primary focus for outcomes also lies in the emotional realm, such as the development of empathy (Hoffman, 2000a,b) or helping behaviours (Knafo, Steinberg, & Goldner, 2011).

#### 1.3.2.1 Emotion Contagion

The earliest expression of the influence of others' emotional mental states is <u>emotion contagion,</u> in which an infant will cry at the sound of other infants' cries (Hatfield & Cacioppo, 1994; Hoffman, 2000b; Zahn-Waxler & Radke-Yarrow, 1990) or more generally mimic the emotional expressions of others (for a review, see Dimberg & Oehman, 1996). This effect is

strongest when the cries come from other humans, suggesting a role of either implicit understanding of, or connection to, other human beings (Sagi & Hoffman, 1976). Sagi and Hoffman (1976) tested this by playing a tape recorded sound of either an infant crying or a synthetic newborn cry of the same intensity as a newborn baby's cry. Infants cried in response to both sounds, but infants exposed to the actual human baby's cry cried significantly more than those exposed to the synthetic cry. Although infants at this stage are believed to not be separating their own minds or feelings from those of others (a necessary component for affective perspective taking), this reaction to another human is a precursor to later affective perspective taking abilities and empathy (Hoffman, 2000b).

#### 1.3.2.2 Basic Emotion Recognition

Infants' ability to discriminate between basic emotions (i.e., happiness, sadness, anger, fear, disgust, and surprise; Ekman, Friesen, & Ellsworth, 1982) displayed in others occurs around 4- to 7-months of age (Caron, Caron, & Myers, 1982; Kestenbaum & Nelson, 1990; Ludemann & Nelson, 1988; Nelson, 2001). This early ability is reflected in the neurological development of facial processing (de Haan, Pascalis, & Johnson, 2002). Specifically, infants at 6 months of age show the same event-related potential component as adults when identifying the specificity of human faces, suggesting some elements of our facial processing abilities are online at or shortly after birth. This ability to discriminate between the various emotions is critical for children's social learning as children use social referencing, the use of others as a guide for how to react to unfamiliar stimuli or situations, to navigate an unknown world (e.g., Feinman, 1982; Feinman & Lewis, 1983). Social referencing, which starts to appear around 6 to 9 months but continues to develop into toddlerhood (Walden & Ogan, 1988), builds upon the ability to identify basic emotional states in others. Infants (and children) look to others for emotional responses

that suggest how they should react. This ability involves the implicit understanding that not only will others have emotional reactions to situations, but that there is valuable information or knowledge to be gleaned from their reaction (for a review, see Baldwin & Moses, 1994). Social referencing also provides a ripe example of how affective and cognitive perspective taking combine to allow infants to learn about and make sense of the world around them.

Although the ability to identify basic emotions is functional in infancy, it is thought not to be nearly fully developed until ages 3 to 4 (Pons, Harris, & de Rosnay, 2004); that is, by age 3-4, it is believed that children are performing nearly as well as adults on basic emotion identification. Research is somewhat mixed in support for this trajectory, however, with some research finding only small improvements in identifying most of the basic emotions from age 5 or 7 into adulthood (although all children performed well above chance in discriminating or identifying emotions) (Durand, Gallay, Seigneuric, Robichon, & Beaudoin, 2007) whereas others have found small improvements between 6- and 8-years of age and then a plateau until adolescence when another improvement is seen in more difficult tasks (Kolb, Wilson, & Taylor, 1992). However, the fact that the improvements come in more difficult tasks may be reflective of cognitive demands instead of emotion recognition.

#### 1.3.2.3 Complex Emotion Recognition

Complex emotion recognition can refer to both the identification of more nuanced emotions (e.g., embarrassment, jealousy, shame, guilt, and pride; although for evidence of pride as a basic emotion, see Tracy & Robins, 2004, 2008) as well as the identification of mixed emotions (such as feeling happy and sad simultaneously). The identification of nuanced emotions has been found to develop later than that for basic emotions, with preschool children showing some proficiency (e.g., Bosacki & Moore, 2004), but their understanding continues to

develop into early childhood (e.g., around 7 years of age; Russell & Paris, 1994). Nuanced emotions may take longer to develop as they may require different perceptual abilities than basic emotions. Basic emotions may be readily identified using features of the face (such as a smile, a frown, etc.) without taking the entire face into consideration (for a discussion, see Carey & Diamond, 1977). Configural face processing, on the other hand, is likely required to identify more nuanced emotions and involves looking at the features, the relationship between the features, and consideration of the face as a whole (for a review, see Maurer, Le Grand, & Mondloch, 2002). Furthermore, recognizing complex emotions often also involves not just facial processing, but body processing too (e.g., pride involves a head up, shoulders back, and standing tall bodily pose; Tracy & Robbins, 2004). Although children have been found to be able to use configural face processing at age 5 (Durand et al., 2007), much of the research suggests they struggle to do this in certain tasks, like face matching (Bruce, Campbell, Doherty-Sneddon, Import, Langton, et al., 2000; Carey & Diamond, 1994; Kolb et al., 1992; Mondloch, Le Grand, & Maurer, 2002); therefore, this skill may be present but difficult for children and thus underutilized or only used in specific situations.

A second category of complex emotions includes mixed emotions. Although young children can express mixed emotions, an understanding of mixed emotions does not occur until later in development. The origins of this ability seem to begin in preschool when children as young as 4 years of age are able to select mixed-emotion options to describe faces that reflect these mixed emotions (e.g., happy and sad) (Kestenbaum & Gelman, 1995); however, the ability to fully discuss and predict mixed emotions seems to be more fully developed later on, around 7-8 years of age (Pons et al., 2004). It is unclear exactly how much younger children understand with respect to mixed emotions and clearly a certain degree of cognitive perspective taking is

needed as mixed emotions tend to be situational, but the antecedent is present in preschool even if the full understanding does not emerge until later in childhood.

#### 1.3.2.4 Emotional Context Understanding

The final component of affective perspective taking to be reviewed is the use of context to understand emotional displays or experiences. It is also one of the last elements of affective perspective taking to develop. This component of affective perspective taking is influenced greatly by the development of cognitive perspective taking abilities and reflects the overlap between the two components of social perspective taking. For example, as toddlers start to understand other people's desires and preferences, they also start to understand how these can influence what another person feels (Pons et al., 2004); that is, they understand that someone who doesn't get something they want will feel sad even if they receive something the child himself wants.

The more complete understanding of desires and beliefs that comes about in late preschool or early childhood also allows for children to start understanding the influence of false beliefs on emotions (de Rosney, Pons, Harris, & Morrell, 2004; Pons et al., 2004). At about the same age, children are able to use context to identify mixed emotions when the basis for acknowledging a mixed emotion is in the context of a story (Kestenbaum & Gelman, 1995). This contextual understanding of emotions continues to develop into adulthood and is subject to many of the same cognitive limitations as cognitive perspective taking.

#### 1.3.2.5 Outcomes Associated with Affective Perspective Taking

Deficits or problems in affective perspective taking have been linked with aggressive behaviour (Dodge, 1980; Hughes, Dunn, & White, 1998; Sharp, 2008), social competence (for a

review, see Trentacosta & Fine, 2010), prosocial behaviours (Knafo et al., 2011), and academic competence (Izard, Fine, Schultz, Mostow, Ackerman, et al., 2001). Problems identifying certain emotions have been linked to psychopathic traits in adults (for a review, see Dawel et al., 2012) and adolescents (Blair & Coles, 2000; Blair et al., 2001; Dadds, Perry, Hawes, Merz, Riddell, et al., 2006). Additionally, many mental disorders include deficits or biases in affective perspective taking such as learning disabilities (for a review, see Tur-Kaspa, 2002), depression (Gur, Erwin, Gur, Zwil, Heimberg, & Kraemer, 1992; Schepman, Taylor, Collishaw, & Fombonne, 2011; van Wingen, van Eijndhoven, Tendolkar, Buitelaar, Verkes, & Fernández, 2011), Turner Syndrome (Lawrence, Kuntsi, Coleman, Campbell, & Skuse, 2003), and schizophrenia (for a review, see Kohler et al., 2000), amongst others. On the positive side, better emotion understanding has been linked to secure attachment in children aged 3 to 6 years of age (de Rosnay & Harris, 2002) and to the ability to cooperate in middle childhood (Johnson, 1975).

In addition to the aforementioned relationships, affective perspective taking is linked to empathy more generally, with the ability to identify others' emotions predicting a significant amount of variance in the display of empathy (Hooker, Verosky, Germine, Knight, & D'Esposito, 2008). Affective perspective taking is also thought to be the mechanism behind the finding that young children will offer more assistance and comfort to an individual who has been the victim of a transgression (versus a neutral event) even if the victim does not show any emotional signal of distress (Vaish, Carpenter, & Tomasello, 2009). These interrelationships (between empathy, social competence, aggression, and affective perspective taking) may explain some of the social competence and aggressive behaviours found in certain populations, such as those with learning disabilities (for a review, see Tur-Kaspa, 2002) and those high on psychopathic traits (Blair, 2003; DeMatteo, Heilbrun, & Marczyk, 2005; Loney, Frick, Clements,

Ellis, & Kerlin, 2003; Mayberry & Espelage, 2007). In sum, the ability to perceive, infer, and reason about the emotions of others is critical to socialization and the failure to do so accurately is associated with several problematic outcomes.

### 1.4 The *Reading the Mind in the Eyes* Task

One commonly-used measure of perspective taking was developed by Simon Baron-Cohen and his colleagues and is known as the *Reading the Mind in the Eyes* task (Baron-Cohen, Joliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Wheelwright, Hill, et al., 2001; Baron-Cohen, Wheelwright, Spong, et al., 2001). It is the measure of interest herein due in large part to its popularity and widespread use as a measure of perspective taking, but also because it allows for the examination of the methodological considerations of interest herein. In the *Reading the Mind in the Eyes* task, participants are presented with several photographs of the eye-region and asked to choose which of four words best describes the person's mental states. This was conceived of as a measure of "how well the participant can put himself into the mind of the other person, and 'tune in' to their mental state" based on minimal cues (Baron-Cohen et al., 1997, p. 241). The *Reading the Mind in the Eyes* task has proved effective in discriminating adults with Asperger's syndrome or high-functioning autism (of normal intelligence) from controls (Baron-Cohen et al., 1997; Baron-Cohen, Wheelwright, Hill, et al., 2001). In addition to work with those with Autism Spectrum Disorders, the *Reading the Mind in the Eyes* task has been used to examine perspective taking deficits in other clinical populations, such as those with schizophrenia (Craig, Hatton, Craig, & Bentall, 2004), patients who have had amygdalectomy or prefrontal cortical lesions (Stone, Baron-Cohen, Young, & Calder, 1998), and depression (Lee, Harkness, Sabbagh, & Jacobson, 2005). A children's version of the *Reading the Mind in the Eyes* task was developed by simplifying the response options and reducing the number of items

and has been shown to discriminate between children with Autism Spectrum Disorders and those without (Baron-Cohen, Wheelwright, Spong, et al., 2001).

The *Reading the Mind in the Eyes* task has been conceived of as a generalized measure of social perspective taking (Baron-Cohen, Wheelwright, Spong, et al., 2001). Confusingly, although some researchers have used the *Reading the Mind in the Eyes* task as a measure of perspective taking broadly speaking (Craig et al., 2004; Lee et al., 2005), some have utilized it as a cognitive perspective taking task (Gregory, Lough, Stone, Erzinclioglu, Martin, et al., 2002; Sharp, 2008), and others still have utilized it solely as an affective perspective taking task (Richell, Mitchell, Newman, Leonard, Baron-Cohen, et al., 2003; Tonks, Williams, Frampton, Yates, & Slater, 2007a,b)—sometimes with questionable success. This dual-usage is reflective of the first measurement issue in assessing perspective taking: The degree to which a measure assesses one or both components of social perspective taking.

Measures can and do assess social perspective taking broadly speaking with elements of both cognitive and affective perspective taking included. The problem is that research on the overlap or correlation between performance on the two components has found them to be distinct abilities with small (accounting for < 10% of variance; see Hinnant & O'Brien, 2007) or no (Kurdek, 1977; Kurdek & Rodgon, 1975) correlations between them and thus any measure needs to provide individual scores for each. In the case of the *Reading the Mind in the Eyes* task, the fact that it has been used for different purposes suggests that researchers are not clear on what component(s) exactly is (or are) being measured. The face validity of the task would suggest it is best conceived of as an affective perspective taking task given the use of the eye region that is expressing various emotions; however, the request that individuals identify what the individual is "thinking or feeling" encompasses the cognitive component of social perspective taking. If

elements of both affective and cognitive perspective taking are used to provide a single score, as is done in the *Reading the Mind in the Eyes* task, important distinctions or even nuances between the two components will be ignored, providing a less-than-complete picture of each individual component and a muddied picture of social perspective taking more generally. The *Reading the Mind in the Eyes* task, by including questions for both cognitive and affective perspective taking provides a global score that may be influenced differentially for different people. For instance, if individual A does well on the cognitive perspective taking questions, but not the affective ones, and individual B does well on the affective perspective taking questions, but not the cognitive ones, a general score will not reflect this difference as they may end up with very similar scores. Furthermore, when compared to specific outcomes, the results will be muddied by the fact that relationships vary for the different components of social perspective taking, as previously discussed. For example, although there is a wealth of research finding an affective perspective taking deficit in individuals higher on psychopathic tendencies (e.g., Blair et al., 2005; Blair & Coles, 2000), when the *Reading the Mind in the Eyes* task was used to assess the same relationship, no deficit was found (Richell et al., 2003; Sandvik, Hansen, Johnsen, & Laberg, 2014).

The <u>second methodological issue</u> is the effect of response format on performance and the degree to which compensatory strategies may be used to solve the task at hand. The *Reading the Mind in the Eyes* task is subject to the use of compensatory strategies because of the potential to use non-perspective taking strategies with a force-choice response format. For example, Peterson and Miller (2012) found that the *Reading the Mind in the Eyes* task was *unrelated* to face processing abilities in a sample of adults. One would assume, given the stimuli used in the *Reading the Mind in the Eyes* task (i.e., eye regions taken from images of faces), that face

processing would be an influential component in the degree of success on it *if* one solved the task as expected, but if alternate means are used (e.g., process of elimination), face processing may not be necessary. Similarly, as noted above, the *Reading the Mind in the Eyes* task has been used to examine affective perspective taking in relation to psychopathic tendencies in adults, revealing no deficits (Glass & Newman, 2006; Richell et al., 2003; Sandvik et al., 2014), contrary to findings with adults (e.g., Dolan & Fullam, 2006) and children (e.g., Blair et al., 2001) with related tendencies that did not rely on the *Reading the Mind in the Eyes* task as the measure of affective perspective taking (for a review, see Blair, 2003). Interestingly, research has found that adults higher on psychopathic tendencies recruit different areas of the brain when doing an affective perspective taking task compared to those with lower levels of these tendencies (Gordon et al., 2004), supporting the possibility that these individuals employ different neural circuitry because of the use of compensatory strategies or abilities.

The <u>third and final methodological issue</u> of interest herein is how participants' answers are coded and whether different coding schemes may tap differences in performance that are influenced by other abilities, such as verbal or cognitive abilities. Coding can be done on a specific basis or a more general basis. That is, when coding for content, one can consider only specific terms as correct or consider if someone is able to get the 'gist' of the answer; this choice will have implications for performance. For example, the *Reading the Mind in the Eyes* task, which uses term-specific coding, may require specific vocabulary (e.g., knowing the meaning of the emotional label 'Cross') to get the correct answer and some individuals, especially children, may not have the appropriate vocabulary knowledge. Evidence for the effect of coding scheme on the *Reading the Mind in the Eyes* task specifically comes from Tonks and colleagues (2007a) who looked at affective perspective taking in 9- to 15-year olds using both the *Florida Affect*

*Battery* and the *Reading the Mind in the Eyes* task. The *Florida Affect Battery* had different task types, including an affect naming task comparable to the *Reading the Mind in the Eyes* task and they were both forced-choice tasks. However, the *Florida Affect Battery* has the same terms to select from for each question (happy, sad, angry, frightened, neutral), making the coding more valence-based which should not be dependent on verbal ability, whereas the *Reading the Mind in the Eyes* task utilizes specific terms which differ for each question. Performance on the *Florida Affect Battery* was generally better with scores in the 70-85% range whereas scores on the *Reading the Mind in the Eyes* task were in the 60-75% range. There was also far more variability in the *Reading the Mind in the Eyes* task with variances that were often double what they were for the same age group for the *Florida Affect Battery*.

One possibility is that the more strict and specific coding of the *Reading the Mind in the Eyes* task required greater verbal ability, thus resulting in lower scores. Although this would seem to be most prominent in research with children, research with adults has found that the *Reading the Mind in the Eyes* task is strongly linked to verbal ability. For example, Peterson and Miller (2012) found that 28% of the variance in performance on the *Reading the Mind in the Eyes* task in a sample of adults was due to variance in verbal ability, suggesting a far larger role of verbal ability than expected in adulthood. This type of task demand imposed by the type of coding or response categories researchers employ could negatively influence performance for individuals whose verbal abilities are on the lower end whereas it could positively influence performance for those on the higher end of the spectrum, especially when combined with the ability to use compensatory strategies. Importantly, such effects on performance are likely unnecessarily skewing (i.e., masking or enhancing) participants' *actual* affective perspective taking abilities.

The presence of research that highlights these specific issues with respect to the *Reading the Mind in the Eyes* task makes it an ideal measure for an examination of these three specific measurement considerations or limitations, despite the fact that these limitations (which will be discussed in more detail below) are relevant for multiple measures, not just the *Reading the Mind in the Eyes* task. The overarching goal of the current dissertation is to systematically examine these issues in the *Reading the Mind in the Eyes* task as this has not been done on such a widely-used measure, and also as an example of how these limitations may manifest in social perspective taking research more generally and what impact they may have on the results. This research will allow researchers to better understand how their methodological choices influence their findings (past, present, and future) and aims to provide researchers with better tools to study perspective taking.

## 1.5 Three Methodological Issues

### 1.5.1 Cognitive Vs. Affective Perspective Taking: What Kind of Perspective Taking is Being Measured?

The first issue of interest herein is the question of what component of social perspective taking is being assessed when a measure purports to assess 'social perspective taking' generally speaking. Both cognitive and affective perspective taking abilities are critical to an individual's social well-being and both involve an individual attempting to discern something—be it emotions or cognitions—that are part of *another* person's mind, which is difficult for many people and may be seen as contrary to the human mind's 'default' stance of considering oneself first (for a discussion, see Decety & Sommerville, 2003). Indeed, young humans start off quite egotistical in their thoughts and actions towards others, though most develop perspective taking

quite naturally enough, suggesting both components of social perspective taking may be quite natural. Regardless of how difficult or natural the two aspects of social perspective taking may be, their developmental trajectories and outcomes differ and the assumed strong relationship between these components of social perspective taking seems to be unwarranted. Older studies have found that cognitive and affective perspective taking abilities in childhood were discrete entities with the intercorrelation between them nonsignificant (Kurdek, 1977; Kurdek & Rodgon, 1975). A more recent study found that their relationship in children was small, with less than 10% shared variance (Hinnant & O'Brien, 2007). In a review looking at the relationship between perspective taking and morality, Kurdek (1978) discusses how limited our understanding of the outcomes and relationships to these constructs are when research only considers them in the broadest possible sense. This fits with findings from Oswald (1996) who found different outcomes to empathic arousal and altruistic behaviour based on whether individuals were asked to engage in cognitive versus affective perspective taking.

A measure that purports to assess social perspective taking more generally and does not differentiate the subtypes will fail to provide the type of discrimination that is necessary to understand how these different abilities fit in a broader framework of socio-emotional skills and outcomes. It is unknown how we should interpret data from these measures except to say that very low scores would be indicative of problems in both cognitive and affective perspective taking abilities and very high scores would indicate proficiency in both abilities. Scores in the middle, however, are subject to myriad possibilities, likely providing very little useful information in understanding either type of social perspective taking. In addition, the associated outcomes would be quite different for all of these possibilities yet would not be reflected in the

score and any examination of the relationships to outcomes may yield insignificant findings or conflicting findings across studies.

Without consideration of the constituent parts, we miss the important nuances that provide the information needed to truly understand the nature of the relationships and how broadly applicable the findings are, especially when the constituent parts are less clearly related than assumed by some. Clearly then, it is necessary to specify the specific components of perspective taking of interest and to examine how measurement of these specific abilities influences our understanding of them. It is thus important that researchers are clear about which *component* of social perspective taking they are measuring and ideally that any global measure will include component scores to allow comparisons to be made between the two types of social perspective taking and their relationship to different outcomes.

#### 1.5.1.2 Assessment in Different Populations

For many years, the 'gold standard' population used to develop measures of social perspective taking has been individuals with Autism Spectrum Disorders. The *Reading the Mind in the Eyes* task is no exception, with the development of both the adult and child tasks focused on the discrimination of those with Autism Spectrum Disorders from those without (i.e., Baron-Cohen et al., 1997; Baron-Cohen, Wheelwright, & Hill, 2001; Baron-Cohen, Wheelwright, Scahill, et al., 2001). One reason why this population is of interest is because of the severe social perspective taking deficits in this group and the resulting impact on their social well-being. However, with respect to more global measures of social perspective taking (including the *Reading the Mind in the Eyes* task), children and adults with Autism Spectrum Disorders are part of a population in which there is the presence of a dual deficit of both cognitive and affective perspective taking (for a review, see Baron-Cohen, 2000). This allows global measures to detect

these broad social perspective taking deficits relative to other populations. Despite the effectiveness in identifying Autism traits or differentiating individuals with Autism Spectrum Disorders from those without, a global measure of social perspective taking may not be as effective at providing information for other populations, for determining individual differences in the components of social perspective taking, and in identifying nuanced relationships to various social, cognitive, or emotional outcomes. There are other populations, like those with learning disabilities, psychopathic traits, schizophrenia, and depression, who show either less severe deficits or deficits in only one component of social perspective taking.

The population of primary interest herein, due to their prevalence and ease of identification in the general population, is children with specific learning disabilities. There is a wealth of evidence that children with learning disabilities are more likely to suffer socially throughout childhood (for a review, see Tur-Kaspa, 2002). There are myriad reasons for their social difficulties. For example, children with learning disabilities struggle to accurately perceive and interpret social and emotional cues in social situations (Pearl, 1987; for a review, see also Rourke, 1988) and they have a negativity bias in their perceptions of social situations (Weiss, 1984).

Although the struggle to perceive social and emotional cues may suggest a more global social perspective taking deficit, research examining the two components separately (cognitive vs. affective) has strongly favoured a deficit in affective perspective taking specifically. That is, many studies have shown a deficit in affective perspective taking in children with learning disabilities when compared to typically-developing peers (e.g., Axelrod, 1982; El-Haddad & Laouris, 2011; Holder & Kirkpatrick, 1991; Jackson, Enright, & Murdock, 1987; for a review, see Tur-Kaspa, 2002). In contrast, research looking at cognitive perspective taking has not found

consistent differences between children with learning disabilities and typically-developing children (e.g., Horowitz, 1981; Waterman, Sabesky, Silvern, Aoki, & McCauley, 1981). This deficit in affective perspective taking (and specifically the emotion recognition element of interest herein) has also been found in adults with learning disabilities (Carvajal, Fernández-Alcaraz, Ruede, & Sarrión, 2012; Davies, Frude, Jenkins, Harding, & Hill, 2014; Owen, Browning, & Jones, 2001; for a review, see Wood & Kroese, 2007), suggesting long-term difficulties. Therefore, any measure that accurately assesses affective perspective taking should identify individuals with learning disabilities, even though their deficits may not be as 'severe' as those with, for example, Autism Spectrum Disorders. Thus, by including a group of children with learning disabilities, I provide an additional way of addressing the question of whether or not the *Reading the Mind in the Eyes* task provides a valuable measure of individual differences in the affective components of social perspective taking (as part of the larger objective of assessing its value as a measure of either or both of the affective and cognitive subcomponents).

### 1.5.2 Forced-Choice Vs. Open-Ended Response Formats: What Affect Do They Have on Measurement and Performance?

A second issue of interest herein is how the response format can impact performance. Of particular interest, forced-choice response formats (much like multiple choice memory measures) may enable individuals to use alternate strategies to 'solve' the task. As previously mentioned, Peterson and Miller (2012) compared performance on the *Reading the Mind in the Eyes* task with face processing in a group of typical adults. As the task itself requires individuals to determine the mental or emotional state of an individual by looking at the eye region, it was assumed that it should have a strong relationship to face processing. However, they found that face processing was *un*related to the *Reading the Mind in the Eyes* task, accounting for less than one tenth of a

percent of the variance in performance. One possibility for this result is that the task is set up such that alternate strategies may be employed instead of specific perspective taking abilities that would utilize face processing.

Many of the measures that are utilized in social perspective taking research today are forced-choice in nature, particularly for affective perspective taking. That is, individuals are presented with options for answers and asked to select the correct one. This is in contrast to open-ended tasks that involve posing a question (e.g., "What is he feeling?" or "Why is he thinking that?"), allowing participants to *generate* their own answer, and coding the content of the answer. The differences between the two response formats may seem trivial, yet there are reasons to believe they may significantly influence findings. In many ways the arguments parallel those in the literature for distinguishing between recall and recognition memory. A vast body of work has repeatedly shown that open-ended (i.e., free recall) memory tasks tap distinct processes from those elicited by forced-choice (i.e., recognition) memory tasks, with the latter usually being less sensitive for detecting subtle, underlying memory deficits (Breen, 1993; Calev, 1984). Forced-choice memory tasks provide response options that may serve as cues to elicit the answer and allow for alternative strategies for solving the task that do not tap memory processes, such as 'process-of-elimination' (e.g., 'I know it's not A or B it must be C.'). Following similar logic, a forced-choice perspective taking task may recruit less 'perspective taking specific' processes than an open-ended format. Moreover, some of the populations that are studied with these measurement tools (e.g., those with psychopathic tendencies) are the very individuals who are most apt to employ compensatory strategies.

More evidence that alternate strategies may inadvertently influence the assessment of perspective taking (affective perspective taking specifically in this case) comes from Gordon and

colleagues (2004) who measured emotion recognition abilities (i.e., affective perspective taking) in typical college students and assessed neurological activation using fMRI during the task. The students were split into two groups based on a median split for callous-unemotional (CU) tendencies (the primary component for psychopathy, also linked to a deficit in affective perspective taking). Although there were no behavioural differences between the two groups, there were significant neurological differences. Specifically, those who scored *lower* on CU tendencies activated areas of the brain associated with face processing and emotion whereas those who scored *higher* on these tendencies activated areas of the brain associated with perception and cognition. One possible hypothesis for this result is that these individuals were able to use alternate strategies to solve the affective perspective taking task, hiding any deficits they may have had.

Arguably, one could say that if these individuals can solve the task, *regardless of how they do it*, then they do not show a deficit. However, the ability to solve it using cognitive means may be a product of the task that cannot always be replicated in real life (nor under other task conditions); this would explain why we can observe affective deficits in those high on psychopathic traits, even if they fail to demonstrate them in some test settings. In child samples where there are no other factors of influence (such as heightened psychopathic traits) children with better cognitive skills may be able to employ alternate strategies to solve social perspective taking tasks when the format is forced-choice, even if they lack the associated perspective taking abilities. Thus, the second question of interest lies in how response format (i.e., forced-choice vs. open-ended) influences performance and if the use of alternate strategies in forced-choice tasks may be able to explain any performance differences.

### 1.5.3 Specific Vs. Valence-Based Coding Schemes: What Affect Do They Have on Measurement and Performance?

Cognitive and verbal abilities naturally play a role in the assessment of various social, emotional, and cognitive skills. Simply being able to answer questions requires a level of verbal proficiency (e.g., understanding the question); additionally, cognitive abilities like short-term memory and inhibitory control can be necessary for certain types of questions. The goal, however, should be to ensure that social perspective taking tasks only assess these cognitive skills insofar as they are *necessary* to obtain an accurate measure of social perspective taking. With respect to both cognitive and affective perspective taking, there is a wealth of research suggesting that these abilities are highly correlated with the assessment of cognitive – most often specifically verbal – abilities in children and adults (e.g., Bosacki & Astington, 1999; Jenkins & Astington, 1996; Nilsen & Graham, 2009; Peterson & Miller, 2012; for a meta-analysis, see Milligan, Astington, & Dack, 2007), but it is less clear how *necessary* they are to the assessment of social perspective taking skills.

Work with young children has found strong correlations between cognitive perspective taking abilities and verbal ability with greater verbal ability associated with greater cognitive perspective taking (Bosacki & Astington, 1999; Jenkins & Astington, 1996; Meins, Fernyhough, Wainwright, Das Gupta, Fradley, et al., 2002). Notably, Jenkins and Astington (1996) found that the relationship to verbal ability in their 3- to 5-year-old sample was not strictly linear, but rather that a certain threshold was needed before children could 'pass' the false belief task, even after the effects of age were partialled out. In a preadolescent sample, Bosacki and Astington (1999) found that general vocabulary was correlated with a perspective taking measure that assessed a generalized understanding of mental states and emotional states in others using vignettes.

Notably, vocabulary accounted for almost 10% of the variance in social perspective taking, which was similar to the amount of variance social perspective taking accounted for in social competence (15%, or 9% when controlling for vocabulary). In another study, performance on a variety of cognitive perspective taking tasks in a group of preschool-aged children was predicted by both maternal mental state language use in infancy (accounting for 11% of the variance) and verbal ability (accounting for 16% of the variance) (Meins et al., 2002). Astington and Jenkins (1999) used a longitudinal methodology to demonstrate that language development was a requirement for cognitive perspective taking development (at least in 3-year-olds), not that they co-occurred or that cognitive perspective taking facilitated language development. Futhermore, Harris, de Rosnay, and Pons (2005), in a review of the literature, found in favour of a strong *causal* (not correlational) role of language in mental state understanding. It is therefore expected that some degree of verbal ability will be necessary for the assessment of cognitive perspective taking, but the question remains, "How much?"

