THE DEVELOPMENT OF MORAL EVALUATIONS IN CHILDREN AND ADULTS

by

KARLENA DIANE OCHOA

A DISSERTATION

Presented to the Department of Psychology
and the Division of Graduate Studies of the University of Oregon
In partial fulfilment of the requirements of
Doctor of Philosophy

September 2021

DISSERTATION APPROVAL PAGE

Student: Karlena Diane Ochoa

Title: The Development of Moral Evaluations in Children and Adults

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Psychology by:

Dr. Louis Moses	Co-chair
Dr. Kathryn Mills	Co-chair
Dr. Dare Baldwin	Core Member
Dr. Nicole Giuliani	Institutional Representative

and

Andrew Karduna	Interim Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded September 2021

DISSERTATION ABSTRACT

Karlena Diane Ochoa

Doctor of Philosophy

Department of Psychology

September 2021

Title: The Development of Moral Evaluations in Children and Adults

Although the influence of intent understanding on children's moral development has been long studied, limited research has examined the influence of belief understanding on that development. The purpose of this dissertation was to further investigate the interplay between children's mental state understanding, or Theory of Mind, and moral development during childhood. In two studies we presented children with morally-relevant belief vignettes to examine the extent to which they incorporate both intent and belief information in their moral judgments (judgments of moral intent, deserved consequences, praise or blame). We also examined how children's moral judgments compared to adults' judgments and whether individual differences in executive function and empathy are related to those judgments. In Study 1 (*N* = 109 children, *N* = 42 adults), 4- and 5-year-olds with false belief understanding, but not those without false belief understanding, were able to make appropriate intent judgments in situations in which the agent's intent did not align with the outcome. Yet, all children had difficulty making consequence judgments based on intent. In Study 2 (*N* = 61 children, *N* = 36 adults), 5- and 7-year-olds with false belief understanding were again able to make appropriate intent judgments and 7-year-olds did so at adult-like levels. Nonetheless children still did not differentiate their consequence judgments based on the agent's intent. That

said, children did assign blame and praise based on intent. Children's moral judgments differed from those of adults in several respects, indicating that moral reasoning develops substantially beyond the early school years. In assessing individual differences, we did not find evidence of a relation between executive function and moral judgments in Study 1, but there was an intriguing relation between children's empathy and their moral judgments in Study 2. Children with higher parent-reported empathy seemed to be influenced more by the outcome than by intent because they assigned a positive consequence to the agent when the outcome was good and a negative consequence when the outcome was bad. Overall, our findings suggest that integrating theory of mind and moral judgment is a multi-faceted developmental achievement that unfolds only gradually over childhood.

CURRICULUM VITAE

NAME OF AUTHOR: Karlena Diane Ochoa

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
California State University San Marcos, San Marcos, California

DEGREES AWARDED

Doctor of Philosophy, Psychology, University of Oregon, 2021
Master of Arts, Psychology, California State University San Marcos, 2016
Bachelor of Arts, Psychology, California State University San Marcos, 2014

AREAS OF SPECIAL INTEREST:

Developmental Psychology
Social Cognitive Development
Theory of Mind
Moral Development
Executive Function
Quantitative Methods

PROFESSIONAL EXPERIENCE:

Graduate Student Consultant, Center for Translational Neuroscience, University of Oregon, 2020-2021

Graduate Research Fellow, Developing Brains in Context Lab, University of Oregon, 2019-2021

Graduate Student Representative, Committee for an Inclusive Community, University of Oregon, 2019-2021

Graduate Teaching Fellow, Department of Psychology, University of Oregon, 2017-2021

Graduate Research Fellow, Developing Mind Lab, University of Oregon, 2016-2021

Graduate Student Outreach Coordinator, Team Duckling, University of Oregon. 2017-2019

Graduate Research Fellow, Research on Children's Knowledge of Social Thinking and Reasoning, California State University San Marcos, 2014-2016

GRANTS, AWARDS, AND HONORS:

Committee for an Inclusive Community Training Award, University of Oregon, 2021

Graduate Education Committee Research Award, University of Oregon, 2021

Carolyn M. Stokes Memorial Scholarship, University of Oregon, 2019

Committee for an Inclusive Community Travel Award, University of Oregon, 2019

People's Choice Award, Three Minute Thesis Competition, University of Oregon, 2017

Promising Scholar Award, Department of Psychology, University of Oregon, 2016

First Year Fellowship Award, Department of Psychology, University of Oregon, 2016

Research Initiative for Scientific Enhancement (RISE*R25) Graduate Fellowship Recipient, California State University San Marcos, 2014

Magna Cum Laude Honors, California State University San Marcos, 2014

PUBLICATIONS:

Ochoa, K. D., Rodini, F. J., & Moses, J. L. (invited revision). False Belief Understanding and Moral Judgment in Young Children. *Developmental Psychology*. https://doi.org/10.1111/bjdp.12247

Ochoa, K. D., & Vanderbilt, K. E. (invited revision). Children's selective trust based on sources' relative knowledge or intentions. *Infant and Child Development*.

Vanderbilt, K. E., & **Ochoa, K. D.**, Heilburn, J. (2018). Consider the source: Children link the accuracy of text-based sources to the accuracy of the author. British Journal of Developmental Psychology, 36, 634-651.

ACKNOWLEDGEMENTS

The completion of this dissertation would not have been possible without the incredible mentors, colleagues, friends, and family in my life. First and foremost, I would like to express my deepest appreciation for my co-advisors, Dr. Lou Moses and Dr. Kate Mills. Thank you both for your constant support and guidance, and for always believing in me. I owe many thanks to Dr. Louis Moses, whose expertise was invaluable in formulating the research questions and methodology for my dissertation. Dr. Kate Mills, I will be forever grateful that you took me on as your graduate student during my third year. Your enthusiasm and mentorship over the past few years is what allowed me to complete my PhD. I would also like to extend my sincere thanks to Dr. Dare Baldwin and Dr. Nicole Giuliani for your feedback and guidance on my dissertation. To Lori Olsen, thank you for keeping me on track, the yearly birthday chocolates, and for your personal support throughout my time in the program. To the members of the Committee for an Inclusive Community, thank you for engaging in meaningful conversations and for your commitment to improve our department in various ways.

I am also grateful to all the members of the Developing Mind Lab, Developing Brains in Context Lab, and the Acquiring Minds Lab, for their invaluable feedback and continuous support. A special thanks to my wonderful research assistants at the University of Oregon -none of this research could have been conducted without you. Finally, many thanks to all the children and families who spent time participating in our research at the University of Oregon, Eugene Science Center, and virtually during the pandemic.

For my friends and family who have seen it all, thank you for your unwavering support. A shout out goes to Sarah Donaldson, Sarah Horn, Melissa Moss, and Stephanie Gluck, for their invaluable friendship throughout this program.

A very special thanks goes to:

Mike, my success would not have been possible without you by my side.

Tara, for being the best friend anyone could ask for.

Alex, for being you and never letting me forget my worth.

Dad, for teaching me the value of education.

Mom, your courage and strength continue to live on through me.

Midnight and Bullseye, for unconditional love and snuggles.

This dissertation is dedicated to my brother.

Alex, this is for you - you are my inspiration every day.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Sometime during the preschool years, before formal schooling, most children begin to display understanding of their own and others' mental states such as desires, beliefs, intentions, and emotions. This Theory of Mind (ToM) allows them to understand, explain, and predict others' and their own behavior (Premack & Woodruff, 1978) and is critical for other domains of social functioning (Astington, 2003; Baron-Cohen et al., 1985; Leslie, 1988; Wellman, 2020). Without understanding others' mental states, it would be very difficult to understand the reasons for others' behavior (Astingon, 2003); in the absence of such understanding it would, for example, be very difficult to create and maintain friendships. In addition, ToM ability has been associated with peer acceptance (Slaughter et al., 2002), school adjustment (Brock et al., 2019), and more sophisticated pretend play (Taylor & Carlson, 1997). Further, deficits in ToM have been associated with developmental disorders, such as autism (Baron-Cohen et al., 1985; Leslie, 1988).

One domain in which ToM is critically important is that of moral reasoning and moral judgment (Baird & Astington, 2005; Smetana et al., 2012; Wainryb, & Ford, 1998). Research on children's moral judgments often involves asking a child to judge an action as right or wrong when the protagonist's desires and/or intentions are known, and this has been an area of interest at least since Piaget (1932). For example, a child may be told a story in which a protagonist intended to steal an apple from the grocery store, but it rolled out of the cart before leaving the store (Cushman et al., 2013). Then children are asked to make a judgment of "wrongness"

and deserved punishment for the protagonist. In order to reason about moral rightness or wrongness in a scenario like this, a child would need to consider the intention to steal the apple rather than focusing solely on the neutral outcome. Assessing mental states like intentions may thus allow children to make more sophisticated moral judgments.

Understanding others' beliefs, including false beliefs, is also important for making moral judgments. Research assessing children's understanding of false belief within a moral domain requires children to judge, for instance, whether a protagonist, who made a false statement, should be rewarded or punished. For example, in early research on this topic, Wimmer et al. (1984) presented children with a story in which a protagonist holds a false belief about the location of a desired object. Another character then asks the protagonist about the location of the object; because of the false belief she gives an incorrect answer, even though she was not mal-intentioned. The other character then expresses sadness about not being able to find the desired object. Children are then asked whether the protagonist was lying, and whether she should receive a reward or be punished (Wimmer et al., 1984). Being able to assess the well-intentioned protagonist's belief is important in this scenario in order to recognize that she should not be punished.

Although there has been a long history of research examining the relation between children's understanding of intention and moral judgments, more recently researchers have been examining the interplay between children's understanding of others' intention and knowledge and moral judgments. The purpose of my dissertation is first to review and highlight research on ToM and moral development

during the preschool years. In doing so, I briefly discuss relevant literature and themes for each of these topics separately, and then provide a more in-depth review of these two areas of research in relation to each other. Second, I present two empirical studies that I conducted to further examine the interplay between ToM and moral judgments during childhood and how these judgments compare to those of adults. Lastly, I discuss my findings and contributions to the literature.

Theory of Mind During Early Childhood

In what follows I briefly describe some central aspects of ToM development, including intention and belief understanding. It is important to note that some researchers suggest that an implicit ToM develops much earlier than the preschool years. For example, Hamlin (2013) suggests that 8-month-olds have a rudimentary understanding of intentions, which informs their social evaluations. Yet, there is still debate on whether measures of implicit ToM are truly assessing mental states (see Ruffman, 2014). For the purpose of this dissertation, however, I am primarily focusing on explicit ToM, which is defined as conscious and controlled thoughts or judgments of others' mental states (Low & Perner, 2012). I am focusing on explicit ToM because this reasoning, and moral reasoning skills, develop rapidly during the preschool years, influence each other, and relate to other important facets of development. That said, I acknowledge that there are likely important underpinnings to both ToM and morality earlier in development.

Understanding Others' Intentions

Children's explicit understanding of intentions is important for many reasons, including being able to predict and understand actions (Baldwin & Baird,

2001). In reviewing research on children's understanding of intentionality, it's important to note that there are several aspects that can be considered when deeming something as intentional (Malle, 1999; Malle & Knobe, 1997). Malle (1999) describes five conditions that need to be met in order for adults to judge an actor's behavior as intentional: a) the person must have a desire for the outcome, b) a belief that their behavior will lead to that outcome, c) an intention to perform the behavior, d) the ability to perform the behavior, e) and awareness of going through with the intention while performing the behavior.

Shultz and Wells' (1985) work suggests that children as young as three may use some of these conditions, such as desire for an outcome and an intention to perform the outcome, when judging intentionality. For instance, if an actor's intentions and the outcome match, young children deem it purposeful; but if the intention and outcome do not match, they deem it unintentional. In contrast, 11-year-olds take additional factors into account when judging intentionality; for example, an outcome is deemed as intentional if the actor is aware and presumably in control of the situation (monitoring rule), and that intentions may be discounted if there is enough external reason for the outcome (discounting rule) (Shultz & Wells, 1985).

Further, understanding the distinction between desires and intentions is important because, in contrast to mere desires, intentions are used to actually achieve a desired goal or action (Baird & Astington, 2005). By 4- or 5-years of age children can differentiate intentions from desires (Feinfield et al., 1999). Feinfield et al. presented children with stories in which characters' intentions differed from their

preference/desire and the outcome of the situation. For example, in one story a child decides to go play football because his mom wants him to, even though he wanted to go to the mountains because he likes the snow; yet, he ends up at the mountains because the bus driver got lost. In order to assess children's understanding of intentions and desires, they were asked where the character liked to go (mountains), where the character intended to go (football), and where the character thought they were going (football). Three-year-olds confronted with this scenario conflated intention with desire or outcome. Most 3-year-olds did not discriminate the characters' intentions (saying the child intended to go to the mountains) from the desires and outcome of the situation, while most 4-year-olds did.

There is also evidence suggesting that preschoolers understand that the same action can be motivated by different intentions. Baird and Moses (2001) conducted a series of studies in which 4- and 5-year-olds heard stories in which one action was motivated by different intentions for different protagonists. For example, they were told a story about a boy who was running home to be on time for dinner, and another story about a boy who was running to be healthy and strong. Children were asked about the intention of each boy. Five-year-olds understood that the same action could be motivated by different intentions in a variety of scenarios, whereas 4-year-olds did not. Therefore, by age 5, most children have come to understand several core facets of intentionality.

Understanding Others' Beliefs

There has been extensive research on children's understanding of others' beliefs, with the false belief task often being used as a litmus test of ToM (Wellman et al., 2001). False belief understanding requires a child to understand that a person may hold a belief that is different from reality. It relies upon the understanding that the mind and physical world are distinct and that actions stem from mental representations rather than actual states of the world (Wellman et al., 2001). In a typical false belief task, children are asked to reason about a person's belief which may not be in line with reality. For example, children may be asked what they think is in a familiar container, like a band-aid box, and then they are shown that the box contains a different object, like a toy pig. Children are then asked to predict what a character, who hasn't seen inside the box, will say is in the box. For children to pass this task they must answer that the new character will say that band-aids are in the box (Wellman & Liu, 2004). Children who successfully pass this task are making a distinction between something internal and not tangible (someone's thoughts) and something external and tangible (actual contents of the band-aid box).

False belief reasoning improves markedly during the preschool years. Wellman et al. (2001) conducted a large meta-analysis of false belief studies in an effort to quantitatively synthesize divergent false belief findings. They found that: a) children's false belief performance increases with age; and b) it was not until around 4 years of age that children performed significantly better than chance. Wellman et al. drew the conclusion that children's improved performance on false belief tasks reflects a genuine conceptual change. That is, they argue that young children's difficulties with false belief tasks are not solely the result of a performance deficit

wherein children have false belief understanding but lack ancillary skills needed to express that understanding.

To the extent that children's conceptual understanding of false belief, and other aspects of ToM, is indeed emerging during the preschool period, this should have important implications for their reasoning about moral aspects of others' behavior. To explore this issue, I turn next to reviewing the existing body of work on children's moral development during this same age range.

Moral Development During Early Childhood

Much attention has been paid to the development of moral judgments and reasoning at least since Freud (1959) and Piaget (1932). Freud viewed morality through a psychoanalytic perspective focusing on guilt and what lies in our unconsciousness (Freud, 1959; Turiel, 2007). Piaget was also keenly interested in children's understanding of intentionality in moral judgments and punishment. For example, Piaget presented children with a contrasting pair of stories and then asked children to make moral judgments about the actors. In one story, a boy was called for dinner and as he walked into the dining room he accidentally knocked over a tray with 15 cups that was behind the door and all the cups broke. In the other story, a child was secretly trying to get into the cupboard to get jam while his mother was gone. While reaching up to get the jam he knocked a single cup over and broke it. Children were asked which boy was naughtier. In Piaget's research, children up to 8- or 9-years of age said that the child that broke more cups was naughtier, and delivered blame based on the degree of damage and not the actor's intent. Piaget argued that it wasn't until at least age 10 that children were able to take into

account others' intentions and motives. He suggested that social relationships and experiences are central to morality and that our moral judgments form from interactions with rules, laws, authority, social institutions,and relationships (Piaget, 1932; Turiel, 2007). It is important to note, however, that the reliability and validity of some of Freud and Piaget's methods for examining children's moral judgments have been heavily contested due to the use of projective tests, parent self-report, or artificial experimental stories (Lickona, 1969; Turiel, 2007).

Building on Piaget's work, Kohlberg's stage theory of morality has been highly influential (Arnold, 2000). One of Kohlberg's main contributions to the field was a six-stage developmental theory of moral judgments based on boys' (ages 10-16) responses to moral dilemmas. For example, in one story they are told that a man's wife was dying from cancer, and that doctors believe they found a drug to cure her. A local chemist develops the drug but is charging 10 times what the drug costs, and the man can't afford it. He was unable to raise enough money for the drug and the chemist would not cut him a break. The man decided to break in and steal the drug from the chemist in order to try and save his wife's life. Children were then asked if it was okay that the man stole the drug and whether he should be arrested. Based on children's responses, Kohlberg suggested that children initially make judgments based on punishment avoidance, or obedience (preconventional level), then make judgments based on role expectations, respect for authority and the social system (conventional level), and eventually some come to reason based on mutual respect and understanding of justice and rights (post conventional level) (Turiel, 2007).

A more recent theory is the social domain theory of moral judgment which separates reasoning about moral rules from social conventional rules; this reasoning is typically constructed through children's varied social interactions and experiences (Smetana, 2013). Rules that are always evaluated as wrong even in the absence of rules are moral rules; transgressions or actions that are judged based on rules and relative to the social situation and have fewer stern offenses are considered social-conventional rules (Smetana, 2006). Children are less willing to break moral rules compared to social conventional rules (Chernyak & Kushnir, 2014; Smetana, 2006). For example, almost all children between the ages of 2-9 years judged moral events, like a child shoving another child, as being wrong and worthier of punishment than social conventional violations, like a child not participating in show and tell (Smetana, 2013). These results further support the social domain theory that children's judgments and actions are heterogeneous across situations.

Haidt (2001) discusses that in the more cognitive or rationalist views of morality moral judgments are a result of reasoning. Yet, Haidt argues that moral judgments should be examined as an interpersonal process. This view, the social intuitionist model, suggests moral judgments are a result of moral intuition, and emphasizes the importance of social and cultural influences. Other researchers such as Hoffman (2000) suggest that emotions, like empathy, drive moral actions which further influences moral judgments. Hoffman also suggests that internalization of empathy is scaffolded by parents (Hoffman & Saltzstein, 1967). All the theories discussed so far suggest that, in one way or another, understanding of others' mental states is critical in making moral judgments.

Theory of Mind and Moral Development during Childhood

ToM and moral judgments both rely on the consideration of others' mental states and appear to be interrelated. Indeed, there is a strong conceptual connection between moral judgment and ToM in the ethical codes and legal systems of western cultures. In these cultures punishment is not assigned solely based on outcome. We take into consideration intention, outcome, and agent knowledge. In the legal system, *Mens rea (*"guilty mind"*)*, or a person's intention to commit a crime, is often a necessary component in judging criminality (Kneer & Bourgeois-Gironde, 2017). Take for example, the following scenario from Young and colleagues (2007) in which Grace and her friend are touring a chemical plant. Grace goes to get coffee and her friend asks for sugar in hers. In one situation Grace thinks the white powder by the coffee station is sugar, but unbeknownst to her the powder is toxic and her friend dies. Here she has a benign intention but the outcome is disastrous. In another scenario, Grace thinks the powder is toxic, yet it is actually sugar and her friend is fine. Here she has a malevolent intention but the outcome is positive, or at least neutral. Yet, Grace could be convicted in the court of law in both of these situations, in the first situation because of her possible negligence or carelessness and in the second due to her malicious and purposeful intent. In Young et al.'s study, adults judged this harm as just as wrong as if Grace purposefully and successfully killed her friend.

In more mundane situations, understanding others' desires, intentions, beliefs, and emotions remains vital in making judgments of action rightness and wrongness, permissibility, blame and praise, and deserved punishment or reward.