When considering the role of verbal ability in affective perspective taking, there is no theoretical reason to assume that verbal ability should predict performance outside of a very basic level of verbal understanding (needed to generate basic terms or descriptions). However, despite this lack of a theoretical link, the influence of verbal ability is found when assessing affective perspective taking, particularly in children. For example, de Rosnay and colleagues found strong evidence of a link between language understanding and various elements of emotion understanding (de Rosnay, Pons, Harris, & Morrell, 2004; Pons, Lawson, Harris, & de Rosnay, 2003). Pons and colleagues (2003) examined the role of verbal ability (and age) on affective perspective taking measured as a hierarchy of understanding (starting with simple facial recognition of emotions and moving up to include elements of cognitive perspective taking such

as others' desires, beliefs, etc. and how they interact with emotion understanding) and found that age and verbal ability combined to account for 72% of the variance in affective perspective taking, with verbal ability uniquely accounting for 27% of the variance. One possibility, however, is that the relationships to verbal ability may be reflective of the cognitive perspective taking element of the task (as later understanding included cognitive perspective taking components), and not the affective element. However, Bruce and colleagues (2000) examined different facial emotion processing tasks in young children and found that task demands (including verbal, along with other cognitive demands) *did* influence developmental trajectories. Specifically, when the demands were low, by 6 years of age children had near-perfect accuracy; however, when the task demands were high, performance was poor until children were around 10 years of age; yet both tasks are believed to assess the same emotion recognition element of affective perspective taking. This supports the view of early mastery of basic emotion identification with the caveat that task type can greatly influence performance.

It would seem that the assessment of cognitive perspective taking (and possibly even more so affective perspective taking) is likely being influenced by task demands on one's verbal ability, one being that some tasks may require knowledge of specific vocabulary terms. For example, in the children's *Reading the Mind in the Eyes* task, options include terms like "ashamed", "annoyed", and "disgusted" and specific questions ask children to differentiate between "angry" and "bossy" (from visual cues) as well as "excited" and "happy", for example. Requiring specific vocabulary knowledge, may inhibit the expression of actual social perspective taking skills, or may mask important developmental information.

In the research herein, comparisons will be made between versions of the same task (and thus same stimuli) that vary the verbal demands to determine the exact influence of these

demands on perspective taking abilities. Specifically, I will compare and contrast two coding schemes for both the open-ended and forced-choice versions of the *Reading the Mind in the Eyes* task: One that codes for specific answers versus one that codes for answers that include the 'gist' of what is correct.

I believe that when there is only one specific answer that is coded as correct, individuals who have some social perspective taking understanding but may lack the exact words, are unable to display their level of knowledge. This can happen in both forced-choice and open-ended tasks. In most tasks, there may be a *best* answer, but it does not mean an answer of the same valence or category is 'incorrect', just not *as* correct. In real-life situations, the other answer may suffice in terms of how an individual responds to the situation and thus would not be associated with the negative outcomes linked to poor affective perspective taking abilities (see Section 1.3.2.5). This is somewhat similar to the prototype theory for nouns (Rosch, 1975) and the idea of distance between concepts (Wittgenstein, 1953): Although a single emotional term may be the prototype (or 'correct' answer) for a given question, other emotional terms within a certain distance of the prototype may be valid for the purposes of the social interaction. For example, a task that asks individuals to identify the emotional expression of an individual may be best described as "nervous", but an individual who says "scared" or "overwhelmed" or even "sad" will likely be close enough to the correct answer that they would respond in a socially appropriate manner.

Open-ended tasks can also be subject to influences of verbal ability depending on the amount or specificity of the information an individual must provide. For example, a coding scheme that requires the participant to generate a specific term in their answer (e.g., "nervous") will fall prey to the same demand problem as force-choice tasks that utilize specific terms.

Coding of open-ended tasks invariably includes a level of subjectivity on the part of the coder which can allow for a more nuanced look at what answers are provided. There will be clearly 'wrong' answers and those that are clearly 'correct', with individual raters deciding, using standardized criteria for inter-rater reliability, the acceptability of the in-between answers. Regardless, the use of a broader coding scheme means that individuals who are not hitting the exact mark are still able to demonstrate a level of perspective taking and can be coded as such, as has been done successfully with young preschool-age children (Widen & Russell, 2003). For example, Widen and Russell (2008) found that children's development of emotion understanding begins in a valence-based manner, narrowing to more specific understanding with age. Coding that required specific terms found children's understanding almost completely lacking, but when valence-based coding could be employed, performance was much higher and reflective of their *actual* understanding. However, if the coding is too broad it may fail to make the important distinctions between what is generally correct or incorrect. Depending on the type of understanding one is interested in different coding may provide different information.

## 1.6 The Current Research

### 1.6.1 Questions of Interest

The current research was designed to examine the three measurement issues highlighted above by assessing the efficacy of the *Reading the Mind in the Eyes* task and modified versions to address the following specific questions of interest:

1. a) Given the *Reading the Mind in the Eyes* task claims to assess social perspective taking more generally, yet there is no significant overlap between cognitive and affective perspective taking (e.g., Kurdek & Rodgon, 1975), what component is the *Reading the Mind*

*in the Eyes* task measuring? The face validity would suggest affective perspective taking, yet research using the measure as such (e.g., Richell et al., 2003) would suggest otherwise.

b) Do the modifications made to the task (on response format and coding) alter what type of social perspective taking is measured?

2. What affect does the response format have on perspective taking performance? Does the *Reading the Mind in the Eyes* task allow individuals to use compensatory measures to 'solve' the task? Does an open-ended format avoid this problem and provide a better, and arguably more ecologically valid, measure?
3. Does the type of coding employed by the *Reading the Mind in the Eyes* task – namely the use of term-specific coding – necessitate verbal or vocabulary demands that may inhibit performance for certain individuals? Does a more valence-based coding option improve measurement?

To these ends, a modified version of the *Reading the Mind in the Eyes* task was created by simply changing the response format to one of an open-ended format. This open-ended version of the *Reading the Mind in the Eyes* task (herein referred to as the "open-ended *Eyes Task*") is believed to be a more ecologically valid task, particularly for assessing affective perspective taking.

In addition to modifications to the response format (i.e., changing the task from forced-choice to open-ended), different coding schemes were utilized and compared across both the original *Reading the Mind in the Eyes* task (herein referred to as the "forced-choice *Eyes Task*") and the open-ended *Eyes Task*. As mentioned, the *Reading the Mind in the Eyes* task uses term-specific coding; however, it is possible to create a forced-choice version that employs valence-based coding. Similarly, the open-ended *Eyes Task* can be coded using either term-specific

coding (i.e., requiring that the individual give a specific term, or synonym thereof) or valence-based coding (i.e., requiring that the individual give a term that is of the same valence as the 'correct' answer). Both of these coding schemes were included for both the forced-choice *Eyes Task* and the open-ended *Eyes Task*, providing a comparison across four different scores (two task types x two coding schemes), as is seen in Table 1. For an example of how the coding would influence what is deemed 'correct', see Table 2.

**Table 1.** The Scores of the *Reading the Mind in the Eyes* Task to be Compared.

| | **Forced-Choice *Eyes Task*** | **Open-Ended *Eyes Task*** |
|---|---|---|
| **Term-Specific Coding** | Forced-choice, term-specific *Eyes Task* | Open-ended, term-specific *Eyes Task* |
| **Valence-based coding** | Forced-choice, valence-based *Eyes Task* | Open-ended, valence-based *Eyes Task* |

**Table 2.** Examples and Comparisons of Coding Across *Eyes Task* Scores.

| | ***Eyes Task* Version** | | | |
|---|---|---|---|---|
| **Example** | **Forced-choice, term-specific** | **Forced-choice, valence-based** | **Open-ended, term-specific** | **Open-ended, valence-based** |
| Options:<br>Hate<br>Surprised<br>Kind<br>Cross | Kind | Kind | Kind, caring, sympathetic, nice, thoughtful, benevolent, compassionate | All of the answers in the open-ended, term-specific *Eyes Task* and: happy, glad, joyful, pleased, interested, and any other positive valence |
| Options:<br>Shy<br>Guilty<br>Daydreaming<br>Worried | Worried | Worried, Guilty | Worried, concerned, anxious, nervous, troubled, upset | All of the answers in the open-ended, term-specific *Eyes Task* and: sad, frustrated, depressed, bored, guilty, and any other negative valence |

The first question of interest—what type of social perspective taking is being assessed in each task—will be addressed by comparing performance on each task to: a) a known cognitive perspective taking task; and b) a self-report empathy measure (given the link between affective perspective taking and empathy). Included in these samples of children is one group of children with learning disabilities who have affective perspective taking deficits to allow for a comparison of performance between the versions in identifying deficits. If one version better discriminates between the at-risk group and the typically-developing group of children (controlling for verbal ability and age) then it would suggest that version is a better measure of affective perspective taking specifically.

The second question of interest—whether or not the response format influences measurement by allowing for compensatory abilities to 'solve' the task—will be assessed by comparing performance across the forced-choice and open-ended *Eyes Task* using both a between-subjects and within-subjects methodology. If the two response formats utilize the same skills and abilities, there should be no difference between or within subjects (or a very minimal difference); however, if one task allows the use of compensatory strategies – especially if the task is difficult without these compensatory strategies – one would expect to see significant differences between the two tasks types (specifically better performance on the forced-choice version). Furthermore, if there are differences, these differences in performance will be compared to performance on a cognitive ability measure to determine if the hypothesis involving the employment of alternate strategies is plausible.

The third question of interest—how coding influences the task demands, specifically vocabulary demands—will be assessed by comparing performances both within- and between-subjects. If vocabulary demands are too high, especially for younger children, then the scores

when term-specific coding is used should be much lower than when valence-based coding is used, both between- and within-subjects. In addition to simply comparing performance, relationships to verbal ability—specifically vocabulary—will be conducted by type of coding.

### 1.6.2 Hypotheses

Four specific hypotheses were made with regard to my overarching research questions:

**Hypothesis #1: The open-ended *Eyes Task* will be a better measure of affective perspective taking than the forced-choice *Eyes Task*.** As previously mentioned, the forced-choice *Eyes Task* has the face validity of an affective perspective taking task, given the stimuli used; however, the forced-choice nature of the task that allows for cognitive strategies to circumvent affective processes and the inclusion of non-emotional state terms raises questions about its efficacy as an affective perspective taking measure. The open-ended *Eyes Task* should eliminate this concern, making the task a better measure of affective perspective taking . In this vein, I expect that the open-ended *Eyes Task* will be significantly correlated with self-reported empathy in children (Study 1) whereas I do not expect such a correlation with the forced-choice *Eyes Task*. I also expect the open-ended *Eyes Task* to better differentiate a sample with affective perspective taking deficits from a typically-developing one (Study 3).

**Hypothesis #2: The forced-choice *Eyes Task* will assess more cognitive abilities (including cognitive perspective taking) than the open-ended *Eyes Task*.** The forced-choice *Eyes Task's* proven ability to distinguish typically-developing individuals from those with Autism Spectrum Disorders (Baron-Cohen, Wheelwright, Hill, et al., 2001; Baron-Cohen, Wheelwright, Spong, et al., 2001) suggests it measures some form of social perspective taking. I hypothesize that the forced-choice *Eyes Task*, and especially the forced-choice, term-specific *Eyes Task*, primarily assesses cognitive perspective taking *and* other cognitive abilities. This

will be evidenced by a significant relationship to another known measure of cognitive perspective taking and a strong relationship to verbal abilities (Studies 2 and 3).

**Hypothesis #3: The forced-choice *Eyes Task* allows the use of alternate strategies to 'solve' the task, resulting in inflated scores.** I predict that performance on the forced-choice *Eyes Task* will consistently be higher than that for the open-ended *Eyes Task* (regardless of coding scheme) (Studies 1-3). Furthermore, when compared to a measure of non-verbal cognitive ability, I expect that performance on the forced-choice *Eyes Task* will be related whereas performance on the open-ended *Eyes Task* will not (again, regardless of coding scheme), reflecting the ability of people to use cognitive strategies when solving the forced-choice *Eyes Task*.

**Hypothesis #4: The vocabulary demands of term-specific coding will result in lower scores than when valence-based coding is used.** Specifically, valence-based coding should minimize verbal demands, meaning scores should be higher when valence-based coding is used over term-specific coding (regardless of response format) (Studies 1-3). In addition, the relationship to verbal ability should be higher when term-specific coding is used relative to valence-based coding across all samples and ages (Studies 2 and 3). The potential task demand of vocabulary knowledge may mask the true abilities of individuals (e.g., children) with lower verbal ability or vocabulary knowledge when term-specific coding is employed. This masking would thus influence our understanding of the developmental trajectory of social perspective taking abilities as well as the relationships to social perspective taking skills, especially in populations where task demands may result in undue influence.

# 2 Study 1

## 2.1 Introduction

Study 1 was designed to compare the forced-choice *Eyes Task* and the open-ended *Eyes Task*, using both types of coding, in a between-subjects analysis in a typically-developing child sample. Specifically, the goals were to determine how the measures compare to dispositional empathy and to compare performance based on both response format and type of coding. To this end, participants completed *either* the forced-choice *Eyes Task* or the open-ended *Eyes Task*. Each participant was given two scores reflecting the two coding schemes (i.e., valence-based coding and term-specific coding, refer back to Table 1 for the four scores). Performance on these tasks was compared to dispositional empathy, which was selected in order to provide a measure of an outcome primarily associated with affective perspective taking (as opposed to cognitive perspective taking) as there is evidence showing a positive relationship between affective perspective taking and empathy (Carr & Lutjemeier, 2005; Chambers & Davis, 2012; Chapman, Baron-Cohen, Auyueng, Knickmeyer, Taylor, & Hackett, 2006; Hinnant & O'Brien, 2007; Lamm, Batson, & Decety, 2007; Lui, 2014; McIlwain, 2003; Oswald, 1996, 2002; for a review of earlier findings, see Eisenberg, Shea, Carlo, & Knight, 1991) and a lack of such a relationship with cognitive perspective taking (Davis & Stone, 2003; Oswald, 1996; Paal & Bereczkei, 2007; Repacholi et al., 2003; Sutton, 2001).

Given the established relationship between empathy and affective perspective taking (for specific evidence with facial emotion recognition which is utilized herein, see Besel & Yuille, 2010; Carr & Lutjemeier, 2005), this study provides information for the first question of interest, namely, are the forced-choice and open-ended *Eyes Tasks* assessing affective perspective taking?

If, as has been hypothesized by the creators of the *Reading the Mind in the Eyes* task, the measure assesses a mix of both cognitive and affective perspective taking, I would expect to see a mild or moderate relationship between performance on the forced-choice *Eyes Task* and dispositional empathy. If, however, the hypotheses outlined herein are supported and the forced-choice nature of the forced-choice *Eyes Task* inhibits the assessment of affective perspective taking, there should be no relationship between the forced-choice *Eyes Tasks* and dispositional empathy. Furthermore, if the open-ended response format allow for a better measure of affective perspective taking, there should be a moderate relationship between the open-ended *Eyes Task* and dispositional empathy, in line with my Hypothesis #1. In addition to addressing Hypothesis #1, Study 1 also addresses, in part, Hypotheses #3 and #4 by allowing a within-subjects comparison of performance based on different response formats and coding styles.

## 2.2 Methods

### 2.2.1 Ethics Statement

All children provided verbal assent and parents provided written consent prior to our arrival at the program. The study was approved by the Behavioural and Research Ethics Board of the University of British Columbia.

### 2.2.2 Participants

One hundred and eighteen participants were recruited from after-school care programs in Vancouver, Canada. Participants ranged in age from 4 years to 8 years (72 boys; mean age = 5.84 years, SD=.97 years). An additional 16 participants (11.9% of the full sample) were excluded due to their failure to complete either the measure of dispositional empathy (9) or their version of the *Eyes Task* (7). In the opinion of the research assistants running these children, the

failure to complete the dispositional empathy measure was most often due to failure to understand the questions whereas failure to complete any version of the *Eyes Task* was more often due to parental interruption (i.e., picking the child up). Notably, if children found their version of the *Eyes Task* difficult and included "Don't know" as answers, these data were included in the analyses.

### 2.2.3 Measures

#### 2.2.3.1 Perspective Taking

*Reading the Mind in the Eyes, Children's Version* (Baron-Cohen, Wheelwright, Spong, et al., 2001). This task requires participants to choose the most appropriate term to describe the thoughts or feelings of others based upon still photographs of the eye-region of the face. Participants choose from four possible terms that researchers read aloud. For example, item 3 has the options of "friendly", "surprised", "sad", and "worried" (the correct answer is "friendly"). See Appendix A for examples and coding. Regardless of item type, children are always asked, "What is this person feeling?". Half of the participants completed the forced-choice *Eyes Task* whereas the other half completed the open-ended *Eyes Task* (see below). Two scores were computed based on coding scheme: The typical term-specific coding (a.k.a. forced-choice, term-specific *Eyes Task*) or a valence-based coding in which any similarly-valenced answer was coded as correct (a.k.a. forced-choice, valence-based *Eyes Task*).

*Open-Ended Eyes Task* (Cassels & Birch, 2014; Carey & Cassels, 2013). This task was created using the stimuli from the standard task. It is identical except that it requires participants to generate their own answer when asked "What is this person feeling?". Although the question was phrased as dealing with feelings, children often provided non-emotive terms such as

"nothing", as well as ones that reflected cognitive states, such as "thinking" or "remembering", suggesting thinking of emotions and non-emotional mental states as similar or of the same construct is automatic. For the open-ended, valence-based *Eyes Task*, answers were coded as correct if the participants' answer matched the emotional valence category (i.e., positive, neutral, negative, or hostile) of the correct response from the forced-choice format. For the open-ended, term-specific *Eyes Task*, answers were coded as correct if the participants' answer matched the term (or a synonym of the term) from the forced-choice format[1]. See Appendix A for the coding scheme and examples.

#### 2.2.3.2 Dispositional Empathy

*Bryant Index of Empathy* (Bryant, 1982). This is a 22-item self-report measure of dispositional empathy, validated for ages 5 through 13. To make the measure easier to understand the items were presented verbally in question form (instead of statement form). For example, the item "People who kiss and hug in public are silly" became "Do you think that people who kiss and hug in public are silly?" Options for answers are "True" or "Not true", although as the measure was administered verbally in question form, I accepted "Yes" and "No" as answers. The measure provides one total score for dispositional empathy and some questions involve the perspective taking components of empathy. The inclusion of the perspective taking component of empathy is critical as this component is linked directly to affective perspective taking, or the tendency to consider the emotional state of others. For example, the question, "It's

[1] I completed all coding for this task, but unfortunately due to unforeseen circumstances, there was no reliability coding for this particular Study.
[2] The author of the original measure and members of her lab have used one of each type of story in her research as well (Booth, personal communication 2006).
[3] The forced-choice, valence-based *Eyes Task* and open-ended, valence-based *Eyes Task* were also compared but the

hard for me to see why someone else gets upset" (negatively coded) reflects the perspective taking aspect of empathy.

The measure has good test-retest reliability (ranging from $r$=.74 to .81) and convergent and discriminant validity, as assessed using other measures of empathy (convergent) and reading-achievement scores (discriminant) (Bryant, 1982). Not all children answered all questions and in line with traditional missing data procedures, participants who had more than 10% missing on this measure (> 2 items) were removed from any analysis with this measure (for a total of n=6). For those participants with < 10% missing (1 or 2 items), logistic regression was used to predict the missing data (for a total of n=13). In our sample, the alpha was .58 if considering those who completed all questions or .50 with missing data predicted, in line with the reported .54 for the 5-6 year-old group in the validation study (Bryant, 1982).

### 2.2.4 Procedure

All participants took part at their after-school care program. Children were first given an overview of the study and provided verbal assent; any child who did not want to participate did not have to, even if parents provided consent. Participants were first given the *Bryant Index of Empathy* followed by either the open-ended *Eyes Task* or the forced-choice *Eyes Task*. The measure of dispositional empathy was given first in order to avoid performance on perspective taking influencing the measure of empathy. Although answering questions about dispositional empathy may increase accuracy in the perspective taking task, it would equally influence both versions of the *Eyes Task*.

## 2.3 Results

### 2.3.1 Preliminary Analyses

In the final sample, 65 participants completed the forced-choice *Eyes Task* and 53 completed the open-ended *Eyes Task*. Independent samples t-tests were conducted to ensure that there were no differences between the two groups on sex and age. Neither sex, $t(116)=1.040$, $p=.311$, nor age, $t(116)=.826$, $p=.411$, were significant thus they were not included as covariates in further analyses comparing the two groups, but were included as independent or dependent variables in further analyses. Additionally, to ensure that there were no empathy differences between the two groups, an independent samples t-test was conducted with dispositional empathy as the dependent variable. This too was non-significant, $t(113)=.080$, $p=.936$, and thus any differential relationships between the forced-choice or open-ended *Eyes Task* cannot be due to inherent group differences on empathy. However, there were significant differences by sex on dispositional empathy, with females scoring higher than males, $t(113)=3.50$, $p=.001$, in line with previous findings using self-report measures of empathy (for a review, see Eisenberg & Lennon, 1983), but no significant correlation with age, $r=-.059$, $p=.534$. When examining the relationships between dispositional empathy and the forced-choice or open-ended *Eyes Task*, further analyses by sex, but not age, were conducted in addition to the main analyses. See Table 3 for information on variables of interest between the two groups.

**Table 3.** Comparisons of the Two Groups on the Variables of Interest.

| Variable | Forced-Choice Response Format Group | Open-Ended Response Format Group |
|---|---|---|
| Mean Age (SD) in Years | 5.92 (1.02) | 5.76 (0.95) |
| % Boys | 50.77% | 60.38% |
| Mean Score (SD) on Dispositional Empathy | 12.61 (3.06) | 12.66 (3.29) |
| Mean (Range) for term-specific coding | 13.82 (3 to 22) | 2.53 (0 to 5) |
| Mean (Range) for valence-based coding | 16.55 (5 to 23) | 8.04 (2 to 13) |

Note: For the forced-choice, valence-based *Eyes Task*, only 9 of the 28 items included same-valence foils in the answer options, but all items were included in the performance score.

### 2.3.2 Comparing the Versions of the *Eyes Task* by Age and Sex

In order to compare performance of the forced-choice or open-ended *Eyes Task* with age, Pearson correlation coefficients were calculated. Age was found to be highly correlated with performance on the forced-choice, term-specific *Eyes Task*, $r$=.438, $p$<.001, and the forced-choice, valence-based *Eyes Task*, $r$=.480, $p$<.001, suggesting improvement with age. In contrast, there was no relationship with age for the open-ended, valence-based *Eyes Task*, $r$=-.061, $p$=.662, nor the open-ended, term-specific *Eyes Task*, $r$=-.011, $p$=.937, suggesting that performance was equal across the ages tested.

Independent sample t-tests were used to compare performance by sex across the various *Eyes Task* scores. Despite slight female advantages across perspective taking tasks, there were no statistically significant differences on performance by sex for any of the individual scores, with all p's > .12. See Table 4.

**Table 4.** Relationships Between the *Eyes Tasks* and both Age and Sex.

| | ***Eyes Task* Scores** | | | |
|---|---|---|---|---|
| | **Forced-choice, term-specific** | **Forced-choice, valence-based** | **Open-ended, valence-based** | **Open-ended, term-specific** |
| Age (r) | .438*** | .480*** | -.061 | -.011 |
| Sex (% correct) | | | | |
| Males | 47.62 | 57.79 | 27.01 | 8.48 |
| Females | 51.12 | 60.49 | 31.29 | 9.86 |

Note: *$p$<.05; **$p$<.01; ***$p$<.001

### 2.3.3 Correlations Between Dispositional Empathy and the Versions of the *Eyes Task*

To test the relationships between dispositional empathy and the different versions of the *Eyes Task*, bivariate correlations were conducted. First, as hypothesized, there was a difference in relationships based on response format (i.e., open-ended versus forced-choice): The forced-choice, term-specific *Eyes Task*, $r$=.074, $p$=.567, and forced-choice, valence-based *Eyes Task*, $r$=-.007, $p$=.954, were not significantly related to dispositional empathy. In contrast, the open-ended, valence-based *Eyes Task*, $r$=.272, $p$=.048, and the open-ended, term-specific *Eyes Task*, $r$=.281, $p$=.042, were both significantly related to dispositional empathy, accounting for approximately 8% of the variance. When comparing the correlation coefficients, Fisher's r-to-z transformation was used as the samples were independent from each other. When the coding was term-specific (i.e., forced-choice, term-specific *Eyes Task* vs. open-ended, term-specific *Eyes Task*), the difference between the correlations was non-significant, $z$=1.13, $p_{1\text{-}tail}$=.129, but when the coding was valence-based (i.e., forced-choice, valence-based *Eyes Task* vs. open-ended, valence-based *Eyes Task*), the difference was bordering significance, $z$=1.52, $p_{1\text{-}tail}$=0.064.

When the correlations were examined by sex, the resulting sample sizes were quite small, thus both statistical and practical significance are discussed. For boys, the relationship between dispositional empathy and both open-ended *Eyes Task* scores remained practically significant (and statistically significant for the open-ended, term-specific *Eyes Task*), with $r$=.335, $p$=.061 (open-ended, valence-based *Eyes Task*) and $r$=.456, $p$=.009 (open-ended, term-specific *Eyes Task*), accounting for 11 or 21% of the variance in dispositional empathy, respectively. The relationships to the forced-choice *Eyes Task* scores remained non-significant, though the magnitude of the relationship was larger than in the full sample, particularly for the forced-choice, term-specific *Eyes Task*. For girls, none of the measures were significantly related to dispositional empathy, all *p's* > .14, and, notably, the direction of every relationship was negative. See Table 5 for full results. When comparing the coefficients for boys (controlling for coding), again Fisher's r-to-z transformation was used. For the boys, regardless of coding, the difference between correlation coefficients did not meet statistical significance, both *p's*>.15, one-tailed; however, the low sample sizes in each group (30 and 32 for the forced-choice and open-ended response format, respectively) meant the power to detect such a difference was very low. Notably, the differences in correlations were greater than the aforementioned difference that was nearly statistically significant when all participants were included, thus the likelihood is that power is the primary issue here.

**Table 5.** Perspective Taking Performance and Relationship to Dispositional Empathy by Sex.

| | ***Eyes Task* Scores** | | | |
|---|---|---|---|---|
| | **Forced-choice, term-specific** | **Forced-choice, valence-based** | **Open-ended, valence-based** | **Open-ended, term-specific** |
| Boys % (SD) | 47.62 (15.89) | 57.79 (14.57) | 27.01 (10.54) | 8.48 (4.69) |
| Boys $r_{ab}$ | .220 | .086 | .335† | .456** |
| Girls % (SD) | 51.12 (12.94) | 60.49 (12.10) | 31.29 (8.52) | 9.54 (4.36) |
| Girls $r_{ab}$ | -.266 | -.256 | -.021 | -.107 |

Note: †$p$<.10; *$p$<.05; **$p$<.01

If the variances were severely restricted for either task, it would inhibit the ability of the tasks to detect a significant difference; however, the variances were relatively equal across sexes for all perspective taking scores and dispositional empathy, thus variance cannot likely account for the differences found herein.

### 2.3.4 Comparing Performance by Response Format and Coding Scheme

There were two task types and two coding schemes, resulting in four scores to be compared. In order to compare the effects of both response format (forced-choice vs. open-ended) and coding scheme (term vs. valence-based), a mixed-model ANOVA was conducted with coding scheme (term-specific vs. valence-based) as the repeated measure, child sex and response format (forced-choice vs. open-ended) as the between-subjects variables, and child age as a covariate due to its significant relationship with the forced-choice *Eyes Task*. The repeated-measures ANOVA was selected as the initial test over independent t-tests to help control for Type I error and to examine possible interactions between response format, coding, and sex.

Age was a significant covariate, $F(1,113)$=13.145, $p$<.001, with average performance increasing with age, but there was no main effect of sex, $F(1,113)$=2.312, $p$=.131, or any interaction with sex, all $p$'s>.14. There were main effects of both response format, $F(1,113)$=303.996, $p$<.001, and coding scheme, $F(1,113)$=14.853, $p$<.001. When examining the

effect of response format, the forced-choice *Eyes Task* resulted in significantly higher scores than the open-ended *Eyes Task*. The main effect of coding was that, regardless of response format, valence-based coding led to significantly higher scores than term-specific coding. The interaction between response format and coding scheme was also significant, $F(1,113)=66.766$, $p<.001$. Examination of the means showed an ordinal interaction in that the direction of the effects of coding were the same for each response format (valence-based coding being associated with significantly higher scores than term coding; $M_{\text{forced-choice,term-specific}}=49.1\%$, $M_{\text{forced-choice,valence-based}}=58.9\%$, $M_{\text{open-ended,term-specific}}=9.5\%$, $M_{\text{open-ended,valence-based}}=29.5\%$), but the degree of the effect of coding was *greater* for the open-ended *Eyes Task*.