For example, say you are asked to assign a punishment (or no punishment) to someone who threw a ball at another person's head; you would likely want to know whether the person intended to hit the person maliciously, whether it was an accident, and perhaps whether the person was acting recklessly or carefully. It would be unfair to harshly punish someone who was acting carefully, but accidently threw the ball and hit their friend. Yet, we may want to punish someone who purposely threw the ball at their friends' head in order to cause damage, because we take into account their intention. Or, we may want to punish someone who accidentally hit their friend but should have known that that was a likely outcome in the situation because we take into account their knowledge and belief states. Therefore, it's critical to incorporate intention, belief, and outcome information in our moral judgments. Being able to reason about others' intentions and beliefs allows us to make appropriate moral judgments and further to make and create positive social relationships.

Associations Between Theory of Mind and Moral Development during Childhood

The interplay between ToM and moral development is already evident during preschool. Both concurrent and longitudinal associations between ToM and moral development have been found during preschool and early childhood (Astington, 2004; Killen et al., 2011; Leslie et al., 2006; Wainryb & Ford, 1988). For example, Killen et al. (2011) found that children (ages 3-8) who failed false belief tasks were more likely to attribute negative intentions to an accidental transgressor (i.e., a protagonist with good intention who accidentally caused a bad outcome because of a false belief), than children who passed the false belief tasks.

There is also evidence for longitudinal association between ToM and moral judgments. Smetana and colleagues (2012) tested children three times across a period of a year. At all three time points, children completed a battery of five ToM tasks (Wellman & Liu, 2004) assessing diverse desires, diverse beliefs, false beliefs, and belief emotion. Children were also asked about moral and conventional transgressions in the *Social rules interview* (Smetana et al. 2012). The results revealed bidirectional longitudinal associations between ToM and different criteria for moral judgments. For example, children who viewed moral transgressions as less permissible than social conventional transgressions had more advanced ToM 6 months later, and children who viewed moral transgressions as wrong (regardless of whether an authority figure said it was okay) had more advanced ToM 6 months later. There was also evidence that ToM influences moral judgments but, surprisingly, the effects were not all in the expected direction. Children between 2.5 – 5.5 years with more mature ToM understanding viewed moral transgressions as more alterable (i.e., said it was okay to commit a moral transgression if a teacher said it was okay) at the two successive waves (Smetana et al., 2012).

Further, Sodian et al. (2016) found a predictive relation between infants' understanding of agents' intentions and their reasoning about moral intentions at 5-years-old. The researchers assessed goal encoding at 7 months, measuring implicit ToM using a habituation task, and moral understanding at age 5 years through a moral vignette that included an accidental transgressor. Children were asked about the transgressor's beliefs and intentions. Sodian et al. found that infants who judged an agents' actions in terms of goals, suggesting an understanding of intention, were

more likely than infants who failed the task, to correctly attribute positive intentions to the accidental transgressor at age 5 years. Similar to Sodian et al., Lane et al. (2010) found that more advanced ToM at 3.5 years, measured by an understanding of others' beliefs and desires, was predictive of children's moral reasoning about others' physical and psychological needs at age 5.5 years.

Collectively, this research suggests that the development of moral judgments and ToM are bidirectional; having social interactions or experiences where moral transgressions occur may allow children to learn and to reason about others' mental states, which in turn enables children to reason about more complex moral situations. Next, I discuss how and when children incorporate specific aspects of ToM, like intention and belief, into moral judgments.

Children's Moral Judgments – Outcome to Intent Shift

When making evaluations of what is "right" and "wrong", it is important to consider a protagonist's intention in addition to the outcome of the situation. Piaget (1932/1965), as well as subsequent research, demonstrated that when children judge the permissibility of acts, they initially focus on the outcome of the action, rather than the intentions of the actor (Yuill & Perner, 1988). As discussed previously, in Piaget's work most children under the age of 10 years judged the actor with the good intention to be naughtier because of the bad outcome, whereas adults judged based on intentions (Piaget, 1932/1965).

In contrast to Piaget's, more recent research has found that young preschoolers can sometimes take intention into account in making moral judgments. Nelson-Le Gall (1985) examined children's (3- to 4-year-olds) and adults' judgments

of intentionality and moral judgments of characters in situations in which outcomes were foreseeable or unforeseeable, and positive or negative. For example, in the positive and foreseeable story, children were told a story in which Nick throws a ball to his friend Pat, and Pat catches it. In the negative but unforeseeable story, children were told that Nick is angry with Pat so he is throwing the ball along the fence roughly, and Pat steps out from behind the fence and gets hit in the head. Children were then asked to rate the character from very bad to very good as well as being asked about the character's intentions. Overall, children, as compared to adults, were more likely to attribute intentionality to characters, especially in the unforeseeable stories. In addition, (a) characters in the positive outcome condition were rated more positively by both children and adults in the foreseeable compared to the unforeseeable story, and (b) characters in the negative outcome condition were rated more positively by both children and adults in the unforeseeable compared to the foreseeable story. These results suggest that young children can use information about outcome foreseeability to make moral judgments. Yet, a sizeable number of children, unlike adults, judged characters' actions as more intentional in the unforeseeable in comparison to the foreseeable story (Nelson-Le Gall, 1985). Hence, although children have some understanding of intentionality and motive, this understanding is not as well-entrenched as adults' understanding, especially when foreseeability is involved.

Subsequent research has found age-related improvement in the ability to incorporate intent information into moral judgments (Cushman et al., 2013; Nobes et al., 2016; Zelazo et al., 1996). Cushman et al. examined 4- to 8-years olds'

judgments of accidental harm and attempted harm. In the accidental harm story, children were told a story about a child accidentally tripping on a rock and pushing someone over, while the attempted harm story was about a child trying to push someone over but tripping on a rock and missing. There were significant age-related trends in acceptability/naughtiness judgments. Children 4- to 8-years-old increasingly judged attempted (but failed) harm as naughtier than accidental harms. Further there was a steep decrease in naughtiness and punishment judgments for the accidental harm situation; at 4 years of age, the majority of children viewed the act as naughty and punishable, while almost all children by age 8 did not rate the act as naughty or punishable. This suggests that, at younger ages, children focus on outcome and as they age they incorporate intention information into acceptability judgments. Similarly, Zelazo and colleagues (1996) found that 3-year-olds focus almost entirely on outcome information when making acceptability judgments in response to both canonical responses to causal acts (e.g. an animal is petted and feels good) and noncanonical responses to acts (e.g. an animal gets petted and hurts and cries). Yet, 5-year-olds and adults incorporate intention and outcome information for judging acceptability. This research suggests that there is a continuous developmental progression in integrating outcome and intention information in order to judge the acceptability versus naughtiness of an act.

Similar to judging the acceptability and naughtiness of an act, it is also important to consider intention and outcome when determining deserved reward or punishment. Research suggests a similar developmental trend of incorporating intention and outcome when judging punishability (Cushman et al., 2013; Margoni &

Surian, 2017; Nobes et al., 2016; Zelazo et al., 1996). For example, while some 4- and 5-year-olds appropriately used both intention and outcome information when assigning a punishment rating in which intention (good vs. bad) and outcome information (bad vs. good) were not in line, almost all 3-year-olds used only outcome *or* intention information (Zelazo, et al., 1996). Nobes and colleagues (2016) more recently replicated these results. They also found that when the acceptability question was rephrased to be agent-focused, asking about the morality of the actor rather than the morality of the act, it substantially increased the percent of intention-based punishments that 5- and 6-year-olds made. Taken together, this suggests that older children increasingly incorporate intent into moral judgments.

Gummerum and Chu (2014) investigated how intention and outcome information influences older children (8-year-olds), preadolescents (12-year-olds), and adolescents (15-year-olds) second-hand punishment (i.e., participants have the option to punish someone who wronged them) and third-hand punishment (i.e., participants have the option to punish someone who wronged another individual). In the third-party punishment condition, adults primarily punished based on intention and outcome, whereas children and adolescents punished primarily on the outcome. Yet, they found that adolescents' could incorporate intention and outcome into second-hand punishment.

Although the influence of intention and outcome on judgments of deserved punishment has been examined extensively, less research has addressed the use of this information for deserved *reward* for helping behavior. It is possible that children and adults may not make these judgments in a similar fashion because helping

doesn't create a moral violation in the same way that hurting does. For example, "do good" is more vague and open to interpretation than, "don't cause harm". Further, social domain theory emphasizes the importance for fairness, but it may be more difficult to judge fairness in "good situations", such as charitable giving. Margoni and Surian (2017) examined judgments of deserved reward and goodness judgments for helping behaviors. The method was similar to Cushman et al. (2013) but, in this case, the purpose was to see whether the same outcome to intent shift happens for helping behaviors. Children ages 4-8 years judged goodness and deserved reward for accidental helping and attempted help. Unlike in Cushman et al., children heard how the recipient felt after the vignette. The results revealed evidence of an outcome to intent shift of goodness judgments in accidental help and attempted help stories (Margoni & Surian, 2017). Goodness attributions for accidental help started to decrease around age 5-6 years. In addition, judgments of deserved reward, compared to deserved punishment, seemed to be more outcome-based than intent-based. One possibility is that hearing how a recipient feels increases the salience of the outcome. An alternative explanation is that children are still learning how to reason about reward.

False Belief Understanding and Moral Judgments

As previously discussed, beliefs as well as intentions are relevant to moral judgments. False belief understanding appears to lag behind understanding of others' intentions and desires. Therefore, it's important to assess at what age children begin to incorporate belief, especially false belief information, into moral judgments. In contrast to intentions, however, there has been very little research

assessing how and when children integrate belief information into their moral judgments. Although some earlier work indirectly examined belief and knowledge understanding in relation to moral development (Wimmer et al., 1984; Yuill & Perner, 1988), it is only recently that a systematic analysis of the relation has been undertaken. Killen et al. (2011) assessed 3- to 8-year-old children's understanding in a moral transgression task embedded within a false belief story. In this story, a well-intentioned boy accidentally causes a negative outcome because he acts on the basis of a false belief about a container's contents. Specifically, as the boy was helping a teacher clear tables, he threw out a paper bag which, unbeknownst to him, contained another child's cupcake. Children then responded to a range of questions including those assessing their false belief understanding (i.e., what did the boy believe was in the bag?), their moral appraisal of the actor's intention (i.e., whether the boy thought he was doing something all right or not all right), and, in a second study, whether punishment was warranted (i.e., whether the boy should get in trouble). Killen et al. found that children without false belief understanding were more likely to attribute negative intentions and to assign punishment to the accidental transgressor than children with false belief understanding. Moreover, it was not until children were 7- or 8-years-old that they attributed positive intentions to the actor at high levels. This suggests it is well beyond 4- or 5-years of age, the typical age in which children acquire false belief understanding, that children begin to shift their moral judgments and understand the act was not intentional in these circumstances

Taken together, most children seem to follow a similar progression of incorporating mental state understanding into moral judgments. First, children focus on outcome when making moral judgments, later they start incorporating intent information. Subsequently, children begin to incorporate false belief information into judgments. Lastly, children include mental state understanding into judgments of deserved punishment.

Examining the Relation between Theory of Mind and Moral Development

Why is it that, at least in Westernized cultures, children initially focus on outcomes and only later incorporate intent information into decisions about punishment, act acceptability, and even judging individuals as good or bad? Initially, Piaget proposed that moral judgments develop from interactions with rules, laws, authority, relationships, and social institutions (Turiel, 2007). Piaget suggested that younger children make rigid moral judgments based on laws and rules (Piaget, 1932). Typically, young children would judge a behavior as "bad" if the observable consequence was bad, regardless of intent. As children grow older, they become more "autonomous" in their judgments and do not solely rely on outside rules or laws. Children are eventually able to judge based on their own view and rules of morality; further, they become able to consider others' intentions and use that information along with outcome information to make moral judgments (Piaget, 1932).

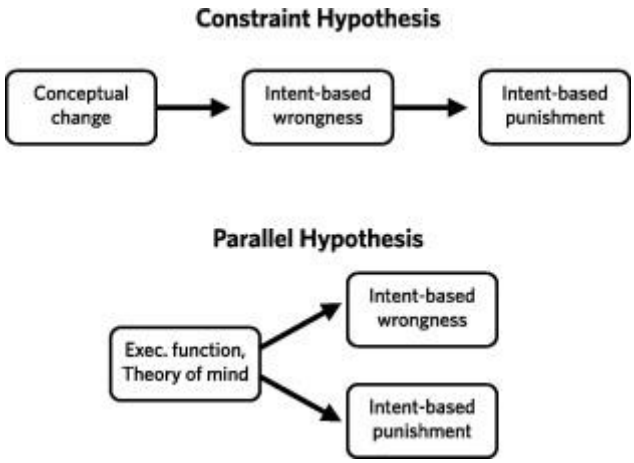
A central question is whether a conceptual reorganization of children's understanding of morality makes possible the outcome-to-intent shift, or whether changes in other domains, such as executive function or ToM, are responsible.

Cushman and colleagues (2013) examined this question in research that tested two hypotheses. The Constraint Hypothesis, presented in Figure 1.1, suggests that children undergo a conceptual change that allows them to first incorporate intent information into judgments of wrongness and then after that into intent-based punishment judgments. This hypothesis suggests that first children will make judgments of wrongness and deserved punishment based on an agent's causal responsibility/outcome of the act. Then, children go through the outcome-to-intent shift in which they start judging moral wrongness based on an agent's mental states, but do not do the same for assigning punishment. Ultimately, children come to judge punishability based on what is viewed as morally wrong, not just the outcome. In other words, the ability to judge wrongness based on intent constrains judgments of punishability, then both judgments become more contingent on intent. It is the reorganization of concepts, including integration of outcome and intent information into wrongness and punishment judgments, within the moral domain that generates the outcome-to-intent shift.

In contrast, Cushman et al. (2013) also articulated the Parallel Hypothesis, which suggests that changes in domains outside of morality—such as improvements in executive function—drive moral development. If so, then appropriate intent-based wrongness and intent-based punishment judgments should occur simultaneously. If the changes are driven, for example, by changes in executive function or ToM, then both judgments should be equally affected.

Figure 1.1

The Constraint and Parallel Hypothesis (Cushman et al., 2013)



The findings from Cushman et al. support the constraint hypothesis: older children (6-8 years) used intent-based information more heavily for "wrongness" judgments compared to punishment judgments. Only later do children acquire a new concept of moral wrongness which then influences judgments of deserved punishment. If the change were driven by changes in executive function or ToM, both judgments would be equally affected by intent information. Nonetheless, although ToM may not be the underlying mechanism for this shift, children must of course be able to reason about others' intentions and knowledge to make these moral judgments (Killen et al., 2011).

Margoni and Surian (2017) further examined the Constraint Hypothesis by assessing children's judgments of "goodness" and deserved reward for accidental helping and attempted helping. They found support for the outcome-to-intent shift for helping: increases with age emerged in children's attributions of goodness to the character who attempted to help (but failed). Yet, there was no relationship between punishment judgments and age, such that punishment judgments were

more outcome-based at all ages. This finding contrasts with those of Cushman et al., and does not support the Constraint Hypothesis, because there were no developmental changes for judgments for deserved reward or punishability for accidental help. Yet, no other social or cognitive abilities were tested so these findings unfortunately are uninformative with respect to predictions of the Parallel Hypothesis.

Individual differences in the Relationship between Theory of Mind and Moral Development

Other researchers emphasize the importance of executive function in making moral judgments. Smetana, Rote, et al. (2012) examined the association between individual difference factors and moral judgments in children across the span of a year (2.5 – 4 years). Supporting their hypothesis, they found that children rated as higher in effortful control generally rated moral transgressions as wrong at higher levels than those with lower effortful control. They also suggested that effortful control may be important in helping children to avoid participating in immoral behavior (Smetana, Rote, et al., 2012). Further, Cowell et al. (2017) found that children's (ages 5-12) executive function skills predicted their generosity in a resource allocation task. These studies support the importance of examining domains outside of the moral domain, such as executive function, that may be associated with moral judgments.

Another aspect of social cognition that may be helpful in making moral judgments is emotion understanding and empathy. Even though most research focuses on the influence of behavioral and cognitive processes, some research

suggests that affective processes, such as empathy, are also important for more complex moral judgments (Hoffman, 2000; although see Cowell et al., 2017 for conflicting findings). Ball and colleagues (2017) examined the influence of affective and cognitive processes on moral judgments with a sample of socioeconomically diverse 3.5-year-olds. They presented children with two prototypical social conventional transgression stories, and six prototypical moral transgression stories; they asked children about deserved punishment and for a severity rating of the transgression. As predicted, greater empathy was associated with increased severity in judgments of moral transgression (specifically psychological harm), but not social conventional transgressions. The researchers suggest that empathy and ToM may help children understand the consequences of moral transgressions on others. ToM understanding allows children to interpret others' mental states while empathy may allow the child to understand what it may be like to be the victim of the transgressions. Interestingly, Ball and colleagues found that children with low ToM and high empathy made the most mature moral criterion judgments; these judgments asked children if it would be okay to act immorally if an authority figure said it was okay, or if there were no rules about the act. The researchers suggest that children with lower ToM may rely more on affective information; for example, they consider the negative outcome when judging immoral acts, and do not necessarily incorporate intent information. It's thus important to consider both cognitive and affective processes as they both may influence preschoolers' moral judgments.

In sum, research suggests that during the preschool years, children's ToM development and moral judgments improve substantially. Then, sometime between

the ages of 4 and 8, children begin to incorporate information about belief and intent into judgments of act acceptability and the moral intentions of a character. Incorporating intent information into judgments of deserved punishment and reward is more challenging and may emerge at an even later age. What's not entirely clear is the underlying mechanism that allows for the integration of this information to allow for more sophisticated moral judgments. It is possible that children undergo a conceptual reorganization of moral concepts, but individual differences, such as executive function and empathy, may also influence the development of children's moral judgments.

Dissertation Aims

This dissertation is designed to examine the interplay between theory of mind and moral development during childhood. The primary goals are to 1) examine and compare the extent to which children during early childhood (4- and 5-year-olds) and middle childhood (7-year-olds), and adults incorporate belief and intent understanding into moral judgments, 2) to further explore how children and adults incorporate these understandings into their judgments of deserved reward and punishment, and 3) to investigate whether individual differences in executive function and empathy are related to moral judgments across age.

In Chapter 2, I review preliminary research that formed the basis for this dissertation (Ochoa, Rodini, & Moses, 2020), in Chapters 3 and 4 I present two empirical studies in which I examined the relationship between false belief understanding and moral judgment for children and adults, and lastly in chapter 5 I discuss the overall findings and implications of this research.

Unanswered Questions

Although Killen et al. (2011) established a link between belief understanding and moral judgments, finding that children without false belief understanding were more likely to attribute negative intentions and to assign punishment to the accidental transgressor than children with false belief understanding, many questions were left unanswered. First, Killen and colleagues administered only one morally-relevant belief story with only one combination of agent's belief (false) and intention (good). It remains to be seen whether there are developmental changes across other combinations of belief (true vs. false) and intention (good vs. bad), and whether belief and intention interact in some way. Children may have had similar difficulty even in a true belief context and may have responded differently had the intentions of the actor been negative.

Second, in the Killen et al. (2011) study, in order to perform well on the task, the children were required to make assumptions about the agent's intention (i.e., that the boy cleaning tables would not have thrown out the cupcake had he known it was in the bag). Hence, the younger children may have performed poorly either because they lacked false belief understanding or because they did not hold a default assumption of benign intent. They might show greater ability to incorporate belief understanding with moral evaluations if agents' intentions are made more explicit.

Third, when children initially attain false belief understanding, they may still have difficulty applying that knowledge to moral judgments. There may be a lag such that their ability to integrate their newly acquired belief understanding with their

moral understanding is delayed. Interestingly, a similar lag has been found with respect to integrating belief information with emotion understanding. De Rosnay et al. (2004) told children stories in which, for example, a character mistakenly believed that a container held a preferred food when in actuality it contained a disliked food. Although 4- and 5-year-olds understood the character's mistaken belief, they nonetheless incorrectly predicted that the character would feel sad upon seeing the container. It was not until 6-years-old that children made correct emotion predictions (see also, Wellman & Liu, 2004). This finding, that children cannot immediately use belief information to inform judgments about another mental state (emotion), demonstrates difficulty in integrating concepts within the ToM domain. It remains to be seen whether similar difficulty, and a similar developmental lag, would be found in integrating belief reasoning with understanding in a different domain, that of moral judgment. Preliminary evidence for such a lag comes from the Killen et al. (2011) study. They found that while 7-year-olds rated the accidental transgressor as having positive intentions, 5-year-olds (most of whom correctly answered the false belief question) gave a neutral rating of the agent's intention. Whether that finding is replicable for positive intentions and extendable to negative intentions is not known.