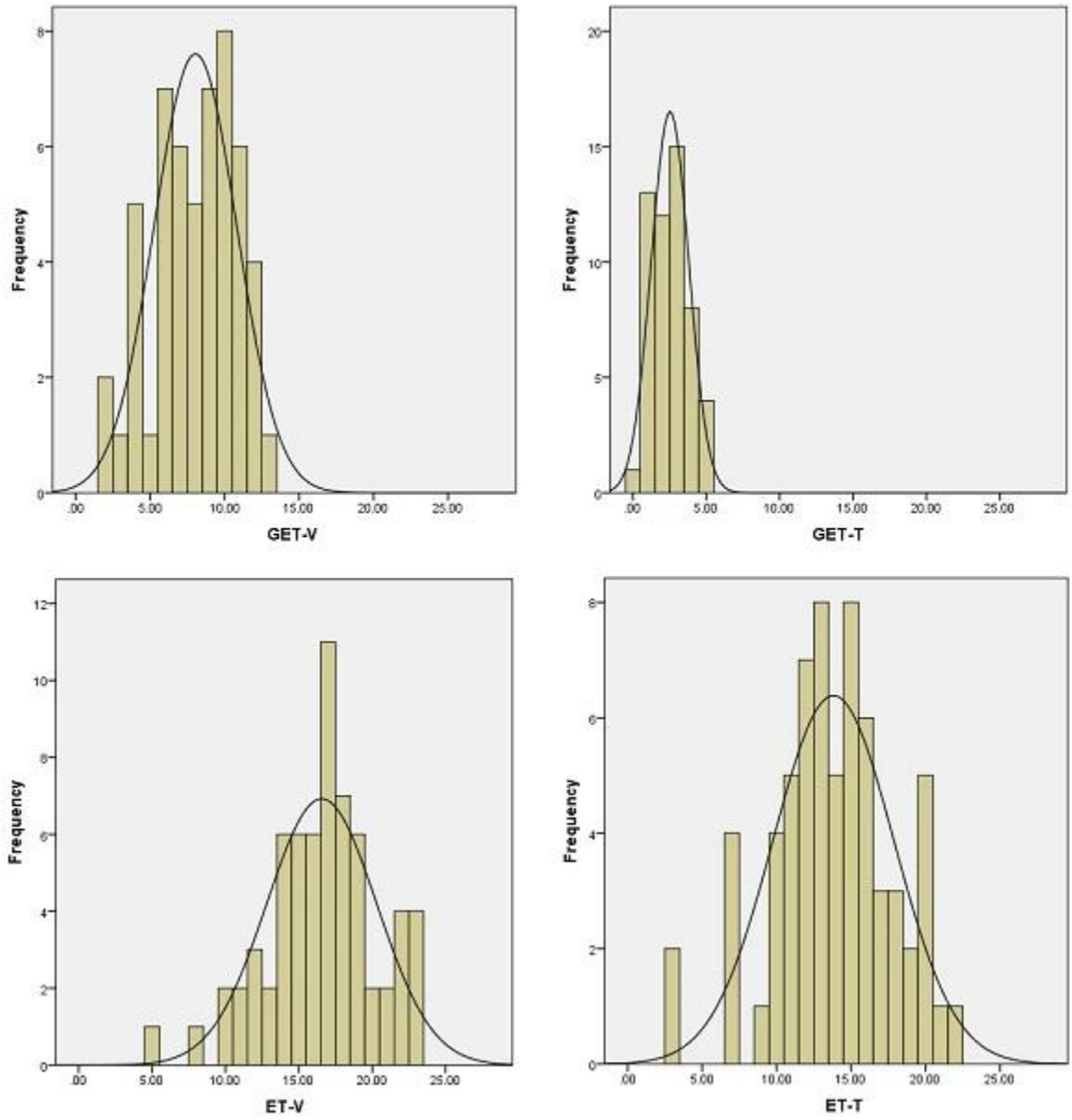
Of note, the scores for the open-ended, term-specific *Eyes Task* were incredibly low and this may have influenced the specific analyses in which they were included, thus follow-up t-tests were run to ensure that the main effects were not due to these low scores. In comparing response format (controlling for coding), the forced-choice *Eyes Task* resulted in significantly higher scores than the open-ended *Eyes Task*, regardless of coding, $t(79)=21.153$, $p<.001$ (term-specific coding, Case 4 t-test used due to unequal variances) and $t(116)=13.745$, $p<.001$ (valence-based coding). In comparing coding (controlling for response format), scores for valence-based coding were significantly higher than those for term-specific coding, $t(64)=16.648$, $p<.001$ (forced-choice *Eyes Task*) and $t(52)=17.006$, $p<.001$ (open-ended *Eyes Task*). Thus, the main effects for response format and coding do not seem to be unduly influenced by the low open-ended, term-specific *Eyes Task* scores. See Table 6 for scores.

**Table 6.** Percentage Correct by Response Format and Coding Scheme.

| | **Forced-Choice *Eyes Task*** | **Open-Ended *Eyes Task*** |
|---|---|---|
| **Term-Specific** | 49.1% | 9.0% |
| **Valence-Based** | 58.9% | 29.5% |

Importantly, the distributions on the forced-choice, term-specific *Eyes Task*, open-ended, valence-based *Eyes Task*, and forced-choice, valence-based *Eyes Task* were very similar, with no ceiling or floor effects and all showed a relatively normal distribution; however, the distribution of the open-ended, term-specific *Eyes Task* was clustered near the bottom suggesting a floor effect despite a somewhat normal distribution of the small range of scores. This implies that term coding combined with an open-ended response format is impractical to use with children of this age. See Figure 1 for details. There was also significantly greater variance for the forced-choice *Eyes Task*; statistically, this means there was a greater likelihood of finding significant relationships with other variables using the forced-choice format.

**Figure 1.** Distributions of the Four *Eyes Task* Scores.




Note: GET-V=open-ended, valence-based *Eyes Task*; GET-T=open-ended, term-specific *Eyes Task*; ET-V=forced-choice, valence-based *Eyes Task;* ET-T=forced-choice, term-specific *Eyes Task*

## 2.4 Discussion

The primary goals of the current study were to: a) compare and contrast the ability of the open-ended and forced-choice *Eyes Task* to assess affective perspective taking by comparing performance to dispositional empathy, b) compare and contrast scores by response format, and c) compare and contrast scores when coding is term-specific versus valence-based.

The first goal was to determine if either the open-ended or forced-choice *Eyes Task* better predicted dispositional empathy and thus might be considered a better measure of affective perspective taking, given the established link between empathy and affective perspective taking (e.g., Besel & Yuille, 2010; Chambers & Davis, 2012), instead of, or in addition to, cognitive perspective taking. It was hypothesized that the open-ended *Eyes Task* (regardless of coding) would better serve as a measure of affective perspective taking due to the face validity of the task as an affective perspective taking measure. The results supported this hypothesis. Independent of coding scheme, the forced-choice *Eyes Task* was not significantly related to dispositional empathy whereas the open-ended *Eyes Task* was significantly related to dispositional empathy, despite the higher variance in the forced-choice version that should have resulted in a higher correlation if all else was equal. Notably, scores on the open-ended *Eyes Task* accounted for approximately 8-9% of the variance in dispositional empathy, which is similar in magnitude to previous studies examining the relationship between affective perspective taking and empathy (which have been in the range of 10% to 12%; Hinnant & O'Brien, 2007; Hooker et al., 2008).

This relationship, however, was moderated by participant sex, with boys showing strong relationships between dispositional empathy and the open-ended *Eyes Task* whereas the relationships for girls were non-significant. Indeed, the largest relationship for girls was with the forced-choice, term-specific *Eyes Task* and it was negative, suggesting better performance on the

forced-choice, term-specific *Eyes Task* was associated with *lower* dispositional empathy, though this remained statistically non-significant. This counters previous research that *has* found a relationship between the *Reading the Mind in the Eyes* task and dispositional empathy (Chapman et al., 2006). One difference between the current research and the Chapman and colleagues study was the measurement of dispositional empathy; whereas Chapman and colleagues used a parent-report version of dispositional empathy, the research herein utilized a self-report version and it is unclear, currently, which measure would be more accurate – parent or child report.

One possibility for these sex differences is that the variance on the tasks was different for girls than boys; however, as noted above no differences in variance were found between the two sexes on any measure. A more likely possibility is that females are more likely to respond to social demands when reporting dispositional empathy and thus their dispositional empathy scores are not reflective of their *actual* empathy. Some support for this comes from research that has found that although self-report dispositional empathy measures often find a female advantage, behavioural measures of empathy, particularly when anonymous, fail to find this sex advantage (for a review, see Eisenberg & Lennon, 1983). Another possibility is that the relationship between empathy and affective perspective taking is limited or absent in females. One possible mechanism for this is via parenting. It has been shown repeatedly that mother-child discussions of emotions lead to greater emotional understanding and social competence, including displays of empathy, cooperation, and prosociality (Dunn et al., 1991; Dunn, Bretherton, & Munn, 1987; Denham, Renwick-DeBardi, & Hewes, 1994; Harris et al., 2005); however, it is possible that sex differences in these discussions result in differing relationships between affective perspective taking and empathy between the two sexes. There have been some sex differences found in this type of emotion socialization (Denham, Bassett, & Wyatt, 2010;

Root & Denham, 2010) that may account for the different patterns in young childhood. For example, Denham and colleagues (2010) found that although parents' styles do not often differ based on the sex of their child, daughters are more susceptible to parental influence of emotion socialization than sons; therefore, given the same degree of emotion socialization, girls are more likely to internalize this type of socialization, possibly leading to greater dispositional empathy *regardless* of affective perspective taking abilities. A final possibility is that this is a spurious result that would not emerge in replication. Future research is needed to confirm the presence of sex differences in this relationship and, if present, to examine the mechanisms behind it, though this remains outside the scope of the current research as it is subsidiary to the objectives of this dissertation.

The second goal of this study was to examine the effect of response format. There was a main effect of response format with the forced-choice *Eyes Task* being significantly easier than the open-ended *Eyes Task*, regardless of the type of coding. This provides evidence that the response format influences performance and, in line with previous memory research (e.g., Breen, 1993), the forced-choice nature of the task may ease the burden on the participant. One possibility for this is that the participant can utilize other mechanisms to 'solve' the task; however, the data in this study cannot speak directly to the mechanisms behind these results, only confirm that they do indeed exist, as the changes were tested in a between-subjects methodology.

The final goal of this study was to examine the effect of coding. Results support a main effect of coding; specifically, valence-based coding provided higher scores than term-specific coding when compared within the same response format. This should come as no surprise as the valence-based coding provided for more answers to be deemed "correct" than a term-specific

coding scheme. Term-specific coding in the open-ended version included not only the correct term on the forced-choice, term-specific *Eyes Task* but any synonym as well, leaving open a broad range of terms, but only a narrow range of meaning; conversely, the valence-based coding allowed for greater range in meaning to be deemed correct and, as argued in section 1.5.3, this is likely more reflective of day-to-day interactions. The hypothesis herein is that the term-specific coding results in the recruitment of verbal ability, specifically vocabulary knowledge, that is unrelated to social perspective taking, not simply that more answers are deemed 'correct'. The current study does not speak to this specific hypothesis, but provides support for further exploration by looking directly at the role of verbal ability or vocabulary knowledge as there was a statistically significant difference between scores.

Looking at the second two goals in conjunction, and in line with the hypotheses that both forced-choice response format and term-specific coding influence task performance in different ways, there was a hierarchy across the four scores, suggestive of an additive effect of response format and coding. Per the original hypotheses, a forced-choice response format should *ease* the task burden by providing individuals with additional means to 'solve' the task and thus result in higher scores. Additionally, term-specific coding is assumed to *limit* performance, or increase the task burden, by placing additional task demands on the participant, most notably in the form of verbal ability or vocabulary knowledge. Different combinations of the response format and coding would thus be expected to result in different scores, as was shown herein: Forced-choice response format coupled with valence-based coding resulted in the highest average score followed by the forced-choice response format coupled with term-specific coding then the open-ended response format coupled with valence-based coding and finally the open-ended response format coupled with term-specific coding.

The significant ordinal interaction between coding and response format does suggest that the effects of coding were also more prominent in the forced-choice response format. This should not be surprising as the more options that are correct, the more likely it is that individuals will select a correct answer, intentionally or not. Although only a few items had same-valence foils, when looking at those, 50-75% of the answers were deemed correct for the forced-choice, valence-based *Eyes Task*, significantly increasing the likelihood that even a guess would result in a correct answer. Therefore, although the ordinal interaction was significant, it is likely not of equal practical value as the main effects of coding and response format. That is, the main effects do inform on the effects of coding and response format; however, the significant ordinal interaction may simply reflect the higher likelihood of individuals selecting a correct answer (even unintentionally) when the choices are limited. Taken together, the effects were in line with Hypotheses 3 and 4; however, as response format was assessed using a between-subjects design whereas coding was assessed using a within-subjects design, it is difficult to directly compare the effects. For that, it would be preferable to see how these interact using a within-subjects design for all elements, something that is employed in Study 2.

# 3 Study 2

## 3.1 Introduction

The first study provided preliminary results to help answer the three main questions of interest herein. With respect to the first question regarding affective vs. cognitive perspective taking, we saw that the forced-choice *Eyes Task* was unrelated to dispositional empathy though the open-ended *Eyes Task* was significantly correlated to dispositional empathy, at least for boys. This provides some support for the open-ended *Eyes Task* as a measure of affective perspective taking. However, by design, Study 1 did not provide any information on the degree to which any of the measures assessed *cognitive* perspective taking. Regarding the second question on the differences in response format, Study 1 provided evidence that response format significantly impacted measurement with the forced-choice format associated with greater performance than the open-ended format, though no conclusions could be made regarding why this was the case. Finally, with respect to the effect of coding on performance, valence-based coding provided higher scores than term-specific coding, regardless of response format. However, it was unclear if this was simply due to more answers being deemed correct or, as hypothesized, it was related to verbal ability.

Study 2 was designed to build upon Study 1 using a within-subject methodology to maximize power and remove any concerns arising from a between-subjects design. Being able to make within-subjects comparisons will help remove any questions of possible demographic or participant-level variables influencing the results as well as increase the power to detect relationships.

With respect to the first question of interest—what are the measures assessing—Study 2 included a measure of *cognitive* perspective taking by which to compare performance of the forced-choice and open-ended *Eyes Task* and to determine which may be considered a cognitive perspective taking task specifically. Given both the face validity of the open-ended *Eyes Task* as an affective perspective taking task (i.e., the participants are identifying emotional states using facial stimuli) and the relationship to dispositional empathy in Study 1, I hypothesized that the open-ended *Eyes Task* would *not* be related to cognitive perspective taking. In contrast, I hypothesized that the forced-choice *Eyes Task* would be related to cognitive perspective taking as it has previously distinguished children with Autism Spectrum Disorders from those without and as it was unrelated to dispositional empathy (Study 1), it seems it was not differentiating these groups based on affective perspective taking. Given the dual-deficit of perspective taking in Autism Spectrum Disorders, it is thus likely that it is the assessment of cognitive perspective taking that is discriminating between children with and without Autism Spectrum Disorders (e.g., Baron-Cohen, Wheelwright, Hill, et al., 2001; Baron-Cohen, Wheelwright, Spong, et al., 2001; Happé, 1994).

The second question of interest—the effect of response format on performance—can also be best assessed using a within-subjects design. Individuals' performance on the open-ended *Eyes Task* can be directly compared to their performance on the forced-choice *Eyes Task*, as well as their relationships to the other variables of interest, including a proxy for cognitive ability. It is hypothesized that the forced-choice response format enables alternate means of solving the task, which would be indicative of other cognitive processes. In order to further assess this, a measure of verbal analogies will be used as a proxy for cognitive ability and whether

performance based on response format is linked to verbal analogies will either support or challenge this hypothesis.

The addition of a measure of verbal ability (and specifically vocabulary knowledge) in Study 2 will also provide additional insight into the third question of interest—the effect of coding. The hypothesis is that term-specific coding not only restricts performance via limiting the number of correct answers, but also by requiring a certain level of vocabulary knowledge. Term-specific coding not only requires individuals to understand the very specific options, but also may require a certain threshold of vocabulary knowledge more generally. The threshold element refers to the fact that understanding the nuances between similarly-valenced words may require a greater vocabulary understanding. For example, understanding the distinction between "afraid" and "nervous" requires a greater understanding of the degree of fear which comes from a more general vocabulary understanding. To assess this, a measure of verbal ability will be used (which includes a vocabulary subscale) and relationships to the measures based on the type of coding (i.e., valence-based versus term-specific) will be made within individuals to determine if the hypothesis is tenable.

Taken together, Study 2 will expand upon the findings in Study 1. Study 2 will provide greater information about what type(s) of perspective taking is being assessed, the influence of response format, and whether coding relates to verbal ability.

## 3.2 Methods

### 3.2.1 Ethics Statement

All children provided verbal assent and parents provided written consent. The study was approved by the Behavioural and Research Ethics Board of the University of British Columbia.

### 3.2.2 Participants

Seventy-seven typically-developing children (43 boys) from Vancouver, Canada participated in the K.I.D. Studies Centre at the University of British Columbia. The children ranged in age from 5- to 12-years (M=8.04 years, SD=1.59 years). Data from sixteen additional children were excluded due to the participants' low English proficiency because English was not their native language (4), failure to complete the tasks (3), or missing data for coding purposes (9).

### 3.2.3 Measures

#### 3.2.3.1 Perspective Taking

*Reading the Mind in the Eyes, Children's Version* (Baron-Cohen, Wheelwright, Spong, et al., 2001). See Study 1. In the current study, only half of the items were given in this format so that the other half could be given in the open-ended format.

*Open-Ended Eyes Task* (Cassels & Birch, 2014; Carey & Cassels, 2013). See Study 1. In the current study, only half of the items were given in this format so the other half could be given in the forced-choice format. One quarter of responses were coded by another experimenter who was blind to the participant's performance on other measures (κ=.90, combined with Study 3).

*Happé's Strange Stories* (Happé, 1994). This task involves answering questions about the non-literal statements made by characters in short stories. Participants are asked two questions for each story: "Is what the person said true?" and, "Why did they say it?" Correctly answering these questions requires that the participant understands the character's mental states. Answers are coded based on how well the child's answer reflects an awareness of the mental state behind the statement. The 12 different story types include: lie, white lie, joke, pretend,

misunderstanding, persuade, appearance/reality, figure of speech, sarcasm, forget, contrary emotions, and double bluff. Only one of the stories pertains to emotional states, and it involves reasoning about the situations that lead to contrary emotions; it does not involve emotion recognition. To reduce the burden on participants, only one of each type of story was used (12 of the original 24)[2]. The stories have been found to reliably discriminate amongst children with known mental-state identification deficits, such as those associated with Autism Spectrum Disorders (Happé, 1994). The child's answers for each story were scored on a 3-point scale with 0 being awarded for an incorrect answer, 1 point for a partially-correct answer, and 2 points for a fully-correct answer. The participant's total score (i.e., the sum of the score for each story, out of 24) was used for analyses. One-fifth of responses were coded by another experimenter who was blind to the participant's performance on other measures ($\kappa$=.81, combined with Study 3).

#### 3.2.3.2 Verbal Ability

*Verbal Ability.* To assess verbal ability, the Verbal Comprehension test of the Woodcock-Johnson III (WJ-III) (Woodcock, McGrew, & Mather, 2001) was administered. Despite its label, the Verbal Comprehension test is comprised of four subtests measuring expressive lexical knowledge (e.g., Can you tell me what [picture] is called?), synonyms (e.g., Tell me another word for _____), antonyms (e.g., What is the opposite of _____?), and analogies (e.g., Finish the sentence – A is to B as C is to _____). In the current study, the overall score ("Verbal Ability", which is comprised of the participant's performance on the four subtests) will be used as will two of the subtest scores: *Expressive lexical knowledge* will be used as the measure of vocabulary knowledge. *Analogies* will serve as a proxy for cognitive abilities independent of vocabulary

[2] The author of the original measure and members of her lab have used one of each type of story in her research as well (Booth, personal communication 2006).

knowledge as the two subscales only shared 10% variance and analogies often involve other cognitive abilities above and beyond verbal ability.

### 3.2.4 Procedure

Participants were first administered either the two versions of the *Eyes Task* or the Strange Stories, counterbalanced, as there was no reason to believe that either measure would influence performance on the other. Within the versions of the *Eyes Task*, the open-ended *Eyes Task* always preceded the forced-choice *Eyes Task* so that participants could not simply echo terms provided in the forced-choice format. Previous work has shown that performance on forced-choice tasks predicts performance on open-ended tasks when administered first (Harrigan, 1984) and thus it seemed prudent to ensure that that the forced-choice *Eyes Task* would not influence the open-ended *Eyes Task*. Participants were then given the Verbal Comprehension test of the WJ-III.

## 3.3 Results

### 3.3.1 Preliminary Analyses

As this was a within-subjects design, there was no need to run any preliminary group comparisons, however, analyses examining the relationships between age and sex and all other variables were conducted to determine if they needed to be controlled for in other analyses. Similarly, comparisons between verbal ability and the outcome measures (i.e., the four scores of the *Eyes Task* and cognitive perspective taking) were of interest to determine if verbal ability needed to be controlled for in other analyses.

There were no sex differences on any of the variables of interest, all *p's* = .122 to .942, nor were there any sex by age interaction effects on the variables of interest, all *p's* = .236 to

.520. Age was significantly correlated with three of the four components of the verbal ability subscale and the overall verbal ability score, *r's*=.551 to .706, *p's* < .001. Surprisingly it was the vocabulary subscale that age was not significantly related to, *r*=.187, *p*=.106. Age was also significantly related to the measure of cognitive perspective taking (*Happé's*), the forced-choice, term-specific *Eyes Task*, and the forced-choice, valence-based *Eyes Task*. Age was not significantly related to both the open-ended, valence-based *Eyes Task*, and the open-ended, term-specific *Eyes Task*. See Table 7 for the correlations. As age was significantly related to several variables of interest, it was included in further analyses to control for any age-related changes. Verbal ability and all subscores of the Verbal Comprehension test were significantly related to cognitive perspective taking, *r*=.359 (vocabulary), .402 (analogies), .457 (antonyms), .510 (synonyms), and .548 (combined), all *p's* ≤ .001. Therefore verbal ability and age were controlled for in all analyses including cognitive perspective taking.

**Table 7.** Correlations Between Age and the Perspective Taking Scores.

| | **Pearson Correlation Coefficient with Age** | **Significance (p-value)** |
|---|---|---|
| *Happé's Strange Stories* (cognitive perspective taking) | .424 | <.001 |
| Forced-choice, term-specific *Eyes Task* | .319 | .005 |
| Forced-choice, valence-based *Eyes Task* | .287 | .012 |
| Open-ended, term-specific *Eyes Task* | .094 | .419 |
| Open-ended, valence-based *Eyes Task* | .036 | .757 |

### 3.3.2 Relationships Between Cognitive Perspective Taking and the Versions of the Eyes Task

To determine if cognitive perspective taking is being assessed in any of the measures, bivariate correlations with *Happé's Strange Stories*, the measure of cognitive perspective taking, were conducted. The forced-choice *Eyes Task* was significantly correlated with cognitive perspective taking, regardless of coding, $r$=.338, $p$=.003 (term-specific) and $r$=.331, $p$=.004 (valence-based), whereas the open-ended *Eyes Task* was not, $r$=.165, $p$=.153 (term-specific) and $r$=.089, $p$=.442 (valence-based). This lends support to the idea that the forced-choice *Eyes Task* was assessing cognitive perspective taking and the open-ended *Eyes Task* was not; however, cognitive perspective taking was highly related to age and verbal ability and the forced-choice *Eyes Task* was also significantly correlated with age. Although a certain degree of verbal ability is expected to be present in the assessment of cognitive perspective taking (see Section 1.5.3), a measure that only overlaps with verbal ability or age could not be said to be assessing cognitive perspective taking.

To control for the influence of age and verbal ability on the cognitive perspective taking scores, a linear regression was run predicting cognitive perspective taking from age and verbal ability (using the combined verbal ability score) and the unstandardized residual was saved and served as the measure of cognitive perspective taking *unrelated* to age or verbal ability. As expected, this variable was not significantly related to any subscore of verbal ability ($r$'s=.028 to .096, $p$'s=.407 to .810), the full verbal ability score ($r$=.000), and age ($r$=.000), but remained highly correlated with the original cognitive perspective taking score, $r$=.832, $p$<.001. However, when correlated with the forced-choice *Eyes Task*, the relationship was non-significant (regardless of coding), $r$=.092, $p$=.427 (term-specific) and $r$=.127, $p$=.273 (valence-based),

suggesting the significance above was due to shared variance with age and verbal ability. See Table 8 for a summary of results. Based on these analyses, it would seem neither the forced-choice *Eyes Task* nor the open-ended *Eyes Task* accurately assesses cognitive perspective taking.

**Table 8.** Correlations Between Cognitive Perspective Taking and the Scores for the *Eyes Tasks*.

| | ***Eyes Task* Scores** | | | |
|---|---|---|---|---|
| | **Forced-choice, term-specific** | **Forced-choice, valence-based** | **Open-ended, term-specific** | **Open-ended, valence-based** |
| ***Happé's Strange Stories*** | r=.338 (p=.003) | r=.331 (p=.004) | r=.165 (p=.153) | r=.089 (p=.442) |
| ***Happé's Strange Stories*, verbal ability and age removed** | r=.092 (p=.427) | r=.127 (p=.273) | r=.143 (p=.217) | r=.074 (p=.526) |

### 3.3.3 Performance by Response Format

In order to determine if response format influences performance, comparisons of performance based on response format were conducted using paired-samples t-tests. When performance by response format was compared (controlling for coding scheme), the forced-choice *Eyes Task* was consistently easier than the open-ended *Eyes Task*, $t(75)=19.376$, $p<.001$ (valence-based) and $t(75)=31.410$, $p<.001$ (term-specific). See Table 9 for scores. This provides initial support for the idea that forced-choice response formats are easier than their open-ended counterparts.

**Table 9.** Percentage Correct by Response Format.

| **Term-Specific Coding** | | **Valence-Based Coding** | |
|---|---|---|---|
| **Forced-Choice** | **Open-Ended** | **Forced-Choice** | **Open-Ended** |
| 62.88% | 10.24% | 68.80% | 35.71% |

### 3.3.4 Testing the Role of Alternate Cognitive Abilities

Given the significant difference between response formats, regardless of coding, the question is *why* it is easier. Although there is likely some influence of response format, in that having terms available to choose from negates the need to generate terms, I hypothesized that the better performance on the forced-choice *Eyes Task* is also because a force-choice format, unlike an open-ended format, allows for the use of other cognitive abilities to 'solve' the task (e.g., process of elimination). To determine if the hypothesis that other cognitive abilities influenced performance on the forced-choice *Eyes Task* is tenable, bivariate correlations with verbal analogies (as the proxy for cognitive ability, see Section 3.2.3.2) were conducted and compared between the two response formats (controlling for coding scheme).

The bivariate correlations revealed that the measure of verbal analogies was significantly related to the forced-choice *Eyes Task*, but not the open-ended *Eyes Task*. See Table 10 for the correlations. In order to compare the correlation coefficients, Steiger's (1980) method is used with dependency included as the sample is dependent. The correlations between the forced-choice and open-ended *Eyes Task* were significantly different from each other regardless of coding, $z$=2.62, $p_{1\text{-}tail}$=.004 (term-specific) and $z$=2.26, $p_{1\text{-}tail}$=.018 (valence-based). This lends support to my hypothesis that cognitive abilities other than perspective taking abilities may be used in a forced-choice response format, resulting in the potential for greater, but arguably inflated, scores that could mask participants' deficits in perspective taking.

**Table 10.** Correlations Between the *Eyes Tasks* and Verbal Analogies.

| **Term-Specific Coding** | | **Valence-Based Coding** | |
|---|---|---|---|
| **Forced-Choice** | **Open-Ended** | **Forced-Choice** | **Open-Ended** |
| r=.407[a] | r=.055[b] | r=.361[a] | r=.066[b] |
| p<.001 | p=.637 | p<.001 | p=.572 |

Note: [ab]denotes significant differences between the correlations

### 3.3.5 Performance by Coding Scheme

To determine if coding influences performance, paired-samples t-tests comparing scores by type of coding were conducted (controlling for response format). There were significant differences based on coding scheme, with the valence-based coding resulting in higher scores than the term-specific coding, $t(75)=10.320$, $p<.001$ (forced-choice *Eyes Task*) and $t(75)=20.979$, $p<.001$ (open-ended *Eyes Task*). See Table 11 for scores. This is in line with the results of Study 1, but as mentioned in Study 1, it was impossible to tell if valence-based coding increases performance artificially by allowing more correct answers or if, as hypothesized, it is because it is less tied to a specific vocabulary knowledge requirement.

**Table 11.** Percentage Correct by Coding Scheme.

| **Forced-Choice *Eyes Task*** | | **Open-Ended *Eyes Task*** | |
|---|---|---|---|
| **Term-Specific** | **Valence-Based** | **Term-Specific** | **Valence-Based** |
| 62.88% | 68.80% | 10.24% | 35.71% |

### 3.3.6 Testing the Role of Vocabulary Knowledge

In order to determine if the differences found by coding scheme were due to differences in vocabulary knowledge or verbal ability, further analyses were conducted looking at the relationships between performance by coding scheme (controlling for response format) and both vocabulary knowledge and the overall verbal ability subscale. If vocabulary knowledge or verbal ability plays a role in term-specific coding, one would expect to see significantly greater relationships between these variables and scores for term-specific coding than for valence-based coding in the *Eyes Tasks*.

To determine if vocabulary or verbal ability is linked to coding scheme, two types of analyses were performed. First, bivariate correlations were conducted and compared to

determine if the relationships to vocabulary knowledge and verbal ability were greater for term-specific coding than valence-based coding. Second, both types of coding were entered into a linear regression predicting vocabulary or verbal ability to determine if one type of coding was a greater predictor of these abilities, above and beyond the effect of the other.

For the forced-choice *Eyes Task*, both types of coding were related to both vocabulary knowledge and verbal ability. See Table 12 for correlations. Term-specific coding had a greater association with both vocabulary knowledge and verbal ability and the difference between the correlations was statistically significant, $z$=.165, $p_{1\text{-}tail}$=.049 (vocabulary knowledge) and $z$=1.917, $p_{1\text{-}tail}$=.028 (verbal ability). For the open-ended *Eyes Task*, both types of coding were not significantly related to vocabulary knowledge and verbal ability. See Table 12. The relationships with vocabulary are in the hypothesized direction, with the term-specific coding leading to a more positive relationship, but the relationship is still not significant or significantly different from that when valence-based coding is used, $z$=1.036, $p_{1\text{-}tail}$=.150. The relationship to verbal ability more generally is not in line with the current hypothesis that it is the coding scheme that would lead to significant relationships with verbal ability as the relationships are nearly identical, regardless of coding scheme.

**Table 12.** Correlations Between the *Eyes Tasks* and both Vocabulary Knowledge and Verbal Ability.

| ***Eyes Task* Score** | **Vocabulary Knowledge** | **Verbal Ability** |
|---|---|---|
| Forced-choice, term-specific | r=.355**[,a] | r=.474***[,d] |
| Forced-choice, valence-based | r=.293***[,b] | r=.406***[,e] |
| Open-ended, term-specific | r=.013[c] | r=.066[f] |
| Open-ended, valence-based | r=-.122[c] | r=.055[f] |

Notes: *p<.05, **p<.01, ***p<.001, [a,b,c,d,e,f]denotes significant difference between correlations

When predicting vocabulary knowledge scores from the different scores for the forced-choice *Eyes Task*, betas could not be compared due to the high collinearity between the two

coding schemes; therefore, two regressions were conducted alternating which variable was included in Step 1 and Step 2 to determine if one significantly predicts vocabulary knowledge above and beyond the other by looking at $R^2$ change. The first regression put the forced-choice, term-specific *Eyes Task* in Step 1 and the forced-choice, valence-based *Eyes Task* in Step 2; the forced-choice, term-specific *Eyes Task* was a significant predictor as expected, but the forced-choice, valence-based *Eyes Task* did not add any significant variance. See Table 13 for full results. In contrast, when the forced-choice, valence-based *Eyes Task* was entered in Step 1 and the forced-choice, term-specific *Eyes Task* was added in Step 2, the forced-choice, term-specific *Eyes Task* predicted a significant amount of variance in vocabulary knowledge above and beyond that of the forced-choice, valence-based *Eyes Task*. The pattern was identical for verbal ability. When the open-ended response format was tested, nothing was significant, regardless of which coding scheme was entered first.