Finally, a lag may also be present between children's appreciation of moral intentions and their accurate assignment of consequences to actors. For example, Cushman et al. (2013) found that around 5-years-old children first incorporate intent information into judgments of wrongness and only later, between 6-8 years, into intent-based punishment judgments. They argue that this pattern is not driven by

changes in relation to ToM but instead represents the reorganization of concepts within the moral domain itself, specifically in how intent relates to wrongness and punishment. While Killen et al. (2011) questioned children regarding both moral intentions and moral consequences, their data are not broken down in such a way as to clearly determine whether the latter were more difficult to appreciate than the former.

CHAPTER II

PRELIMINARY RESEARCH

Ochoa, Rodini, and Moses (2020) is a two study paper in which we further examine the link between belief understanding and moral judgment in childhood. This chapter summarizes the first study (completed by Joseph Rodini and Louis Moses). That study developed the methodology that formed the basis of this dissertation. Hence, here I provide an overview of the study; the detailed method, procedure and analysis can be found at: https://psyarxiv.com/3ysef/. The second study from Ochoa et al. (completed by me) is presented fully in chapter 3 and is Study 1 of my dissertation.

The first study of Ochoa et al. (2020) aimed to address the limitations described above by creating stories in which motive/intention (good or bad) was crossed with outcome (good or bad) in situations in which the protagonists held either a true or false belief (see Table 2.1). For example, in the good motive/bad outcome story, the protagonist discovers two containers, one with a skunk and another with a kitten. The protagonist explicitly states he wants to make his friend happy by giving him the kitten and knows that skunks make his friend really upset. The protagonist leaves the scene, and while he is gone the two animals switch places. When the protagonist comes back, he gives his friend the container that contains the skunk, due to his false belief, even though he intended to give him the kitten. Participants (4- and 5-year-olds) then answered questions about the agents' intention, belief, and deserved reward/punishment. Further, individual differences in

executive function were assessed to examine their relationship with moral judgments.

Table 2.1

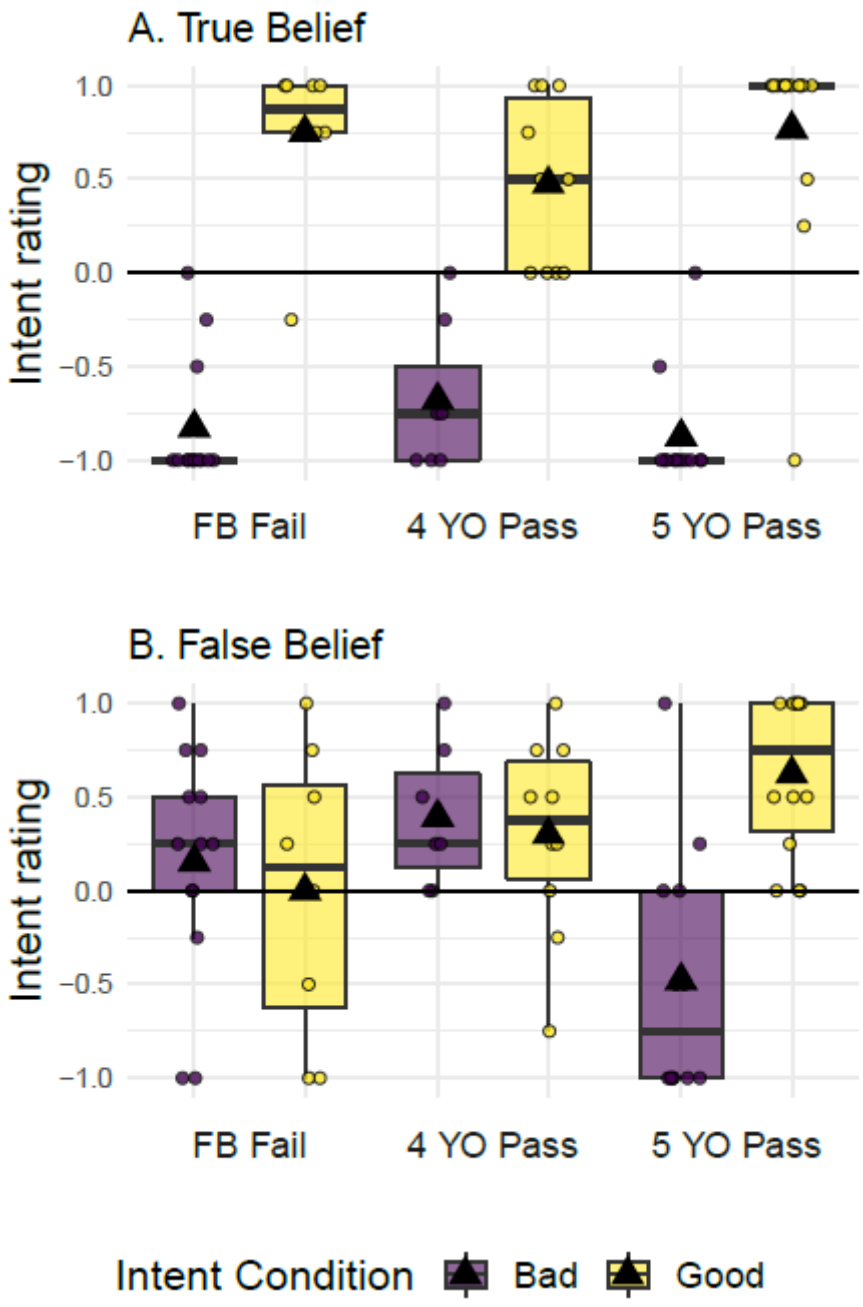
Experimental Manipulation of Intent and Belief Factors to Yield Morally-Relevant Belief Vignette Outcomes

Story Stimuli	Intent Condition	X Belief manipulation	= Outcome	
Kitten	Good	True/False	Good & True: Kitten shared (intended positive outcome)	Good & False: Skunk Shared (accidental harm)
Skunk	Bad	True/False	Bad & True: Skunk shared (intended harm)	Bad & False: Kitten shared (accidental positive outcome)

Children were categorized into three groups based on their false belief understanding: false belief failers (those who failed one or both of the morally-relevant false belief vignettes), 4-year-old false belief passers, and 5-year-olds false belief passers. When agents held true beliefs, almost all children successfully reasoned about their intentions (see 2.1, Panel A) and deserved reward (in the good motive condition) or punishment (in the bad motive condition) (See Figure 2.2, Panel A). However, when agents held false beliefs, only 5-year-olds were able to appropriately reason about intentions, (See Figure 2.1 Panel B). Yet, even 5-year-olds had difficulty reasoning about punishment and reward in the false belief condition (See Figure 2.2, Panel B).

Figure 2.1

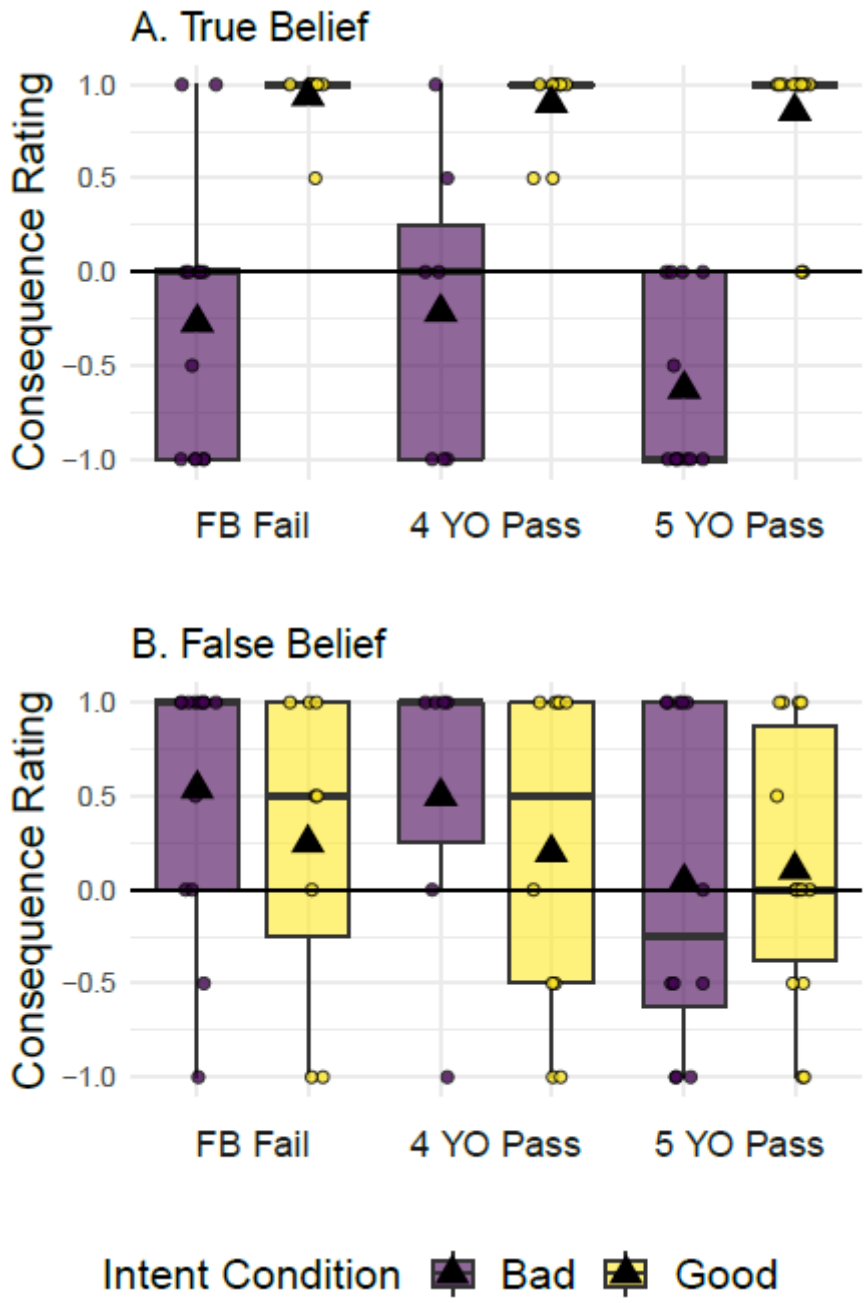
Intent Judgment Separated by False Belief Group and Intent Condition (Ochoa, Rodini, & Moses, 2020)



Note. The black triangles represent the mean of the group and dots represent individual responses.

Figure 2.2

Consequence Judgment Separated by False Belief Group and Intent Condition (Ochoa, Rodini, & Moses, 2021).



Note. The black triangles represent the mean of the group and dots represent individual responses.

There were thus two developmental lags between recognizing that agents held false beliefs and using that information in making moral judgments: a lag between understanding false beliefs and correctly attributing moral intentions based on those beliefs, and a lag between correctly assigning moral intentions and accurately assigning moral consequences. Five-year-olds who recognized false beliefs performed similarly to the 7-year-olds in Killen et al. (2011): they attributed good intentions to agents with good motives (with a bad outcome) and bad intentions to agents with bad motives (with a good outcome) in the false belief conditions. However, 4-year-olds who passed the false belief questions were unable to incorporate that information in making moral judgments. Similarly, 5-year-olds who assigned intentions accurately still performed no better than chance when assigning consequence.

CHAPTER III

STUDY 1

We had five primary aims for this study. First, we wanted to see whether the developmental lags found in the first study of Ochoa et al. (2020) would replicate when the processing demands of the task and length of study were reduced. We simplified the design of the moral vignettes in two ways. Half of the vignettes in the first study of Ochoa et al., involved three items to track, such as two food items and an animal that moved the food items while others involved just two animals who moved of their own accord. In my dissertation studies we used only animals (and not food items) so that children would only have to direct their attention to the animals of interest rather than a third animal entering and moving the food items from one box to another. In addition, we used only a single box such that children would not have to track the movements of two animals across two separate boxes (see procedure).

To shorten the testing session, we also eliminated the standard belief and moral tasks used in the first study, and we reduced the number of test questions for each moral vignette by combining the agent evaluation and agent intention questions into a single question and dropping the friend question (see first study materials at https://psyarxiv.com/3ysef/). In addition, because children performed so well on true belief trials in the first study, we only administered false belief trials for this study.

Second, to gain deeper insights into children's thinking about assigning consequences, we asked open-ended questions probing the reasons behind their

judgment of deserved reward or punishment. In this way, we hoped to uncover underlying moral principles that might guide children's reasoning (Cushman et al., 2006).

Third, because even 5-year-olds in the first study of Ochoa et al. did not appropriately assign reward/punishment on false belief trials (and neither did some of the 7-year-olds in Killen et al., 2011), we included an adult sample for comparison purposes to determine a developmental endpoint for consequence reasoning.

Fourth, because some of the cell sizes were low for the false belief groups in the first study, we increased the sample size for this study so that the central tests of our hypotheses would be better-powered.

Lastly, we attempted to replicate the relationship between individual differences in executive functioning (EF) and moral judgments. In the first study, we found that EF was related to false belief understanding: children without false belief understanding performed significantly worse on EF tasks compared to children with false belief understanding. Further, EF was related to children's moral judgments. We aimed to see if we could replicate these findings with a larger sample and a different battery of executive function measures.

Study 1 Hypotheses

Following from the findings from the first study of Ochoa et al. we hypothesized that:

H1: children who did not show false belief understanding, would not be able to appropriately attribute agent intentions, but would instead tend to answer

intention questions based on outcome, and would do so more often than false belief passers.

H2: there would be a developmental lag for use of false belief information in response to intention questions, such that adults and 5-year-old false belief passers would make more appropriate intent-based judgments than 4-year-old passers.

H3: 4-year-old passers would have more difficulty assigning punishment/reward than 5-year-old passers and adults.

H4: punishment/reward judgments would again be harder than intent judgments, particularly for 4- and 5-year-olds.

Our hypotheses and analysis plan were pre-registered prior to analyzing the data on the Open Science Framework (OSF) at https://osf.io/3p5m9/.

Methods

Participants

One-hundred and nine children participated, 63 4-year-olds (34 girls; M_{age} = 53.80 months, *SD* = 3.56) and 46 5-year-olds (21 girls; M_{age} = 65.70 months, *SD* = 3.33). An additional 5 children were tested but excluded from analyses due to experimenter error (3), inattention (1), and family interference (1). The sample was representative of the population from which it was drawn. Eighty-six children were White, 5 reported being two or more races/ethnicity, 3 were Asian, 2 were Hispanic, Latino, or Spanish, 2 were Middle Eastern, and 1 was Native American/Native Alaskan. Ten parents did not report their child's race or ethnicity. Seventy-five percent of families reported making at least $40,000 a year.

Eighty-nine children participated in the lab and were recruited from a participant database at a large research university in the Western United States. Parents were provided compensation in the form of $10, and children received a small toy. A further 25 children participated in a quiet space at a local children's museum—parents of these children did not receive compensation, but children received stickers.

Forty-two adults (69% female, M_{age} = 19.71 years, *SD* = 2.50 years) from the Psychology and Linguistics Human Subjects Pool at the University of Oregon also participated. The majority (26) reported being White, 8 were Asian or Asian-American, 2 were Black, and 6 reported being of two or more race/ethnicities or other. Participants received compensation in the form of class credit.

Design

Participants were randomly assigned to one of two intent conditions (good or bad) in which they responded to two morally-relevant belief vignettes. An *a priori* analysis indicated at least 30 false belief passers would be required in each age group (4-year-olds, 5-year-olds, and adults) for the study to be adequately powered. Because we anticipated that many children would fail the false belief (false belief) task, we oversampled in order to achieve our targeted numbers of false belief passers. Ultimately, we recruited 63 4-year-olds (30 false belief passers), 46 5-year-olds (31 false belief passers), and 42 adults (40 false belief passers).

Procedure

Children who participated in the lab were tested individually in a single 30 minute videotaped session. Children who participated at the local children's

museum were tested in a quiet space in a single session. Adults completed all tasks in the lab on an iPad.

Measures

Morally-relevant Belief Vignette

Similar to the first study of Ochoa et al. (2020), participants in each age group were randomly assigned to the two intent conditions until the required sample size of 30 was achieved: for children, good intent (N = 46, 28 girls, M_{age}= 57.70 months) or bad intent (N = 59, 26 girls, M_{age} = 60 months) and, for adults, good intent (N = 22, 14 women, M_{age}= 19.86 years) or bad intent (N = 20, 15 women, M_{age}= 19.75 years). Each participant heard two morally-relevant belief vignettes(one with a boy agent and one with a girl agent) in which the agent held a false belief, with vignette order counterbalanced across conditions. Good intent stories involved an agent who wanted to make a friend happy by sharing a desirable animal (kitten/butterfly). Because of a false belief the agent ends up sharing an undesirable animal (skunk/spider). In the bad intent stories, the agent wanted to make their friend upset, but because of a false belief, ended up sharing a desirable animal. Unlike in the first study of Ochoa et al., the vignettes involved only one box as follows. The agent first opened the box, stating what animal was in it and whether it would make their friend happy or sad. The next animal then appeared on the scene with the agent stating whether it would make the friend happy or sad. At this point the agent left, after which the animal in the box jumped out and also left. The other animal then jumped in the box. The agent then returned and, when their friend arrived, they gave them the box that now contained the unintended animal (see Table 3.1).

Table 3.1

Study 1: Experimental Manipulation of Intent and Belief Factors to Yield Morally-Relevant Belief Vignette Outcomes

Story Stimuli	Intent Condition	X Belief	= Outcome
Kitten/Skunk	Good	False	Agent intended to share a kitten but due to a false belief a skunk was shared. (accidental harm)
Butterfly/Spider	Bad	False	Agent intended to share a spider but due to a false belief a butterfly was shared. (accidental positive outcome)

Children were asked three comprehension questions over the course of the story: two of these concerned whether each of the items would make the recipient feel good or bad (e.g., "Do kittens make [RECIPIENT] feel good or bad?") and one concerned whether the agent was present or absent for the switch (e.g., "Was [AGENT] there to see the skunk get in the box and the kitten leave?").

Following each vignette, participants responded to six test questions in a fixed order:

(1) intention evaluation ("When [AGENT] handed [RECIPIENT] the box, was [AGENT] trying to be nice, mean, or just okay?"

(2) agent belief ("What does [AGENT] think is in the container?");

(3) reality check ("What is really in the container?")

(4) consequence evaluation ("Should [AGENT] get in trouble, like a timeout, a treat, like a trip to the zoo, or nothing? If [AGENT] gets nothing, he will not get a timeout or a trip.")

(5) open ended consequence explanation ("Why should [AGENT] get [assigned consequence)?]

(6) ("How will [RECIPIENT] feel when he opens the container?").

Participants' responses to the open-ended punishment/reward questions were coded in three ways: (1) content of the justification, (2) correctness of the justification, and (3) whether the justification matched the assigned punishment/reward. The content of the justification was coded as referencing: Mental state, such as, intention, knowledge, or belief (e.g., "she didn't know the spider got in the box", "trying to make her friend happy"); Outcome (e.g., "there's a butterfly in the box"); or Undifferentiated/Uninterpretable (e.g., "I don't know", or "just because"). The justification was also coded for correctness (i.e., judging whether it matched facts from the story): Correct (e.g., "she was trying to be nice" in the good intent condition), Incorrect (e.g., "she was trying to be mean" in the good intent condition), or Undifferentiated/Uninterpretable. Lastly, the justification was coded for whether it matched the assigned punishment/reward (e.g., assign a reward to an agent because "she was trying to do something nice") or Incorrect (e.g., assign a reward to the agent even though they state that "She was being mean"). In most cases assigning nothing, (no treat or reward) was coded as a match because that would be considered acceptable in both scenarios. If the initial justification was coded as undifferentiated, it remained undifferentiated in this coding. Two

independent coders rated all observations and attained high inter-rater reliability (Cohen's K = .82), with 90.3% agreement across 792 observations. All discrepancies were resolved in discussion between the two coders and the first author.