**Table 13.** Regressions Predicting Vocabulary Knowledge and Verbal Ability.

| **Predicting Vocabulary Knowledge** | | |
|---|---|---|
| | **$\Delta R^2$** | **p-value** |
| Step 1: Forced-choice, term-specific *Eyes Task* | **.126** | .002 |
| Step 2: Forced-choice, valence-based *Eyes Task* | .016 | .251 |
| Step 1: Forced-choice, valence-based *Eyes Task* | **.086** | .010 |
| Step 2: Forced-choice, term-specific *Eyes Task* | **.056** | .032 |
| Step 1: Open-ended, valence-based *Eyes Task* | .015 | .293 |
| Step 2: Open-ended, term-specific *Eyes Task* | .004 | .586 |
| Step 1: Open-ended, term-specific *Eyes Task* | .000 | .910 |
| Step 2: Open-ended, valence-based *Eyes Task* | .019 | .241 |

| Predicting Verbal Ability | | |
|---|---|---|
| | **$\Delta R^2$** | **p-value** |
| Step 1: Forced-choice, term-specific *Eyes Task* | **.225** | <.001 |
| Step 2: Forced-choice, valence-based *Eyes Task* | .015 | .235 |
| Step 1: Forced-choice, valence-based *Eyes Task* | **.165** | <.001 |
| Step 2: Forced-choice, term-specific *Eyes Task* | **.075** | .009 |
| Step 1: Open-ended, valence-based *Eyes Task* | .003 | .666 |
| Step 2: Open-ended, term-specific *Eyes Task* | .004 | .604 |
| Step 1: Open-ended, term-specific *Eyes Task* | .006 | .519 |
| Step 2: Open-ended, valence-based *Eyes Task* | .001 | .838 |

Note: $\Delta R^2$ values that are in bold are significant at the <.05 level.

### 3.3.7 Testing the Discrete or Continuous Effect of Vocabulary and Verbal Ability

A final analysis was conducted to help determine if the effects of vocabulary knowledge and verbal ability in the forced-choice *Eyes Task* seen in Section 3.3.5 are continuous or discrete in nature. That is, does an increase in vocabulary and verbal ability continue to influence performance for either coding scheme or is it a matter of reaching a particular cut-off? Although the correlational analyses suggest a continuous relationship, it may be driven by individuals on the low-end of the spectrum. If a threshold is needed, then once it is reached there should be no effect of vocabulary or verbal ability on scores by coding scheme. Supportive of this, when the difference between the forced-choice, term-specific *Eyes Task* and forced-choice, valence-based *Eyes Task* scores is calculated and correlated with vocabulary knowledge and verbal ability, the correlation is significant and negative, $r$=-.320, $p$=.005 (vocabulary knowledge) and $r$=-.366, $p$=.001 (verbal ability). The negative correlation suggests that as vocabulary or verbal ability

increases, the difference in scores between the forced-choice, term-specific *Eyes Task* and forced-choice, valence-based *Eyes Task* decrease.

This is further supported by chi-square analyses on the breakdown of the difference score (which was one of three possibilities: 0, 1, or 2) and whether the participant was above or below the median for vocabulary knowledge or verbal ability. Both chi-squares for vocabulary knowledge, $\chi^2(2)=6.106$, $p=.047$, and verbal ability, $\chi^2(2)=12.119$, $p=.002$, were significant, with children who showed no difference in scores more likely to be above the median whereas any difference was associated with a higher likelihood of being below the median. See Table 14 for the full breakdown of scores. Of note, children who had no difference in their score by coding scheme also performed the best on the forced-choice, term-specific *Eyes Task*, scoring significantly better than those with differences ($M_{0\text{-diff}}=10.41$ vs. $M_{1\text{-diff}}=8.16$ or $M_{2\text{-diff}}=7.54$, as the max difference by coding scheme was 2 points ).

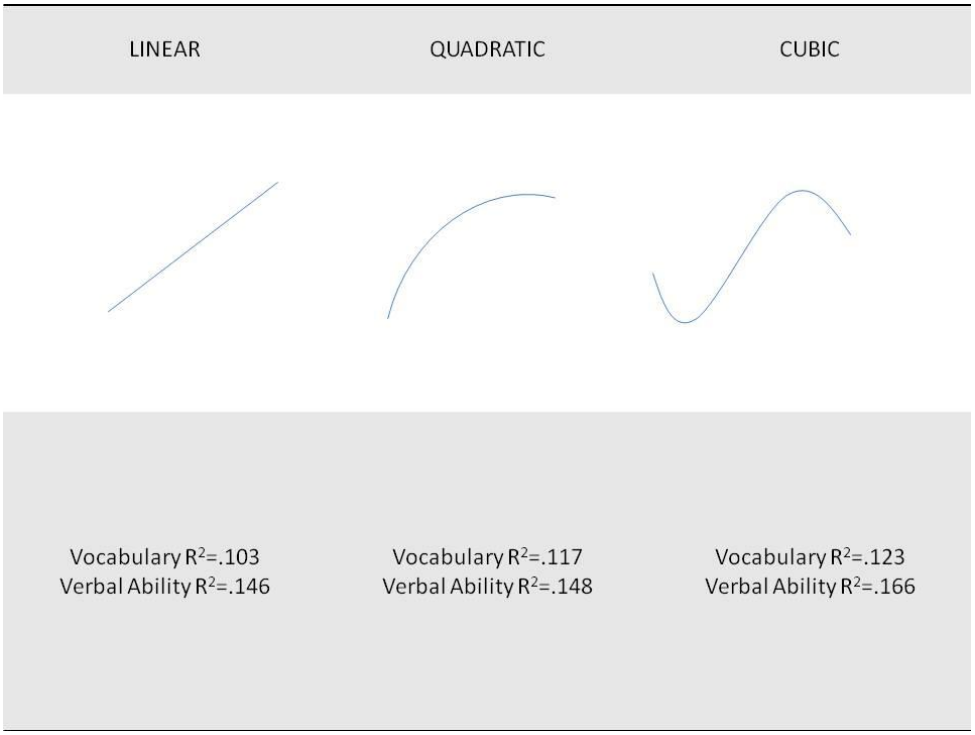
**Table 14.** Vocabulary Knowledge and Verbal Ability Median Splits by Difference Score by Coding Scheme for the Forced-Choice *Eyes Task*.

| | **Difference = 0** | **Difference = 1** | **Difference = 2** |
|---|---|---|---|
| **Vocabulary** | | | |
| Below Median | 10 | 25 | 8 |
| Above Median | 17 | 12 | 5 |
| **Verbal Ability** | | | |
| Below Median | 7 | 23 | 10 |
| Above Median | 20 | 14 | 3 |

The chi-square results would suggest that the relationship is discrete, with some form of a threshold effect for the role of vocabulary knowledge or verbal ability on the relationship between coding scheme and performance in the forced-choice *Eyes Task*; that is, once children attain a particular level of vocabulary knowledge or verbal ability (here, the median proved to be a significant cut-off, though it is possible the cut-off is lower), the effects of coding scheme are

diminished. Further support for this comes when the relationship to the difference scores by coding scheme are compared with vocabulary knowledge and verbal ability using quadratic and cubic functions instead of linear ones. The amount of variance accounted for by vocabulary knowledge and verbal ability in the difference scores when using linear relationships is 10.3% and 14.6%, respectively. When a quadratic fit is used, the amount of variance explained increases to 11.7% and 14.8%, respectively, and the curve suggests a rapid increase for low levels of performance and then a leveling off effect. When a cubic fit is used, there is a further increase to 12.3% and 16.6%, respectively, and the leveling off occurs, but there is also the potential for a decrease in the difference at very high levels of vocabulary. See Figure 2 for the various fits and to see the shape fit to the data.

**Figure 2.** Linear, Quadratic, and Cubic Fits to the Difference between the Forced-Choice, Term-Specific *Eyes Task* and the Forced-Choice, Valence-Based *Eyes Task*.



## 3.4 Discussion

The primary goal of Study 2 was to expand upon the findings of Study 1 by further exploring the three questions of interest—what type of perspective taking is being assessed, how does response format influence performance, and how does coding influence performance? With respect to the first question of interest, the forced-choice and open-ended *Eyes Task* were compared to an open-ended measure of cognitive perspective taking. This measure was chosen for its open-ended nature and use of 'gist' coding (akin to valence-based coding), which is different from the *Reading the Mind in the Eyes* task, ensuring any relationship was not due to overlapping methodological variance. The open-ended *Eyes Task* was not significantly related to the measure of cognitive perspective taking. Taken with the findings in Study 1, in which the open-ended *Eyes Task* was linked to dispositional empathy, a known outcome associated with affective perspective taking, it seems that the open-ended *Eyes Task* is likely measuring a form of affective, but not cognitive, perspective taking in children.

Precisely what the forced-choice *Eyes Task* is assessing in children, however, remains an open question. Although the forced-choice *Eyes Task* was related to cognitive perspective taking in a bivariate analysis, when verbal ability and age were controlled for, the relationship was not only non-significant, but fell close to zero, raising questions of the validity of the task as a measure of purely cognitive perspective taking. Given that the *Reading the Mind the Eyes* task has been found to discriminate between typically-developing children and children with Autism Spectrum Disorders (Baron-Cohen, Wheelwright, Hill, et al., 2001; Baron-Cohen, Wheelwright, Spong, et al., 2001), one must wonder if that relationship is due to other, non-perspective taking variables, such as verbal or cognitive abilities. Regardless, the results thus far do not support the forced-choice *Eyes Task* as a pure measure of either affective or cognitive perspective taking.

The second question of interest pertained to the issue of response format and the possibility that alternate strategies could be used to 'solve' a forced-choice task. Using a within-subjects design, Study 2 provided a stronger comparison of performance based on response format than Study 1. As with Study 1, regardless of coding scheme, the forced-choice *Eyes Task* resulted in significantly higher scores than the open-ended *Eyes Task*. This suggests that either the forced-choice *Eyes Task* can be solved using alternate methods, or the open-ended *Eyes Task* restricts performance. It is difficult to imagine how the open-ended *Eyes Task* restricts performance, but the forced-choice format may elicit the use of other methods. Results looking at the relationship between response format and a proxy for cognitive ability (verbal analogies) found significant relationships between the proxy and performance on the forced-choice *Eyes Task*, but not the open-ended *Eyes Task*. Furthermore, the relationships to cognitive ability were statistically different between the forced-choice and open-ended *Eyes Task* (regardless of coding scheme). These findings lend support to the hypothesis that alternate cognitive processes or strategies are being used in the forced-choice *Eyes Task*. It is therefore possible that scores from the forced-choice *Eyes Task* may be artificially inflated as individuals are using skills unrelated to perspective taking to solve them. This would also fit with the findings thus far that the forced-choice *Eyes Task* does not seem to be assessing affective or cognitive perspective taking. If children are using other means to solve the task, even the face-validity of the task as a measure of affective perspective taking would be negated.

Finally, the question of how coding influences performance was assessed by comparing performance by coding scheme to vocabulary knowledge and overall verbal ability. The results were only partially in line with the current hypothesis. The hypothesis did not hold up for the open-ended *Eyes Task*, as vocabulary and verbal ability were unrelated to performance,

regardless of coding. One possible reason for this is that the age range of children tested was such that the level of vocabulary knowledge required to detect a difference was too high. This is somewhat supported by the fact that performance on the open-ended, term-specific *Eyes Task* was so low that it bordered on a floor effect. Children simply may not have the type of vocabulary needed at the tip of their tongue to accurately assess the nuanced emotions presented in the task without help in the form of given options; however, when given the opportunity to match valence, they are capable of demonstrating their knowledge, in line with results by other researchers (Widen & Russell, 2008).

In the forced-choice *Eyes Task*, the results were in line with expectations, with vocabulary knowledge and verbal ability being significantly more related to performance when coding was term-specific than when coding was valence-based. Notably, scores using valence-based coding were still significantly related to both vocabulary knowledge and overall verbal ability, but this may be explained by the forced-choice response format used. Specifically, the testing herein was conducted using the actual terms in the question, with just the coding reflecting the acceptance of same-valence foils. The term-specific response format likely requires a certain level of vocabulary knowledge to read and minimally understand the four choices. If valence-based options were provided instead of term-specific ones, I would expect the relationship with vocabulary knowledge and generalized verbal ability to decrease, possibly to non-significance.

Additionally, the results suggest that this effect is partially due to a threshold effect in which children with higher levels of vocabulary are less likely to demonstrate differences by coding scheme and thus a certain vocabulary knowledge or verbal ability threshold can remove their effects. Overall, the results support the hypothesis that coding scheme – specifically term-

specific coding – is related to verbal ability or vocabulary knowledge, but only when the response format is forced-choice. I argue that this suggests that the use of valence-based coding would be more likely to support the measurement of social perspective taking ability in those with lower vocabulary knowledge or verbal ability.

# 4 Study 3

## 4.1 Introduction

Participants in Studies 1 and 2 were typically-developing. The original development of the *Reading the Mind in the Eyes* task (Baron-Cohen et al. 1997; Baron-Cohen, Wheelwright, Spong, et al., 2001) was driven by the need to identify *differences* between clinical and typically-developing populations. Although much of the work on deficits in social functioning has targeted individuals with Autism Spectrum Disorders, who arguably have a dual deficit in affective and cognitive perspective taking and possibly other global cognitive impairments as well, there are other groups who have demonstrated *social* deficits without either more global cognitive impairment or a dual deficit in social perspective taking abilities, such as those with learning disabilities. The hypotheses thus far have been based on several premises that would influence the ability of these tasks to identify individual differences rather than the more severe deficits of those with Autism Spectrum Disorders. Furthermore, the more global cognitive impairment associated with many types of Autism Spectrum Disorders may influence performance on these tasks and the interest herein is in the assessment of the social perspective taking abilities, not the task demands or possible uses of alternate strategies or abilities.

What each task measures, and the success with which it measures it, is paramount to its ability to accurately identify typically-developing from various clinical or at-risk populations. For example, when looking at individuals with Autism Spectrum Disorders, both cognitive and affective perspective taking is impaired and thus measures that assess one or both components may be able to accurately identify these individuals. As mentioned in Section 1.5.1.2, one population for whom affective perspective taking has specifically been found to be at-risk is

individuals with specific learning disorders (both children and adults), with research showing they are impaired in identifying emotional states relative to typically-developing individuals (e.g., Owen, Browning, & Jones, 2001; for a review on children see Tur-Kaspa, 2002); however, no associated deficit in cognitive perspective taking has been found (though for evidence of role playing deficits, which often include both cognitive and affective perspective taking, see, Clegg, Hollis, Mawhood, & Rutter, 2005; Dickstein & Warren, 1980; Wong & Wong, 1980). The current study therefore includes the assessment of a group of children with learning disabilities and who are at-risk for affective perspective taking deficits to examine how the different tasks are able to reliably discriminate between the this group and the typically-developing group from Study 2 after controlling for other variables that might distinguish these two groups.

In addition to being at-risk for affective perspective taking deficits, the group selected herein was deemed to have learning disabilities that included deficits in verbal ability, either diagnosed as a specific learning disorder (which precludes a diagnosis of widespread cognitive disability; American Psychiatric Association, 2013) or via educational assessment at other schools. However, the school itself requires that there is no global cognitive impairments for registration in the school and thus no fear that this may influence findings. The preclusion of any widespread cognitive disability would suggest that these children should be able to use other cognitive means to solve tasks, and as such, the hypothesis that response format can recruit alternate means can be tested herein. The presence of a verbal ability deficit allows for an examination of the role of vocabulary knowledge or verbal ability in coding by seeing how a deficit in verbal ability may influence performance. Thus, this group of children provides further means to test the three main questions of interest and the associated hypotheses.

To these ends, children from a local school for those with learning disabilities were recruited to take part in the study. The exact same measures as in Study 2 were included. Based on the original hypotheses, the open-ended *Eyes Task*, but not the forced-choice *Eyes Task*, should reliably discriminate between the group herein and the typically-developing group from Study 2, as the open-ended *Eyes Task* is hypothesized to be a better measure of affective perspective taking specifically (with the results from Studies 1 and 2 lending support to this hypothesis). Additionally, the open-ended *Eyes Task* is hypothesized to be unrelated to cognitive perspective taking in the current sample, as it was in the typically-developing sample. I had initially hypothesized that the forced-choice *Eyes Task* would be a better measure of cognitive perspective taking, but the results from Study 2 did not fully support this and thus I am unsure what to expect with this sample.

The lack of widespread cognitive disability would suggest that the current sample should perform as well on the forced-choice *Eyes Task* as the typically-developing sample, age- and/or verbal ability-related changes notwithstanding. Differences in response format are hypothesized to be due to the use of alternate strategies that this sample *should* be able to employ alternate strategies as there are no widespread cognitive deficits in this sample; thus, the expectation is to see the same effect of alternate strategies as was seen in Study 2. The findings from Study 2 also suggested that verbal ability and vocabulary knowledge are related to performance when a forced-choice response format is used (regardless of coding) and as the current sample is known to have verbal deficits, this can be explored further as we might expect worse performance in this group based on verbal deficits. Finally, assuming verbal ability or vocabulary knowledge differences between the clinical and typically-developing groups, if the original hypothesis linking vocabulary knowledge and type of coding is supported, differences in performance based

on coding should be explained by differences in verbal ability, as was found for the forced-choice response format in Study 2.

## 4.2 Methods

### 4.2.1 Ethics Statement

All children provided verbal assent and parents provided written consent prior to participation in the study. The study was approved by the Behavioural and Research Ethics Board of the University of British Columbia.

### 4.2.2 Participants

Fifty-four students (6- to 13-years old; $M_{age}$=11.22 years, SD=1.68 years; 29 boys) from a local school for children with learning disabilities were recruited. For inclusion children had to have average intelligence (a requirement of the school), have not had any brain injuries (a requirement of the school), have no severe behavioural problems, but have a verbal learning disability and show deficits in social-emotional processing as assessed by school officials. For the purpose of the study herein, the school identified the students that fit these criteria without providing the specific information, such as the details of any specific learning disorder or results of educational assessments. However, school officials informed that for the labeling of the learning disability, children were assessed either at a previous school as part of an educational assessment or by a clinical psychologist privately (thus a diagnosis of a specific learning disorder) and thus the learning disability could not be due only to poor grades in a given subject. Similarly, although there was an awareness of no severe behavioural problems, it is possible that children had comorbid internalizing problems, such as depression or anxiety, which was not assessed herein.

The identification of social-emotional processing deficits was done using a non-standardized measure developed by school officials and administered to all children in the school upon entry and continuously as they worked on these skills. The ratings for the children in this study were their most recent assessment, not their assessment at intake. The measure required the children to write about multiple static images depicting emotional states and scenarios. The children were asked to describe: a) how the person is feeling, and b) what might cause the individual to feel this way (using elements of the image to explain). Children's answers were coded for the correctness of the emotion label given and the ability to provide logical and rational contexts for the particular emotions. A final score pertaining to the degree of deficit was based on a coding scheme devised by school officials and provided children with a qualitative score ranging from "no deficit" to "severe deficits". In the current study, no participant was rated as having 'no deficit', and the participants with the *lowest* level of social weaknesses were rated as having "mild-moderate" deficits (n=7), with the majority being rated as having between moderate and severe socio-emotional deficits. Although I was given the chance to look at the measure, I was not given permission to replicate the measure herein. As the measure was not standardized, the children herein will only be referred to as being "at-risk" for affective perspective taking deficits and thus labeled the "at-risk group".

### 4.2.3 Measures

#### 4.2.3.1 Perspective Taking

*Reading the Mind in the Eyes, Children's Version* (Baron-Cohen, Wheelwright, Spong, et al., 2001). This was identical to the measure in Study 2, except for half of the participants, the

stimuli used was switched so that the half normally used for the open-ended version was used here for the forced-choice version and vice versa.

*Open-Ended Eyes Task* (Cassels & Birch, 2014; Carey & Cassels, 2013). This was identical to the measure in Chapter 3, except for half of the participants, the stimuli used was switched so that the half normally used for the open-ended version was used here for the forced-choice version and vice versa.

*Happé's Strange Stories* (Happé, 1994). See Study 2.

#### 4.2.3.2 Verbal Ability

*Verbal Ability*. See Study 2.

### 4.2.4 Procedure

The procedure was identical to that of Study 2.

## 4.3 Results

### 4.3.1 Preliminary Analyses

In order to ensure that performance did not differ based on which stimuli set was used for which measure, independent samples t-tests were conducted on all four possible scores (the forced-choice, term-specific *Eyes Task*; the forced-choice, valence-based *Eyes Task;* the open-ended, valence-based *Eyes Task*; and the open-ended, term-specific *Eyes Task*) with the stimuli version as the grouping variable. There were no significant differences in either of the forced-choice *Eyes Tasks* or the open-ended, valence-based *Eyes Task*, all $p's > .08$, but there was a significant difference for the open-ended, term-specific *Eyes Task*, $t(52) = 3.087$, $p = .003$, with participants who received the new, switched stimuli performing worse than those who received

the original stimuli. Notably, however, the scores for the open-ended, term-specific *Eyes Task* were so low, regardless of condition, that a floor effect was likely (M=5.8% or 11%). Because of this, analyses for the open-ended, term-specific *Eyes Task* will be included for completion's sake, but with the caveat that the results are unlikely to be stable and the measure is likely too difficult a task to use with children, particularly these children at-risk for affective perspective taking deficits and with a verbal deficit.

When examining the variables by sex, there were no sex differences on any variable (including all perspective taking measures and all verbal ability subscales and full-scale score), all *p's*>.13. Age was significantly correlated with vocabulary ($r$=.418, $p$=.002) and verbal ability ($r$=.336, $p$=.013), and was nearly significantly correlated with the open-ended, valence-based *Eyes Task* ($r$=-.233, $p$=.089), but was not significantly related to all other variables.

The measure of cognitive perspective taking was related to all verbal ability subscales (all *p's*<.004) and full-scale score ($r$=.639, $p$<.001). These relationships were very strong, with the global score of verbal ability accounting for over 40% of the variance in cognitive perspective taking, yet the relationship between age and cognitive perspective taking within this group was non-significant, $r$=-.003, $p$=.982, in line with previous research that a certain level of verbal ability, not age, is necessary to successfully pass cognitive perspective taking tasks (Happé, 1995). Verbal ability is thus controlled for when looking at relationships to cognitive perspective taking.

The final set of preliminary analyses compared this at-risk group to the typically-developing group from Study 2 as several later analyses will include comparisons between the two groups. The two groups differed significantly on age, with the at-risk group being significantly older than the typically-developing group, $t(129)$=10.709, $p$<.001, and thus age was

controlled for in all further analyses. Despite this significance, the two groups did not differ on vocabulary, $t(129)=.844$, $p=.400$, verbal analogies, $t(129)=.098$, $p=.922$, or verbal ability more generally, $t(129)=1.038$, $p=.301$, confirming a degree of verbal deficit in this at-risk group. The distribution of males and females was similar between the two groups, $\chi^2=.034$, $p=.854$.

### 4.3.2 Relationships Between Cognitive Perspective Taking and the Versions of the *Eyes Task* Within the At-Risk Group

The first analysis was comprised of correlational analyses between the four *Eyes Task* scores and the measure of cognitive perspective taking. As previously mentioned, the cognitive perspective taking task was strongly and significantly related to verbal ability and thus a score factoring out the variance predicted by verbal ability was used as the measure of cognitive perspective taking. When these scores were correlated in the at-risk group, the measure of cognitive perspective taking was significantly correlated with all four scores. See Table 15 for a summary.

**Table 15.** Correlations Between the *Eyes Tasks* and Cognitive Perspective Taking.

| | ***Eyes Task* Version** | | | |
|---|---|---|---|---|
| | **Forced-choice, term-specific** | **Forced-choice, valence-based** | **Open-ended, term-specific** | **Open-ended, valence-based** |
| *Happé's Strange Stories* | $r=.338$, $p=.013$ | $r=.341$, $p=.012$ | $r=.357$, $p=.008$ | $r=.439$, $p=.001$ |

For this at-risk group, then, the forced-choice *Eyes Task* may be measuring some form of cognitive perspective taking. However, the relationship to the open-ended *Eyes Task* was unexpected and one statistical possibility for this is that it is due to shared method variance. The cognitive perspective taking measure is both open-ended and coding is akin to valence-based coding (i.e., the gist is all that is necessary to 'solve' the problem), thus the correlation may be

due to the overlap in methodology; yet, as this was not identified in the typically-developing population, it is unclear why it would present itself here. Thus, a non-statistical explanation is more likely.

### 4.3.3 Comparing Performance on the *Eyes Task* Scores between Groups

The second analysis compared performance on the perspective taking tasks between the two groups (i.e., the at-risk group and the typically-developing group from Study 2) using independent samples t-tests. For the forced-choice *Eyes Task*, there were no significant differences between the groups, $t(128)=.317$, $p=.752$ (term-specific), $t(128)=.755$, $p=.451$ (valence-based). In contrast, there were significant differences on the cognitive perspective taking measure, $t(128)=2.748$, $p=.007$, and the open-ended, valence-based *Eyes Task*, $t(128)=4.410$, $p<.001$, though not the open-ended, term-specific *Eyes Task*, $t(128)=1.406$, $p=.162$. See Table 16 for group means.

**Table 16.** Means for the At-Risk and Typically-Developing Groups on the Perspective Taking Tasks.

| | **At-Risk Group** | **Typically-Developing Group** |
|---|---|---|
| Forced-choice, term-specific *Eyes Task* | 8.93 | 8.99 |
| Forced-choice, valence-based *Eyes Task* | 9.89 | 9.76 |
| Open-ended, term-specific *Eyes Task* | 1.19 | 1.43 |
| Open-ended, valence-based *Eyes Task* | 3.76** | 5.00** |
| *Happé's Strange Stories* | 13.57* | 11.78* |

Note: *p<.01, **p<.001

Notably, for cognitive perspective taking, the at-risk group *outperformed* the typically-developing group, which is likely due to their increased age relative to the typically-developing group. This provides further support for the idea that the at-risk group has no cognitive

perspective taking deficits. On the other hand, the difference for the open-ended, valence-based *Eyes Task* favoured the typically-developing group, in line with the hypothesis that the open-ended *Eyes Task* assesses affective perspective taking (and not cognitive perspective taking) and can identify any affective perspective taking deficits in the at-risk group.

To confirm that the difference in performance in the open-ended, valence-based *Eyes Task* was significant above and beyond any other group difference, a hierarchical binary logistic regression was conducted predicting group membership while controlling for other variables of interest. As age and cognitive perspective taking were significantly different between the groups, they were entered in the first step to control for their predictive power. The forced-choice, term-specific *Eyes Task* and open-ended, valence-based *Eyes Task*[3] were entered in the second step to compare their ability to predict group membership. If, as hypothesized, the open-ended, valence-based *Eyes Task*, but not the forced-choice, term-specific *Eyes Task*, is assessing affective perspective taking, then it is expected that only the open-ended, valence-based *Eyes Task* will predict group membership above and beyond the effects of age and cognitive perspective taking.

Prior to any variable being entered, 58.5% of cases were classified correctly into their respective groups. After the inclusion of age and cognitive perspective taking, 84.6% of cases were classified correctly. Age was the only significant predictor, with every month of increased age increasing the likelihood that a participant will be in the at-risk group by 8.6% ($p<.001$). Cognitive perspective taking was non-significant ($p=.994$) suggesting the differences in performance are due to the age-related differences in the two groups. When the two *Eyes Tasks*

[3] The forced-choice, valence-based *Eyes Task* and open-ended, valence-based *Eyes Task* were also compared but the results were nearly identical. When the open-ended, term-specific *Eyes Task* was included instead of the open-ended, valence-based *Eyes Task*, it was non-significant, but it is difficult to interpret given the floor effect and version differences highlighted above.

scores were included, 88.5% of cases were classified correctly, a small increase (only 3.9%), but a significant change to the model nonetheless, $\chi^2(2)=10.082$, $p=.006$. Looking at the predictive power of the individual variables, controlling for the others, age remained significant with every month increase in age increasing the likelihood of being in the at-risk group by 9% ($p<.001$) (i.e., a one-year age difference would result in a 108% increase in the likelihood of being in the at-risk group). Of the remaining variables (cognitive perspective taking, the forced-choice, term-specific *Eyes Task*, and the open-ended, valence-based *Eyes Task*), only the open-ended, valence-based *Eyes Task* was also significant, with a unit *decrease* on the open-ended, valence-based *Eyes Task increasing* the likelihood an individual is in the at-risk group by 62.3% ($p=.009$) (or a unit *increase* on the open-ended, valence-based *Eyes Task decreases* the likelihood an individual is in the at-risk group by 39.4%). This supports the hypothesis that the open-ended, valence-based *Eyes Task* is assessing affective perspective taking whereas the forced-choice, term-specific *Eyes Task* is not and that this is not due to other group differences in age or cognitive perspective taking.

#### 4.3.4 Performance by Response Format Within the At-Risk Group

To determine if there was an effect of response format in this within-subjects methodology, paired samples t-tests were conducted comparing scores by response format (controlling for coding scheme). In line with findings from Studies 1 and 2, the forced-choice response format was associated with greater performance than the open-ended response format, regardless of coding scheme, $t(53)=23.736$, $p<.001$ (term-specific coding) and $t(53)=18.930$, $p<.001$ (valence-based coding). See Table 17 for scores.

**Table 17.** Percentage Correct by Response Format.

| Term-Specific Coding | | Valence-Based Coding | |
|---|---|---|---|
| **Forced-Choice** | **Open-Ended** | **Forced-Choice** | **Open-Ended** |
| 63.76% | 8.47% | 70.64% | 26.85% |

### 4.3.5 Testing the Role of Alternate Cognitive Abilities Within the At-Risk Group

In order to see if the differences found in Section 4.3.4 were a result of alternate strategies being used, performance on the forced-choice and open-ended *Eyes Task* were compared to the measure of verbal analogies, as was done in Study 2. In this at-risk sample, unlike that of the typically-developing sample in Study 2, the verbal analogies score was not significantly related to all *Eyes Tasks* scores, all *p's*>.26. See Table 18.

**Table 18.** Correlations Between the *Eyes Tasks* and Verbal Analogies.

| Term-Specific Coding | | Valence-Based Coding | |
|---|---|---|---|
| **Forced-Choice** | **Open-Ended** | **Forced-Choice** | **Open-Ended** |
| r=.141 | r=.139 | r=.141 | r=.156 |
| p=.310 | p=.315 | p=.308 | p=.261 |

It is possible, however, that the relationship is not linear and rather a threshold applies for the application of alternate strategies in this group. In order to determine if a threshold applies, participants scoring in the top 25% and bottom 25% on verbal analogies were compared on the difference score between the forced-choice and open-ended *Eyes Task* (controlling for coding scheme; e.g., the forced-choice, valence-based *Eyes Task* minus the open-ended, valence-based *Eyes Task*). Independent samples t-tests were conducted and no difference was found between the top and bottom groups (regardless of coding scheme), $t(45)$=.184, $p$=.855 (valence-based) and $t(45)$=.291, $p$=.773 (term-specific). Similarly, when difference scores were correlated with verbal analogies in the entire at-risk sample, they were non-significant, $r$=.015, $p$=.916 (valence-

based) and $r$=.079, $p$=.569 (term-specific). Thus it would seem that either a verbal analogies measure is not a good proxy of the ability to use alternate strategies in this group or they simply are not using alternate strategies.

#### 4.3.6 Performance by Coding Scheme Within the At-Risk Group

In order to compare scores by coding scheme, paired-samples t-tests were conducted within the at-risk group. Coding scheme did have a significant effect on performance, with valence-based coding resulting in greater scores than term-specific coding, regardless of response format, $t(53)$=10.118, $p$<.001 (forced-choice *Eyes Task*) and $t(53)$=13.967, $p$<.001 (open-ended *Eyes Task*). See Table 19.

**Table 19.** Percentage Correct by Coding Scheme.

| **Forced-Choice *Eyes Task*** | | **Open-Ended *Eyes Task*** | |
|---|---|---|---|
| **Term-Specific** | **Valence-Based** | **Term-Specific** | **Valence-Based** |
| 63.76% | 70.64% | 8.47% | 26.85% |

#### 4.3.7 Testing the Role of Vocabulary Task Demands Within the At-Risk Group

In order to examine the role of vocabulary knowledge or verbal ability on coding, correlations between the four scores of the *Eyes Tasks* and both vocabulary knowledge and verbal ability were conducted and compared. For the forced-choice *Eyes Task*, verbal ability was significant correlated with both coding schemes, $r$=.295, $p$=.031 (term-specific) and $r$=.274, $p$=.045 (valence-based) whereas correlations between vocabulary knowledge and both coding schemes approached significance, $r$=.239, $p$=.082 (term-specific) and $r$=.239, $p$=.082 (valence-based). However, the correlations by coding scheme were not significantly different from each other, all $p$'s>.29. For the open-ended response format, there were no significant correlations, all $p$'s>.23. See Table 20 for correlations with vocabulary knowledge. When the difference

between the types of coding (controlling for response format; e.g., the forced-choice, valence-based *Eyes Task* minus the forced-choice, term-specific *Eyes Task*) were correlated with vocabulary knowledge and verbal ability, the correlations were non-significant across response format, all *p's*>.19. These findings suggest that verbal ability is playing a role in *response format*, but not *coding scheme* in this at-risk group.

**Table 20.** Correlations Between the *Eyes Tasks* and Vocabulary Knowledge.

| **Forced-Choice *Eyes Task*** | | **Open-Ended *Eyes Task*** | |
|---|---|---|---|
| **Term-Specific** | **Valence-Based** | **Term-Specific** | **Valence-Based** |
| r=.239 | r=.239 | r=.014 | r=.127 |
| p=.082 | p=.082 | p=.921 | p=.361 |

## 4.4 Discussion

The results of the current study answered some outstanding questions, but also raised others. With respect to the question of what type of social perspective taking (cognitive vs. affective) is being assessed, the results herein strongly support the idea of the open-ended *Eyes Task* (specifically the open-ended, valence-based score) assessing affective perspective taking specifically, in line with the stated hypothesis. This was evidenced by the significant group differences between the at-risk and typically-developing groups and the ability of the open-ended, valence-based *Eyes Task* to predict group membership above and beyond age and cognitive perspective taking. Notably, the forced-choice *Eyes Task* did not identify any group differences, lending further support to the notion that the forced-choice *Eyes Task* (regardless of coding) does not seem to measure affective perspective taking.

The forced-choice *Eyes Task* may measure cognitive perspective taking in this at-risk group, or at least some element thereof, based on the significant relationship with cognitive perspective taking and the lack of differences found between the at-risk and typically-developing

groups. However, the fact that this relationship did not arise in the typically-developing sample in Study 2 raises the question of *why* it appeared herein and not previously. The fact that all four scores of the *Eyes Task* (regardless of response format or coding) correlated with the measure of cognitive perspective taking suggests that the relationships may reflect different overlapping variance; however, one would expect method variance to appear regardless of the sample, and this did not appear in the typically-developing sample of Study 2. Another non-statistical possibility is that because of the presumed affective perspective taking deficits in this at-risk sample, these children may utilize certain cognitive perspective taking skills in an attempt to compensate for an affective deficit. As such, the overlap with the measure of cognitive perspective taking may reflect this compensatory use. On the other hand, the shared variance with the forced-choice *Eyes Task* may better reflect shared variance of cognitive perspective taking, especially as verbal ability was controlled for.

The findings regarding response format and coding scheme were somewhat contrary to the results in Study 2. Although the same effects on performance were demonstrated in the at-risk group, with both the forced-choice response format and valence-based coding resulting in better performance, the potential mechanisms driving this effect were less clear. Despite the fact that there were no group differences in vocabulary knowledge or verbal analogies, the at-risk group did not show the same relationships between performance and verbal analogies or vocabulary knowledge based on response format or coding scheme, respectively.

One possibility that immediately comes to mind is that the variances of these measures in the at-risk group were too small to detect a difference. Although the variances for the at-risk group on these measures were smaller than those of the typically-developing group, Levene's test for equality of variances was non-significant for all three variables, therefore the variance in

the at-risk group was not *significantly* smaller, and thus should not have had an effect. A second, more interesting, possibility is that the nature of the learning disability in the at-risk children influences the way they utilize these abilities. Starting with the knowledge that these children do have a verbal deficit (as part of their learning disability but also as identified relative to the typically-developing group in Study 2), it stands to reason that the verbal ability they demonstrate may be learned more through direct instruction than more naturally occurring or less effortful observational learning. In turn, their use of verbal skills may be more effortful than automatic. Contrast this with the typically-developing group, whose use of verbal skills is likely less effortful, and I would hypothesize there to be differences in both the way in which these abilities are used and how readily they are used.

In line with this, the use of *verbal* analogies as a proxy for alternate cognitive strategies may not have been a good proxy in this group because of these verbal deficits. This may be akin to children who naturally understand math and utilize it every day without much thought versus those children who have to actively think about what they are doing when faced with a math problem. As such, any measure of alternate strategies that includes a verbal component (such as the use of verbal analogies herein) may not be an accurate proxy for cognitive skills in a sample with verbal deficits as the verbal demands may be too high. A related possibility is that the verbal demands were high enough that the participants 'maxed out' their cognitive resources in their effortful use of verbal ability. This would mean there was little to no room for them to be able to call upon other cognitive or verbal abilities to help them solve any of the tasks, resulting in a failure to see any differences in the relationships to verbal analogies or vocabulary knowledge in this at-risk group.

In sum, this study provided more evidence that the open-ended *Eyes Task* is a good measure of affective perspective taking whereas the forced-choice *Eyes Task* is not. The forced-choice *Eyes Task* was, however, related to cognitive perspective taking, even after controlling for verbal ability. As this at-risk group did not have any cognitive perspective taking deficits, the forced-choice *Eyes Task* did not distinguish between the groups. The hypotheses regarding the effects of response format and coding scheme on performance were supported, though the hypothesized mechanisms behind them were not. This is potentially due to the nature of the verbal ability deficits in this at-risk group, and as such modifications would be necessary to better address these questions in future research. Although I had anticipated that the deficits in verbal ability would have resulted in even greater effects of response format and coding scheme within this particular sample, this was not the case. I conjecture that this may be due to differences in the ease with which verbal abilities are executed between the typically-developing and at-risk groups, though future research is needed to test this claim directly.

# 5 General Discussion

## 5.1 Summary of Results

The overarching goal of this dissertation was to systematically examine how measurement and methodological issues can constrain or alter the assessment of social perspective taking in children using a well-known and oft-used measure, the *Reading the Mind in the Eyes* task. Furthermore, to provide researchers with additional measurement tools to assess social perspective taking in future research. Three specific questions served as the impetus for the research at hand: 1) As social perspective taking is comprised of two distinct components—affective and cognitive—what exactly is the *Reading the Mind in the Eyes* task assessing when it claims to measure social perspective taking more generally?; 2) How does the response format influence performance? Specifically, does a forced-choice response format recruit alternate abilities?; and 3) How does coding influence the assessment of social perspective taking? Namely, does the degree of specificity required in the answer increase verbal task demands thus influencing our understanding of an individual's social perspective taking abilities?

To address these questions, three studies were developed and conducted with children ranging from 4 to 13 years of age and included both typically-developing children and a at-risk group of children with social-emotional skills deficits (in addition to verbal learning disabilities). The widely-used measure of social perspective taking—the children's version of the *Reading the Mind in the Eyes Task*—was used to address these questions as it fulfilled several criteria important to the analyses at hand. First, the measure includes a child and adult version allowing researchers to assess developmental changes in social perspective taking. Although the current studies were conducted with children, work with adults would allow for a better understanding of

how these measurement or methodological issues influence assessment of social perspective taking across development (my preliminary work with adults will be discussed below and can be read up on in more detail in Appendix B).

Second, the *Reading the Mind in the Eyes* task is one of the most popular measures of social perspective taking, making the implications of the current studies highly relevant to much of the research on social perspective taking. As this task has informed on social perspective taking skills for a variety of individuals (e.g., those with depression, schizophrenia, psychopathic tendencies; e.g., Craig et al., 2004; Lee et al., 2005; Richell et al., 2003), it is imperative that researchers understand how the aforementioned methodological issues may have influenced these findings and how they may influence future findings so they can take appropriate care to choose the best methods for their particular research purposes.

Third and finally, the *Reading the Mind in the Eyes* task is structured in such a way as to allow the three main questions of interest to be systematically explored. The task purports to be a measure of social perspective taking more generally, with no differentiation between affective and cognitive perspective taking. As such, the task has been used for both purposes (e.g., Gregory et al., 2002; Sharp, 2008; Tonks et al., 2007a,b), despite the fact that the two components are distinct from one another, even if there is overlap conceptually (Hinnant & O'Brien, 2007; Kurdek & Rodgon, 1975; Oswald, 1996, 2002). The task is also forced-choice in nature, but the stimuli (images of the eye region expressing an emotion or mental state) can be easily used with an open-ended task by asking the child to generate an answer instead of selecting from pre-set options. Changing the response format and administering the tasks as either between-subject (Study 1) or within-subject (Studies 2 and 3) allowed me to examine the direct influence of response format on performance as the stimuli remained the same and only

the response format changed. Finally, the task utilizes term-specific coding but in both formats (forced-choice and open-ended), valence-based coding is possible. In the open-ended version, both term-specific and valence-based coding are easy to employ and in the forced-choice version, either the specific emotion term can be labeled correct or any answer of the same emotional valence as the correct term, allowing for a direct examination of the influence of coding on performance.

### 5.1.1 Question #1: What Type of Perspective Taking is Being Measured?

In order to address this question, scores on the forced-choice and open-ended *Eyes Task* were compared to a measure of dispositional empathy (Study 1), compared to another measure of cognitive perspective taking (Studies 2 and 3), and used to differentiate children at-risk for affective perspective taking deficits from typically-developing children. For a summary of the findings, see Table 21. First, with respect to the assessment of affective perspective taking and in line with my hypotheses, the forced-choice *Eyes Task* was *unrelated* to dispositional empathy in the typically-developing sample whereas the open-ended *Eyes Task* was significantly correlated with dispositional empathy, though only for boys. There were no differences for girls, potentially due to the effect of stereotype threat in that girls are expected to be more empathic, especially when using self-report measures. Furthermore, the open-ended *Eyes Task* (specifically the open-ended, valence-based *Eyes Task*), but not the forced-choice *Eyes Task*, was able to significantly predict group membership between the typically-developing and at-risk groups, above and beyond age and cognitive perspective taking.

Second, with respect to assessing cognitive perspective taking, the results varied by group (i.e., typically-developing and at-risk): For typically-developing participants, the forced-choice *Eyes Task* was correlated with the measure of cognitive perspective taking in a bivariate analysis;

however, when age and verbal ability were controlled for, the relationship ceased to be significant, raising questions about the validity of the task as a measure of cognitive perspective taking. The open-ended *Eyes Task*, on the other hand, was not related to the measure of cognitive perspective taking in a typically-developing sample. For the at-risk group, results were different, with both the forced-choice and open-ended *Eyes Task* being significantly correlated with cognitive perspective taking, even when verbal ability was factored out (age was not significantly related to cognitive perspective taking in this sample and was thus not partialled out). Therefore, in this at-risk group, both the forced-choice *Eyes Task* and the open-ended *Eyes Task* were seen as potential measures of cognitive perspective taking. One possible explanation for the findings in the at-risk group is that any deficit in affective perspective taking results in these children using cognitive perspective taking skills in a compensatory manner; that is, the children are attempting to use their cognitive perspective taking skills to solve an affective perspective taking task.

**Table 21.** Summary of Findings for Research Question 1.

| | ***Eyes Task* Score** | | | |
|---|---|---|---|---|
| | **Force-choice, term-specific** | **Forced-choice, valence-based** | **Open-ended, term-specific** | **Open-ended, valence-based** |
| **Relationship to Dispositional Empathy (boys only) (Study 1)** | X | X | r=.456<br>p=.009 | r=.335<br>p=.061 |
| **TD-AR Differences (Study 3)** | X | X | X | t(128)=4.410, p<.001 |
| **Predict Group Membership (Study 3)** | X | X | X | Unit ↓ = 62.3% ↑ risk clinical group |

Notes: TD=typically-developing; AR=at-risk

In sum, the research herein supports the hypothesis that the open-ended *Eyes Task* is a better measure of affective perspective taking than the forced-choice *Eyes Task*. The question of what the forced-choice *Eyes Task* measures remains up for debate as the task did not overlap with a known measure of cognitive perspective taking in the typically-developing sample, but did in the at-risk group.

### 5.1.2 Question #2: How Does Response Format Influence Performance?

Response format has already been demonstrated to influence the type of social perspective taking assessed, as evidenced above with the open-ended *Eyes Task* (combined with valence-based coding) being a better measure of affective perspective taking specifically than the forced-choice *Eyes Task*. An additional concern was that a forced-choice *Eyes Task* could enable participants to use alternate strategies to solve the task thereby potentially masking mild-moderate perspective taking deficits. As demonstrated throughout all three studies, the forced-choice *Eyes Task* did elicit better performance than the open-ended *Eyes Task*, regardless of the coding scheme or participant sample (i.e., typically-developing or at-risk). See Table 22 for a summary of the results across the studies.

**Table 22.** Summary of Findings for Research Question 2.

| | ***Eyes Task* Score** | | | |
|---|---|---|---|---|
| | **Force-choice, term-specific** | **Open-ended, term-specific** | **Forced-choice, valence-based** | **Open-ended, valence-based** |
| **Performance (Study 1, TD)** | 49.1% | 9.5% | 58.9% | 29.5% |
| **Performance (Study 2, TD)** | 62.88% | 10.24% | 68.80% | 35.71% |
| **Relationship to Cognitive Proxy (Study 2, TD)** | r=.407[a] p<.001 | r=.055[b] p=.637 | r=.361[a] p=.001 | r=.066[b] p=.572 |

| | *Eyes Task* Score | | | |
|---|---|---|---|---|
| | **Force-choice, term-specific** | **Open-ended, term-specific** | **Forced-choice, valence-based** | **Open-ended, valence-based** |
| **Performance (Study 3, AR)** | 70.64% | 8.47% | 63.76% | 26.85% |
| **Relationship to Cognitive Proxy (Study 3, AR)** | r=.141<br>p=.310 | r=.139<br>p=.315 | r=.141<br>p=.308 | r=.156<br>p=.261 |

Note: [a,b]Denotes significantly different correlations; TD = typically-developing group; AR = at-risk group

A key question was whether or not the superior performance in the forced-choice *Eyes Task* was due to the use of alternate strategies or was simply because the task, for whatever other reason, was easier while still assessing social perspective taking specifically. Although there was no way to assess how children came to their answer, comparing performance on the tasks to a measure of verbal analogies—a measure that was more cognitive than vocabulary in nature, as evidenced by the low correlation between vocabulary knowledge and verbal analogies and served as a proxy for alternate cognitive abilities—allowed for an assessment of the plausibility of the alternate strategy hypothesis. In typically-developing children, the hypothesis is tenable, with performance on the forced-choice *Eyes Task* significantly correlated with verbal analogies whereas performance on the open-ended *Eyes Task* was not. When response formats were directly compared (controlling for coding scheme) in their relationships to verbal analogies, there were significant differences, suggesting that the effect of using alternate cognitive abilities is stronger for the forced-choice *Eyes Task* than the open-ended *Eyes Task*, in line with the hypothesis that the forced-choice versions may be 'solved' using alternate strategies.

For the at-risk group, the results differed from those of the typically-developing group, with no measure significantly related to verbal analogies. One possible reason is that the use of *verbal* analogies in a group that has deficits in *verbal* ability may have precluded the verbal

analogies as being a good substitute for alternate cognitive strategies. A second possible reason is that the nature of the learning disabilities in this group precludes them from using the same strategies as typically-developing individuals when solving social perspective taking tasks. That is, there is something unique about having a learning disability, or specifically either verbal or affective perspective taking deficits, that alters the use of cognitive strategies for tasks that include social perspective taking or verbal abilities. A third possibility, in line with both previous possibilities, is that the forced-choice *Eyes Task*, and its use of specific vocabulary terms, requires the effortful use of verbal ability in this at-risk group (due to their relative deficit in verbal ability). This effortful use creates additional cognitive demands, leaving no room for the child to utilize *other* cognitive abilities to solve the task and thus any differences in performance are simply due to more options being coded correct. That is, if these verbal demands were not a part of the task, these children may be able to use alternate strategies to solve the task and differences may relate more to the measure of cognitive ability.

In sum, systematically changing the response format resulted in significant changes both in terms of what type of social perspective taking is being assessed in the *Reading the Mind in the Eyes* task and in performance. Specifically, the forced-choice *Eyes Task* is positively associated with verbal ability and age in children, but it is unclear how it is related to cognitive perspective taking (it is unrelated in typically-developing participants but related in the at-risk sample) and it does not seem to be a good measure of affective perspective taking. An open-ended methodology allows for a better assessment of affective perspective taking and also may assess cognitive perspective taking to a degree in children with learning disabilities who are at-risk for affective perspective taking deficits, though it may simply reflect shared method variance or the use of cognitive perspective taking skills to compensate for affective deficits.

Additionally, the forced-choice *Eyes Task* seems to recruit cognitive abilities that may allow for alternate strategies to be employed by typically-developing individuals whereas it is less clear if children with learning disabilities (at least those of the type included in the current sample) are able to employ these alternate strategies. The open-ended *Eyes Task* does not seem to allow for the use of alternate strategies, regardless of population.

### 5.1.3 Question #3: How Does Coding Influence Performance?

Across the three studies, valence-based coding was associated with greater performance, which is not surprising as valence-based coding allows for a greater number of options to be deemed 'correct'. The key question here, however, was how differences in coding might be influenced by vocabulary knowledge or verbal ability. I hypothesized that term-specific coding would place unnecessary vocabulary demands on younger children and may limit their performance unnecessarily. The first notable finding was that the combination of an open-ended task and term-specific coding led to a floor effect in all three studies. Given the plethora of words that children can choose from, requiring children to be so specific seems to inhibit their performance and is likely not a good way to assess any component of social perspective taking for children of this young age. Allowing children more freedom in explaining their answers (as was allowed in the cognitive perspective taking task, *Happé's Strange Stories*, and the open-ended, valence-based *Eyes Task*) can provide a better assessment of their real capacity.

For both typically-developing children and children at-risk for affective perspective taking deficits, the relationships to vocabulary knowledge and verbal ability were significant for the forced-choice *Eyes Task*, regardless of coding. That is, scores for both the forced-choice, term-specific *Eyes Task* and forced-choice, valence-based *Eyes Task* were related to vocabulary knowledge and verbal ability. This was not the case for the open-ended *Eyes Task*, with both the

open-ended, valence-based *Eyes Task* and open-ended, term-specific *Eyes Task* being *unrelated* to vocabulary knowledge and verbal ability in both groups. Of note, however, is that in the typically-developing group, the degree to which there was an association with vocabulary knowledge and verbal ability was significantly greater for the term-specific coding than the valence-based coding in the forced-choice *Eyes Task*, suggestive of an even greater use of vocabulary or verbal skills when combining the forced-choice response format with term-specific coding. This finding was not replicated in the at-risk group. For all children there does seem to be an additional reliance on vocabulary knowledge when the task is forced-choice, and for typically-developing children, this reliance is even greater when the coding is term-specific. See Table 23 for a summary of findings.

**Table 23.** Summary of Findings for Research Question 3.

| | ***Eyes Task* Scores** | | | |
|---|---|---|---|---|
| | **Force-choice, term-specific** | **Forced-choice, valence-based** | **Open-ended, term-specific** | **Open-ended, valence-based** |
| **Performance (Study 1, TD)** | 49.1% | 58.9% | 9.5% | 29.5% |
| **Performance (Study 2, TD)** | 62.9% | 68.8% | 10.2% | 35.7% |
| **Relationship to Vocabulary Knowledge (Study 2, TD)*** | r=.355[a]<br>p=.002 | r=.293[b]<br>p=.101 | r=-.122[c]<br>p=.293 | r=.013[c]<br>p=.910 |
| **Performance (Study 3, AR)** | 63.8% | 70.6% | 8.5% | 26.9% |
| **Relationship to Vocabulary Knowledge (Study 3, AR)*** | r=.239<br>p=.082 | r=.239<br>p=.082 | r=.014<br>p=.921 | r=.127<br>p=.361 |

Notes: [a,b,c]Denotes significantly different correlations; TD=typically-developing; AR=at-risk; *Results for verbal ability were more significant than that for vocabulary knowledge, thus only vocabulary knowledge is included herein as it is the main variable of interest.

As this was contrary to the hypothesis that coding specifically would be related to vocabulary knowledge, it is paramount to understand why: a) there was no effect for coding in the open-ended versions, b) there was still a significant relationship between the valence-based coding in a forced-choice response format, and c) the at-risk group did not show an additional effect of coding (above and beyond the effect of response format). With respect to the first question, one possibility is that the verbal ability needed to generate an answer is so minimal that no relationship to vocabulary knowledge or verbal ability was found. One would expect a greater requirement for vocabulary knowledge in the term-specific coding, but this may have been masked by the floor effect seen for that particular version (i.e., the open-ended, term-specific *Eyes Task*). That is, as performance was so low in the open-ended, term-specific *Eyes Task*, there was no variance available to correlate with vocabulary knowledge or verbal ability more generally.

With respect to the second question—why there was a significant relationship between vocabulary knowledge or verbal ability and the valence-based coding in a forced-choice response format?—it is almost certainly a product of the fact that the forced-choice *Eyes Task* required a specific level of vocabulary knowledge to simply provide an answer. That is, from the child's point-of-view, the task was the same, with the same verbal demands involved in understanding and selecting from the four term-specific options. The difference was in the coding afterwards. The incremental findings with the typically-developing group highlight this nicely with a certain degree of vocabulary knowledge or verbal ability needed to get *any* right answer and an additional layer of vocabulary knowledge or verbal ability needed to get *the term-specific* right answer. It is akin to the difference between understanding an emotional category versus the

nuances therein and how the nuanced understanding requires *additional* vocabulary knowledge of the terms therein.

With respect to the third and final question—why did the at-risk group not show an additional effect of coding (above and beyond the effect of the forced-choice response format)?—it is likely that the verbal demands of the forced-choice *Eyes Task* were no different for this group as the verbal demands of the task remained the same (the same reason why there was an effect of vocabulary knowledge on the valenced-based coding for the forced-choice *Eyes Task*). As this group had verbal deficits in addition to social-emotional deficits (i.e., their learning disability was verbal and their verbal mental age, as calculated herein, was not different from that of the typically-developing group although they were significantly older), they may not automatically recruit verbal abilities, instead their use may be effortful. This effortful use may create more cognitive demands for the task and thus they may be 'maxing out' their verbal abilities for each question, leaving no room for variability in vocabulary knowledge or verbal ability to be influenced by coding scheme, only response format.

### 5.1.4 Overall Summary

Taken together, this dissertation found that methodological factors in the *Reading the Mind in the Eyes* task not only influence our assessment of social perspective taking in children by altering their performance, but can also influence what *type* of perspective taking is being measured. Two response formats (forced choice vs open-ended) and two coding schemes (term specific vs. valenced) were systematically compared to assess their impact on performance, relations to vocabulary and general cognitive abilities, and efficacy at measuring cognitive versus affective perspective taking. Overall, there was a significant effect of response format, but less of an effect of coding scheme.

The effect of response format was seen in both performance (i.e., how well children did on a given version of the *Eyes Task*) and in what type of social perspective taking was being measured. Children scored higher when a forced-choice response format was used relative to the use of an open-ended response format; however, the forced-choice format, unlike the open-ended format, was related to both cognitive and verbal abilities in typically-developing children. These relationships to non-social perspective taking abilities suggests that performance on the forced-choice format is more apt to be influenced, unnecessarily, by the participants' vocabulary knowledge or use of alternate strategies (e.g., process of elimination). Furthermore, the forced-choice format was unrelated to dispositional empathy and cognitive perspective taking; it also failed to differentiate typically-developing from children at-risk for affective perspective taking deficits, leaving up for debate the question of *what* the forced-choice *Eyes Task* is measuring. In contrast, the open-ended *Eyes Task* was significantly related to dispositional empathy and was able to differentiate the at-risk group from the typically-developing group.

Coding scheme, however, did not exert such a clear influence. Although there was an influence of coding scheme on performance across all groups, with term-specific coding resulting in lower scores than valence-based coding, the relationship to vocabulary knowledge was not fully borne out. Only for the forced-choice *Eyes Task* did vocabulary knowledge predict differences in performance based on coding scheme, and again, only for typically-developing children. Thus, the hypothesis that vocabulary knowledge would influence performance via coding scheme was largely unsupported.

## 5.2 Broader Significance of the Research

The research herein has broad implications for both the specific literature using the *Reading the Mind in the Eyes* task specifically as well as for the field of social perspective taking more generally. Each will be discussed in turn below.

### 5.2.1 Significance for *Reading the Mind in the Eyes* Task Research

In fitting the current results into the broader research using the *Reading the Mind in the Eyes* task with children, there are four considerations of particular interest: 1) how the task has been used to identify Autism Spectrum Disorders, 2) how it has been used as a cognitive perspective taking task, 3) how it has been used as an affective perspective taking task, and 4) its relationships to other cognitive and verbal abilities.

Regarding the first question, although the current research did not include a sample of children with Autism Spectrum Disorders, the findings do have relevance to that particular line of research. As mentioned in Section 1.5.1.2, children with Autism Spectrum Disorders are one population in which a dual deficit of social perspective taking is present, with children (and adults with Autism Spectrum Disorders) showing deficits in *both* affective and cognitive components of social perspective taking. The *Reading the Mind in the Eyes* task, which is reportedly assessing social perspective taking more generally (i.e., both cognitive and affective), could arguably differentiate children with Autism Spectrum Disorders from those without even if subcomponent scores are unavailable. However, the current findings raise questions about the validity of the *Reading the Mind in the Eyes* task as a general measure of social perspective taking, particularly if the interest is in individual differences or the assessment of components of social perspective taking. The failure to find any relationship to dispositional empathy, to differentiate between an at-risk group with known affective perspective taking deficits and

typically-developing children, and to show little-to-no relationship with another measure of cognitive perspective taking raises concerns about the measure as one of general social perspective taking abilities. If, however, the task is not measuring generalized social perspective taking, what could it be measuring? It is possible that the *Reading the Mind in the Eyes* task is actually measuring yet another ability that differentiates Autism Spectrum Disorders from controls. Based on Item Response Theory and Confirmation Factor Analysis findings, the *Reading the Mind in the Eyes* task seems to represent a singular ability that best differentiates children on the low-end of the perspective taking spectrum (Carey & Cassels, 2013). That is, it will do well at identifying children who score poorly on the task (like those with Autism Spectrum Disorders), but will provide little in the way of discrimination for those who score higher, even if they are scoring below the mean.

One possibility is that it assesses a more general aversion to eye gaze. Monitoring eye-gaze and the eye-region of the face is typically necessary to identify cognitive or affective mental states (for a review, see Itier & Batty, 2009) and as the *Reading the Mind in the Eyes* task asks children to identify these mental states from an image of the eye region, an aversion to eye gaze would impede performance. Children with Autism Spectrum Disorders have been found to avoid eye gaze and other facial information (Dawson, Toth, Abbott, Osterling, Munson, et al., 2004; Klin, Jones, Schultz, Volkmar, & Cohen, 2002), making affective and cognitive perspective taking nearly impossible. This would also explain why other research using the *Reading the Mind in the Eyes* task and other measures of cognitive perspective taking (including *Happé's Strange Stories*, the measure of cognitive perspective taking herein) has found very little overlap outside of information-processing requirements (Brent, Rios, Happé, & Charman, 2004; Kaland, Callesen, Møller-Nielsen, Mortensen, & Smith, 2008).