Adults answered the same questions as children. The only difference was that adults were asked the three comprehension questions at the conclusion of each vignette because we were not concerned about memory demands for them, whereas children were asked during each vignette.

Individual Difference Measures for Children

Executive Function

The following four executive function measures were used to make an executive function composite. These measures are different from the measures that children completed in the first study of Ochoa et al. (2020). We wanted to use the NIH Toolbox because they could be completed on an iPad and were automatically scored and standardized and are appropriate for children ages 3 years and older.

NIH Toolbox Dimensional Change Card Task. Children completed the dimensional change card sort task (DCCT), from the NIH Toolbox Kit, on an iPad. This task is designed to measure children's executive functioning and attention by asking children to sort a set of cards by one dimension (e.g., shape) and then by a different dimension (e.g., color). If children did well on the first two blocks (passing at least 4 out of 5 trials for each block), they were asked to sort by shape and color within the same block (30 trials). Children received a score from 0-40 depending on how many trials they correctly answered across all blocks. (Gershon et al., 2013).

NIH Toolbox Flanker Inhibitory Control and Attention Task. Children completed the NIH Toolbox Kit Flanker Inhibitory Control Task (Flanker) on an iPad (Gershon et al., 2013). The Flanker is a measure of attention and inhibitory control. Children are asked to focus on a fish in the center of the screen and choose the arrow that points in the direction which the fish is pointing. The center fish has several other fish flanking on each side. Children are reminded to look at the middle (where the star is) before each trial. Children receive familiarization trials and corrective feedback before moving on to the 20 test trials. Children received a score between 0-20 based on how many of the test trials they got correct.

Corsi Block Tapping Test. In order to measure working memory, another aspect of executive function, children completed the Corsi Block Tapping Test forwards and backwards (Farrell et al., 2006). Children are asked to watch a researcher tap a pattern of blocks and then repeat the same pattern (or in a backwards manner). Children complete practice trials with corrective feedback, if needed. The pattern starts with one block and can increase up to 9 blocks. The block stops as soon as two consecutive errors are made. Children complete two blocks of each forward and backwards tests. Working memory, for forwards and backwards, was measured by the highest number of blocks which children can correctly reproduce. The scores spans from 0-9.

Head-Toes-Knees-Shoulders. The Head-Toes-Knees-Shoulders (HTKS) task measures different aspects of executive function, including inhibitory control, working memory, and attention (Ponitz, 2008). Children play a game in which they do the opposite of what an experiment asks them to do. For example, if the

experimenter says head they should touch their toes. Children are given familiarization trials and corrective feedback. If children do well on the initial head/toes block, then shoulders/knees are included in the next block. The task ends when children make 3 consecutive errors, or complete all 40 trials. Each item is scored from 0-2 (0 = incorrect, 1 = self-correct, or 2 = correct). Scores range from 0-52 and are calculated based on the total of the 6 practice trials and 20 test items.

Verbal Ability

NIH Toolbox Picture Vocabulary Task. Children completed the NIH Toolbox Kit Picture Vocabulary Task (PVT) on an iPad (Gershon et al., 2013). The PVT is an adaptive vocabulary task that measures receptive vocabulary. Four images are displayed and then children are asked to pick the image that represents the spoken word. If children get a word correct the vocabulary gets harder, and if they get it incorrect the vocabulary gets easier. Generally, children completed about 25 trials. Each child receives a *Theta score* (which is the overall performance of an individual, similar to a Z score, but calculated via item response theory).

Results

Analyses

All analyses were conducted in *R* (R Core Team, 2017) and figures were produced using the *ggplot2* package (Wickham, 2009). We conducted a series of 2 (intent: bad or good) X 4 (false belief group: false belief failers, 4-year-old passers, 5-year-old passers, and adult passers) ANOVAs on children's responses to most questions, and appropriate post-hoc or pairwise comparisons were conducted to examine interaction effects.

Individual difference measures were examined for outliers and deviation from the assumption of normality. Only extreme outliers were winsorized. In order to be considered an extreme outlier a data point must be beyond an outer fence (lower or upper) of the boxplot.

No main effects of gender, story order (girl vs. boy story first), or experimenter were found. Therefore, these factors were collapsed in subsequent analyses. We used the recipient emotion prediction question as a comprehension check. The majority of children and adults appropriately judged that the recipient would feel sad in the good intent condition (M = -.88, SD = .42, and M = -.96, SD = .21, respectively) and happy in the bad intent condition (M = .69, SD = .67, and M = .60, SD = .68, respectively).

Agent Belief

A 2 (intent: good vs. bad) x 4 (age: 4- vs. 5-year-olds) ANOVA was conducted in response to the agent belief question. Adults were not included in this analysis because they were essentially performing at ceiling in both conditions: good intent (M = 1, SD = 0); bad intent condition (M = .98, SD = 0.11). A main effect of age revealed 5-year-olds (M = 0.74, SD = 0.41) ascribing the appropriate belief more often than 4-year-olds (M = 0.58, SD = 0.46), $F(1,105) = 4.41$, $p = .04$, $\eta_p^2 = .03$. The main effect of intent and the interaction were not significant. As in Study 1, the sample was then divided into groups as a function of belief understanding: participants who failed one or both belief questions (N = 50, 48 children, M_{age} = 57.80 months, SD = 6.76 months, and 2 adults), 4-year-olds who passed both (N = 30,

M_{age} = 53.60 months, *SD* = 3.19), 5-year-olds who passed both (*N* = 31, M_{age} = 65.6 months, *SD* = 3.39), and adults who passed both (*N* = 40).

Agent Intention

Figure 3 (top panel) depicts performance on the intention question. Adults performed at ceiling: reporting that the agent in the bad intent condition was trying to be mean and the agent in the good intent condition was trying to be nice. For children, a 3 (false belief group) x 2 (intent) factorial ANOVA revealed a main effect of intent, $F(1,102) = 28.95$, $p < .001$, $\eta_p^2 = .22$, but not a main effect of false belief group, $F(1,102) = 1.67$, $p = 0.19$. Overall, children appropriately rated agents in the good intent condition (*M* = 0.32, *SD* = .70) as significantly better intentioned than those in the bad intent condition (*M* = -0.34 , *SD* = .68). However, this main effect was qualified by an interaction between the false belief group and intention, $F(2,102) = 9.13$, $p = .004$, $\eta_p^2 = .10$. Follow-up simple effects comparisons of intent for each false belief group revealed that false belief failers did not significantly differentiate intention ratings across the good and bad intent conditions, $p = .40$. In contrast, 4- and 5-year-old passers appropriately rated agents in the good intent condition as significantly better intentioned than agents in the bad intent condition, *ts*(29 and 30) = -3.61 and -5.15, *ps* = .006 and < .001, respectively. Further, 4- and 5-year-old passers performed significantly better than chance in all conditions (*ps* < .02), except for 4-year-olds in the good intent condition who were only marginally different from chance, $t(13) = 1.86$, $p = .09$.

Agent Consequence

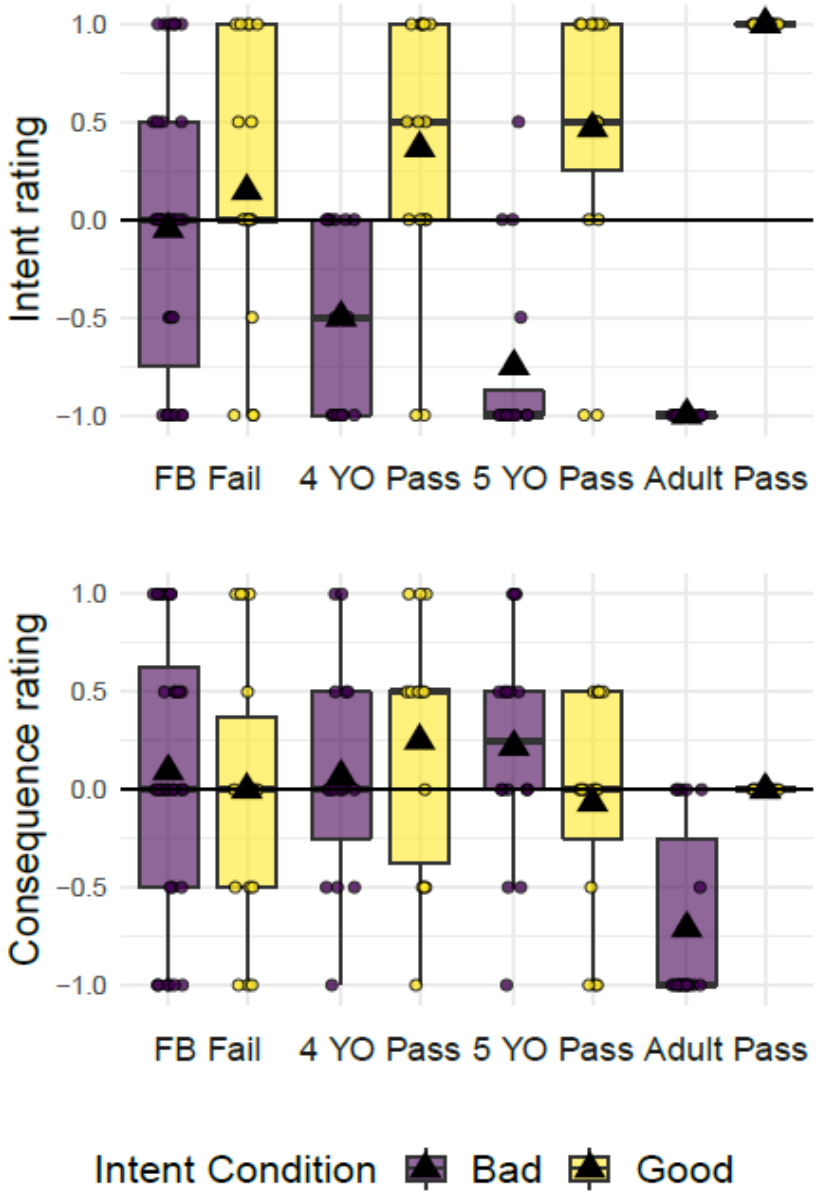
Figure 3 (bottom panel) depicts performance on the agent consequence question. A 4 (group: false belief failers, 4-year-old passers, 5-year-old passers, and adult passers) x 2 (intent: good vs. bad) factorial ANOVA revealed a main effect of false belief group, $F(3, 142) = 5.92$, $p < .001$, $\eta_p^2 = .11$, and a significant interaction between false belief group and condition, $F(3, 142) = 5.37$, $p = 002$, $\eta_p^2 = .10$. The main effect of intent was not significant, $F(1, 142) = 2.19$, $p > .05$, $\eta_p^2 = .02$. Follow-up simple effects comparisons revealed that only adults appropriately rated agents in bad intent conditions as deserving more punishment than those in good intent conditions, $t(39) = -3.92$, $p = .003$. No other false belief group distinguished punishment ratings for agents in the good versus bad intent conditions (*ps* > .05).

We conducted the same follow-up analysis as in the first study of Ochoa et al. to examine whether false belief groups differed in whether their consequence ratings were correct or incorrect. Specifically, in the bad intent condition, assigning trouble is a clear correct response but assigning nothing could also be considered reasonable as there is no negative outcome. Assigning a treat in this condition is clearly incorrect. Conversely, in the good intent condition, assigning a treat is a clear correct response but assigning nothing could also be considered reasonable as there is no positive outcome. Assigning trouble in this condition is clearly incorrect. Children thus received a score from 0-2 based on their correct/incorrect responses across the two trials coded as just described. Adults were excluded because their ratings in both conditions were at ceiling. For children, a 3 (false belief group) x 2 (intent) ANOVA on correct consequence ratings did not reveal any main effects or an interaction between false belief group and intent condition (*ps* > .05). False belief

failers (52% correct), 4-year-old passers (53% correct), and 5-year-old passers (48% correct) had similar difficulty assigning consequences in the false belief context, with no group performing better than chance, *ts*(47, 28, 30) = -0.17, 0.90, -0.31, *ps* > .37.

Figure 3.1

Intent and Consequence Judgment Separated by False Belief Group and Intent Condition (Ochoa, Rodini, & Moses, 2021)



Note. The black triangles represent the mean of the group and dots represent individual responses.

Justification for Punishment or Reward

Finally, we examined the proportion of justifications for punishment/reward broken down by false belief group for the content of the justification, correctness of the justification, and whether the justification matched the assigned punishment/reward. Because 4- and 5-year-old false belief passers revealed essentially the same pattern of justifications across the three codes, we collapsed them for this analysis. For children, a large portion of justifications were coded as undifferentiated/uninterpretable (43% vs. 10% for adults). Of the justifications that were interpretable, the majority for all false belief groups matched the assigned punishment/reward: 92% false belief failers, 97% Child passers, and 100% Adult passers.

With respect to the content of the justifications, mental states were referred to more often than outcomes, and that was true even for false belief failers (see Table 3.2). In a further analysis, however, we examined whether the mental state reference was consistent with the story. Even though false belief failers referenced mental states more than outcomes, they often did so incorrectly. Only 42% of false belief failers who referenced a mental state did so correctly, versus 85% for child false belief passers and 97% for adult passers. The other 58% of false belief failers referenced mental states that were inconsistent with the vignettes (e.g., stating that "he wanted to give his friend something nice" in the bad intent condition or "because she did something mean" in the good intent condition).

Table 3.2

Study 1: Proportion of Justification Types for Punishment/Reward by False Belief Group

False Belief Group	Number of Justifications	Content of Justification		
		Mental State	Outcome	Undiff.*
False Belief Failers	100	.39	.15	.46
Child Passers	122	.39	.22	.39
Adult Passers	80	.88	0	.12

*Undiff = Undifferentiated

Executive Function and Moral Evaluations

Only children who participated in-lab and completed at least 3 out of the 5 EF measures are included in subsequent analyses *(N = 83)*. Descriptive statistics for individual difference measures by false belief group are displayed in Table 3.3. An EF composite measure was created by z-scoring the average performance on the 5 tasks: DCCT, Corsi Block Forwards and Backward, HTKS, and Flanker. See Table 3.3. There were two outliers for the Corsi Block Forwards measure. After winsorizing that measure and rerunning all analyses the interpretation remained the same. Therefore, for all subsequent analyses we used the raw, non-winsorized, EF composite measure.

Table 3.3

Study 1: Means and Standard Deviations for Individual Difference Measures by False Belief Group

	False Belief Group		
	False Belief Failers *N* = 39	4-Year-Old Passers *N* = 23	5-Year-Old Passers *N* = 21
Variable	*M (SD)*	*M (SD)*	*M (SD)*
Dimensional Change Card Task (EF)	21.90 (16.03)	20.40 (14.93)	27.05 (14.13)
Corsi Block Tapping Test - Forward (EF)	3.33 (1.15)	3.65 (0.88)	4.05 (0.67)
Corsi Block Tapping Test - Backward (EF)	2.25 (1.02)	2.40 (0.94)	2.62 (0.86)
Head Toes Knees Shoulders (EF)	21.87 (17.69)	33.30 (13.59)	39.76 (11.76)
Flanker Inhibitory Control and Attention Task (EF)	30.90 (11.43)	26.94 (13.34)	35.56 (7.15)
Aggregate of EF tasks, *z*-scored	-.28 (.82)	-.02 (.62)	.40 (.42)

Note. *M* = mean, *SD* = standard deviation.

First, we wanted to examine whether false belief group and age could predict the EF composite score for children. A 2(age group: 4 vs. 5) x 2(false belief group: pass vs. fail) ANOVA revealed significant main effect of age ($F(1,79) = 12.22$, $p < .001$, $\eta_p^2 = .13$, 95% CI [.03, .27]) and false belief group on the EF composite ($F(1,79) = 10.46$, $p = 0.002$, $\eta_p^2 = .12$, 95% CI [.02, .25]). However, there was no interaction between age and false belief group ($F(1,79) = 0.62$, $p > 0.05$). Five-year-olds ($M = 0.33$ $SD = 0.47$), had significantly higher EF composite scores than 4-year-olds (M = -

0.27, *SD* = 0.77). Further, false belief passers (*M* = 0.18, *SD* = 0.57) had higher EF composite scores than false belief failers (*M* = -0.28, *SD* = 0.82).

Table 3.4

Study 1: Raw Correlations Among Executive Function Measures

Executive Function Measure	1.	2.	3.	4.	5.
1. Dimensional Change Card Task	1	-	-	-	-
2. Corsi Block Tapping Test - Forward	.18	1	-	-	-
3. Corsi Block Tapping Test - Backward	.22~	.24*	1	-	-
4. Head Toes Knees Shoulders	.53***	.50***	.38**	1	-
5. Flanker Inhibitory Control and Attention Task	.24~	.48***	.50***	.56***	1

Note. ***p < .001, **p <.01, * p <.05, ~p <.10

Second, we conducted a logistic regression to examine whether the EF composite could predict children's intent judgments (correct vs. incorrect) while controlling for age and verbal ability. Children's responses were coded as correct if they answered that the agents in the good intent condition were being nice and agents in the bad intent condition were being mean for both stories. EF did not significantly predict intent judgments (see Table 3.5).

Table 3.5

Study 1: Logistic Regression of Executive Function Composite Predicting Intent Judgments

Variable	*B*	*SE*	*z*	*p*
Constant	-4.43	3.21	-1.38	.17
Age (months)	0.07	0.05	1.53	.13
Verbal ability	-0.14	0.21	-0.68	.50
EF composite	0.56	0.37	1.50	.14

Finally, we wanted to see whether EF was related to children's consequence judgments. Children's responses were coded as correct if they answered that the agents in the good intent condition should get either a treat or nothing, and incorrect if they assigned trouble. In the bad intent condition, children's responses were coded as correct if they reported that the agents should get in trouble or nothing, and incorrect if they assigned treat. EF was not related to children's consequence judgments (see Table 3.6)

Table 3.6

Study 1: Logistic Regression of Executive Function Composite Predicting Consequence Judgments

Variable	*B*	*SE*	*z*	*p*
Constant	1.92	2.96	0.65	.52
Age (months)	-0.004	0.04	-0.08	.94
Verbal ability	0.39	0.21	1.85	.06
EF composite	-0.06	0.37	-0.16	.87

Discussion

In this study we examined whether the developmental differences found in the first study of Ochoa et al. (2020) would persist when information processing demands were reduced. In contrast to Study 1, both 4- and 5-year-old false belief passers appropriately rated agents in the good intent condition as significantly better intentioned than agents in the bad intent condition. In the first study only 5-year-old passers had made this distinction. It appears that reducing the task processing demands may have helped younger children focus on the relevant information needed to make accurate moral judgments.

We had also hypothesized that 4-year-old passers would have more difficulty assigning appropriate consequences than 5-year-old passers. However, we again found that children of all ages, regardless of false belief understanding, had difficulty doing so. Despite the reduced processing demands, children did not significantly differentiate consequence ratings for agents in the good versus bad intent conditions.

In addition, we hypothesized that consequence judgments would be harder than intent judgments for children. As in the first study of Ochoa et al., this hypothesis was confirmed: 4- and 5-year-old false belief passers were often able to properly assign agent intent, but generally did not make use of this information when assigning a punishment/reward. With respect to consequence justifications, we found that all groups tended to reference mental states more often than outcomes. This initial finding was surprising because previous research suggests that younger children, and especially those without false belief understanding, often

focus more on outcomes when assigning consequences (Cushman et al., 2013; Zelazo et al., 1996). However, when examining whether participants' references to mental states were accurate, children without false belief understanding often incorrectly referenced agents' intentions by aligning those intentions with their mischaracterizations of agents' beliefs. For example, in the good intent condition, they appeared to think that the agent knew there was an unpleasant object in the container, and therefore they thought he or she must have a bad intention. As a result, they recommended punishment for the agent. In that sense, these children were perhaps still outcome-focused — they assigned intentions to the agent in their justifications that matched the outcome of the vignette.

Further, we found that 4- and 5-year-old false belief passers' moral judgments differed from those of adults in several respects. Both children and adults significantly distinguished good intentions from bad intentions. However, whereas adults were at ceiling in doing so, children were much more variable. Moreover, while children did not significantly distinguish good versus bad intent in their consequence judgments, adults had little difficulty in doing so.

As in study one of Ochoa et al. (2020), we found that older children had higher executive composite scores than younger children, and that individual differences in executive function predicted false belief understanding. Yet, we did not find a relationship between executive function and moral judgments. This could be in part due to reducing the executive demands of the moral vignettes. Further research should aim to better understand the relationship between executive function and moral judgments.

CHAPTER VI

STUDY 2

Although we found revealing interactions between false belief understanding and moral judgment in Study 1 of my dissertation, some important questions were left unanswered. First, as just discussed, our research shows that by age 5 children's moral judgments are not yet comparable to those of adults; even the oldest children in our sample failed to assign consequences differentially as a function of intent in the false belief context. Therefore, in Study 2 we extended the age range from 5-year-olds to 7-year-olds to examine whether adult levels of understanding are achieved by this age.

Second, we included a more fine-grained scale of punishment and reward rather than a simpler trichotomy. In Study 1 of my dissertation, children were asked to pick between either assigning trouble, assigning a treat, or nothing. In contrast, in Killen et al. (2011), children were additionally given the option to decide whether the consequence should be small or large. In Killen et al., responses were coded as 2 for a lot of punishment, 1 for a little punishment, and 0 for no punishment. 3-to 4-year-olds viewed it as very acceptable to punish (*M* = 1.75/2, SD = .29), 5 to 6-year-olds on average leaned more towards a little punishment (*M* = .80, SD = .21), and 7- to 8-year-olds viewed it as least acceptable to punish (*M* = .64, *SD* = .25). Using a graded scale appeared to uncover developmental increments in understanding that may have been harder to detect using the “all or nothing” response options used in Study 1.

Therefore, in Study 2 we asked children to assign either a little trouble or a lot of trouble, assign a small treat or big treat, or nothing. It may be that the majority of children in Study 1 chose nothing in the bad intent condition because they felt that the given options were not appropriate given that the bad intention did not match the positive outcome. Children may be more willing to assign a little punishment in the bad intent condition rather than a lot of punishment since the outcome was not negative, even though the intent was. Similarly, children may feel it is more acceptable to give a well-intentioned agent a small treat, but not a large one because the outcome was still negative.

Third, although we didn't find an effect of intent on consequence judgments, it's possible that children's judgments of *praise and blame* would be more closely linked to intention (and belief) than reward and punishment. Malle (2013) suggests that when we assign punishment we are evaluating the act, but when we are judging someone as blameworthy or praiseworthy we are evaluating agents for their role in that event. While we may hesitate to give rewards when outcomes are inadvertently bad, we may feel increased freedom to offer praise for good intentions in those cases; conversely, while we may not punish when good outcomes are fortuitously brought about by ill-intentioned agents, we may nonetheless be quite willing to blame those agents in such cases. In that respect, praise and blame may represent something of a way station on children's road to incorporating mental states, and specifically intent understanding, into consequence judgments.

Further, more recent research with adults suggests that individuals show an asymmetry when assigning praise and blame, such that they are more likely to blame

a person for bad side effects than praise for good side effects (Knobe, 2003; Sarin et al., 2017). For example, Helen hates her sister and she wants her to look bad at prom, so she agrees to make her sister a dress. She made an ugly dress but the dress ended up fitting perfectly and she was chosen as Prom Queen (good side-effect). In a separate situation, Helen loves her sister and she wants her to look good at prom, so she agrees to make her a dress. The dress ended up fitting Helen's sister very poorly and she got bullied at school for it (bad side-effect). In this case, a side effect is generally brought on as a result of something that may not have been intentional. Knobe (2003) notes that individuals are more likely to say that a negative side effect was brought on intentionally (and is therefore blameworthy) compared to a side effect that is positive. Similarly, Sarin and colleagues (2017) found an asymmetry in adults' judgments: agents who were well-intentioned, but brought about a negative outcome were held more responsible and blameworthy than an ill-intentioned agent who brought about a positive outcome. Hence, in Study 2 we examined participants' blame/praise judgments and whether children would do better assigning these judgments compared to consequence judgments.

Fourth, it's possible that individual differences in empathy may also be related to moral judgments in these contexts. Most research focuses on the influence of behavioral and cognitive processes, but affective processes, such as empathy, are also important for more complex moral judgments (Hoffman, 2000; Turiel, 2007). For example, Ball et al. (2017) examined the influence of affective processes, in this case empathy, and cognitive processes on moral judgments with a sample of socioeconomically diverse 3.5-year-olds. Children heard social

conventional and moral transgression stories and were asked to rate the severity of the transgression (i.e., “Is it OK or not OK for (transgressor) to X?” If the child responded that it was not OK, they then were asked, “Is it a little bit bad or very bad?”) and were asked if the transgressor should be punished. As previously discussed, greater empathy was associated with increased severity in judgments of moral transgression (specifically psychological harm), but not social conventional transgressions. Further, children with greater empathy assigned more punishment than children with less empathy. The researchers suggest that empathy and theory of mind may help children understand the consequences of moral transgressions on others. While theory of mind understanding may allow children to interpret others’ mental states, empathy may allow them to understand what it may be like to be the victim of the transgressions. While this research suggests a link between empathy and moral judgments, we do not know whether that link persists when the intention is not in line with the outcome, as in our false belief scenarios. One possibility is that children with high empathy focus on the outcome and therefore have a harder time incorporating intent information when assigning reward or punishment because they are empathizing with the victim of the transgression. Alternatively, at least in the good intent condition, children with higher empathy may also empathize with the transgressor who made a mistake. This is an open question that we plan to examine by assessing child empathy in study 2.

A final point of difference from Study 1 was that Study 2 was conducted online from February-May 2021 because of the COVID-19 pandemic. As detailed in

the procedure, interactions with children thus necessarily took a somewhat different form although the content remained similar to the earlier study.

Our hypotheses and analysis plan were preregistered on the Open Science Framework prior to analyzing the data. The preregistration can be found at https://osf.io/yprwk/.

Study 2 Hypotheses

Do 5- and 7- year-olds integrate belief and intent information when making moral judgments? Specifically we looked at whether children differentiate their moral judgments based on agent intent (positive vs negative) in situations in which the agent holds a false belief.

Hypothesis 1a: We predicted that there would be a main effect of age, such that older children would be more likely than younger children to be false belief passers.

Hypothesis 1b. We also predicted that there would be an age-related change in making intent judgments, such that overall older children would be more likely than younger children to correctly make intent judgments in false belief contexts.

How do 5-year-olds, 7-year-olds, and adults make consequence judgments in scenarios in which agent intent and outcome of the situation do not align, due to a false belief? We examined whether children differentiated their consequence judgments based on agent intent (positive vs negative) in situations in which the agent holds a false belief. As previously mentioned, we examined consequence judgments in two different ways. First, they were asked whether the agent should get in trouble (a little or a lot), a treat (small or big), or nothing. Second, using a 5-

point scale comparable to the consequence scale, we asked whether the agent should be praised, blamed, or neither.

Hypothesis 2a. For the consequence evaluation, we hypothesized an interaction between age group and agent intent. As in Study 1, we expected that 5-year-olds would show little understanding of consequence evaluation in false belief contexts. Yet, we expected that 7-year-olds would differentiate their consequence rating based on agent intent. Whether the older children would do as well as adults was an open question. Further, we believed that older children and adults, compared to younger children, would be more likely to assign trouble in the bad intent condition. Yet, we did not expect a difference in the good intent condition, because we thought that the majority of participants would respond that the appropriate consequence would be nothing, as adults did in Study 1.

Hypothesis 2b. When assessing blame and praise judgments, we predicted that, unlike for punishment/reward, younger children would be able to appropriately blame the agent in the bad intent condition. Therefore, we did not expect a main effect of age, but we did hypothesize a main effect of condition such that participants would blame the agent in the bad intent condition and offer a neutral response or praise in the good intent condition.

Hypothesis 2c. For children, we predicted that assigning blame and praise would be easier than assigning consequences. For 5-year-olds, we did not think children would differentiate their consequence judgments based on intent, but we did think they would when making blame/praise judgments. For 7-year-olds, we expected to see more of an adult-like pattern of responses for consequence

judgments, such that they would often punish in the bad intent condition, but give nothing in the good intent condition. Further, we expected 7-year-olds to blame the agent for the bad intention, saying it was a bad/mean thing to do, and likely either say nothing or praise in the good intent condition. For adults, we expected a similar pattern of judgments of consequence as in Study 1 - mostly assigning punishment in the bad intent condition, but assigning nothing in the good intent condition. As for assigning blame/praise, we predicted that in the bad intent condition most adults would assign blame, and sometimes assign praise in the good intent condition or choose to say nothing.

Is there a relationship between children's empathy and moral judgments?

Hypothesis 3: In the good intent condition, we predicted that the relation between empathy and consequences could manifest in two different ways.

Children with higher empathy may ignore the agent's good intent and punish the agent because they empathize with the recipient, or victim in this case. Alternatively, children with higher empathy may empathize with the well-intentioned agent even though the outcome is negative. They may feel badly for the agent, recognizing that mistakes can happen, and offer the agent a reward (or at the very least not punish).

We did not have specific hypotheses about the bad intent condition, because we were not sure there would be a clear relation between empathy and consequence assignment. As adults, we would presumably not empathize with an agent who didn't successfully fulfill their ill-intentioned goal. On the other hand, we

might feel happy for the recipient who receives a pleasant gift and that may lead some to reward the agent. This was an open question which we explore in this study.

Study 2 Overview

Study 2 examined children's (5- and 7-year-olds) and adults' moral judgments using the pre-recorded morally relevant vignettes from Study 1. As in Study 1 we asked participants to make belief, intent, and consequence judgments. The belief and intent judgments remained the same but we adapted the consequence judgment to include a 5- point graded scale. Further, we added a 5-point blame and praise judgment to assess whether children and adults make these judgments based on belief and intent. Lastly, we examined whether empathy is related to moral judgments. All participants reported their online empathy for the agent and recipient. In addition, we asked parents to report on their children's empathy. We were particularly interested as to whether empathy judgments are related to consequence and blame and praise judgments.

Methods

Participants

Sixty-two children were recruited via a developmental database from the University of Oregon, and 4 children were recruited from a developmental science listserv (Parent and Researcher Collaborative, 2021). Families were provided compensation in the form of $10 Amazon or Powell's City of Books electronic gift cards.

Sixty-six children participated (24 female, 34 male, and 8 parents did not report a gender): 35 5-year-olds (M_{age} = 64.10 months, *SD* = 2.81) and 31 7-year-olds

(M_{age} = 89.40 months, *SD* = 3.65). An additional 3 children were tested but excluded from analyses because the parent reported their child as having developmental delays.

The sample was representative of the population from which it was drawn. Fifty-two parents reported their child's race as White or Caucasian, 3 parents reported their child was of Hispanic, Latino, or Spanish origin, 3 reported children as being two or more races, and 11 parents did not report their child's race. Sixty-one percent of families reported making at least $75,000 a year, and 62% of children were in school or daycare at least part-time.

Thirty-six adults (75% female, M_{age} = 19.92 years, *SD* = 3.14 years) from the Psychology and Linguistics Human Subjects Pool at the University of Oregon also participated. The majority (78%) reported being White or Caucasian, and 53% reported being Freshman students. Adult participants received compensation in the form of class credit for a psychology course.

Design

Within each age group, participants were randomly assigned to one of two intent conditions (good or bad) in which they responded to two morally-relevant belief vignettes. An *a priori* power analysis indicated at least 30 false belief passers would be required in each age group (5-year-olds, 7-year-olds, and adults) for the study to be adequately powered. We aimed to collect up to 45 participants per age group, or when we had 30 false belief passers per age group, because it was likely that a few participants per age group would not pass the false belief task. Ultimately, we recruited 35 5-year-olds, 31 7-year-olds, and 36 adults.

Procedure

Children were tested individually in a single 15-minute live video session with a researcher. Children were engaging in remote learning and had the technology available needed to participate in our study. We asked that the caregiver was present at the beginning of the session in case of any technical difficulties. It was up to the caregiver whether they wanted to sit next to the child or not. The researcher shared their computer screen and played the recorded video vignettes, and then asked children the test questions. Adults completed all tasks online through Qualtrics on their own time.

Measures

Morally-relevant Belief Vignettes

The pre-recorded vignettes were the same videos that adults watched in Study 1. Each video was approximately 2 minutes long, and participants watched two stories with vignette order counterbalanced across conditions. As in Study 1, good intent stories involved an agent who wanted to make a friend happy by sharing a desirable animal but because of a false belief the agent ends up sharing an undesirable animal. In the bad intent stories, the agent wanted to make their friend upset, but because of a false belief, ended up sharing a desirable animal.

Following each vignette, participants were asked the same three comprehension questions as in Study 1. The first two concerned whether each of the items would make the recipient feel good or bad and the last concerned whether the agent was present or absent for the switch.

Participants then responded to the eight test questions described below. An image of the relevant character (agent/recipient) was paired with each question to direct children's attention to the character we are asking about.

A Likert-type happy face scale was used to accompany the blame/praise question and the two empathy questions, and the consequence scale was used for the consequence evaluation question. Training trials for the scales were conducted before using the scales for the first time.

1. Agent intention evaluation ("When [AGENT] handed [RECIPIENT] the box, was [AGENT] trying to be nice, mean, or just okay?").

2. Agent belief ("What does [AGENT] think is in the container?").

3. Reality check ("What is really in the container?").

4. Consequence evaluation ("What should happen to [AGENT]?"Should he/she get in:
    - a lot of trouble
    - a little trouble
    - nothing
    - a small treat
    - or should he/she get a big treat?)

The consequence question was accompanied by the following pictorial scale.

5. Open-ended consequence explanation ("Why should [AGENT] get [assigned consequence]?)

6. Blame/praise evaluation ("What should we say to [AGENT]?"

    - You tried to do a really, really nice thing. That was good!
    - You tried to do a nice thing.
    - Nothing.
    - You tried to do a mean thing.

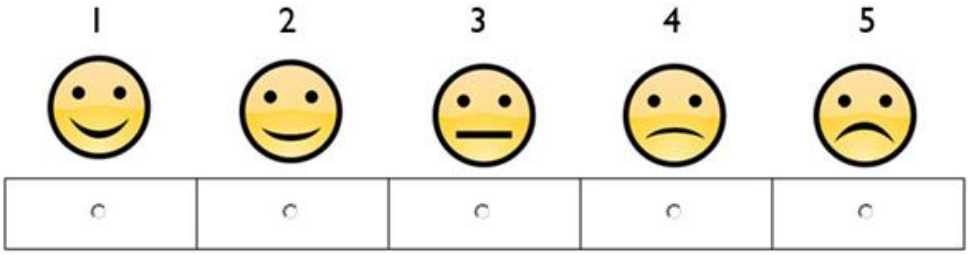
- You tried to do a really, really mean thing. That was bad!

Before asking the following two empathy questions, the researcher said: "Now [RECIPIENT] opens the box and sees a [actual animal] inside!"

7. Empathy - Agent ("How do you feel for [AGENT]?").
    - Really happy for her/him.
    - A little happy.
    - Not happy or sad.
    - A little sad.
    - Or, really sad for her/him?

8. Empathy - Recipient ("How do you feel for [RECIPIENT]?").
    - Really happy for her/him.
    - A little happy.
    - Not happy or sad.
    - A little sad.
    - Or, really sad for her/him?

The Likert-type happy face scale is shown below:



The order of the blame/praise and consequence questions was counterbalanced between-subjects. Likewise, the order of the agent empathy and recipient empathy questions was counterbalanced.

As in Study 1, participants were credited with false belief understanding if they correctly answered the agent and reality check question in both stories. Scoring for the test measures fell into three schemes: Agent Belief and Reality Check (correct or incorrect), Agent Intention Evaluation (a three-point scale based on answer valance (-1 = mean, 0 = just okay, and 1 = nice), and the remainder of the forced-choice questions (a continuous, five-point scale based on answer valence, with the

most negatively-valenced answers coded -2, neutral answers coded 0, and most positively-valenced answers coded 2). The open-ended Consequence Evaluation will be coded at a later date and is not discussed further here.

Caregiver Measures

***Demographics and Family Questionnaire*.** Parents/guardians were asked to complete a 13-item questionnaire that asks about race and ethnicity, and socioeconomic status. See Appendix A.

Griffith Empathy Measure (Dadds et al., 2008). The Griffith Empathy Measure (GEM) is a parent-report measure of child empathy that was adapted from the Bryant (1984) empathy scale. The measure contains both affective and cognitive components of empathy. The scale contains 23 items; for each item parents use a likert-type scale (-4 strongly disagree to +4 strongly agree) to report how much they agree or disagree with each statement. See Appendix B.

In addition, we asked parents to report on their childs' mental state understanding (Children's Social Understanding Scale; Tahiroglu et al., 2014) and their own parenting in different discipline situations (The Parenting Scale; Arnold et al., 1993). These measures were not the primary focus of this study and are not discussed further here.

Adult Measures.

Demographics Questionnaire. Adults filled out a 6-item questionnaire about race and ethnicity, family socioeconomic status, class standing, and age. See Appendix C.

Further, we asked adult participants to complete the Empathy Quotient (Baron-Cohen & Wheelright, 2004), Adult Autism Spectrum Questionnaire (Baron-Cohen et al., 2001), Reflective Functioning Questionnaire (Fonagy et al., 2016), Moral Foundations Questionnaire (Graham et al., 2013), and the Interpersonal Reactivity Index (Davis, 1980). These measures were not the primary focus of this study and are not discussed further here.

Results

Analyses

We only included participants who passed both false belief questions (N = 97) in the moral vignettes. A total of 61 children passed the false belief questions in both vignettes; 30 5-year-olds (15 in the bad intent condition, M_{age} = 64.00 months, SD = 2.88) and 31 7-year-olds (17 in the bad intent condition, M_{age} = 89.40, months, SD = 3.65). Thirty-six adults passed the false belief questions (18 in the bad intent condition). There were five 5-year-olds (3 in the bad intent condition) and three adults (2 in the bad intent condition) who did not pass false belief questions and were excluded from further analyses.

All analyses were conducted in *R* (R Core Team, 2017) and figures were produced using the *ggplot2* package (Wickham, 2009). We conducted a series of 2 (intent: bad or good) X 3 (age group: 5, 7, adult) ANOVAs on children's responses to most questions, and appropriate post-hoc or pairwise comparisons were conducted to examine interaction effects. Individual difference measures were examined for outliers and deviation from the assumption of normality. Only extreme outliers were

winsorized or removed. In order to be considered an extreme outlier a data point needed to be beyond an outer fence (lower or upper) of the boxplot.

Means and standard deviations for the main variables of interest can be found in **Table 4.1.**

Agent Intention

Figure 4.1 depicts performance on the intention question. Almost all adults and all 7-year-olds performed at ceiling: reporting that the agent in the bad intent condition was trying to be mean and the agent in the good intent condition was trying to be nice. Since there was no variability in 7-year-olds' responses, and close to zero in adult responses, it was not appropriate to conduct an ANOVA as planned.

Even though 5-year-olds' responses were more variable than 7-year-olds and adults, they still appropriately differentiated their intent judgments based on intent condition $F(1,28) = 60.83$, $p < .001$, $\eta_p^2 = .69$, and performed significantly better than chance in both the good ($t(14) = 10.46$, $p < .001$, $d = 2.70$) and bad intent condition ($t(14) = -3.52$, $p = .003$, $d = -.91$).

Agent Consequence

Figure 4.2 depicts performance on the agent consequence question. A 3 (age group) x 2 (intent) factorial ANOVA revealed a main effect of intent on the consequence judgments, $F(1, 91) = 23.87$, $p < .001$, $\eta_p^2 = .21$, but no main effect of age group ($F(2, 91) = 0.10$, $p = .91$ $\eta_p^2 = .01$) or significant interaction ($F(2, 91) = 1.53$, $p = .22$, $\eta_p^2 = .03$). Overall, participants deemed agents in the bad intent condition as deserving of a negative consequence ($M = -.75$, $SD = 1.02$) and agents in the good intent condition as deserving of a positive consequence ($M = .34$, $SD = 1.16$).