The second area of significance is how the *Reading the Mind in the Eyes* task has been used as a measure of cognitive perspective taking specifically. I had hypothesized that the *Reading the Mind in the Eyes* task would be better seen as a measure of cognitive perspective taking than affective perspective taking based on a) its ability to distinguish between controls and those with Autism Spectrum Disorders (Baron-Cohen, Wheelwright, Spong, et al., 2001), and b) the knowledge that the measure is one of a unitary ability (Carey & Cassels, 2013); however, the results herein did not support my hypothesis. The findings that the *Reading the Mind in the Eyes* task was only minimally related to a measure of cognitive perspective taking, in line with previous findings (Brent et al., 2004; Kaland et al., 2007), suggests that at the very least, it is not measuring the same type of cognitive perspective taking as *Happé's Strange Stories.* As it does not involve goal or intention understanding, desire or preference understanding, knowledge attribution, beliefs or false beliefs, the *Reading the Mind in the Eyes* task cannot be a measure of any of these types of cognitive perspective taking. However, even these findings must be noted with the caveat that the results with the at-risk sample could offer some circumstantial support for the *Reading the Mind in the Eyes* task as a measure of cognitive perspective taking. First, in this at-risk sample, the *Reading the Mind in the Eyes* task was correlated with cognitive perspective taking (but so was the open-ended format), even after verbal ability was partialed out. Second, the current sample did not have cognitive perspective taking deficits (indeed they outperformed the control group on the measure of cognitive perspective taking, likely due to their older age) and there was comparable performance between the two groups on the *Reading the Mind in the Eyes* task. However, this would not be enough to claim the measure is a good measure of cognitive perspective taking per se, but rather it may be more complicated and nuanced than any black-and-white statement about what the *Reading the Mind in the Eyes* task

purportedly measures. It is important to note here that the original development of the *Reading the Mind in the Eyes* task was to differentiate between typically-developing children (and adults) and those with Autism Spectrum Disorders. As such, it has proven to be highly effective. The concerns regarding the *Reading the Mind in the Eyes* task herein stem from *others'* use of the measure as a measure of social perspective taking more broadly in addition to its use as a specific measure of cognitive or affective perspective taking. Given the initial development criteria—to differentiate a clinical from a typically-developing population—it may come as no surprise that it is in this clinical realm that the measure functions best.

Third, the *Reading the Mind in the Eyes* task has been used as a specific measure of affective perspective taking, with questionable results. Although the face validity of the task could suggest a measure of affective perspective taking, the current results counter this possibility. First, the failure to find a relationship between dispositional empathy and the *Reading the Mind in the Eyes* task would suggest the measure does not assess affective perspective taking. Although this counters previous research that *has* found a relationship between the *Reading the Mind in the Eyes* task and dispositional empathy (Chapman et al., 2006), there were differences in how dispositional empathy was assessed (parent- versus child-report) which may have influenced findings. If one focused solely on this singular finding herein, it would be unclear which assessment of dispositional empathy would be more accurate (parent or child report) and which finding should hold more weight; however, the current research also found that the *Reading the Mind in the Eyes* task was unable to differentiate children at-risk for affective perspective taking deficits from typically-developing children (whereas the open-ended *Eyes Task* was able to). Taken together, this supports the notion that

the *Reading the Mind in the Eyes* task is *not* a good measure of affective perspective taking per se.

If the *Reading the Mind in the Eyes* task is not a good measure of affective perspective taking then previous research utilizing it as such needs to be re-evaluated. For example, Sharp (2008) used the *Reading the Mind in the Eyes* task as a measure of affective perspective taking and found that performance on it was diminished in non-clinical children with higher levels of conduct problems (controlling for overall IQ), linking the findings to those looking at deficits in individuals high on psychopathic traits or tendencies. Although affective perspective taking deficits are well-documented in children with higher psychopathic tendencies (e.g., Blair, 2003), the measure employed by Sharp made no such distinctions between conduct problems and psychopathic tendencies. This is a critical distinction as individuals with conduct problems and *low* levels of psychopathic tendencies do not show the same deficits in affective perspective taking (e.g., Blair et al., 2001). A second study that utilized the *Reading the Mind in the Eyes* task as a measure of affective perspective taking in children with traumatic brain injury, Tonks and colleagues (2007b), found deficits in these children compared to controls. Although they included a measure of general executive functioning, they did not include other cognitive or verbal measures. In both of these studies, the alternate findings could be due to either relationships to specific cognitive or verbal abilities that were not controlled for or a deficit in use of eye gaze to determine facial expressions.

The final consideration with respect to the *Reading the Mind in the Eyes* task specifically is the findings that the task is related to various cognitive and verbal abilities—vocabulary knowledge, verbal ability more generally, and verbal analogies (used as a proxy for cognitive abilities)—and age. This is in contrast to some research that found no relationships to various

cognitive abilities (Chapman et al., 2006; Sharp, 2008; Tonks et al., 2007b) but in line with others (Brent et al., 2004). One explanation for the discordant findings is that, with the exception of Chapman and colleagues (2006), these studies did not include specific measures of verbal ability with which to make comparisons. Instead, they included broader intelligence measures that included verbal ability as only one component (i.e., general IQ; Sharp, 2008) or they did not include any assessment of verbal ability (i.e., only general executive functioning; Tonks et al., 2007b). If performance on the *Reading the Mind in the Eyes* task is predominantly driven by verbal ability, then the failure to find these relationships in other measures of cognitive ability would be expected. The general implications of these relationships herein, however, is that performance on the *Reading the Mind in the Eyes* task in typically-developing populations may be in part due to *other* abilities, and not social perspective taking per se. This is in line with research that failed to find a relationship between the *Reading the Mind in the Eyes* task and facial processing, but instead found a strong relationship to verbal ability in adults (Peterson & Miller, 2012) and research that found lowering the verbal load led to significant relationships between performance on the *Reading the Mind in the Eyes* task and false-belief understanding in children and adults (Peterson & Slaughter, 2009). Moving forward, any study that utilizes the *Reading the Mind in the Eyes* task should include a measure of verbal ability, preferably vocabulary knowledge and more general verbal ability, to control for their effects on performance.

### 5.2.2 Significance for Social Perspective Taking Research

The second area of significance for the findings herein is for social perspective taking research broadly speaking. Although the studies herein utilized one specific measure of social perspective taking, many of the findings are applicable to research on social perspective taking

more generally. First, the importance of specifying and understanding which component of social perspective taking is being measured cannot be overstated. The findings herein suggest that measures that purport to assess both components without providing individual scores may fail to assess either appropriately. This is particularly important as different populations may have different weaknesses that can be masked by the use of a generalized measure. When researching specific populations, it is imperative that researchers are aware of which component of social perspective taking is hypothesized to be deficient and to make sure their measures provide a score for that particular component. Evidence of this may be seen in the failure of the *Reading the Mind in the Eyes* task to identify deficits in affective perspective taking in those with higher levels of psychopathy, especially the interpersonal or affective component (Richell et al., 2003; Sandvik et al., 2014), with this failure not replicating well-known and established findings regarding emotion recognition and psychopathy (for a review, see Blair, 2003).

The findings herein also raise concerns about the use of individuals with Autism Spectrum Disorders as the gold standard for the development of social perspective taking tasks. As individuals with Autism Spectrum Disorders have a dual deficit in social perspective taking (i.e., they show deficits in both cognitive and affective perspective taking), measures that do not make this distinction may still identify those with Autism Spectrum Disorders, but fail to provide useful information for those who do not show such dual deficits. There is considerable merit then in a measure that both assesses *degrees* of social perspective taking skills as well as identifying deficits in populations that do not show a dual deficit but only demonstrate a deficit in one component of social perspective taking. The aforementioned findings by Carey and Cassels (2013) speak to the inability of the *Reading the Mind in the Eyes* task to do this. However, findings from Carey and Cassels also suggest that the open-ended *Eyes Task*, which

has support as a measure of affective perspective taking specifically, can provide more information at differing levels of performance. This ability to provide information at different levels shows that the open-ended *Eyes Task* is a measure that can possibly provide information on group distinctions as well as degrees of ability.

It is paramount that researchers who are interested in social perspective taking and its components utilize appropriate measures for their questions at hand. For instance, researchers interested in individual differences may want to avoid the use of the *Reading the Mind in the Eyes* task and instead focus on another measure that provides more information about those without Autism Spectrum Disorders or other clinical impairments. Furthermore, if one is interested in affective perspective taking specifically, the use of the open-ended *Eyes Task* or another measure, such as the *Florida Affect Battery* (Bowers, Blonder, & Heilman, 1991), would be preferable to the use of the *Reading the Mind in the Eyes* task. Similarly, if one is interested in the assessment of cognitive perspective taking, a measure such as *Happé's Strange Stories* (Happé, 1994) may be preferable. Regardless of the specific measure chosen, it is up to individual researchers to ensure they have identified their construct of interest and are using a measure that enables them to tap that specific contract. Finally, unless the goal is to identify individuals with dual deficits, generalized social perspective taking measures should be avoided as the two components—affective and cognitive—may be conceptually related, but do not necessarily overlap practically (Hinnant & O'Brien, 2007; Kurdek, 1977; Kurdek & Rodgon, 1975) and thus any global measure may fail to capture important information.

The third area of general relevance is with respect to response format. This has been an area that has been identified by other researchers as being relevant to the proper identification of skills (e.g., Widen & Russell, 2003), particularly in children, with open-ended tasks providing

ample information about social perspective taking abilities (e.g., Happé, 1994; Widen & Russell, 2003, 2008). The findings herein support the notion that performance on social perspective taking tasks may be influenced by response format. Specifically, forced-choice tasks may allow for alternate strategies to be used, artificially increasing performance relative to open-ended formats. Arguably, the aforementioned findings regarding the failure to find a relationship between the *Reading the Mind in the Eyes* task and facial processing (Peterson & Miller, 2012) may reflect individuals' tendencies to use cognitive abilities if they are easier or more readily available than social perspective taking skills. Any task that enables these alternative strategies (which are also counter to real-life situations, reducing the ecological validity of the tasks) might inadvertently be assessing cognitive abilities (including verbal abilities) instead of the social perspective taking abilities of interest. As such, previous research using forced-choice measures may not be as accurate as those that have utilized open-ended measures.

There is also the issue of the number of response options available and how this may influence the use of alternate strategies. Theoretically, the more response options there are, the harder it may be to use alternate strategies, which may be why an open-ended task with limitless response options better assesses certain constructs. In line with this, Peterson and Slaughter (2009) found that a two-option forced-choice response format for the *Reading the Mind in the Eyes* task with easier terms enabled the task to more accurately assess cognitive perspective taking, as assessed by the relationship to false belief understanding. Although one may think this speaks to better assessment, false belief understanding is critically tied to verbal ability (see Happé, 1995 for a discussion) and thus the removal of additional options and the use of easier terms may simply have reflected the level of verbal ability of the group. Further evidence comes from Bindemann, Burton, and Langton (2008) who found that assessment of affective

perspective taking was inversely related to the number of options available, with fewer options resulting in better emotion discrimination, likely due to the reduction of competing terms. Notably, these findings are in line with the results herein that found the open-ended task (i.e., many response options available) had lower performance than the forced-choice task (i.e., four options available). It also suggests that the verbal demands of multiple response options may pose difficulties for some populations or samples and that the use of alternate strategies may be inversely related to the number of response options as well.

I argue that researchers should consider the use of open-ended measures when assessing social perspective taking abilities. The goal of research in the realm of social perspective taking is often to understand its relationships to various other constructs and its developmental trajectory. If one response format changes what is measured due to the assessment of other cognitive abilities, this negatively influences researchers' ability to make claims about social perspective taking. As such, the response format with the least external influences is the one that should be preferred.

The specific influence of coding was also found to hold across response format, with term-specific coding resulting in worse performance than valence-based coding; however, I could not claim that this effect is due to anything more than simply allowing more options as 'correct'. Relevant for future research, term-specific coding in the open-ended *Eyes Task* resulted in a floor effect with young children and thus may be inappropriate for use with that particular population as the verbal demands and emotion term vocabulary knowledge (particularly for nuanced emotions) are likely not as developed as that of older individuals. This is in line with research from Widen and Russell (2008) who found emotion understanding starts out in a valence-based manner and becomes term-specific with age. Valence-based coding in an

open-ended task, however, yielded a range of scores that seemed to measure the individual differences in affective perspective taking while avoiding the task demands of unnecessary vocabulary knowledge that was present in the forced-choice *Eyes Task*, regardless of coding. Valence-based coding can be incorporated into forced-choice tasks in two ways: As I did here (by coding all same-valence terms as correct) or by providing valence-based options for the forced-choice options, as is the case in the *Florida Affect Battery*. Research comparing the *Reading the Mind in the Eyes* task (term-specific options) and the *Florida Affect Battery* (valence-based options) in children (Tonks et al., 2007a) was in line with findings herein; namely, term-specific options were associated with worse performance than valence-based options (although the differences in stimuli may also have accounted for this distinction). I would recommend to researchers that if there is a preference for the use of a forced-choice task, the use of valence-based coding or options be employed in order to avoid inadvertently measuring vocabulary knowledge or verbal ability instead of social perspective taking.

Taken together, the findings have implications in how we understand previous social perspective taking research and what kind of considerations should be made moving forward. In particular, the influence of cognitive and verbal abilities is a concern if the goal is to specifically measure social perspective taking abilities, and these effects are most pronounced in tasks that are forced-choice in nature. Though this concern is magnified when assessing the construct in any population in which these abilities may be compromised, findings herein suggest it is also a concern for typically-developing populations. Some of the cognitive abilities that may influence findings, such as the use of alternate strategies, may be more automatic than social perspective taking skills (either affective or cognitive) and thus measures that enable their use would be unduly influenced. Furthermore, the current findings support the need to ensure tasks are

specific to the type of social perspective taking of interest, as global measures may fail at both identifying individuals with only a specific social perspective taking deficit and at identifying individual differences or degrees of performance in social perspective taking. One final recommendation is to be aware of the population of interest. For example, the development of the *Reading the Mind in the Eyes* task was to differentiate those with Autism Spectrum Disorders and those without and has proven very successful at this task because the deficits associated with Autism Spectrum Disorders are being assessed (in one form or another) using the *Reading the Mind in the Eyes* task. However, not all measures that are developed for one reason can be used appropriately for another, such as using the *Reading the Mind in the Eyes* task to assess individual differences in affective perspective taking. Researchers should be wary of using measures in ways contrary to their development without doing appropriate research on the validity of such uses.

## 5.3 Strengths, Limitations, and Future Directions

### 5.3.1 Strengths and Limitations

The current research is among the first to systematically and directly examine how various methodological factors may influence the assessment, and thus our understanding of, social perspective taking in children in an oft-used measure of generalized social perspective taking. The literature on social perspective taking in children is not without conflicting results (for a review, see Harris, 2006) and yet to date the focus has not been on the specific methodological factors of interest herein, even though the role of methodology in assessment has been readily identified in other areas of social perspective taking, such as false belief understanding (for a discussion, see Saxe, 2013) and emotion recognition (Widen & Russell,

2008). Although one might believe that a limitation to the current results stems from the use of only one measure of social perspective taking broadly speaking, I believe it to simultaneously be one of the strengths of this dissertation. By using one particular measure and systematically altering the response format and manner of coding, the influence of each of these measurement issues on performance could be directly examined. In contrast, the use of multiple measures would have added additional elements that would have needed to be controlled for (such as different stimuli), making direct comparisons between forced-choice and open-ended tasks and valence-based and term-specific coding nearly impossible. For example, the research by Tonks and colleagues (2007a) that included the *Florida Affect Battery* and the *Reading the Mind in the Eyes* task could not make direct comparisons between the two tasks as the stimuli differed. By utilizing the same stimuli herein, I was able to identify the specific effects of response format and coding and how they were linked to non-social perspective taking abilities.

A second strength herein is the use of a wide age-range of children. The children herein ranged from 4- to 12-years of age for the typically-developing samples and 6- to 13-years of age for the at-risk sample. This range allowed for a study of how these various methodological issues relate to assessing performance across development. For example, the lack of an age-related finding for the open-ended *Eyes Task* suggests that understanding of emotions is relatively stable across these age ranges. The significant relationship between age and the forced-choice *Eyes Task* was moderated by verbal ability, suggesting that the age-related changes that have been found previously using the *Reading the Mind in the Eyes* task (Tonks et al., 2007a) may be a reflection of the development of verbal ability, not an improvement in social perspective taking.

A third strength is the use of a sample of children at-risk for affective perspective-taking deficits as a means of comparison instead of the more typical group of children with Autism Spectrum Disorders. Although some may believe the *lack* of an Autism Spectrum Disorder group is a limitation, I have outlined herein why they may not be an ideal group for the assessment of the current questions surrounding methodology (see Section 1.5.1.2). Specifically, their dual deficit in both cognitive and affective perspective taking would have negated the ability to accurately identify the *specificity* of what type of social perspective taking is being measured in the forced-choice or open-ended *Eyes Task* tested herein. Additionally, the verbal and cognitive deficits that often accompany Autism Spectrum Disorders would have prohibited or negatively influenced the assessment of the role of vocabulary knowledge and cognitive abilities on coding and response format. The choice of the current group of children with learning disabilities, with specific verbal deficits and being at-risk for affective perspective taking deficits, allowed for the questions of interest to be examined in the absence of broader cognitive deficits.

This research is not, however, without its limitations. First, although I assume the findings regarding the methodological concerns will hold across other social perspective taking measures, this will need to be empirically tested in future research. Similar to this, it is unclear if the current findings would hold for measures already developed as specific cognitive or affective perspective taking measures. Although there is no a priori reason to believe the findings would be different with respect to the systematic comparison of response format and coding, the fact remains that the research has not been done and there may be elements specific to other tasks that negate the influence of these methodological factors.

A second limitation of this dissertation research is the use of a single measure of cognitive perspective taking (*Happé's Strange Stories*) and a single measure of dispositional empathy (which includes elements of affective perspective taking in the questions; *Bryant Index of Empathy*) to assess how well the different versions of the *Eyes Task* assess cognitive or affective perspective taking. In a typical study looking at construct validity, multiple measures would be included for comparison and I acknowledge that the absence of other measures of cognitive or affective perspective taking may limit the ability to generalize findings. However, there were two main reasons I limited this research to these measures. First, the interest herein was *not* in demonstrating construct validity because the measure of interest (the *Reading the Mind in the Eyes* task) has already been in use for some time, but rather seeing how systematic changes to an existing measure may influence what is assessed. When running studies with children where time is a factor (i.e., keeping children interested for long periods can be difficult and limitations put on the research by the schools and programs where the children were recruited), I chose to focus on only a couple singular measures for the purposes herein. Second, the questions herein pertained to methodological issues in the assessment of social perspective taking and as such, any measure needed to avoid methodological confounds. For example, the use of *Happé's Strange Stories* was intentional in that it is a known measure of cognitive perspective taking that includes an open-ended response format and valence-based coding, the opposite of the forced-choice, term-specific *Reading the Mind in the Eyes* task. Removal of shared variance with the *Reading the Mind in the Eyes* task was essential to ensuring any relationship was not an assessment of method variance.

A third limitation with this dissertation research is the type of coding used for the forced-choice, valence-based *Eyes Task*. As previously mentioned, there are two types of valence-based

coding possible for a forced-choice task: Code all same-valence options as correct or replace specific terms with valence-based terms in the question options. The findings herein consistently found that the forced-choice, valence-based *Eyes Task* was highly related to vocabulary knowledge and it may be that the use of terms in the question itself influenced children's performance and their reliance on vocabulary knowledge. Changing the task to one in which the forced-choice format includes valence-based terms may remove the reliance on vocabulary knowledge, which is exactly what I found in my preliminary work with adults: The forced-choice, valence-based *Eyes Task* was the *only* measure to be completely unrelated to vocabulary knowledge (the open-ended *Eyes Task* was significantly related to vocabulary knowledge in adults, regardless of coding, supporting the notion that open-ended tasks require a certain degree of verbal ability to generate answers, though this counters the findings with children herein) (see Appendix B for a discussion). Importantly, contrary to expectations, performance on the forced-choice, valence-based *Eyes Task* was *worse* than that of the forced-choice, term-specific *Eyes Task* in the adult sample when this type of valence-based coding was used. This could suggest that the heightened vocabulary knowledge of adults acts as a benefit in solving forced-choice tasks, contrary to it being a detriment in children (see both herein and Happé, 1995 for a discussion of the role of verbal ability in solving social perspective taking tasks). Research with children comparing the different types of valence-based coding would be needed to make direct comparisons.

A fifth limitation is the lack of standardization and information on the sample of children with learning disabilities in Study 3. First, the measure by which the school officials assessed social-emotional functioning was not standardized and thus it is difficult to ascertain exactly what deficits the children had. As I did not include a measure of affective perspective taking in

this study, I could not confirm that the stated deficit in affective perspective taking was present, hence the use of the term "at-risk" herein. Second, although I was informed that students were diagnosed with a specific learning disorder, I was not given information about the exact diagnosis or any possible concurrent problems (although the parent information sheet did ask for any other diagnosis or suspected diagnosis and there were none listed outside of social problems and the learning disability). As learning disabilities often co-occur with both externalizing (e.g., Willcutt & Pennington, 2000) and internalizing problems (Gallegos, Langley, & Villegas, 2012; Mammarella, Ghisi, Bomba, Bottesi, Caviola, et al., 2014), it is possible that there were other factors that may have influenced performance in this group. If possible, future research should confirm the diagnosis and include specific, standardized, norm-referenced measures of known comorbidities in order to better ascertain the broader applicability of the findings to other children with learning disabilities.

A final limitation is the lack of an adult sample with which to make comparisons. There is good reason to believe that some of these methodological issues would be either more or less enhanced in an adult sample. For example, adults may be more prone to use alternate strategies to solve forced-choice tasks as they have had more experience with forced-choice formats (e.g., in exams) in which alternate strategies are often encouraged. In contrast, they may be less likely to show significant differences in performance due to vocabulary demands if a threshold for vocabulary has already been met. That is, the relationship between vocabulary knowledge and term-specific coding may only hold when there are large differences in vocabulary knowledge, and these may be reduced in older samples. The work by Peterson and Miller (2012) suggest this is *not* the case as they found large, significant relationships between verbal ability and performance on the *Reading the Mind in the Eyes* task, but more work is needed to confirm this.

### 5.3.2 Future Directions

There are several areas of research that naturally follow from the findings herein. The first, one I have already begun to undertake, is the examination of the same three questions of interest in an adult population in order to more fully assess the role of development. Preliminary research I conducted on adults examining the forced-choice and open-ended *Eyes Task* (and the various types of coding) with reference to the same three methodological questions of interest has found support for some of these hypotheses (see Appendix B for more information on the measures and findings). Given concerns about the online recruitment method employed and the validity of this dataset more generally this study was not included in this dissertation and preliminary findings should be interpreted with caution. In this preliminary adult study, when term-specific coding was used, the forced-choice *Eyes Task* not only resulted in greater performance than in the open-ended *Eyes Task* (using a within-subject methodology), but the difference in performance was highly influenced by non-verbal cognitive ability (as assessed by *Raven's Matrices*). This is in line with the belief that adults who have had more experience using alternate strategies are more prone to use them whenever possible. Furthermore, although there were significant relationships to vocabulary knowledge in adults for the forced-choice, term-specific *Eyes Task* and the open-ended *Eyes Task* (regardless of coding), the *differences* in performance by coding scheme (e.g., between the forced-choice, term-specific *Eyes Task* and forced-choice, valence-based *Eyes Task*) were *unrelated* to the differences in vocabulary knowledge, suggesting that although adults are using vocabulary knowledge to solve nearly *all* of the tasks, the *degree* to which they are using these vocabulary skills does not result in the type of task demands seen in the children herein. Therefore, it is possible that whereas some issues become more prominent with age, others may become less so as various abilities develop and

pass certain thresholds. As the *Reading the Mind in the Eyes* task is one measure that can be used across development—from childhood to adulthood—research with adults will provide a clear trajectory of how task demands and any changes to the task influence performance across this time span.

A second area of future research is to examine these same methodological issues with other measures of social perspective taking to determine if they have the same influence. Making these changes in specific cognitive and affective perspective taking measures will help us understand if these changes are relevant to all social perspective taking measures or only those, like the *Reading the Mind in the Eyes* task, that claim to be generalized social perspective taking measures. Importantly, testing this in other measures should not just be done in those that are forced-choice and term-specific in coding, but in those that may be open-ended and/or include valence-based coding.

A third area of future research is to examine the role of these methodological issues in other clinical populations with known or assumed social perspective taking deficits. The findings herein that children with learning disabilities did not show the same effect of vocabulary knowledge and cognitive ability as typically-developing children suggests that their learning disability (specifically in the verbal realm) may influence the use of these strategies. It is possible that different clinical populations would respond differently to the different methodologies based on the nature of their disorder or disability or even personality traits. For example, individuals high on psychopathic traits may show an even greater effect of using alternate strategies because the use of manipulation (which is well-documented in those with higher psychopathic tendencies) often requires higher cognitive abilities. Additionally, examining the effects in adults with learning disabilities might help determine the degree to

which individuals with these disabilities learn to use alternate strategies for their perspective taking deficits.

A final important area for future research involves going *back* and re-examining the earlier literature bearing in mind the lessons learned from the current research. Of particular interest will be re-examining conflicting findings and systematically testing whether any of the three methodological issues shown to be of importance herein may explain these earlier conflicting findings. For example, as previously mentioned, affective perspective taking deficits have been found in both children and adults with high levels of psychopathic tendencies (e.g., Blair et al., 2005; Loney et al., 2003; for a review, see Blair, 2003), but some other research has failed to replicate these findings (e.g., Richell et al., 2003) which may be due to the researchers' choice of dependent measure. For example, based on the findings from my dissertation I would propose that Richell and colleagues (2003) question of interest (i.e., are there affective perspective taking deficits in those with higher levels of psychopathy?) be re-examined using the open-ended *Eyes Task* instead of the original *Reading the Mind in the Eyes* task.

## 5.4 Conclusions

In sum, the present research demonstrates that certain measurement and methodological issues can have a profound effect on the assessment of social perspective taking skills in children. Specifically, forced-choice response formats may be eliciting skills other than social perspective taking, muddying the waters of how we interpret findings from these particular measures. Additionally, measures that claim to be generalized measures of social perspective taking should not be used as measures of either cognitive or affective perspective taking. Unless these measures have been specifically validated for the assessment of the components of social perspective taking, researchers should not assume that a generalized measure would accurately

assess either cognitive or affective perspective taking. This is particularly true if the measure of interest has been validated by differentiating typically-developing individuals from those with deficits in both cognitive and affective perspective taking, such as those with Autism Spectrum Disorders, as it is often unclear what element of the task is leading to this differentiation. Although there is a logical reason to assume that cognitive and affective perspective taking are so intertwined that they can be combined in a more generalized assessment, this is not necessarily the case.

The gold standard in the development of many social perspective taking tasks has been to be able to identify individuals with Autism Spectrum Disorders due to their known social perspective taking deficits, but the present research indicates that this may need to change. Autism Spectrum Disorders, with a dual deficit in both cognitive and affective components of social perspective taking and limitations in the use of eye gaze, will never provide the type of nuanced information that should be a requirement of social perspective taking tasks. Ideally, these tasks should be able to identify subtle deficits, deficits in only one component of social perspective taking, and also provide important individual differences information in addition to identifying those who are at the lowest end of the social perspective taking spectrum.

When researchers consider the development of social perspective taking tasks, I hope that the concerns raised by the present research will be heeded. The specificity of the task with respect to what type of social perspective taking is measured and the effects of response format on performance, including the additional task demands in the form of vocabulary knowledge to understand the options, are important considerations that have been largely overlooked to-date. It is critical that researchers bear these methodological issues in mind when interpreting findings (both old and new) and when choosing (or developing) measures in order to provide an accurate

picture of the nature, development, and individual differences in social perspective taking and its components. Moving forward, it is my hope that researchers will take these findings to heart so we may improve our understanding of this incredibly important construct that influences our social development and well-being.

## References


Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67,* 635-645.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition.* Washington, DC: American Psychiatric Publishing, Inc.

Apperly, I. (2011). *Mindreaders: The Cognitive Basis of "Theory of Mind".* East Sussex, UK: Psychology Press.

Astington, J. W., & Gopnik, A. (1991). Theoretical explanations of children's understanding of the mind. *British Journal of Developmental Psychology, 9,* 7-31. DOI: 10.1111/j.2044-835X.1991.tb00859.x

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35,* 1311-1320. DOI: 10.1037/0012-1649.35.5.1311

Astington, J. W., Harris, P. L., & Olson, D. R. (1988). *Developing Theories of Mind*. Cambridge, UK: Cambridge University Press.

Atance, C. M., & O'Neill, D. K. (2004). Acting and planning on the basis of a false belief: Its effects on 3-year-old children's reasoning about their own false beliefs. *Developmental Psychology, 40,* 953-964.

Axelrod, L. (1982). Social perception in learning disabled adolescents. *Journal of Learning Disabilities, 15,* 610-613. DOI: 10.1177/002221948201501011

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of normal cognition and moral action. *New Directions for Child and Adolescent Development, 103,* 37-49. DOI: 10.1002/cd.96

Baldwin, D. A. (1993). Early referential understanding: infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29,* 832-843. DOI: 10.1037/0012-1649.29.5.832

Baldwin, D.D. & Moses, L.J. (1994). Early Understanding of Referential Intent and Attentional Focus: Evidence from Language and Emotion. In C. Lewis & P Mitchell (Eds.) *Children's Early Understanding of Mind: Origins and Development.* (pp. 133-156). East Sussex, UK: Lawrence Earlbaum Associates Ltd.

Baron-Cohen, S. (1997). *Mindblindness: An Essay on Autism and Theory of Mind.* MIT Press.

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. In S. Baron-Cohen, H. Tager-Flushberg, D. J. Cohen (Eds.) *Understanding other minds: Perspectives from developmental cognitive neuroscience (2nd ed.).* (pp. 3-20). New York, NY: Oxford University Press.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with Autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry, 38,* 813-822. DOI: 10.1111/j.1469-7610.1997.tb01599.x

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42,* 241-251.

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome. *Journal of Developmental and Learning disabilities, 5,* 47-78.

Bateman, A. W., & Fonagy, P. (2012). *Handbook of Mentalizing in Mental Health Practice.* Arlington, VA: American Psychiatric Publishing, Inc..

Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of mind through the ages: older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Experimental Aging Research, 27,* 481-502. DOI: 10.1080/0361073X.2011.619466

Besel, L. D. S., & Yuille, J. C. (2010). Individual differences in empathy: The role of facial expression recognition. *Personality and Individual Differences, 49,* 107-112. DOI: 10.1016/j.paid.2010.03.013

Bindemann, M., Burton, A. M., & Langton, S. R. H. (2008). How do eye gaze and facial expression interact? *Visual Cognition, 16,* 708-733. DOI: 10.1080/13506280701269318

Birch, S. A. J. (2005). When knowledge is a curse: children's and adults' reasoning about mental states. *Current Directions in Psychological Science, 14,* 25-29.

Birch, S. A. J., & Bernstein, D. M. (2007). What can children tell us about hindsight bias: a fundamental constraint on perspective-taking?. *Social Cognition, 25,* 98-113.

Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science, 14,* 283-286. DOI: 10.1111/1467-9280.03436

Birch, S. A. J., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences, 8,* 254-260. DOI: 10.1016/j.tics.2004.04.011

Blair, R. J. R. (2003). Neurobiological basis of psychopathy. *The British Journal of Psychiatry, 182,* 5-7. DOI: 10.1192/bjp.182.1.5

Blair, R. J. R., & Coles, M. (2000). Expression recognition and behavioural problems in early adolescence. *Cognitive Development, 15,* 421-434.