Table 4.1

Study 2: Means and Standard Deviations for Primary Variables of Interest by Age Group and Intent Condition

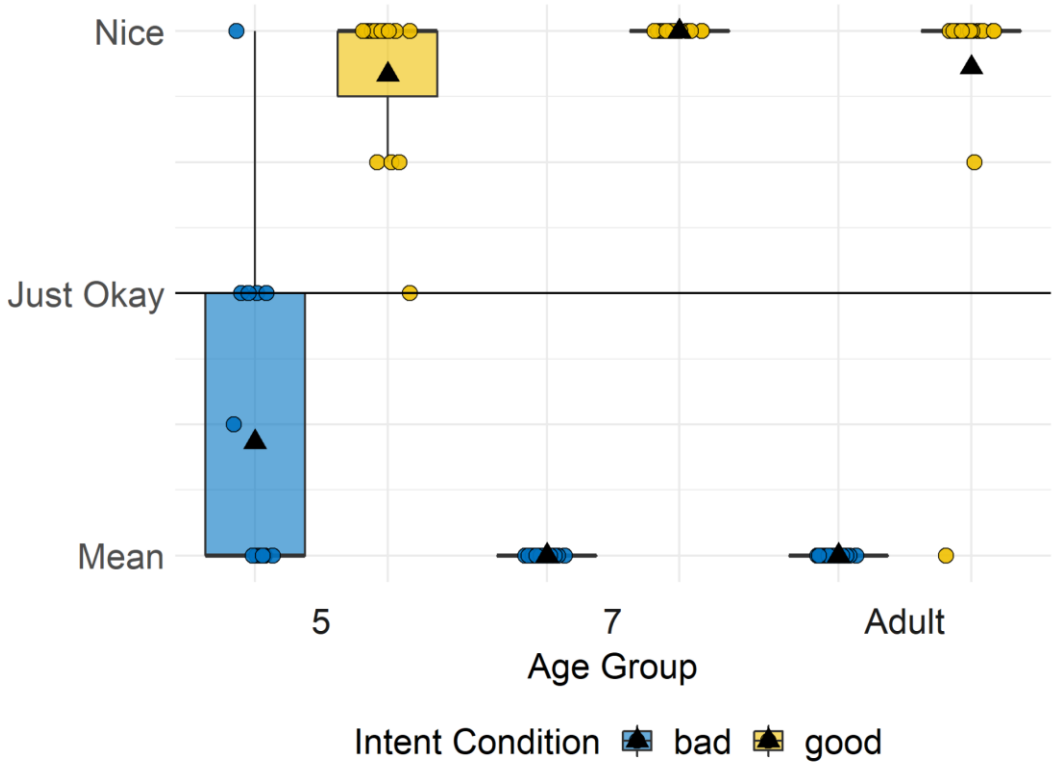
Moral Judgment	Age Group	Intent Condition	*M (SD)*
Intent	5	Bad	-0.57 (0.62)
	7	Bad	-1.00 (0.00)
	Adult	Bad	-1.00 (0.00)
	5	Good	0.83 (0.31)
	7	Good	1.00 (0.00)
	Adult	Good	0.86 (0.48)
Consequence	5	Bad	-0.83 (1.29)
	7	Bad	-0.56 (1.16)
	Adult	Bad	-0.86 (0.54)
	5	Good	0.43 (1.37)
	7	Good	-0.04 (1.31)
	Adult	Good	0.56 (0.78)
Blame/Praise	5	Bad	-1.13 (0.92)
	7	Bad	-1.24(0.53)

Table 4.1 continued

Moral Judgment	Age Group	Intent Condition	*M (SD)*
Blame/Praise	Adult	Bad	-1.39(0.58)
	5	Good	0.43 (1.37)
	7	Good	1.39(0.71)
	Adult	Good	1.42(0.86)
Agent Empathy	5	Bad	-0.13 (1.46)
	7	Bad	-0.21 (0.64)
	Adult	Bad	0.11 (0.50)
	5	Good	-0.43 (1.13)
	7	Good	-0.54 (1.18)
	Adult	Good	-1.39 (0.58)
Recipient Empathy	5	Bad	1.43 (1.15)
	7	Bad	1.71 (0.40)
	Adult	Bad	1.14 (0.89)
	5	Good	-1.17 (1.11)
	7	Good	-1.46 (0.54)
	Adult	Good	-1.35 (0.61)

Figure 4.1

Intent Judgment Separated by Age Group and Intent Condition in Study 2



Note. The black triangles represent the mean of the group and dots represent individual responses.

Follow-up simple effects comparisons revealed that adults $t(91) = -3.89$, $p = .003$) and 5-year-olds ($t(91) = -3.17$, $p = .02$) rated agents in bad intent conditions as deserving more punishment than those in good intent conditions, but 7-year-olds did not ($t(91) = -1.32$, $p = .77$). Because the interaction was not significant, however, we do not give great weight to the failure of 7-year-olds to strongly differentiate reward/punishment on the basis of intent.

We conducted the same follow-up analysis as in Study 1 to examine the extent to which age groups differed in whether their consequence ratings were

correct or incorrect. In the bad intent condition, we coded the following as correct: nothing, little trouble, and a lot of trouble. In the good intent condition, we coded the following as correct: nothing, little treat, and big treat.

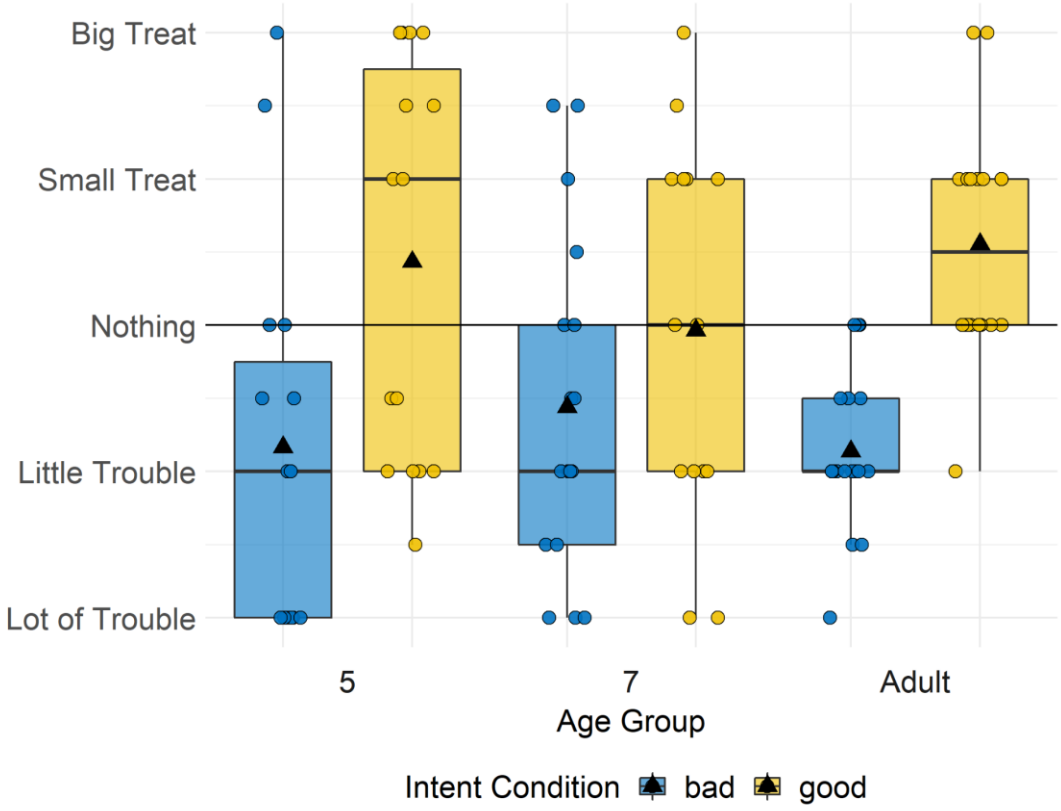
A 3 (age group) x 2 (intent) ANOVA on correct consequence ratings (0-2) revealed a main effect of age group ($F(2, 91) = 6.89$ $p = .002$, $\eta_p^2 = .13$), and intent condition ($F(1, 91) = 4.42$, $p = .05$, $\eta_p^2 = .03$), but not a significant interaction between the two ($F(2, 91) = 0.56$, $p = .57$, $\eta_p^2 = .01$). Overall, participants made more correct consequence judgments in the bad intent condition (56% correct) compared to the good intent condition (44% correct). Further, adults (97.2%) made more correct judgments overall compared to both 5-year-olds (66.7% correct) and 7-year-olds (64.5% correct), *p*s = .01, .003. As in study 1, children did not perform better than chance *(p*s > .15) which is not surprising because only adults consistently differentiated consequence judgments based on intent.

Agent Blame/Praise Judgment

Figure 4.3 depicts performance on the agent blame/praise judgment. A 3 (age group) x 2 (intent) factorial ANOVA revealed a main effect of intent on the blame/praise judgment, $F(1, 90) = 338.27$, $p < .001$, $\eta_p^2 = .79$, but no main effect of age group ($F(2, 90) = 0.60$, $p = .55$, $\eta_p^2 = .02$) or significant interaction ($F(2, 90) = 0.18$, $p = .84$, $\eta_p^2 = .04$). All age groups appropriately rated the agent in the good intent condition as more praiseworthy than the agent in the bad intent condition, who was rated as more blameworthy.

Figure 4.2

Consequence Judgment Separated by Age Group and Intent Condition in Study 2



Note. The black triangles represent the mean of the group and dots represent individual responses.

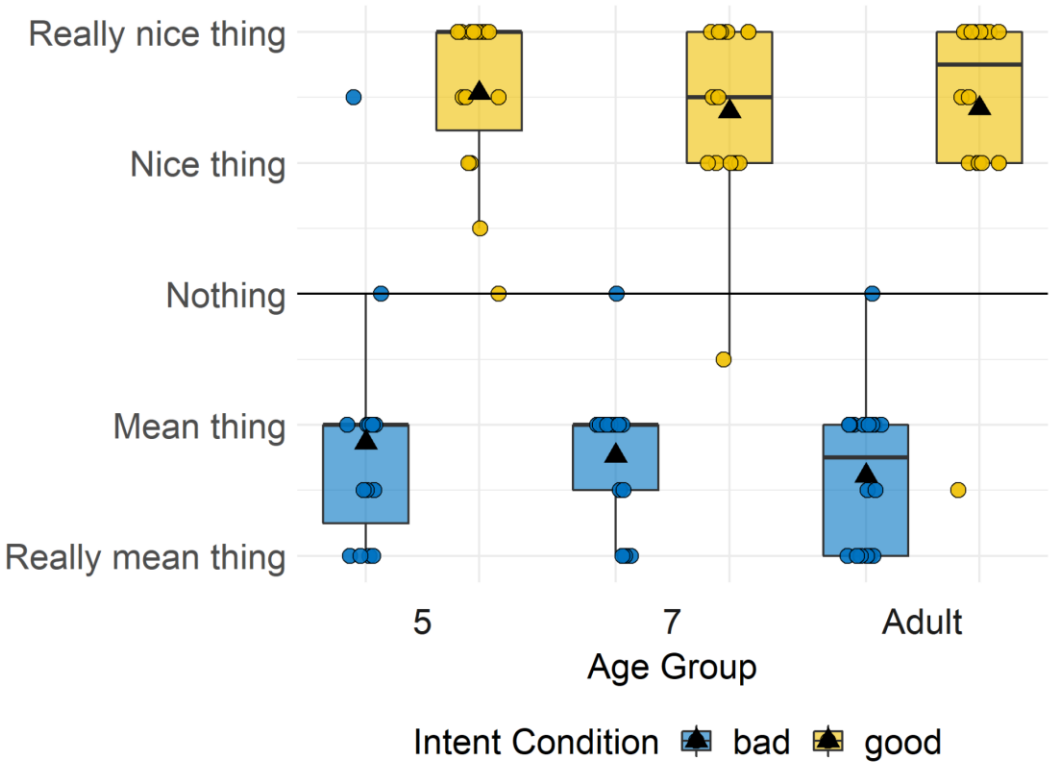
We conducted the same follow-up analysis as we did for consequence judgment to examine whether age groups differed in whether their blame/praise ratings were correct or incorrect. In the bad intent condition we coded the following judgments as correct: nothing, mean thing, really mean thing. In the good intent condition we coded the following as correct: nothing, nice thing, and really nice thing.

A 3 (age group) x 2 (intent) ANOVA on blame/praise judgments (0-2) did not reveal any significant effects. This is likely because all age groups did well, 93.3% of

5- year-olds, 96.8% of 7-year-olds, and 97.2% of adults made correct blame/praise judgments. In contrast to the agent consequence judgment, all age groups performed better than chance on the blame/praise judgments.

Figure 4.3

Blame/Praise Judgment Separated by Age Group and Intent Condition in Study 2



Note. The black triangles represent the mean of the group and dots represent individual responses.

In an exploratory analysis we further examined the extent to which correct consequence and blame/praise judgments were correlated. There was a significant positive correlation between judgments (r = .24, p = .02), although this relationship should be interpreted with caution since only 4 participants did not correctly answer the blame/praise judgments. In addition, when we examined the relation between blame/praise judgments and consequence judgments using the graded scale, they

were not related in either the good intent ($r(45) = .21$, $p = .15$) or bad intent conditions ($r(48) = .08$, $p = .57$).

Agent Empathy

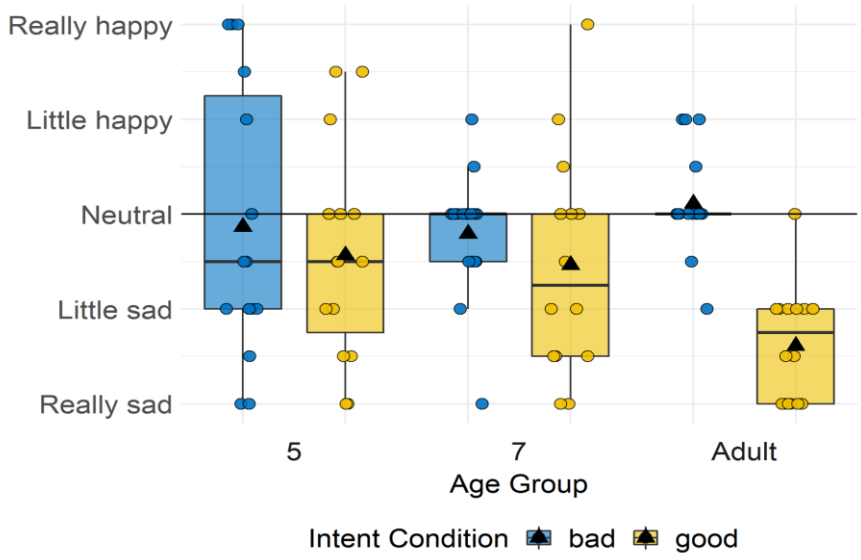
Figure 4.4 depicts performance on the agent empathy question. A 3 (age group) x 2 (intent) factorial ANOVA revealed a main effect of intent on the empathy question, $F(1, 91) = 15.25$, $p < .001$, $\eta_p^2 = .14$, and a significant interaction, $F(2, 91) = 4.38$, $p = .02$, $\eta_p^2 = .09$, but no main effect of age group, $F(2, 91) = 1.23$, $p = .30$, $\eta_p^2 = .03$. Overall, participants felt sadder for the agent in the good intent condition than in the bad intent condition. However, simple effects tests showed that adults were the only group that significantly differentiated their empathy between the agent in the good intent ($M = -1.39$, $SD = 0.58$) versus the bad intent condition ($M = -0.11$, $SD = 0.50$), $t(91) = 4.73$, $p < .001$.

Recipient Empathy

Figure 4.5 depicts performance on the recipient empathy question. A 3 (age group) x 2 (intent) factorial ANOVA revealed only a main effect of agent intent on the recipient empathy question, $F(1, 90) = 264.27$, $p < .001$, $\eta_p^2 = .75$. Overall, participants felt sadder for the recipient in the good intent condition than the bad intent condition, and happier for the recipient in the bad intent condition. The main effect of age ($F(2, 90) = 1.01$, $p = .37$, $\eta_p^2 = .03$) and interaction were not significant ($F(2, 90) = 2.11$, $p = .22$, $\eta_p^2 = .03$).

Figure 4.4

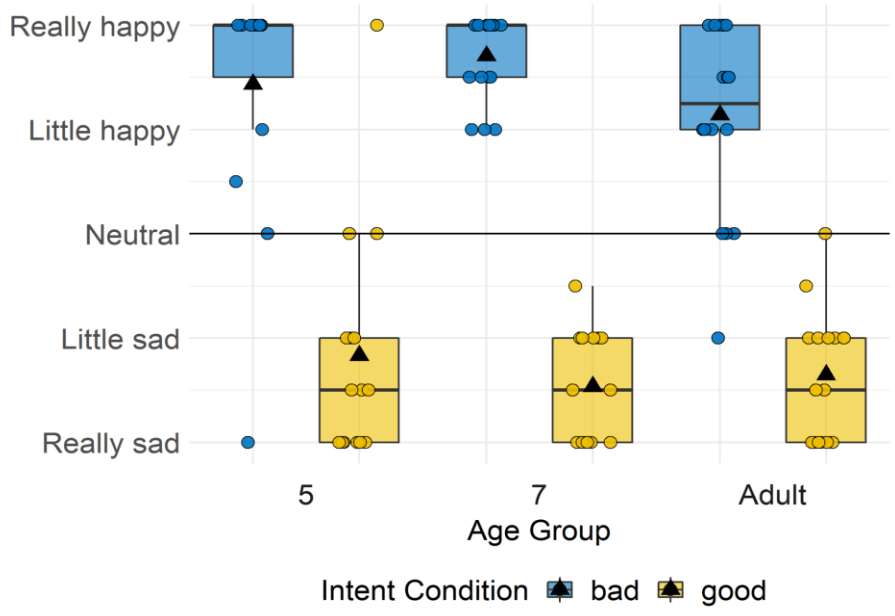
Agent Empathy Judgment Separated by Age Group and Intent Condition in Study 2



Note. The black triangles represent the mean of the group and dots represent individual responses.

Figure 4.5

Recipient Empathy Judgment Separated by Age Group and Intent Condition in Study 2



Note. The black triangles represent the mean of the group and dots represent individual responses.

Correlates of Children's Moral Judgments

We explored the possible relations between children's empathy and moral judgments in two ways. First, we examined whether children's online empathy judgments for the agent and recipient were related to their average consequence and blame/praise judgments.

In the good intent condition, empathy for the agent was marginally correlated with agent consequence ($r(27) = .34$, $p = .07$). Likewise, in the bad intent condition, empathy for the agent was significantly correlated with agent consequence ($r(20) = .54$, $p = .001$). In both conditions, children who felt happier for the agent were more likely to assign a positive consequence, and children who felt sadder for the agent assigned more negative consequences. Yet, empathy for the agent was not related to children's blame/praise judgments in either condition ($ps > .20$). It's important to note that most children did well on the blame/praise judgments, and therefore not leaving much variability for relations with other variables to emerge.

Children's empathy for the recipient was not significantly correlated with their consequence judgments for the agent in either intent condition ($ps > .05$). Empathy for the recipient was not correlated with children's blame/praise judgment in the bad intent condition, but was negatively correlated with those judgments in the good intent condition ($r(27) = -.41$, $p = .03$). This relation was heavily affected by a single child saying that they felt really happy for the recipient in the good intent condition, whereas the rest of the children generally said they felt sad or very sad for

the recipient. The correlation was no longer significant when we removed that child's response ($r(26) = -.17$, $p = .38$).

Table 4.2

Raw Correlations Among Empathy Measures and Consequence and Blame/Praise Judgments for Children in the Good Intent Condition

Measure	1.	2.	3.	4.	5.
1. Agent Empathy	1	-	-	-	-
2. Recipient Empathy	.41*	1	-	-	-
3. Parent-Report Empathy	-.27	-.10	1	-	-
4. Consequence Average	.34~	.15	-.40*	1	-
5. Blame/Praise Average	-.25	-.41*	.32	.15	1

Note. ***p* <.01, * *p* <.05, ~*p* <.10

Second, we examined whether parent-report of empathy, using the Griffith Empathy Measure (GEM) (Dadds et al., 2008), was related to children's consequence judgment and blame/praise judgment. See Table 4.2 for descriptive statistics. We used the total score from the GEM, with higher scores representing more empathy.