Blair, R. J. R., Budhani, S., Colledge, E., & Scott, S. (2005). Deafness to fear in boys with psychopathic tendencies. *Journal of Child Psychology and Psychiatry, 46,* 327-336. DOI: 10.1111/j.1469-7610.2004.00356.x

Blair, R. J. R., Colledge, E., Murray, L., & Mitchell, D. G. V. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal Child Psychology, 29,* 491-498.

Bloom, P. (2000). *How Children Learn the Meanings of Words.* MIT Press.

Bosacki, S. L., & Moore, C. (2004). Preschoolers' understanding of simple and complex emotions: Links with gender and language. *Sex Roles, 50,* 659-675.

Bosacki, S., & Astington, J. W. (1999). Theory of mind in preadolescence: relations between social understanding and social competence. *Social Development, 8,* 237-255. DOI: 10.1111/1467-9507.00093

Bowers, D., Blonder, L. X., & Heilman, K. M. (1991). The Florida affect battery. *Gainesville: Center for Neuropsychological Studies.*

Breen, E. K. (1993). Recall and recognition memory in Parkinson's disease. *Cortex, 29,* 91-102.

Brent, E., Rios, P., Happé, F., & Charman, T. (2004). Performance of children with Autism spectrum disorder on advanced theory of mind tasks. *Autism, 8,* 283-299. DOI: 10.1177/1362361304045217

Brosseau-Liard, P. E., & Birch, S. A. J. (2010). 'I bet you know more and are nicer too!': What children infer from others' accuracy. *Developmental Science, 13,* 772-778. DOI: 10.1111/j.1467-7687.2009.00932.x

Brosseau-Liard, P., Cassels, T., Birch, S. (2014). You seem certain but you were wrong before: Developmental change in preschoolers' relative trust in accurate versus confident speakers. PLoS One, DOI: 10.1371/journal.pone.0108308

Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Langton, S., McAuley, S., & Wright, R. (2000). Testing face processing skills in children. *British Journal of Developmental Psychology, 18,* 319-333. DOI: 10.1348/026151000165715

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development, 53,* 413-425. DOI: 10.1111/j.1467-8624.1982.tb01331.x

Butterworth, G., & Morissette, P. (1995). Onset of pointing and the acquisition of language in infancy. *Journal of Reproductive and Infant Psychology, 14*, 219-231. DOI: 10.1080/02646839608404519

Calev, A. (1984). Recall and recognition in chronic nondemented schizophrenics. *Journal of Abnormal Psychology, 93,* 172-177.

Carey, J. M., & Cassels, T. G. (2013). Comparing two forms of a childhood perspective-taking measure using CFA and IRT. *Psychological Assessment, 25,* 879-892.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195,* 312-314.

Carey, S., & Diamond, R., (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1,* 253-274. DOI: 10.1080/13506289408402302

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*, 315-330.

Carpenter, M., Call, J., Tomasello, M. (2002). Understanding "prior intentions" enables two-year-olds to imitatively learn a complex task. *Child Development, 73,* 1431-1441. DOI: 10.1111/1467-8624.00481

Carpenter, M., Call, J., Tomasello, M. (2005). Twelve- and 18-month olds copy actions in terms of goals. *Developmental Science, 8,* F13-F20. DOI: 10.1111/j.1467-7687.2004.00385.x

Carr, M. B., & Lutjemeier, J. A. (2005). The relation of facial affect recognition and empathy to delinquency in youth offenders. *Adolescence, 40,* 601-619.

Carvajal, F., Feráandez-Alcaraz, C., Rueda, M., & Sarrión, L. (2012). Processing of facial expressions of emotions by adults with Down syndrome and moderate intellectual disability. *Research in Developmental Disabilities, 33,* 783-790. DOI: 10.1016/j.ridd.2011.12.004

Cassels, T. G., & Birch, S. A. J. (2014). Comparisons of an open-ended vs. forced-choice 'mind reading task: Implications for measuring perspective-taking and emotion recognition. *PLoS One.* doi:10.1371/journal.pone.0093653

Chambers, J. R., & Davis, M. H. (2012). The role of the self in perspective-taking and empathy: Ease of self-simulation as a heuristic for inferring empathic feelings. *Social Cognition, 30,* 153-180. DOI: 10.1521/soco.2012.30.2.153

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: Evidence from the Empathy Quotient (EQ) and the "Reading the Mind in the Eyes" test. *Social Neuroscience, 1,* 135-148.

Charman, T., Carroll, F., & Sturge, C. (2001). Theory of mind, executive function and social competence in boys with ADHD. *Emotional and Behavioural Difficulties, 6,* 31-49. DOI: 10.1080/13632750100507654

Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders – a follow-up in later adult life. Cognitive, language, and psychosocial outcomes. *Journal of Child Psychology and Psychiatry, 46,* 128-149.

Craig, J. S., Hatton, C., Craig, F. B., & Bentall, R. P. (2004). Persecutory beliefs, attributions and theory of mind: Comparison of patients with paranoid delusions, Asperger's syndrome and healthy controls. *Schizophrenia Research, 69,* 29-33. DOI: 10.1016/S0920-9964(03)00154-3

D'Entremont, B. (2000). A perceptual-attentional explanation of gaze following in 3- and 6-month-olds. *Developmental Science, 3,* 302-311. DOI:10.1111/1467-7687.00124

D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development, 20,* 569-572. DOI: 10.1016/S0163-6383(97)90048-5

Dadds, M. R., Perry, Y., Hawes, D. J., Merz, S., Riddell, A. C., Haines, D. J., Solak, E., & Abeygunawardane, A. I. (2006). Attention to the eyes and fear-recognition deficits in child psychopathy. *The British Journal of Psychiatry, 189,* 280-281. DOI: 10.1192/bjp.bp.105.018150

Davies, B., Frude, N., Jenkins, R., Harding, C., & Hill, C. (2014). A preliminary study into the relationship between emotional perception ability and challenging behaviour in adults with an intellectual disability. *Journal of Intellectual Disabilities, 18,* 382-392. DOI: 10.1177/1744629514554846

Davis, M., & Stone, T. (2003). Synthesis: Psychological understanding and social skills. In B. Repacholi & V. Slaughter (Eds.) *Individual differences in theory of mind. Macquarie monographs in cognitive science* (pp. 305-352). East Sussex, UK: Psychology Press.

Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience & Biobehavioral Reviews, 36,* 2288-2304.

Dawson, G., Toth, K., Abbott, R., Osterling, J., & Munson, J. (2004). Early social attention impairments in Autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology, 40,* 271-283. DOI: 10.1037/0012-1649.40.2.271

De Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience, 14,* 199-209.

De Rosnay, M., & Harris, P. L. (2002). Individual differences in children's understanding of emotion: the roles of attachment and language. *Attachment and Human Development, 4,* 39-54. DOI: 10.1080/14616730210123139

De Rosnay, M., Pons, F., Harris, P. L., & Morrell, J. M. B. (2004). A lag between understanding false belief and emotion attribution in young children: relationships with linguistic ability and mothers' mental-state language. *British Journal of Developmental Psychology, 22,* 197-218. DOI: 10.1348/026151004323044573

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in Cognitive Sciences, 7,* 527-533. DOI: 10.1016/j.tics.2003.10.004

DeMatteo, D., Heilbrun, K., & Marczyk, G. (2005). Psychopathy, risk of violence, and protective factors in a noninstitutionalized and noncriminal sample. *International Journal of Forensic Mental Health, 4,* 147-157. DOI: 10.1080/14999013.2005.10471220

Denham, S. A., Bassett, H. H., & Wyatt, T. M. (2010). Gender differences in the socialization of preschoolers' emotional competence. *New Directions for Child and Adolescent Development, 128,* 29-49. DOI: 10.1002/cd.267

Denham, S. A., Renwick-DeBardi, S., & Hewes, S. (1994). Emotional communication between mothers and preschoolers: Relations with emotional competence. *Merrill-Palmer Quarterly, 40,* 488-508.

Dickstein, E. B., & Warren, D. R. (1980). Role-taking deficits in learning disabled children. *Journal of Learning Disabilities, 13,* 33-37. DOI: 10.1177/002221948001300706

Dimberg, U., & Öhman, A. (1996). Behold the wrath: Psychophysiological responses to facial stimuli. *Motivation and Emotion, 20,* 149-182.

Dodge, K. A. (1980). Social cognition and children's aggressive behaviour. *Child Development, 51,* 162-170.

Dolan, M., & Fullam, R. (2006). Face affect recognition deficits in personality-disordered offenders: association with psychopathy. *Psychological Medicine, 36,* 1563-1569.

Dorris, L., Espie, C. A. E., Knott, F., & Salt, J. (2004) Mind-reading difficulties in the siblings of people with Asperger's syndrome: evidence for a genetic influence in the abnormal development of a specific cognitive domain. *Journal of Child Psychology and Psychiatry, 45,* 412-418.

Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers and their young children. *Developmental Psychology, 23,* 132-139.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: individual differences and their antecedents. *Child Development, 62,* 1352-1366.

Durand, K., Gallay, M., Seigneuric, A., Robichon, F., & Baudouin, J. Y. (2006). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology, 97,* 14-27. DOI: 10.1016/j.jecp.2006.12.001

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin, 94,* 100-131. DOI: 10.1037/0033-2909.94.1.100

Eisenberg, N., Shea, C. L., Carlo, G., & Knight, G. P. (1991). Empathy-related responding and cognition: A "chicken and the egg" dilemma. In W. M. Kurtines, J. Gewirtz, & J. L. Lamb (Eds.) *Handbook of Moral Behavior and Development: Volume 2: Research* (pp. 63-88). East Sussex, UK: Psychology Press.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1982). What are the similarities and differences in facial behavior across cultures. *Emotion in the Human Face, 2,* 128-144.

El-Haddad, C., & Laouris, Y. (2011). The ability of children with mild learning disabilities to encode emotions through facial expressions. In A. Esposito, A.M. Esposito, R. Martone, V.C. Müller, G. Scarpetta (Eds.), *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: Theoretical and Practical Issues, Volume 6456* (pp. 387-402). Berlin: Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-18184-9_34

Feinman, S. (1982). Social referencing in infancy. *Merrill-Palmer Quarterly, 28,* 445-470.

Feinman, S., & Lewis, M. (1983). Social referencing at ten months: A second-order effect on infants' responses to strangers. *Child Development, 54,* 878-887.

Flavel, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly, 50,* 274-290. DOI: 10.1353/mpq.2004.0018

Frith, U., Happé, F. G. E. (1994). Autism: Beyond "theory of mind". *Cognition, 50,* 115-132. DOI: 10.1016/0010-0277(94)90024-8

Gallegos, J., Langley, A., & Villegas, D. (2012). Anxiety, depression, and coping skills among Mexican school children: A comparison of students with and without learning disabilities. *Learning Disability Quarterly, 35,* DOI: 10.1177/0731948711428772

Glass, S. J., & Newman, J. P. (2006). Recognition of facial affect in psychopathic offenders. *Journal of Abnormal Psychology, 115,* 815-820.

Gopnik, A. (2009). *The Philosophical Baby: What Children's Minds Tell Us About Truth, Love, and the Meaning of Life.* London, UK: Bodley Head.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). *The Scientist in the Crib.* New York: William Morrow and Company.

Gordon, H. L., Baird, A. A., & End, A. (2004). Functional differences among those high and low on a trait measure of psychopathy. *Biological Psychiatry, 56,* 516-521.

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., & Hodges, J. R. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. *Brain, 125,* 752-764.

Gur, R. C., Erwin, R. J., Gur, R. E., Zwil, A. S., Heimberg, C., & Kraemer, H. C. (1992). Facial emotion discrimination: ii. behavioral findings in depression. *Psychiatry Research, 42*, 241-251. DOI: 10.1016/0165-1781(92)90116-K

Hadwin, J., & Perner, J. (1991). Pleased and surprised: Children's cognitive theory of emotion. *British Journal of Developmental Psychology, 9,* 215-234. DOI: 10.1111/j.2044-835X.1991.tb00872.x

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48,* 101-119. DOI: 10.1016/0010-0277(93)90026-R

Happé, F. G. E. (1994). An advanced test of social perspective taking: Understanding the story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24,* 129-154.

Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with Autism. *Child Development, 66,* 843-855. doi:10.1111/j.1467-8624.1995.tb00909.x

Harrigan, J. A. (1984). The effects of task order on children's identification of facial expressions. *Motivation and Emotion, 8,* 157-169.

Harris, P. L. (2006). Social cognition. In W. Damon & R. M. Lerner (Eds), *Handbook of child psychology: Volume 2, Cognition, perception, and language (6th ed.)* (pp. 811-858). Hoboken, NJ, US: John Wiley & Sons Inc. doi:10.1002/9780470147658.chpsy0219

Harris, P. L., De Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science, 14,* 69-73.

Hatfield, E., & Cacioppo, J. T. (1994). *Emotion Contagion.* Cambridge, UK: Cambridge University Press.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107,* 311.

Hinnant, J. B., & O'Brien, M. (2007). Cognitive and emotional control and perspective taking and their relations to empathy in 5-year-old children. *The Journal of Genetic Psychology, 168,* 301-322.

Hinnant, J. B., & O'Brien, M. (2007). Cognitive and emotional control and perspective taking and their relations to empathy in 5-year-old children. *The Journal of Genetic Psychology, 168,* 301-322.

Hoffman, M. L. (2000a). Empathy and prosocial behavior. In M. Lewis, J. M. Iaviland-Jones, L. F. Barrett (Eds.) *Handbook of Emotions* (pp. 440-455). New York, NY: The Guilford Press.

Hoffman, M. L. (2000b). *Empathy and Moral Development*. Cambridge: Cambridge University Press.

Holder, H. B., & Kirkpatrick, S. W. (1991). Interpretation of emotion from facial expression in children with and without learning disabilities. *Journal of Learning Disabilities, 24,* 170-177. DOI: 10.1177/002221949102400305

Hooker, C. I. Verosky, S. C., Germine, L. T., Knight, R. T., & D'Esposito, M. (2008). Mentalizing about emotion and its relationship to empathy. *Social Cognitive and Affective Neuroscience, 3,* 204-217. DOI:10.1093/scan/nsn019

Horowitz, E. (1981). Popularity, decentering ability, and role-taking skills in learning disabled and normal children. *Learning Disability Quarterly, 4,* 23-30. DOI: 10.2307/1510709

Hughes, C., Dunn, J., & White, A. (1998). Trick or treat? Uneven understanding of mind and emotion and executive dysfunction in 'hard-to-manage' preschoolers. *Journal of Child Psychology and Psychiatry, 39,* 981-994.

Ickes, W. J. (1997). *Empathic Accuracy.* New York, NY: The Guilford Press.

Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behaviour and academic competence in children at risk. *Psychological Science, 12,* 18-23.

Jackson, S. C., Enright, R. D., & Murdock, J. Y. (1987). Social perception problems in learning disabled youth: Developmental lag versus perceptual deficit. *Journal of Learning Disabilities, 20,* 361-364. DOI: 10.1177/002221948702000612

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science, 17,* 757-758. DOI:10.1111/j.1467-9280.2006.01778.x

Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology, 32,* 70-78. DOI:10.1037/0012-1649.32.1.70

Johnson, D. W. (1975). Cooperativeness and social perspective taking. *Journal of Personality and Social Psychology, 31,* 241-244.

Johnson, S., Slaughter, V., & Carey, S. (2002). Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. *Developmental Science, 1,* 233-238. DOI:10.1111/1467-7687.00036

Kaland, N., Callesen, K., Møller-Nielsen, A., Mortensen, E. L., & Smith, L. (2008). Performance of children and adolescents with Asperger Syndrome or high-functioning Autism on advanced theory of mind tasks. *Journal of Autism and Developmental Disorders, 38,* 1112-1123.

Kestenbaum, R., & Gelman, S. A. (1995). Preschool children's identification and understanding of mixed emotions. *Cognitive Development, 10,* 443-458. doi:10.1016/0885-2014(95)90006-3

Kestenbaum, R., & Nelson, C. A. (1990). The recognition and categorization of upright and inverted emotional expressions by 7-month-old infants. *Infant Behavior and Development, 13,* 497-511. doi:10.1016/0163-6383(90)90019-5

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11,* 32-38. DOI:10.1111/1467-9280.00211

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with Autism. *Archives in General Psychiatry, 59,* 809-816. DOI:10.1001/archpsyc.59.9.809

Knafo, A., Steinberg, T., & Goldner, I. (2011). Children's low affective perspective-taking ability is associated with low self-initiated pro-sociality. *Emotion, 11,* 194-198.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76,* 1261-1277. DOI:10.1111/j.1467-8624.2005.00849.x

Kohler, C. G., Walker, J. B. Martin, E. A. Healey, K. M., & Moberg, P. J. (2009). Facial emotion perception in schizophrenia: A meta-analytic review. *Schizophrenia Bulletin, 36,* 1009-1019. doi:10.1093/schbul/sbn192

Kolb, B., Wilson, B., & Taylor, L. (1992). Developmental changes in the recognition and comprehension facial expression: Implications for frontal lobe function. *Brain and Cognition, 20,* 74-84. doi:10.1016/0278-2626(92)90062-Q

Kosson, D. S., Suchy, Y., Mayer, A. R., & Libby, J. (2002). Facial affect recognition in criminal psychopaths. *Emotion, 2,* 398-411.

Kurdek, L. A. (1977). Structural components and intellectual correlates of cognitive perspective taking in first-through fourth-grade children. *Child Development, 48,* 1503-1511.

Kurdek, L. A. (1978). Perspective taking as the cognitive basis of children's moral development: a review of the literature. *Merrill-Palmer Quarterly of Behavior and Development, 24,* 3-28.

Kurdek, L. A., & Rodgon, M. M. (1975). Perceptual, cognitive, and affective perspective taking in kindergarten through sixth-grade children. *Developmental Psychology, 11,* 643-650. DOI:10.1037/0012-1649.11.5.643.

Laditan, B. (2013). *The Honest Toddler: A Child's Guide to Parenting.* New York, NY: Scribner.

Lalonde, C. E., & Chandler, M. J. (1995). False belief understanding goes to school: on the social-emotional consequences of coming early or late to a first theory of mind. *Cognition and Emotion, 9,* 167-185. DOI:10.1080/02699939508409007

Lamm, C., Batson, C., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience, 19,* 42-58. DOI:10.1162/jocn.2007.19.1.42

Lawrence, K., Kuntsi, J., Coleman, M., Campbell, R., & Skuse, D. (2003). Face and emotion recogniztion deficits in turner syndrome: a possible role for x-linked genes in amygdala development. *Neuropsychology, 17,* 89-39. DOI:10.1037/0894-4105.17.1.39

Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

Lee, L., Harkness, K., Sabbagh, M., & Jacobson, J. (2005). Mental state decoding abilities in clinical depression, *Journal of Affective Disorders, 86,* 247-258.

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science, 12,* 264-271. DOI:10.1111/k.1467-7687.2008.00758.x

Loney, B. R., Frick, P. J., Clements, C. B., Ellis, M. L., & Kerlin, K. (2003). Callous-unemotional traits, impulsivity, and emotional processing in adolescents with antisocial behavior problems. *Journal of Clinical Child & Adolescent Psychology, 32,* 66-80. DOI:10.1207/S15374424JCCP3201_07

Lucas, A. J., & Lewis, C. (2011). Preschoolers' understanding of expertise: Three-year-olds regain trust in previously inaccurate informants. Poster session presented at the *Biennial Meeting of the Society for Research in Child Development, Montreal, QC.*

Ludemann, P. M., & Nelson, C. A. (1988). Categorical representation of facial expressions by 7-month-old infants. *Developmental Psychology, 24,* 492-501.

Lui, J. H. L. (2014). Callous-unemotional traits and empathy deficits: Mediating effects of affective perspective-taking and facial emotion recognition. *Unpublished Master's Thesis, University of Southern Mississippi.*

Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development, 73,* 1073-1084. DOI:10.1111/1467-8624.00458

Mammarella, I. C., Ghisi, M., Bomba, M., Bottesi, G., Caviola, S., Broggi, F., & Nacinovich, R. (2014). Anxiety and depression in children with nonverbal learning disabilities, reading disabilities, or typical development. *Journal of Learning Disabilities.* DOI: 10.1177/0022219414529336

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience & Biobehavioral Reviews, 32,* 454-465.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255-260. doi:10.1016/S1364-6613(02)01903-4

Mayberry, M. L., & Espelage, D. L. (2007). Associations among empathy, social competence, & reactive/proactive aggression subtypes. *Journal of Youth and Adolescence, 36,* 787-798.

McIlwain, D. (2003). Bypassing empathy: A Machiavellian theory of mind and sneaky power. In B. Repacholi & V. Slaughter (Eds.) *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 39-66). New York, NY, US: Psychology, Press.

Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Development, 73,* 1715-1726. DOI:10.1111/1467-8624.00501

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31,* 838-850. DOI:10.1037/0012-1649.31.5.838

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78,* 622-646.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31,* 553-566. DOI:10.1068/p3339

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13,* 329-336.

Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development, 10,* 3-18. doi:10.1002/icd.239

Nilsen, E. S. & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58,* 220-249. DOI:doi:10.1016/j.cogpsych.2008.07.002

Nilsen, E. S., & Graham, S. A. (2012). The development of preschoolers' appreciation of communicative ambiguity. *Child Development, 83,* 1400-1415. DOI:10.1111/j.1467-8624.2012.01762.x

Ordetx, K. (2012). *Teaching Theory of Mind: A Curriculum for Children with High Functioning Autism, Asperger's Syndrome, and Related Social Challenges.* London, UK: Jessica Kingsley Publishers.

Oswald, P. A. (1996). The effects of cognitive and affective perspective taking on empathic concern and altruistic helping. *The Journal of Social Psychology, 136,* 613-623.

Oswald, P. A. (2002). The interactive effects of affective demeanor, cognitive processes, and perspective-taking focus on helping behavior. *The Journal of Social Psychology, 142,* 120-132.

Owen, A., Browning, M., & Jones, R. S. P. (2001). Emotion recognition in adults with mild-moderate learning disabilities. *Journal of Learning Disabilities, 5,* 267-281.

Paal, T., & Bereczkei, T. (2007). Adult theory of mind, cooperation, Machiavellianism: The effect of mindreading on social relations. *Personality and Individual Differences, 43,* 541-551. DOI:10.1016/j.paid.2006.12.021

Pearl, R. (1987). Social cognitive factors in learning-disabled children's social problems. In S.J. Ceci (Ed.), *Handbook of cognitive, social, and neuropsychological aspects of learning disabilities: Vol 2* (pp. 273-294). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perry, B. (2010). *Born for Love: Why Empathy is Essential—and Endangered.* New York, NY: HarperCollins Publishers.

Peterson, C. C., & Slaughter, V. (2009). Theory of mind (ToM) in children with autism or typical development: links between eye-reading and false belief understanding. *Research in Autism Spectrum Disorders, 3*, 462-473. DOI:10.1016/j.rasd.2008.09.007

Peterson, E., & Miller, S. F. (2012). The eyes test as a measure of individual differences: how much of the variance reflects verbal IQ? *Frontiers in Psychology, 3*, 1-6. DOI:10.3389/fpsyg.2012.00220

Pons, F., & Harris, P. L. (2004). Emotion comprehension between 3 and 11 years: developmental periods and hierarchical organization. *European Journal of Developmental Psychology, 1*, 127-152. DOI:10.1080/17405620344000022

Pons, F., Lawson, J., Harris, P. L., & De Rosnay, M. (2003). Individual differences in children's emotion understanding: effects of age and language. *Scandinavian Journal of Psychology, 44*, 347-353. DOI:10.1111/1467-9450.00354

Povinelli, D. J., Reaux, J. E., Bierschwale, D. T, Allain, A. D., & Simon, B. B. (1996). Exploitation of pointing as a referential gesture in young children, but not adolescent chimpanzees. *Cognitive Development, 12*, 423-461. DOI:10.1016/S0885-2014(97)90017-4

Printz, J. (2005). Imitation and moral development. In Hurley, S., & Chater, N. (Eds), *Perspectives on imitation: From cognitive neuroscience to social science, Volume 2: Imitation, Human Development, and Culture.* Cambridge, MA: MIT Press.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: evidence from 12- and 18-month-olds. *Developmental Psychology, 33*, 12-21.

Repacholi, B., & Slaughter, V. (2003). *Individual Differences in Theory of Mind: Implications for Typical and Atypical Development.* New York: Psychology Press.

Repacholi, B., Slaughter, V., Pritchard, M., & Gibbs, V. (2003). Theory of mind, machiavellianism, and social functioning in childhood. In B. Repacholi & V. Slaughter (Eds) *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 67-97). New York: Psychology Press.

Richell, R. A., Mitchell, D. G. V., Newman, C., Leonard, A., Baron-Cohen, S., & Blair, R. J. R. (2003). Social perspective taking and psychopathy: Can psychopathic individuals read the 'language of the eyes'? *Neuropsychologia, 41,* 523-526.

Root, A. K., & Denham, S. A. (2010). The role of gender in the socialization of emotion: Key concepts and critical issues. *New Directions for Child and Adolescent Development, 128,* 1-9. DOI:10.1002/cd.265

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7,* 532-547. DOI:10.1016/0010-0285(75)90021-3

Rourke, B. P. (1988). Socioemotional disturbances of learning disabled children. *Journal of Consulting and Clinical Psychology, 56,* 801-810. DOI: 10.1037/0022-006X.56.6.801

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology, 80,* 201-224.

Russell, J. A., & Paris, F. A. (1994). Do children acquire concepts for complex emotions abruptly? *International Journal of Behavioral Development, 17,* 349-365. DOI:10.1177/016502549401700207

Sagi, A., & Hoffman, M. L. (1976). Empathic distress in the newborn. *Developmental Psychology, 12,* 175-176.

Sally, D., & Hill, E. (2006). The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. *Journal of Economic Psychology, 27,* 73-97.

Sandvik, A. M., Hansen, A. L., Johnsen, B. H., & Laberg, J. C. (2014). Psychopathy and the ability to read the "language of the eyes": Divergence in the psychopathy construct. *Scandinavian Journal of Psychology, 55,* 585-592. DOI: 10.1111/sjop.12138

Saxe, R. (2013). The new puzzle of theory of mind development. In M. Banaji & S. Gelman (Eds.) *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us* (pp. 107-112). New York, NY: Oxford University Press.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Reviews in Psychology, 55,* 87-124.

Schepman, K., Taylor, E., Collishaw, S., & Fombonne, E. (2011). Face emotion procesing in depressed children and adolescents with and without comorbid conduct disorder. *Journal of Abnormal Child Psychology, 40,* 583-593. DOI:10.1007/s10802-011-9587-2

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development, 80,* 1172-1196. DOI:10.1111/j.1467-8624.2009.01324.x

Sharp, C. (2008). Theory of mind and conduct problems in children: Deficits in reading the "emotions of the eyes". *Cognition and Emotion, 22,* 1149-1158.

Sobel, D. M., & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development, 81,* 669-679. DOI:10.1111/j.1467-8624.2009.01422.x

Song, H-J., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology, 44,* 1789-1795.

Song, H-J., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition, 109,* 295-315. doi:10.1016/j.cognition.2008.08.008

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18,* 587-592. DOI:10.1111/j.1467-9280.2007.01944.x

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245-251. DOI:10.1037/0033-2909.87.2.245

Stone, V., Baron-Cohen, S., Young, A., & Calder, A. (1998). Patients with amygdalectomy show impairments in theory of mind. Cambridge: University of Cambridge.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18,* 580-586. DOI:10.1111/j.1467-9280.2007.01943.x

Sutton, J. (2001). Bullies: Thugs or thinkers? *The Psychologist, 14,* 530-534.

Sutton, J., Smith, P. K., & Swettenham, J. (1999). Bullying and 'theory of mind': A critique of the 'social skills deficit' view of anti-social behaviour. *Social Development, 8,* 117-127. DOI:10.1111/1467-9507.00083

Taylor, M., Cartwright, B. S., & Bowden, T. (1991). Perspective taking and theory of mind: Do children predict interpretive diversity as a function of differences in observers' knowledge? *Child Development, 62,* 1334-1351. DOI:10.1111/j.1467-8624.1991.tb01609.x

Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development, 10,* 201-224.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78,* 705-722.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences, 28,* 1-17. DOI:10.1017/S014525X05000129

Tonks, J., William, W. H., Frampton, I., Yates, P, & Slater, A. (2007a). Assessing emotion recognition in 9-15-year olds: preliminary analysis of abilities in reading emotion from faces, voices, and eyes. *Brain Injury, 21,* 623-629. DOI:10.1080/02699050701426865

Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007b). Reading emotions after child brain injury: A comparison between children with brain injury and non-injured controls. *Brain Injury, 21,* 731-739. DOI:10.1080/02699050701426899

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science, 15,* 194-197. doi:10.1111/j.0956-7976.2004.01503008.x

Tracy, J. L., & Robins, R. W. (2008). The automaticity of emotion recognition. *Emotion, 8,* 81-95.

Tretacosta, C. J., & Fine, S. E. (2010). Emotion knowledge, social competence, and behavior problems in childhood and adolescence: A meta-analytic review. *Social Development, 19,* 1-29. DOI:10.1111/j.1467-9507.2009.00543.x

Tur-Kaspa, H. (2002). Social cognition in learning disabilities. In B.Y.L. Wong, M. Donahue (Eds.), *The Social Dimensions of Learning Disabilities: Essays in Honor of Tanis Bryan* (pp. 11-32). Mawhaw, NJ: Lawrence Erlbaum Associates.

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental Psychology, 45,* 534-543.

Van Wingen, G. A., Van Eijndhoven, P., Tendolkar, I., Buitelaar, J., Verkes, J., & Fernandez, G. (2011). Neural basis of emotion recognition deficits in first-episode major depression. *Psychological Medicine, 41,* 1397-1405. DOI:10.1017/S0033291710002084

Walden, T. A., & Ogan, T. A. (1988). The development of social referencing. *Child Development, 59,* 1230-1240.

Walker, S. (2005). Gender differences in the relationship between young children's peer-related social competence and individual differences in theory of mind. *Journal of Genetic Psychology, 166,* 297-312.

Waterman, J. M., Sobesky, W. E., Silvern, L., Aoki, B., & McCauley, M. (1981). Social perspective-taking and adjustment in emotionally disturbed, learning-disabled, and normal children. *Journal of Abnormal Child Psychology, 9,* 133-148. DOI: 10.1007/BF00917863

Weiss, E. (1984). Learning disabled children's understanding of social interactions of peers. *Journal of Learning Disabilities, 17,* 612-614. DOI: 10.1177/002221948401701008

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35,* 245-275.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655-684. DOI:10.1111/1467-8624.00304

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology, 39,* 114-128.

Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development, 23,* 291-312.

Willcutt, E. G., & Pennington, B. F. (2000). Comorbidity of reading disability and attention-deficit/hyperactivity disorder: Differences by gender and subtype. *Journal of Learning Disabilities, 33,* 179-191.

Wittgenstein, L. (1953). *Philosophical Investigations*. New York, NY: MacMillan.

Wong, B., & Wong, R. (1980). Role-taking skills in normally achieving and learning disabled children. *Learning Disability Quarterly, 3,* 11-18.

Wood, P. M., & Kroese, B. S. (2007). Enhancing the emotion recognition skills of individuals with learning disabilities: A review of the literature. *Journal of Applied Research in Intellectual Disabilities, 20,* 576-579. DOI: 10.1111/j.1468-3148.2006.00355.x

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Brief Intellectual Assessment (WJ-III BIA).* Itasca: Riverside Publishing.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69,* 1-34. DOI:10.1016/S0010-0277(98)00058-4.

Yott, J., & Poulin-Dubois, D. (2012). Breaking the rules: Do infants have a true understanding of false belief? *British Journal of Developmental Psychology, 30,* 156-171. DOI:10.1111/j.2044-835X.2011.02060.x

Yuill, N. (1984). Young children's coordination of motive and outcome in judgements of satisfaction and morality. *British Journal of Developmental Psychology, 2,* 73-81. DOI:10.1111/j.2044-835X.1984.tb00536.x

Zahn-Waxler, C., & Radke-Yarrow, M. (1990). The origins of empathic concern. *Motivation and Emotion, 14,* 107-130.

# Appendices

## Appendix A: Correct Answers of the Forced-Choice, Term-Specific *Eyes Task* and Coding of the Open-Ended, Valence-Based *Eyes Task* (Full 28-Item Version)

Coding of the forced-choice, term-specific *Eyes Task*:

| Item | Term 1 | Term 2 | Term 3 | Term 4 |
|---|---|---|---|---|
| 1 | Hate | Surprised | **Kind** | Cross |
| 2 | Unkind | Cross | Surprised | **Sad** |
| 3 | **Friendly** | Sad | Surprised | Worried |
| 4 | Relaxed | **Upset** | Surprised | Excited |
| 5 | Feeling sorry | **Making somebody do something** | Joking | Relaxed |
| 6 | Hate | Unkind | **Worried** | Bored |
| 7 | Feeling sorry | Bored | **Interested** | Joking |
| 8 | **Remembering** | Happy | Friendly | Angry |
| 9 | Annoyed | Hate | Surprised | **Thinking about something** |
| 10 | Kind | Shy | **Not believing** | Sad |
| 11 | Bossy | **Hoping** | Angry | Disgusted |
| 12 | Confused | Joking | Sad | **Serious** |
| 13 | **Thinking about something** | Upset | Excited | Happy |
| 14 | Happy | **Thinking about something** | Excited | Kind |
| 15 | **Not believing** | Friendly | Wanting to play | Relaxed |
| 16 | **Made up her mind** | Joking | Surprised | Bored |
| 17 | Angry | Friendly | Unkind | **A bit worried** |
| 18 | **Thinking about something sad** | Angry | Bossy | Friendly |
| 19 | Angry | Daydreaming | Sad | **Interested** |
| 20 | Kind | Surprised | **Not pleased** | Excited |
| 21 | **Interested** | Joking | Relaxed | Happy |

| Item | Term 1 | Term 2 | Term 3 | Term 4 |
|---|---|---|---|---|
| 22 | Playful | Kind | Surprised | **Thinking about something** |
| 23 | Surprised | **Sure about something** | Joking | Happy |
| 24 | **Serious** | Ashamed | Confused | Surprised |
| 25 | Shy | Guilty | Daydreaming | **Worried** |
| 26 | Joking | Relaxed | **Nervous** | Sorry |
| 27 | Ashamed | Excited | **Not believing** | Pleased |
| 28 | Disgust | Hate | **Happy** | Bored |

Coding of the open-ended, valence-based *Eyes Task*:

| Category | Definition | Examples |
|---|---|---|
| Positive: | Any emotional term that conveys a pleasant state of affairs, regardless of intensity. | Happy, joyful, excited, eager, calm, relaxed, friendly, interested, surprised, satisfied, kind, grateful, sympathetic, silly, curious |
| Negative: | Any emotional term that conveys a negative state of affairs, excluding hostility or anger. | Sad, worried, upset, nervous, scared, guilty, bored, embarrassed, shy, anxious, unhappy, left out, lonely, depressed, frustrated |
| Hostile: | Any emotional term that denotes a strong, negative state that is oriented towards another individual. | Angry, jealous, mad, envious, suspicious, irritated, cross |
| Neutral: | Any term that denotes a mental state that is free of emotional valence. | Normal, serious, thoughtful, questioning, focused, unsure |

Answers (based on a combination of the correct *Eyes Task* response and the image):

| Item | *Eyes Task* Correct Response | Correct Category (open-ended, valence-based *Eyes Task*) |
|---|---|---|
| 1 | Kind | Positive |
| 2 | Sad | Negative |
| 3 | Friendly | Positive |
| 4 | Upset | Negative |
| 5 | Making somebody do something | Neutral |
| 6 | Worried | Negative |
| 7 | Interested | Positive |
| 8 | Remembering | Neutral |
| 9 | Thinking about something | Neutral |
| 10 | Not believing | Negative |
| 11 | Hoping | Positive |
| 12 | Serious | Neutral |
| 13 | Thinking about something | Neutral |
| 14 | Thinking about something | Neutral |
| 15 | Not believing | Negative |
| 16 | Made up her mind | Neutral |
| 17 | A bit worried | Negative |
| 18 | Thinking about something sad | Negative |
| 19 | Interested | Positive |
| 20 | Not pleased | Hostile |
| 21 | Interested | Positive |
| 22 | Thinking about something | Neutral |
| 23 | Sure about something | Neutral |
| 24 | Serious | Neutral |
| 25 | Worried | Negative |
| 26 | Nervous | Negative |
| 27 | Not believing | Negative |
| 28 | Happy | Positive |

## Appendix B: Preliminary Findings From Work With Adults

*The following is a write up of the participants, methods, and results of a study conducted with adult participants as a follow-up to the studies herein with children. The findings were not included as a full study as there were concerns over the validity of the data collection methods. I utilized Amazon's M-Turk method, which resulted in results (as you can see below) that counter very well-established findings, such as a failure to find a sex difference in levels of Autism Spectrum traits. I believe that this may be due to the way in which data is collected, with individuals rushing through the tasks and possibly not giving it their full attention. Although I took care to recruit only individuals who were proficient English speakers, from English-speaking countries, and who had a good rating with the system as well as removing individuals who either took too short a time to reasonably have completed the task with care and consideration or who failed the validity checks, the concerns remained and as such, the study was not included in the main text.*

### B.1 Methods

#### B.1.1 Ethics Statement

All participants provided consent prior to participation in the study. The study was approved by the Behavioural and Research Ethics Board of the University of British Columbia.

#### B.1.2 Participants

A total of 245 adults took part in the study (152 female; 91 male; 2 declined to answer). The average age was 35.13 years (SD=12.56 years), but the range was large (18- to 72-years of age). The sample was predominantly Caucasian (n=181) followed by African-descent (n=25) and Hispanic (n=20), with the remaining comprised of Asian Pacific Islander, Native American, East Indian, Middle Eastern, and Other. The sample was generally well-educated with only 14.2% reporting their highest level of education as a high school diploma or below and 11.4% had a Master's Degree or higher. A sizeable minority (n=56) have been diagnosed with a psychological disorder or developmental delay. One of the diagnoses was on the Autism Spectrum (and an additional two individuals suspected an Asperger's diagnosis) and nearly all

were mood disorders (e.g., depression, anxiety), though there were also diagnoses of attention deficit (and hyperactivity) disorder (n=5), learning disabilities (verbal and non-verbal) (n=7), obsessive-compulsive disorder (n=1), and personality disorders (n=3). An additional 25 participants took part but were excluded for completing the entire task in too short a time, raising questions about the validity (9), for failing the validity checks during the study (4), or for failing to complete the entire study (12).

### B.1.3 Measures

#### B.1.3.1 Perspective Taking

*Reading the Mind in the Eyes* (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). This task requires participants to choose the most appropriate term to describe the thoughts or feelings of others based upon still photographs of the eye-region of the face. Participants choose from four possible terms; for example, item 3 has the options of 'joking', 'flustered', 'desire', and 'convinced' (correct answer is 'desire'). Half of the participants received the traditional version (herein referred to as the forced-choice, term-specific *Eyes Task*) whereas the other half received a valence-based version (herein referred to as the forced-choice, valence-based *Eyes Task*) in which the specific terms were replaced with valenced options (i.e., 'positive', 'negative', 'neutral', and 'hostile'). This was done as the data with children suggested that the specific terms utilized in the forced-choice task may have influenced the findings even when coding is valenced. Participants received half of the items in this format so that the other half could be given in the open-ended, generative format.

*Open-Ended Eyes Task* (Cassels & Birch, 2014; Carey & Cassels, 2013). This task was created using the stimuli from the standard task. It is identical except that it requires participants

to generate their own answer when asked "What is this person feeling?". Answers were coded as correct if the participants' answer matched the valence (i.e., positive, neutral, negative, or hostile) of the correct response from the forced-choice format (herein referred to as the open-ended, valence-based *Eyes Task*) or based on whether or not the participants' answer matched the term (or a synonym of the term) from the forced-choice format (herein referred to as the open-ended, term-specific *Eyes Task*). The coding for the open-ended *Eyes Task* was completed by two coders who overlapped for 82 cases (33% of the sample). The two were blind to the other person's scores and the participants' scores on every other measure. The interrater reliability was .80 when comparing all codes and improved to .86 when converted to a three-item scale (i.e., didn't match, matched valence, and matched term).

#### B.1.3.2 Dispositional Empathy

*The Interpersonal Reactivity Index (Davis, 1980, 1983).* The Interpersonal Reactivity Index (*IRI*) is a 28-item self-report measure of dispositional empathy. The items are rated on a 5-point scale from "Does not describe me well" to "Describes me very well". The measure is comprised of four 7-item subscales that measure specific aspects of empathy, including Perspective Taking (i.e., the tendency to adopt another person's point of view in life), Fantasy (i.e., the tendency to transpose oneself into the feelings and actions of fictional characters), Empathic Concern (i.e., the tendency to experience compassion and concern for other individuals), and Personal Distress (i.e., the tendency to experience self-oriented feelings of concern or discomfort in reaction to another person's distress). In the current sample, the internal reliability was also satisfactory with alphas of .62 (Fantasy), .77 (Perspective taking), .85 (Empathic Concern), and .75 (Personal Distress). Notably, during development, females scored

higher than males on all subscales and this was replicated in the current sample, thus sex will be an additional factor when examining relationships to dispositional empathy.

#### B.1.3.3 Autistic Traits

*The Adult Autism Spectrum Quotient (AQ) (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001).* The *AQ* is a 50-item self-report measure to assess levels of Autism Spectrum traits and provides a single score (out of 50) with higher scores indicative of greater levels of traits associated with the autism spectrum. Each item is rated on a four-point scale from "Definitely Agree" to "Definitely Disagree" but items are scored based on any level of agreement or disagreement (i.e., individuals who report "definitely agree" and "somewhat agree" would each receive the same score for a given item). A score of 32+ is considered a high score, indicative of many traits associated with the autism spectrum and intermediate levels are measured by scores of 20+. The scale is comprised of questions pertaining to five separate areas known to be affected by autism spectrum traits: social skills, attention switching, attention to detail, communication, and imagination. In the current sample, the internal reliability for the full scale was adequate at $\alpha$=.56.

#### B.1.3.4 Vocabulary Knowledge

*The UBC Word Measure*. The UBC Word Measure is a timed measure of vocabulary knowledge. The measure consists of 100 vocabulary words and four possible choices for the synonym and the difficulty level increases as the individual continues on. Participants are given eight minutes total for the measure. In the current study, the page was timed with the time noted on the top of the page so participants were aware of how much time remained. At eight minutes,

the page closed, saving all completed answers. Participants received a score out of 100, with higher scores indicative of greater vocabulary knowledge.

#### B.1.3.5 Non-Verbal Cognitive Ability

*Raven's Advanced Progressive Matrices, Set 1 (Raven, Raven, & Court, 1998).* Set 1 of the Advanced Progressive Matrices was administered to all participants with the noted time limit of twelve minutes. The items comprise of 12 images for which part of the image is missing and participants must select the missing piece from eight options. The matrices have been validated as measure of non-verbal reasoning, specifically analytic intelligence, and can be used across ages. The maximum possible score is 12.

#### B.1.3.6 Demographics

Participants were also asked for various demographic information including, but not limited to, age, sex, ethnicity, education, employment status, and mental health status.

### B.1.4 Procedure

Participants were recruited using Amazon's M-Turk method. In this, individuals select which studies they would like to partake in for a nominal fee. The study was administered using Inquisit, an online software that is based in Canada. Participants first read through the online consent form and were prompted to either exit if they did not agree or to continue to participate. Participants were then administered the open-ended *Eyes Task*, followed by either the forced-choice, term-specific *Eyes Task* or forced-choice, valence-based *Eyes Task*. The open-ended *Eyes Task* was always administered first as previous research has found that the administration of a forced-choice task prior to an open-ended task can lead to individuals simply echoing terms from the forced-choice version (Harrigan, 1984). This was followed by the *AQ*, then the *IRI*, and

finally the intelligence measures (vocabulary knowledge and non-verbal cognitive ability, respectively). Demographics were asked at the very end. Participants then read a debriefing form which included contact information and once their answers were reviewed, they were awarded $1.75 in compensation.

## B.2 Results

### B.2.1 Preliminary Results

As the design included a between-subjects comparison for the forced-choice, term-specific *Eyes Task* and forced-choice, valence-based *Eyes Task* (with n's of 124 and 121, respectively), a comparison on the remaining variables was necessary, even though the assignment to the groups was randomized. There was one significant difference on *Raven's Matrices*, $t(243)=2.854$, $p=.005$, with the participants who received the forced-choice, valence-based *Eyes Task* outperforming those who received the forced-choice, term-specific *Eyes Task*. There were no other significant differences, all $p$'s$>.28$.

Age and sex were examined with respect to all variables. There were expected sex differences on the components of the *IRI*, with females outperforming males, but no other significant differences by sex. See Table B1. Age was generally unrelated to the variables of interest with some notable exceptions. Age was significantly correlated with the *IRI* Fantasy subscale, $r=-.158$, $p=.013$, suggesting younger individuals are more likely to place themselves in the minds of characters than older individuals. Age was also significantly correlated with the two measures of intelligence, in expected directions; age was positively correlated with verbal ability, $r=.278$, $p<.001$, and negatively correlated with *Raven's Matrices*, $r=-.128$, $p=.045$, in

line with previous research showing a decline in performance on *Raven's Matrices* with age (Raven et al., 1998). No other correlations were significant, all *p's*>.32.

**Table B1.** Comparisons of the Key Variables by Sex.

| | **Males** | **Females** | **t-value** |
|---|---|---|---|
| *IRI* Fantasy | 16.36 | 19.89 | 4.580*** |
| *IRI* Perspective taking | 17.81 | 19.09 | 1.776* |
| *IRI* Empathic Concern | 18.19 | 21.89 | 5.106*** |
| *IRI* Personal Distress | 9.02 | 12.06 | 4.129*** |
| open-ended, valence-based *Eyes Task* | 7.69 | 7.79 | .360 |
| open-ended, term-specific *Eyes Task* | 2.53 | 2.64 | .592 |
| forced-choice, term-specific *Eyes Task* | 12.47 | 13.14 | 1.284 |
| forced-choice, valence-based *Eyes Task* | 7.04 | 7.04 | .007 |
| *AQ* | 19.74 | 18.38 | 1.495 |
| Verbal Ability | 54.22 | 53.45 | .352 |
| *Raven's Matrices* | 8.44 | 8.39 | .116 |

Note: *$p$<.10, **$p$<.05, ***$p$<.001

### B.2.2 Comparisons Between the *Eyes Tasks* and the IRI

In order to examine what type of perspective taking is being measured, comparisons to both the measure of dispositional empathy and to the *AQ* were performed. As with children, the hypothesis was that a relationship to dispositional empathy – especially the empathic concern and personal distress aspects – would reflect an assessment of affective perspective taking. As there was a significant sex effect for scores on dispositional empathy, the analyses were done individually for each sex. Furthermore, age was controlled for in analyses with the *IRI* Fantasy subscale. For males, the subscales of the *IRI* were *unrelated* to the open-ended, valence-based *Eyes Task* and the forced-choice *Eyes Task* (regardless of coding), all p's>.17; however, the open-ended, term-specific *Eyes Task* was significantly correlated with the *IRI* Personal Distress

subscale, $r$=-.227, $p$=.031, suggesting that greater performance on the open-ended, term-specific *Eyes Task* was related to lower levels of personal distress. For females, none of the subscales were correlated with any of the perspective taking measures, all *p's*>.07. This counters the data from children which found significant relationships to dispositional empathy for the open-ended *Eyes Task*.

### B.2.3 Comparisons Between the *Eyes Tasks* and the AQ

When examining the relationships to the *AQ*, two different analyses were conducted: One with the continuous measure and one predicting moderate or severe cut-off scores. As individuals with traits associated with the autism spectrum are known to have both cognitive and affective perspective taking difficulties, any differences could be reflective of the assessment of either type of perspective taking; however, the failure to find significant relationships with dispositional empathy could suggest a relationship to cognitive perspective taking. For the continuous analysis, the *AQ* was unrelated to the forced-choice, term-specific *Eyes Task*, $r$=-.027, $p$=.770, suggesting it does not discriminate levels of autism spectrum traits. The *AQ* was, however, significantly correlated with the open-ended, valence-based *Eyes Task*, $r$=.139, $p$=.029, suggesting that higher scores on the open-ended, valence-based *Eyes Task* were indicative of *higher* levels of autism spectrum traits, yet the overall contribution was very small, accounting for only 1.9% of the variance. The relationships to the forced-choice, valence-based *Eyes Task* and open-ended, term-specific *Eyes Task* were non-significant, both *p's*>.12.

The *Reading the Mind in the Eyes* task was developed by the ability to discriminate those at the higher end of the *AQ* and thus a similar analysis is conducted herein. Individuals were split according to their score on the *AQ* in two ways: Those higher than 20 (moderate level of autism spectrum traits) and those higher than 32 (high level of autism spectrum traits) (the cut-

offs were determined by Baron-Cohen and colleagues in the development of the *AQ*; Baron-Cohen, Wheelwright, Skinner, et al., 2001). For the moderate split, there were 56.7% of the participants below the split, and 43.3% above. For the high split, 93.5% were below the cut-off whereas only 6.5% were above. As was done in Study 3 with children, a binary logistic regression was performed predicting the *AQ* group from the different perspective taking scores. Notably, sex, dispositional empathy, verbal ability, and *Raven's Matrices* were included as covariates when relevant. There is generally a sex difference in higher levels of the *AQ* with males being disproportionately represented (Baron-Cohen, Wheelwright, Hill, et al., 2001) so it was included in all regressions, whereas dispositional empathy, verbal ability, and *Raven's Matrices* were all significantly different between the groups using the high cut-off and dispositional empathy was significantly different between the groups using the moderate cut-off. (Note that four separate regressions were conducted including the individual perspective taking scores as the forced-choice, term-specific *Eyes Task* and forced-choice, valence-based *Eyes Task* were between-subject variables.)

For the regressions predicting groups using a moderate cut-off, only the forced-choice, valence-based *Eyes Task* was nearly significant as an individual predictor above and beyond sex and dispositional empathy, $p$=.091, with a unit increase in the forced-choice, valence-based *Eyes Task increasing* the odds of being above the cut-off by 22.4%. The results were similar when predicting group membership using a high cut-off, with only the forced-choice, valence-based *Eyes Task* being nearly significant as an individual predictor above and beyond sex, dispositional empathy, verbal ability, and non-verbal cognitive ability, $p$=.078, with a unit increase in the forced-choice, valence-based *Eyes Task increasing* the odds of being above the cut-off by 318.1%. For the full results, see Table B2.

**Table B2.** Regressions Predicting Moderate and High Cut-Offs for the AQ.

| Predicting Group Membership Using the Moderate Cut-Off | | | |
|---|---|---|---|
| | **B** | **Sig.** | **Exp(B)** |
| Open-ended, valence-based *Eyes Task* | .107 | .140 | 1.113 |
| Open-ended, term-specific *Eyes Task* | .045 | .674 | 1.046 |
| Forced-choice, term-specific *Eyes Task* | .050 | .509 | 1.051 |
| Forced-choice, valence-based *Eyes Task* | .202 | .091 | 1.224 |
| **Predicting Group Membership Using the High Cut-Off** | | | |
| | **B** | **Sig.** | **Exp(B)** |
| Open-ended, valence-based *Eyes Task* | .271 | .101 | 1.311 |
| Open-ended, term-specific *Eyes Task* | .089 | .723 | 1.093 |
| Forced-choice, term-specific *Eyes Task* | .172 | .554 | 1.188 |
| Forced-choice, valence-based *Eyes Task* | 1.157 | .078 | 3.181 |

Note: Exp(B) refers to the odds-ratio; any number above 1 is indicative of an increased risk of scoring above the cut-off.

### B.2.4 Performance by Response Format and Relationships to Alternate Cognitive Abilities

In order to examine the effect of response format, the scores for the open-ended *Eyes Task* were compared to the forced-choice *Eyes Task* (between-subjects), controlling for coding scheme, using paired-samples t-tests. For the term-specific coding, the findings were in line with the previous studies with kids and the forced-choice version was associated with greater performance, $t(123)=40.597$, $p<.001$. For the valence-based coding, however, the findings were different. Specifically, the open-ended version was easier, $t(120)=3.858$, $p<.001$, but the means

were very similar ($M_{open\text{-}ended,valence\text{-}based}$=7.91, $M_{forced\text{-}choice,\ valence\text{-}based}$ =7.05). Thus, the effect of response format seems most influential when the coding is term-specific.

The original hypothesis was that the forced-choice *Eyes Task* could enable the use compensatory strategies in certain individuals, masking deficits in performance. For the term-specific coding, this hypothesis is tenable in adults; however, the results for a valence-based coding suggest it is not tenable. To determine if the difference in performance is linked to non-verbal cognitive ability, first the individual scores for the open-ended and forced-choice versions were correlated with *Raven's Matrices*. The forced-choice, term-specific *Eyes Task*, as expected, was significantly correlated with *Raven's Matrices*, $r$=.288, $p$=.001, whereas the open-ended, term-specific *Eyes Task* was not, $r$=-.011, $p$=.901. The difference between the correlations was highly significant, $z$=2.564, $p_{1\text{-}tail}$=.005. Additionally, the difference score between the forced-choice, term-specific *Eyes Task* and open-ended, term-specific *Eyes Task* was also significantly correlated with *Raven's Matrices*, $r$=.278, $p$=.002. The difference and the direction of the relationships (i.e., both higher scores on the forced-choice, term-specific *Eyes Task* and greater differences between the forced-choice, term-specific *Eyes Task* and open-ended, term-specific *Eyes Task* associated with higher scores on *Raven's Matrices*) suggest that performance on the forced-choice, term-specific *Eyes Task* may be influenced by the ability to use non-verbal cognitive abilities. When the coding is valenced, however, the opposite pattern emerged. The open-ended, valence-based *Eyes Task* was significantly related to *Raven's Matrices*, $r$=.233, $p$=.010, whereas the forced-choice, valence-based *Eyes Task* was not, $r$=.046, $p$=.617. This difference was significant, $z$=1.707, $p_{1\text{-}tail}$=.044, and the relationship between the difference in scores (i.e., the open-ended, valence-based *Eyes Task* minus the forced-choice, valence-based *Eyes Task*) was nearly significant, $r$=.168, $p$=.065. Thus, when the coding is valenced, it would

seem that non-verbal cognitive ability is associated with better performance in an open-ended response format in adults.

A second analysis, controlling for potential influences of other variables, utilized repeated-measure ANOVAs with the response format as the repeated measure, sex as the between-subject variable, and age, *AQ*, vocabulary knowledge, and *Raven's Matrices* as the covariates. For term-specific coding, the main effect of response format is significant, $F(1,118)=30.799$, $p<.001$ as is the interaction between response format and *Raven's Matrices*, $F(1,118)=6.635$, $p=.011$. This suggests that when term-specific coding is used, response format is influential in performance (with a forced-choice response format resulting in better performance) and the difference is related to non-verbal cognitive ability. For the valence-based coding, contrary to what was found using more simplistic methods, the main effect of response format was non-significant, $F(1,113)=1.835$, $p=.178$, and the interaction with Raven's was non-significant, $F(1,113)=2.202$, $p=.141$. Using more complex multivariate methods thus suggests that the effect of response format matters only for term-specific coding and that the difference is attributable at least in part to the use of non-verbal cognitive abilities.

### B.2.5 Performance by Coding Scheme and Relationship to Vocabulary Knowledge

In order to examine the influence of coding on performance and whether there is a role for vocabulary knowledge, paired-sample t-tests or independent sample t-tests were conducted (depending on type of response format). The original hypothesis was that term-specific coding limits performance as it requires a specific vocabulary threshold that many children may not have. The relevance to adults is thus to determine if a specific threshold is needed or if the majority of adults are well-enough above the threshold to make the issue moot.

For the forced-choice *Eyes Task*, an independent sample t-test was conducted as the administration of the forced-choice, term-specific *Eyes Task* or forced-choice, valence-based *Eyes Task* was done between subjects. Contrary to work with children, the forced-choice, valence-based *Eyes Task* was actually significantly *harder* than the forced-choice, term-specific *Eyes Task*, $t(221)=19.894$, $p<.001$. Additionally, the variances were unequal between the two groups, $F(124,121)=13.326$, $p<.001$, suggesting performance was less variable for the forced-choice, valence-based *Eyes Task* than the forced-choice, term-specific *Eyes Task*. Of note, the forced-choice, valence-based *Eyes Task* was different herein than with children; with children, the forced-choice, valence-based *Eyes Task* was calculated by accepting all answers that were of the same valence whereas herein the forced-choice, valence-based *Eyes Task* only included valenced answers as the options whereas herein the options were changed.

For the open-ended *Eyes Task*, a paired-samples t-test was conducted as all participants had scores for both types of coding. In line with the results with children, the valence-based coding resulted in a significant difference between term-specific and valence-based coding, $t(244)=45.408$, $p<.001$, with term-specific coding resulting in lower scores than for valence-based coding.

To examine the potential effect of vocabulary knowledge, the correlations between the two scores were compared. For the forced-choice response format, there was a significant relationship between vocabulary knowledge and the forced-choice, term-specific *Eyes Task*, $r=.274$, $p=.002$, but not for the forced-choice, valence-based *Eyes Task*, $r=.096$, $p=.293$; this difference was not statistically significant, but was trending towards significance, $z=1.43$, $p_{1\text{-}tail}=.076$. For the open-ended response format, both the open-ended, valence-based *Eyes Task*, $r=.194$, $p=.002$, and the open-ended, term-specific *Eyes Task*, $r=.201$, $p=.002$, were

significantly related to vocabulary knowledge and the difference score (i.e., the open-ended, valence-based *Eyes Task* minus the open-ended, term-specific *Eyes Task*) was unrelated to vocabulary knowledge, $r$=.066, $p$=.304. This suggests that the role of vocabulary knowledge is only truly removed in the forced-choice, valence-based *Eyes Task* for adults in which no specific term needs to be understood or generated.

As with the analyses looking at the effect of response format, I included analyses that examined the effect of vocabulary knowledge while controlling for other variables, including age, sex, *AQ*, and *Raven's Matrices*. For the forced-choice response format, a univariate ANOVA was conducted with perspective taking score as the dependent variable, coding and sex as fixed factors, and age, *AQ*, *Raven's Matrices*, vocabulary knowledge, and an interaction between vocabulary knowledge and coding as the covariates. There was a main effect of coding, $F(1,235)$=16.966, $p$<.001, with the term-specific coding leading to significantly higher scores, and a significant effect of vocabulary knowledge, $F(1,235)$=5.574, $p$=.019, suggesting that vocabulary knowledge does influence performance generally. The interaction between coding and vocabulary knowledge was nearly significant, $F(1,235)$=3.763, $p$=.054, even after controlling for these other variables, and the findings from the correlational analyses support the link between vocabulary knowledge and term-specific coding in the forced-choice response format.

For the open-ended *Eyes Task*, a repeated-measures ANOVA was conducted with the open-ended *Eyes Task* scores included as the repeated measures, sex was a between-subjects factor, and age, *AQ*, *Raven's Matrices*, and vocabulary knowledge were included as covariates. There was a significant effect of coding, $F(1,237)$=31.912, $p$<.001, with valence-based coding resulting in higher scores than term-specific coding. There was a main effect of vocabulary

knowledge, $F(1,237)=10.044$, $p=.002$, but no significant interaction with coding, $F(1,237)=.095$, $p=.759$, suggesting that the effects of vocabulary knowledge are significant for the open-ended task, regardless of coding method. Notably, there was a significant interaction with *Raven's Matrices*, $F(1,237)=5.059$, $p=.025$, yet no main effect, $F(1,237)=.348$, $p=.556$. Follow-up correlational analyses did find that the open-ended, valence-based *Eyes Task* was significantly related to *Raven's Matrices*, $r=.151$, $p=.018$, whereas the open-ended, term-specific *Eyes Task* was not, $r=.039$, $p=.544$, and the difference between the scores (i.e., open-ended, valence-based *Eyes Task* - open-ended, term-specific *Eyes Task*) was also significantly correlated with *Raven's Matrices*, $r=.142$, $p=.026$, suggesting that a greater difference between the open-ended, valence-based *Eyes Task* and open-ended, term-specific *Eyes Task* was associated with greater non-verbal cognitive ability. However, the effect of this relationship was tiny, accounting for 2% of the variance, and thus the practical importance is questionable. The results for the open-ended *Eyes Task*, taken as a whole, suggest that term-specific coding is more difficult, in line with hypotheses, but is not related to differences in vocabulary knowledge, contrary to hypotheses.