Table 4.3

Raw Correlations Among Empathy Measures and Consequence and Blame/Praise Judgments for Children in the Bad Intent Condition

Measure	1.	2.	3.	4.	5.
1. Agent Empathy	1	-	-	-	-
2. Recipient Empathy	-.25	1	-	-	-
3. Parent-Report Empathy	-.01	-.19	1	-	-
4. Consequence Average	.54**	-.30~	.33~	1	-
5. Blame/Praise Average	.10	.04	.08	.01	1

Note. ***p* <.01, * *p* <.05, ~*p* <.10

In the good intent condition, parent-report of empathy was significantly negatively correlated with consequence judgments ($r(25) = -.41$, $p = .04$): children with higher empathy reported more negative consequences for the agent in the good intent condition. This could be because they empathized with the victim who received a negatively valenced animal.

In the bad intent condition, parent-report of empathy was marginally correlated with consequence judgments ($r(27) = .33$, $p = .08$). Children with higher empathy reported more positive consequences for the agent in the bad intent condition and those with lower empathy reported more negative consequences. It's important to note that in the bad intent condition there were two outliers of parent-reported empathy. However, after winsorizing the data, the results and

interpretation remained the same. In sum then, when making consequence judgments, children with higher parent-reported empathy tended to be more swayed by the negative or positive outcome for the recipient than by the agent's intentions. In contrast, parent-report of empathy was not related to children's blame/praise judgments in either the good intent ($r(25) = .32$, $p = .10$) or bad intent ($r(27) = .08$, $p = .66$) condition.

Table 4.4

Study 2: Means and Standard Deviations for Griffith Empathy Measure by Age Group and Intent Condition

Age Group	Intent Condition	n	*M*	*SD*
5	bad	15	27.20	28.70
5	good	15	36.10	15.90
7	bad	17	41.00	16.30
7	good	14	29.20	11.20

Discussion

In this study we extended the age range from 5-year-olds to 7-year-olds to examine whether moral judgments are more similar to adults at this age. In addition, we added a graded scale of consequence to examine whether children would be more likely to appropriately punish/reward agents if given a more nuanced set of response alternatives. Further, we asked participants if they wanted to blame or

praise the agent using a graded scale to test whether they might find assigning blame and praise easier than assigning punishment and reward. Lastly, we examined whether individual differences in children's empathy are related to their moral judgments.

As predicted, older children were more likely than younger children to be false belief passers. Further, 100% of the 7-year-olds correctly answered the intent judgments in both questions, whereas only 70% of 5-year-olds did. Nonetheless, similar to Study 1, 5-year-olds differentiated their judgments based on agent intent.

The second aim of this study was to compare 5-year-olds, 7-year-olds, and adults' consequence judgments in scenarios in which agent intent and outcome of the situation do not align, due to a false belief. Using a graded scale of consequences, we hypothesized an interaction between age group and intent condition for consequence judgments such that most participants would assign nothing in the good intent condition, but only 7-year-olds and adults would assign punishment in the bad intent condition. Although the overall model did not reveal an interaction, there was a significant main effect of intent. Participants deemed agents in the bad intent condition as more deserving of a negative consequence than agents in the good intent condition. As in Study 1, we conducted follow-up simple effects comparisons. We found that 5-year-olds and adults deemed the agents in the bad intent condition as more deserving of punishment compared to the agent in the good intent condition but, surprisingly, the 7-year-olds did not. This is a novel finding because 5-year-olds did not differentiate consequence judgments in Study 1 and may have been due to the more nuanced measure of reward/punishment. We

actually hypothesized that 7-year-olds would assign consequences similar to the way adults did, but instead they did not differentiate based on agent intent. This finding should be interpreted cautiously because the interaction was not significant. Therefore we don't give too much weight to this finding.

Descriptively, compared to Study 1 (see Table 5.1) more 5-year-olds seemed willing to assign punishment to the agent in the bad intent condition, and more adults were willing to suggest a treat for the agent in the good intent condition. It's possible that participants found the graded punishment scale as more appropriate since the intent and outcome were not aligned. For example, assigning a little trouble in the bad intent condition or a small treat in the good intent condition may be more appropriate than just outright punishing (i.e., time out) or rewarding (i.e., trip to the zoo), as in Study 1. Yet, when consequence judgments were coded as correct and incorrect children did not perform better than chance. In contrast, adults were above chance and made significantly more correct consequence judgments compared to children.

Unlike prior research with adults (Knobe, 2003; Sarin et al., 2017), we did not find an asymmetry when assigning blame and praise. As hypothesized, even 5-year-olds blamed the agent in the bad intent condition and praised the agent in the good intent condition and did so to the same extent as older children and adults. Moreover, when scored as correct/incorrect, over 90% of participants in each age group responded correctly. Thus, assigning blame/praise was considerably easier for children than assigning reward/punishment. It seems plausible that making blame

and praise judgments may serve as a stepping stone for making consequence judgments. This is discussed more fully in the general discussion.

A subsequent aim of Study 2 was to examine whether children's individual differences in empathy are related to moral judgments. Children's empathy for the agent was not related to their blame/praise judgments. Yet, their empathy for the agent was related to their consequence judgments. In both conditions (marginally significant for good intent) children who felt more sad for the agent assigned more negative consequences, and those who felt happier for the agent assigned positive consequences. It's not exactly clear why this pattern is evident, but it's possible that the order of the questions influenced children's responses. Children were first asked to make consequences and blame/praise judgments, so it may be that children feel sad that the agent is getting in trouble or happy that the agent is getting rewarded. This is something we plan to examine in future research. A separate explanation is that they are focusing primarily on the outcome of the scenario when making consequence judgments. One way we plan to explore this further is by coding children's open-ended responses to the consequence judgment and examining whether children are referencing outcome more often than intention.

Interestingly, children's empathy for the recipient, which was near ceiling, was not related to children's consequence judgments. Further, empathy for the recipient was not related to children's blame/praise judgments in the good intent condition. Even though agent recipient was related to blame/praise judgments in the good intent condition, this correlation was heavily influenced by an outlier; once removed the correlation was no longer significant.

We also examined the relation between parent-report of children's empathy and moral judgments. Although parent-report of empathy was not related to children's blame/praise judgments, it was correlated with children's consequence judgments. Generally we found that children with higher parent-reported empathy tended to be influenced by the outcome and not the agent's intentions when making consequence judgments. Conversely, those with lower parent-reported empathy generally seemed to assign consequences based on intent and this was particularly true in the bad intent condition.

CHAPTER V

GENERAL DISCUSSION AND CONCLUSIONS

Understanding others' mental states, or Theory of Mind (ToM), is critical for social functioning and maintaining positive social relations (Wellman, 2020). Similarly, moral reasoning and judgments can guide our behaviors and decisions, ultimately to be successful in the social world. ToM understanding is critically important for moral judgments and both understandings are developing quickly during childhood. Although the influence of intent understanding on children's moral development has been long studied (e.g., Piaget, 1932), very little research has examined the influence of belief understanding on that development. In this discussion we review the main findings of each study and then further discuss our results in relation to other research, and lastly we will discuss limitations of this research and provide ideas for future directions.

The purpose of this dissertation was to further investigate the interplay between ToM, specifically belief and intent understanding, and moral development during childhood. We also examined how children's moral judgments compare to adults' and whether individual differences in executive function and empathy are related to moral judgments.

In two studies we presented children with morally-relevant belief vignettes to examine the extent to which they incorporate both intent and belief information in their moral judgments. By sampling within the developmental window during which false belief understanding is acquired and consolidated, we found important

developmental changes in children's ability to integrate mental state understanding with moral reasoning.

In prior work, the first study of Ochoa et al. (2020) presented children with moral vignettes to examine the extent to which they incorporate both belief (true and false) intent and information in their moral judgments. Four- and 5- year-olds made appropriate intent and punishment judgments when an agent held a true belief in both intent conditions (see Table 5.1 for means and standard deviations for primary variables of interest for all studies). In the false belief context, two developmental lags were evident: 4-year-olds with false belief understanding did not apply that information to intent judgments, and 4- and 5-year-olds with false belief understanding did not use their belief and intent understanding to make appropriate consequence judgments.

Building on that work, we questioned children concerning belief, intent, and punishment judgments in both studies and asked about blame and praise judgments in Study 2. We also examined individual differences in executive functioning in Study 1 and individual differences in empathy in Study 2 and how these may relate to children's moral judgments.

In Study 1, we attempted to replicate the developmental lag between children's false belief understanding and integrating that understanding with moral evaluations found in the first study of Ochoa et al. (2020). We also added an adult comparison group to establish a developmental endpoint.

Table 5.1

Means for Primary Variables of Interest by Age Group and Intent Condition for All Studies

	Age/FB Group	Intent Condition	Ochoa et al. (2020)	Study 1	Study 2
Intent			*M*	*M*	*M*
	FB Fail	Bad	0.15	-0.05	-
	FB Fail	Good	0	0.15	-
	4	Bad	0.39	-0.50	-
	4	Good	0.30	0.37	-
	5	Bad	-0.48	-0.75	-0.57
	5	Good	0.62	0.47	0.83
	7	Bad	-	-	-1
	7	Good	-	-	1
	Adult	Bad	-	-1	-1
	Adult	Good	-	1	0.86
Consequence					
	FB Fail	Bad	0.54	0.09	-

Table 5.1 continued

	Age/FB Group	Intent Condition	Ochoa et al. (2020)	Study 1	Study 2
Consequence			*M*	*M*	*M*
	FB Fail	Good	0.25	0.00	-
	4	Bad	0.50	0.07	-
	4	Good	0.20	0.25	-
	5	Bad	0.04	0.22	-0.83
	5	Good	0.11	-0.07	0.43
	7	Bad	-	-	-0.56
	7	Good	-	-	-0.04
	Adult	Bad	-	-0.71	-0.86
	Adult	Good	-	0.00	0.56

Note. Scale range for intent judgments is -1 to 1. Scale range for consequence judgments is -1 to 1 in the first study of Ochoa et al. (2020) and Study 1, but is -2 to 2 in Study 2.

In circumstances with reduced processing demands we found that the developmental lag for intent judgments was no longer evident: 4-year-old, as well as 5-year-old false belief passers, now performed better than chance in attributing

agents' intentions. Nonetheless, children of both ages continued to have difficulty making appropriate punishment/reward judgments, indicative of a developmental lag between assigning moral intent and assigning consequences. In contrast, adults showed no such difficulty, performing almost at ceiling on both intent and consequence questions (when the latter were scored as correct/incorrect).

In Study 2 we included a 7-year-old comparison group and a more fine-grained measure of consequence in addition to a blame/praise judgment. Overall, we found that 7-year-olds performed similarly to adults in making intent judgments, but even at this age children did not distinguish consequences based on intent. Yet, in assessing blame and praise we found that both 5- and 7-year-olds assigned blame/praise based on agent intent to the same extent as adults. It's possible that being able to make intent-based blame/praise judgments is a stepping stone for making consequence judgments. This is just one possibility: There are many other factors such as parents' use of reward and punishment or children's own experience assigning consequences, that may also serve as stepping stones for making consequence judgments. Overall, our results suggest that there is still development around moral reasoning that will occur throughout childhood.

Our findings are both similar to and different from those of Killen et al. (2011), who also found that false belief understanding is related to moral judgment. Specifically, they found that only the oldest children (7- and 8-year-olds) attributed good intentions to an accidental transgressor in a false belief context in which a good intention generated a moral violation. We extended Killen et al 's approach by also including scenarios in which an agent had a negative intention, but due to a false

belief ending up making a friend happy. Like Killen et al., we found that false belief understanding was linked to moral judgment. However, unlike Killen et al., our children made the link at an earlier age, with most 5-year-olds showing an understanding of moral intentions in the false belief context. Nonetheless, it is worth noting that not all 5-year-olds in Study 1 and Study 2 correctly inferred intent whereas all 7-year-olds did in Study 2.

Further, children in Killen et al. (2011) deemed punishing the accidental transgressor as less acceptable with age, whereas generally children in our studies did not make a clear distinction in whether to punish or reward agents in either the good or bad intent condition. The findings across their research and ours are difficult to compare, however, because the question formats and wording differ. That said, their 5-year-olds appeared to have had similar difficulty to children of that age in our sample.

Our findings with respect to the difficulty children have in assigning reward and punishment parallel some findings from Cushman et al. (2013). They found that while children can make appropriate moral wrongness judgments based on intent, their punishment judgments are based more on outcome. Specifically, in Cushman et al.'s research children were asked to make wrongness judgments (presumably based on intent) and punishment judgments for cases of accidental harm (benign intent, bad outcome) versus attempted harm (negative intent, neutral outcome). Children, 5-years-old and older, assigned higher levels of punishment than wrongness in the case of accidental harm, and, conversely, higher levels of wrongness than punishment in the case of attempted harm. In contrast, 4-year-olds did not clearly

distinguish punishment and wrongness in either case. Thus, as in our studies, appropriately incorporating consequences on the basis of intent was difficult for children.

In study 2, 5- and 7-year-olds made adult-like blame and praise judgments: They blamed the agent with the negative intention and praised the agent with the positive intention. This is particularly interesting because Malle (2021) suggests that blame judgments are one of the most complex judgments because we take into account other judgments, such as evaluations, norm judgments, and moral wrongness. For example, when making *evaluations* we judge whether something is good/positive or bad/negative, when making *norm judgments* we judge whether something is permissible or forbidden, and when making *wrongness judgments* we assess whether something is morally wrong or immoral. Malle believes in order to make appropriate blame judgments we assess all the aforementioned judgments. Further, in Piaget's (1932) work, children did not assign blame based on intent until at least 8- or 9-years-old. Therefore, the fact that children are making these complex judgments, that require sophisticated information processing (Malle, 2021), is impressive.

Although there has been much research on blame judgments, there has been less research and attention on moral praise. Anderson and colleagues (2020) suggest that the function and process guiding moral praise is starkly different from that guiding moral blame. Historically, watching who gets blamed can help us to know who to avoid and who shares our character and values while the primary function of praise is to help with relationships and cooperation. Due to the different functions of

blame and praise, we may not provide them equally, yet this was not evident in our study. Unlike Sarin and colleagues (2017), we found that children and adults praised the agent with the good intention even though there was a bad outcome and blamed the agent with a negative intention even though there was a positive outcome. This suggests that children are really focusing on the agent's role in the act and not judging the act itself.

Individual Differences in Executive Function and Empathy in Relation to Moral Judgments

A secondary aim of this dissertation was to examine whether individual differences in executive function and empathy were related to children's moral judgments. We were particularly interested in examining executive function as the relation between executive function and false belief understanding is well-documented (Carlson & Moses, 2001; Carlson et al., 1998; Hala et al., 2003). In addition, we examined empathy because prior research suggests that emotion understanding, an aspect of ToM, and empathy are helpful for making complex judgments (Ball et al., 2017; Hoffman, 2000).

In Study 1 we found that even though executive function predicted false belief understanding, it was not related to intent and consequence judgments. It's possible that only a certain level of executive function is needed, for example in order to track the agent's intention and knowledge, to gain false belief understanding but it may not be the underpinning of more advanced moral judgments. It may also be that a certain subcomponent of executive function, such as emotional control, is particularly important for moral judgments. In Study 2 we

examined whether children's on-line empathy for the agent and recipient, and parent-report of children's empathy were related to moral judgments. Overall, we found that children's empathy for the agent was positively related to agent consequence judgments but not blame/praise judgments. As discussed, in the Study 2 discussion, it's possible that the order of the questions influenced children's responses. For example, in the bad intent condition children who assigned more negative consequences to the agent felt sadder for the agent. This may have been influenced by the order of the questions. It's possible that if the empathy questions were asked first the responses may differ. Further, empathy for the recipient was not related to children's consequence or blame/praise judgments.

In Study 2 we also found that parent-report of empathy was not related to children's blame/praise judgments but was related to children's consequence judgments. We found an intriguing pattern: Children with higher parent-reported empathy seemed to be influenced by the outcome in both conditions because they assigned a positive consequence when the outcome was good and a negative consequence when the outcome was bad. Ball and colleagues (2017) found similar results: Children with higher empathy reported greater severity judgments of moral transgressions compared to children with lower empathy. However, their studies included prototypical moral transgressions stories which did not include a false belief. Yet, children with lower parent-reported empathy tended to assign consequences based on intent. Another factor to consider is how parents judge their child's empathy. For example, if a parent tends to observe their child's reaction to an outcome of the situation (child consoling a crying child), but is not aware of the

intentions behind an action (the crying child was not allowed to have the toy he was stealing from another child), they may report their child as very empathic because their child is always responding to the outcome of the situation.

Our results may support Bloom's (2017) view on empathy, such that empathy may not be beneficial in all situations and can actually lead us astray and make us less rational decision makers. Bloom would go as far to say that empathy actually makes us make poor decisions, and this seems to be true to some extent in our study: children with higher parent-reported empathy are rewarding agents with negative intentions and punishing agents with good intentions. Yet, we can not make any strong claims that this is a causal relation because this work is strictly correlation. At the very least, a longitudinal study would be necessary in order to assume children's empathy is directly influencing their moral judgments.

Revisiting the Underpinnings of Moral Development

Again we must consider what potential mechanisms are driving children's moral development. Even though we found evidence that false belief understanding is related to children's moral judgments, children do not automatically gain the ability to make adult-like intent or consequence judgments. Although we did not see evidence of a false belief understanding-to-intent judgment lag for 4-, 5-, and 7-year-olds, we found that the inability to apply that understanding to consequence judgments persisted at least up to 8-years-old.

These results lend support to the constraint hypothesis suggesting that children are truly undergoing a conceptual change and moral development is not being driven solely by domain-general abilities (Cushman et al., 2013). Our results

support this hypothesis in two primary ways: (1) in our study individual differences in executive function were not related to children's moral judgments and (2) even though children made appropriate intent and blame/praise judgments they did not yet make intent-based punishment judgments. The competing hypothesis, the parallel hypothesis, predicts that intent-based wrongness judgments and intent-based punishment judgments occur simultaneously. Although we did not directly assess children's judgments of wrongness (e.g., Is the [agent] a bad, naughty child?) the blame/praise items asked children to assess whether the agent was doing something nice/good or mean/bad. As previously discussed, being able to make these blame/praise judgments may precede and help facilitate children's intent-based punishment judgments. Although in order to be confident that this is the progression of children's moral judgments we would need to follow the children longitudinally.

Limitations and Future Directions

Our results and discussion should be taken in light of several limitations. First, direct comparisons of Study 1 and Study 2 need to be made cautiously for two reasons. Due to the COVID-19 pandemic, Study 2 was conducted virtually and we used pre-recorded vignettes whereas Study 1 was conducted in-person with live demonstrations. It's possible that the medium in which the vignettes were delivered influenced (positively or negatively) children's responses. Although one may have predicted children performing worse because there wasn't a live experimenter acting out the vignette, children actually seemed to do a bit better in Study 2. This could be a result of using a pre-recorded vignette that may have been more

interesting and engaging for children. In addition, parents' presence may have indirectly affected children's responses. In Study 1, almost all children participated with their parents watching from a different room and out of sight, but in Study 2 most parents were sitting next to their children. Even though we asked parents not to interfere with the session unless children were having technical issues, it's possible that children felt more comfortable having their parents nearby and therefore performed better.

Second, we used a different response scale for the consequence question which limits our ability to directly replicate our results. We used a 3-point scale in Study 1 and a 5-point scale in Study 2. We hope to extend and replicate results from Study 2 in the future; specifically because children seemed to perform somewhat better, or at least were more willing to punish in the bad intent condition using the graded scale of consequence.

Even though we preregistered our hypotheses and analysis plan and had a sufficient sample size, it's important to consider the effect sizes of the findings and acknowledge the importance of replication. We do feel confident that intent understanding precedes children's ability to make appropriate consequence judgments because the effect was replicated across both studies and supports prior research. In contrast, even though the effect of intent on blame and praise judgments was very large, these results should be replicated because we used a novel scale. Lastly, because the correlations between empathy and children's moral judgments ranged from small to medium, it's important to determine whether this is a replicable finding in future research.

Importantly, our sample is not representative of all societies and certainly does not represent individuals from diverse cultures and perspectives. A focus on intent in moral situations is not universal even in adult-level moral reasoning across cultures (McNamara et al., 2019). Further, research suggests that cultures emphasize different moral principles based on religion and ideology; for example, cooperation may be vital to survival in certain societies, whereas other societies may be more accepting of selfish behaviors (Gray et al., 2012). In future work we hope to expand our research to include participants from diverse societies.

Our findings might be followed up in a variety of ways. First, unlike adults, even the oldest children in our sample failed to assign consequences differentially as a function of intent in the false belief context. As a result, we do not know when in development children reach adult levels of understanding. Hence, it is important to include older age groups in future work.

Second, another unanswered question is the extent to which children's moral judgments are related to moral behavior in everyday life. Moral behaviors include acting prosocially versus antisocially and lying/cheating versus truth telling. We study moral judgments because we believe it heavily influences the way people act, specifically in moral situations. One study found that children (5 to 9-year-olds) who exhibited prosocial behaviors assigned less punishment to the agent in moral transgression stories compared to children who exhibited fewer prosocial behaviors (Malti et al., 2010). In addition, children who exhibited more aggressive behaviors assigned more punishment to the transgressors. One interesting way to follow-up on our research would be to examine how empathy and prosocial versus antisocial

behaviors influence moral judgments. More research should examine the link between moral judgments and moral behavior to have a better understanding of the relation.

Lastly, it is possible that our scenarios were perplexing to some children given that no explicit reason was provided as to why agents would want to make their friend happy or sad. Although making a friend happy is perhaps something of a default, it may be harder to take in why someone would want to make a friend sad without a specific reason. While children did not appear to be confused about this fact in the true belief context in the first study of Ochoa et al. (2020), it may have proved harder to integrate that information when a false belief was in the mix. Clarifying the reasons behind agents' actions may therefore be an important addition in future work.

Conclusion

In sum, we found revealing interactions between false belief understanding and moral judgment. Only children with false belief understanding were able to reason about intent (good and bad) in contexts in which an agent holds a false belief. Moreover, moral reasoning develops beyond the preschool years, as 5-year-olds' understanding of moral intent itself lagged behind that of 7-year-olds and adults, and even 7-year-olds did not appropriately assign punishment/reward as a function of intent. Yet, by 5 years old, children are making blame and praise judgments based on intent. Overall our findings suggest that integrating theory of mind and moral judgment is a multi-faceted developmental achievement that unfolds only gradually over childhood.

APPENDIX A

PARENT-REPORT DEMOGRAPHICS AND FAMILY QUESTIONNAIRE

Start of Block: Family information questionnaire

Q1 Please fill out the following information about **yourself.**

Q2 What is your gender?

- ○ Male (1)
- ○ Female (2)
- ○ Other (please specify) (3) _____
- ○ Prefer not to say (4)

Q3 What is your age?

_____

Q4 What is your relationship to the child participating in our study?

- ○ Mother (1)
- ○ Father (2)
- ○ Other (please specify) (3) _____

Q5 What is/are the age and gender of your child/children including the child participating in this study? You can leave rows blank if it's not applicable.

	Child gender			Child age
	Male (1)	Female (2)	Other (3)	In years (1)

Participating child (1)	○	○	○
Sibling 1 (2)	○	○	○
Sibling 2 (3)	○	○	○
Sibling 3 (4)	○	○	○
Sibling 4 (5)	○	○	○

Q6 Marital status:

- ○ Married (1)
- ○ Divorced (2)
- ○ Single (3)
- ○ Separated (4)
- ○ Never married (5)
- ○ Other (please specify) (6)

  _____

- ○ prefer not to say (7)

Q7 Education level (please check highest level attained)

	Less than high school (1)	High school graduate (2)	Some college (3)	2 year degree (4)	4 year degree (5)	Graduate degree (6)
Self (1)	○	○	○	○	○	○
Partner- if applicable	○	○	○	○	○	○

Other (please specify) (7)	Prefer not to say (8)
○	○
○	○

Q8 Occupation (self)

_____

Q9 Occupation (partner, if applicable)

_____

Q10 Which category best describes your family annual income?

- Less than $25,000 (1)
- $25,000-$40,000 (2)
- $40,000-$75,000 (3)
- $75,000-$100,000 (4)
- more than $100,000 (5)

Q11 What languages are spoken in the home? Please list all below.

_____

Q12 Is your child an only child?

- Yes (1)
- No (2)

Display This Question:
If Is your child an only child? = No

Q13 What is your child's birth order?

_____

Q14 How much time does your child spend in school/daycare?

- ○ None (1)
- ○ Part-time (2)
- ○ Full-time (3)

End of Block: Family information questionnaire

APPENDIX B

GRIFFITH EMPATHY MEASURE (Dadds et al., 2008)

My child…

1. My child becomes sad when other children are sad.
2. My child gets upset seeing another child being punished for being naughty
3. My child seems to react to the moods of people around them.
4. My child gets upset when another person is acting upset.
5. My child cries or gets upset when seeing another child cry.
6. My child gets sad when watching sad movies or TV.
7. My child becomes nervous when other children around them are nervous.
8. My child acts happy when another person is acting happy. .
9. My child can continue to feel okay even if people around are upset.
10. My child can't understand why other people get upset.
11. My child rarely understands why other people cry.
12. My child would eat the last cookie, even when they know someone else wants it.
13. My child reacts badly when they see people kiss and hug in public.
14. My child doesn't understand why other people cry out of happiness.
15. My child doesn't seem to notice when I get sad. .
16. My child gets sad to see a child with no one to play with.
17. My child treats cats and dogs like they have feelings
18. My child feels sorry for another child who is upset.
19. My child likes to watch people open presents, even if not one for him/her.
20. My child gets upset when seeing another child being hurt.
21. My child laughs when seeing another child laugh.
22. My child gets upset when seeing an animal being hurt.
23. My child feels sad for people who are physically disabled.

Likert-type scale: -4 to (strongly disagree) +4 (strongly agree)

APPENDIX C

ADULT PARTICIPANTS DEMOGRAPHICS QUESTIONNAIRE

What is your age? (In years)

_____

What is your gender?

- Male
- Female
- Other
- Prefer not to answer

What is your class level in school?

- Freshman
- Sophomore
- Junior
- Senior
- Graduate Student
- Not applicable
- Other (please specify) _____

What is the highest level of education attained by either of your parents?

- Less than high school
- High school graduate or equivalency
- Some college
- 2 year / Associate's degree
- 4 year / Bachelor's degree
- Some graduate school
- Master's degree
- Doctorate or professional degree

Estimate to the best of your ability your family's total annual income

- Less than $25,000
- $25,000 - $39,999
- $40,000 - $74,999
- $75,000 - $99,999
- More than $100,000

Are you of Hispanic, Latino, or Spanish origin?

- Yes
- No
- Prefer not to say

What race(s) do you identify as?

REFERENCES CITED

Anderson, R. A., Crockett, M. J., & Pizarro, D. A. (2020). A theory of moral praise. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2020.06.008

Arnold, D. S., O'Leary, S. G., Wolff, L. S., & Acker, M. M. (1993). The Parenting Scale: A measure of dysfunctional parenting in discipline situations. Psychological Assessment, 5(2), 137–144. https://doi.org/10.1037/1040-3590.5.2.137

Apperly, I. A., Warren, F., Andrews, B. J., Grant, J., & Todd, S. (2011). Developmental continuity in theory of mind: Speed and accuracy of belief–desire reasoning in children and adults. *Child development*, *82*(5), 1691-1703. https://doi.org/10.1111/j.1467-8624.2011.01635.x

Arnold, M. L. (2000). Stage, sequence, and sequels: Changing conceptions of morality, post-kohlberg. *Educational Psychology Review, 12(4),* 365-383. doi:http://dx.doi.org/10.1023/A:1009026230018

Astington, J. W. (2003). Sometimes necessary, never sufficient: False-belief understanding and social competence. In B. Repacholi & V. Slaughter (Eds.), Individual differences in theory of mind: Implications for typical and atypical development (pp. 13-38). New York: Psychology Press.

Astington, J. W. (2004). Bridging the gap between theory of mind and moral reasoning. *New Directions for Child and Adolescent Development, (103),* 63–72. http://doi.org/10.1002/cd.98

Baird, J. A., & Astington, J. W. (2005). The development of the intention concept: From the observable world to the unobservable mind. *The new unconscious*, 256-276. http://dx.doi.org/10.1093/acprof:oso/9780195307696.003.0011

Baird, J., & Moses, L. J. (2001). Do Preschoolers Appreciate That Identical Actions May Be Motivated by Different Intentions? *Journal of Cognition and Development, 2(4),* 413-448. https://doi.org/10.1207/S15327647JCD0204_4

Ball, C. L., Smetana, J. G. and Sturge-Apple, M. L. (2017), Following My Head and My Heart: Integrating Preschoolers' Empathy, Theory of Mind, and Moral Judgments. C*hild Development, 88,* 597-611.https://doi.org/10.1111/cdev.12605

Baldwin, D. A. & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences, 5(4),* 171-178. https://doi.org/10.1016/S1364-6613(00)01615-6

Bloom, P. (2017). *Against empathy: The case for rational compassion*. Random House.

Brock, L. L., Kim, H., Kelly, C. L., Mashburn, A. J., & Grissmer, D. W. (2019). Theory of mind, directly and indirectly, facilitates kindergarten adjustment via verbal ability, executive function, and emotion knowledge. *Psychology in the Schools*, *56*(2), 176-193. https://doi.org/10.1002/pits.22216

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37-46. https://doi.org/10.1016/0010-0277(85)90022-8

Boseovski, J. J. (2010). Evidence for "rose-colored glasses": An examination of the positivity bias in young children's personality judgments. *Child Development Perspectives, 4(3),* 212-218. https://doi.org/10.1111/j.1750-8606.2010.00149.x

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child development*, *72*(4), 1032-1053. https://doi.org/10.1111/1467-8624.00333.

Carlson, S., Moses, L., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92. https://doi.org/10.1002/icd.298.

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child development, 69(3*), 672-691. https://doi.org/10.1111/j.1467-8624.1998.tb06236.x

Cowell, J. M., Lee, K., Malcolm-Smith, S., Selcuk, B., Zhou, X., & Decety, J. (2017). The development of generosity and moral cognition across five cultures. *Developmental Science, 20*(4), e12403. https://doi.org/10.1111/desc.12403

Cushman, F., Sheketoff, R.W., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition, 127*, 6-21. https://doi.org/10.1016/j.cognition.2012.11.008

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological science, 17(12),* 1082-1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x

Dadds, M. R., Hunter, K., Hawes, D. J., Frost, A. D. J., Vassallo, S., Bunn, P., Merz, S., & Masry, Y. E. (2008). A measure of cognitive and affective empathy in children using parent ratings. C*hild Psychiatry and Human Development, 39*(2), 111–122. https://doi.org/10.1007/s10578-007-0075-4

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10,* 85.

De Rosnay, M., Pons, F., Harris, P. L., & Morrell, J. M. (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*(2), 197-218. https://doi.org/10.1348/026151004323044573

Farrell Pagulayan K., Busch R. M., Medina K. L., Bartok J. A., Krikorian R. (2006). Developmental normative data for the Corsi Block-tapping task. J. Clin. Exp. Neuropsychol. 28 1043–1052. 10.1080/13803390500350977

Feinfield, K. A., Lee, P. P., Flavell, E. R., Green, F. L., & Flavell, J. H. (1999). Young Children's Understanding of Intention. *Cognitive Development, 14*(3), wic463-486. https://doi.org/10.1016/S0885-2014(99)00015-5

Freud, S. (1959). *Collected papers. (5 Vols.).* Basic Books.

Farrell Pagulayan, K., Busch, R. M., Medina, K. L., Bartok, J. A., & Krikorian, R. (2006). Developmental normative data for the Corsi Block-tapping task. Journal of clinical and experimental neuropsychology, 28(6), 1043-1052. https://doi.org/10.1080/13803390500350977

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y. W., Warren, F., Howard, S., ... & Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The reflective functioning questionnaire. PLoS One, 11(7), e0158678. https://doi.org/10.1371/journal.pone.0158678

Gray, K., Young, L., & Waytz, A. (2012). Mind Perception Is the Essence of Morality. *Psychological Inquiry, 23(2*), 101-124. https://doi.org/10.1080/1047840X.2012.651387

Gummerum, M., & Chu, M. T. (2014). Outcomes and intentions in children's, adolescents', and adults' second-and third-party punishment behavior. Cognition, 133(1), 97-103. https://doi.org/10.1016/j.cognition.2014.06.001

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology*, 47, 55-130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Hala, S., Hug, S., & Henderson, A. (2003). Executive function and false-belief understanding in preschool children: Two tasks are harder than one. *Journal of Cognition and Development, 4*(3), 275-298. https://doi.org/10.1207/S15327647JCD0403_03

Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition, 128(3),* 451-474. https://doi.org/10.1016/j.cognition.2013.04.004

Hoffman, M. (2000). Empathy and moral development. Implications for justice and caring. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511805851

Hoffman, M. L., & Saltzstein, H. D. (1967). Parent discipline and the child's moral development. *Journal of Personality and Social Psychology, 5*(1), 45 57. https://doi.org/10.1037/h0024189

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition, 119(2)*, 197-215. https://doi.org/10.1016/j.cognition.2011.01.006

Kneer, M., & Bourgeois-Gironde, S. (2017). Mens rea ascription, expertise and outcome effects: Professional judges surveyed. *Cognition, 169*, 139-146. https://doi.org/10.1016/j.cognition.2017.08.008

Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical psychology*, *16*(2), 309-324. https://doi.org/10.1080/09515080307771

Lagattuta, K., & Weller, D. (2014). Interrelations between theory of mind and morality. *Handbook of moral development*, 385-407. https://doi.org/10.4324/9780203581957

Lane, J., Wellman, H., Olson, S., LaBounty, J., & Kerr, D. (2010). Theory of mind and emotion understanding predict moral development in early childhood. *British Journal of Developmental Psychology, 28(4),* 871-889. https://doi.org/10.1348/026151009X483056

Leslie, A. M. (1988). Some implications of pretense for mechanisms underlying the child's theory of mind. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 19-46). New York, NY, US: Cambridge University Press.

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology, 50*(1), 45–85. https://doi.org/10.1016/j.cogpsych.2004.06.002

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting Intentionally and the Side-Effect Effect: Theory of Mind and Moral Judgment. *Psychological Science*, *17*(5), 421–427. https://doi.org/10.1111/j.1467-9280.2006.01722.x

Lickona, T. (1969). Piaget misunderstood: A critique of the criticisms of his theory of moral development. Merrill-Palmer Quarterly of Behavior and Development, 15(4), 337-350.

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology*, *30*(1), 1–13. https://doi.org/10.1111/j.2044-835X.2011.02074.x

Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and social psychology review*, *3*(1), 23-48. https://doi.org/10.1207/s15327957pspr0301_2

Malle, B.F. and Knobe, J. (1997a) The folk concept of intentionality. *Journal of Experimental Social Psychology.* 33, 101–121. https://doi.org/10.1006/jesp.1996.1314

Malti, T., Gasser, L., & Gutzwiller-Helfenfinger, E. (2010). Children's interpretive understanding, moral judgments, and emotion attributions: Relations to social behaviour. *British Journal of Developmental Psychology*, 28(2), 275-292. https://doi.org/10.1348/026151009X403838

Margoni, F., & Surian, L. (2017). Children's intention-based moral judgments of helping agents. *Cognitive Development, 41*, 46-64. https://doi.org/10.1016/j.cogdev.2016.12.001.

McNamara, R. A., Willard, A. K., Norenzayan, A., & Henrich, J. (2019). Weighing outcome vs. intent across societies: How cultural models of mind shape moral reasoning. *Cognition, 182*, 95–108. http://doi.org/10.1016/j.cognition.2018.09.008.

Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child development*, *72*(3), 688-690. https://doi.org/10.1111/1467-8624.00306

Moses, L., & Tahiroglu, D. (2010). Clarifying the relation between executive function and children's theories of mind. In B. Sokol, U. Muller, J. Carpendale, A. Young, & G. Iarocci (Eds.), *Self- and social-regulation: Exploring the relations between social interaction, social understanding, and the development of executive functions* (pp. 218–233). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195327694.003.0009.

Nelson-le Gall, S. A. (1985). Motive–outcome matching and outcome foreseeability: Effects on attribution of intentionality and moral judgments. *Developmental Psychology, 21*(2), 332-337. doi:10.1037/0012-1649.21.2.332

Nobes, G., Panagiotaki, G., & Bartholomew, K. J. (2016). The influence of intention, outcome and question-wording on children's and adults' moral judgments. *Cognition, 157,* 190-204. https://doi.org/10.1016/j.cognition.2016.08.019

Piaget, J. (1965/1932). The moral judgment of the child. New York: Free Press.

Ponitz, C. E. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. Early Childhood Research Quarterly, 23(2), 141-158. https://doi.org/10.1016/j.ecresq.2007.01.004

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(4), 515-526. https://doi.org/10.1017/S0140525X00076512

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review, 34*(3), 265–293. https://doi.org/10.1016/j.dr.2014.04.001

Sarin, A., Lagnado, D. A., & Burgess, P. W. (2017). The intention-outcome asymmetry effect. *Experimental Psychology. 64*(2), 124–141. https://doi.org/10.1027/1618-3169/a000359

Shultz, T.R., & Wells, D. (1985). Judging the intentionality of action-outcomes. *Developmental Psychology, 21,* 83-89. https://doi.org/10.1037/0012-1649.21.1.83

Smetana, J. G. (2006). Social domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–153). Mahwah, NJ: Erlbaum.

Smetana, J. G. (2013). *Moral development: The social domain theory view.* In P. D. Zelazo (Ed.), *Oxford library of psychology. The Oxford handbook of developmental psychology (Vol. 1): Body and mind* (p. 832–863). Oxford University Press.

Smetana, J., Jambon, M., Conry-Murray, C., & Sturge-Apple, M. (2012). Reciprocal associations between young children's developing moral judgments and theory of mind. *Developmental Psychology, 48(4),* 1144-1155. https://doi.org/10.1037/a0025891

Smetana, J. G., Rote, W. M., Jambon, M., Tasopoulos-Chan, M., Villalobos, M., & Comer, J. (2012). Developmental changes and individual differences in young children's moral judgments. *Child development, 83(*2), 683-696. https://doi.org/10.1111/j.1467-8624.2011.01714.x

Sodian, B., Licata, M., Kristen, A. S., Paulus, M., Killen, M., & Woodward, A. (2016). Understanding of Goals, Beliefs, and Desires Predicts Morally Relevant Theory of Mind: A Longitudinal Investigation. *Child Development, 87(4),* 1221–1232. https://doi.org/10.1111/cdev.12533

Tahiroglu, D., Moses, L. J., Carlson, S. M., Mahy, C. E. V., Olofson, E. L., & Sabbagh, M. A. (2014). The Children's Social Understanding Scale: Construction and validation of a parent-report measure for assessing individual differences in children's theories of mind. *Developmental Psychology, 50*(11), 2485-2497. http://dx.doi.org/10.1037/a0037914

Taylor, M., & Carlson, S. M. (1997). The relation between individual differences in fantasy and theory of mind. *Child development*, *68*(3), 436-455. https://doi.org/10.1111/j.1467-8624.1997.tb01950.x

Turiel, E. (2007). The development of morality. *Handbook of child psychology*, *3*.

Wainryb, C., & Ford, S. (1998). Young children's evaluations of acts based on beliefs different from their own. *Merrill-Palmer Quarterly,* 484-503.

Wellman, H. M. (2020). Reading Minds: How Childhood Teaches Us to Understand People. Oxford University Press, USA.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child development*, *75*(2), 523-541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*(3), 245-275. https://doi.org/10.1016/0010-0277(90)90024-E

Wickham, H. (2009) ggplot2: elegant graphics for data analysis. Springer New York.

Wimmer, H., Gruber, S., & Perner, J. (1984). Young children's conception of lying: Lexical realism—moral subjectivism. *Journal of Experimental Child Psychology*, *37*(1), 1-30. https://doi.org/10.1016/0022-0965(84)90055-9

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235-8240. https://doi.org/10.1073/pnas.0701408104

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental psychology*, *24*(3), 358. https://doi.org/10.1037/0012-1649.24.3.358

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, Act, and Outcome in Behavioral Prediction and Moral Judgment. *Child Development, 67(5),* 2478–2492. https://doi.org/10.2307/1131635