YOUNG CHILDREN’S SOCIAL AND MORAL EVALUATIONS OF THIRD-PARTY HELPERS AND HINDERERS

by

JULIA W. VAN DE VONDERVOORT

B.A., University of Waterloo, 2012
M.A.Sc., University of Waterloo, 2013

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

January 2020

© Julia W. Van de Vondervoort, 2020
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Young children’s social and moral evaluations of third-party helpers and hinderers

submitted by Julia W. Van de Vondervoort in partial fulfillment of the requirements for
the degree of Doctoral of Philosophy
in Psychology

Examining Committee:

Dr. Kiley Hamlin, Psychology
Supervisor

Dr. Andy Baron, Psychology
Supervisory Committee Member

Dr. Ara Norenzayan, Psychology
Supervisory Committee Member

Dr. Matt Bedke
University Examiner

Dr. Azim Shariff
University Examiner
Abstract

The judgment of others’ actions as good and praiseworthy versus bad and blameworthy is fundamental to humans’ sociomoral functioning. The ability to produce such moral judgments was traditionally characterized as a relatively lengthy process, requiring years of experience and cognitive maturation. However, more recent research has demonstrated that even infants are sensitive to morally relevant interactions amongst third parties. What remains unclear is whether infants’ implicit sociomoral evaluations align with young children’s explicit moral judgments.

This dissertation explores the maturity of preschoolers’ explicit moral judgments when presented with third-party social interactions similar to those used to demonstrate infants’ implicit moral sense. Chapter 1 provides relevant background information regarding the study of moral judgments across infancy and early childhood. Chapter 2 investigates whether infants’ preferences for those who help rather than hinder others’ unfulfilled goals align with preschoolers’ explicit social and moral judgments. Specifically, two experiments explore whether 3- to 5-year-olds selectively prefer helpers, judge helpers as “nicer” than hinderers, and selectively allocate punishment to hinderers across two different scenarios. Chapter 3 examines the extent to which preschoolers’ judgments are sensitive to individuals’ mental states: Three experiments investigate whether 3- and 4-year-olds’ social and moral judgments privilege others’ intentions to help versus hinder a third party or whether judgments are tied to the outcomes achieved. Chapter 4 explores preschoolers’ ability to consider the context in which helpful and unhelpful actions are performed. Specifically, two experiments investigate whether 3- and 4-year-olds differentially evaluate those who help versus hinder previously prosocial or antisocial others. Finally, Chapter 5 discusses the main findings of this dissertation and several open questions regarding the nature of young children’s moral judgments. Together, this dissertation
represents a significant advancement in the understanding of young children’s social and moral cognition. By adapting paradigms developed to study implicit sociomoral evaluations, these studies document important similarities and dissimilarities between infants’ implicit moral sense and young children’s explicit morality.
Lay Summary

Adults readily produce moral judgments; that is, we evaluate certain people and actions as right and worthy of praise and condemn others as wrong and deserving of punishment. Given the importance of moral judgments to our daily lives, it is perhaps unsurprising that decades of research has investigated the development of this moral sense throughout childhood. The present dissertation fills a gap in the existing literature by exploring 3- to 5-year-olds’ moral evaluations of helpers and hinderers. These studies demonstrate that preschoolers positively evaluate helpers over hinderers, that these judgments are sensitive to the characters’ intentions to help or harm (rather than just the outcomes achieved), and that preschoolers judge helping positively and judge hindering negatively across contexts. Together, this work represents a significant step forward in understanding the developmental trajectory of humans’ moral sense.
Preface

I am the primary author of the work presented in this dissertation, which was completed in collaboration with my supervisor, J. K. Hamlin. I designed the studies, collected the data, conducted data analyses, and drafted each manuscript. JKH contributed to study design, data interpretation, and provided critical feedback on each manuscript. The research presented in this dissertation was approved by the UBC Behavioural Ethics Board under certificate H10-01808.


Chapter 3: A version of this chapter has been published: Van de Vondervoort, J. W., & Hamlin, J. K. (2018). Preschoolers focus on others’ intentions when forming sociomoral judgments. Frontiers in Psychology, 9, 1851.

Chapter 4: A version of this chapter is being prepared for publication: Van de Vondervoort, J. W., & Hamlin, J. K. (in prep). Young children’s consideration of contextual information.
# Table of Contents

Abstract ................................................................................................................................. iii
Lay Summary ........................................................................................................................... v
Preface ..................................................................................................................................... vi
Table of Contents ................................................................................................................... vii
List of Tables .......................................................................................................................... ix
List of Figures ........................................................................................................................ x
Acknowledgments .................................................................................................................. xi

**Chapter 1: General Introduction** ......................................................................................... 1
  - Children’s Explicit Moral Judgments ................................................................................ 1
  - Children’s Implicit Moral Evaluations ............................................................................. 6
  - The Current Dissertation ............................................................................................... 11

**Chapter 2: Children’s Judgments of Third-Party Helpers and Hinderers** ......................... 15
  - Introduction .................................................................................................................... 15
  - Experiment 1 ................................................................................................................ 19
  - Experiment 2 ................................................................................................................ 31
  - General Discussion ....................................................................................................... 36

**Chapter 3: Children’s Consideration of Others’ Intentions** ............................................. 44
  - Introduction .................................................................................................................... 44
  - Experiment 1 ................................................................................................................ 52
  - Experiment 2A ............................................................................................................ 63
  - Experiment 2B ............................................................................................................ 67
  - Experiment 3 ................................................................................................................ 71
  - General Discussion ....................................................................................................... 75

**Chapter 4: Children’s Consideration of Contextual Information** ..................................... 84
  - Introduction .................................................................................................................... 84
  - Experiment 1 ................................................................................................................ 92
  - Experiment 2 ................................................................................................................ 102
  - General Discussion ....................................................................................................... 110

**Chapter 5: General Conclusion** ......................................................................................... 116
  - Summary of Main Findings .......................................................................................... 116
  - Limitations and Unanswered Questions ..................................................................... 121
References................................................................................................................................. 137
Appendix A..................................................................................................................................... 161
Appendix B ...................................................................................................................................... 164
List of Tables

Table 2.1 Summary of verbal explanations in Experiment 1 and 2.................................................. 43
Table 3.1 Descriptive and test statistics for t-tests in Experiments 1 – 3 ........................................ 82
Table 3.2 Summary of verbal explanations in Experiments 1 – 3..................................................... 83
Table 4.1 Summary of verbal explanations in Experiment 1............................................................... 114
Table 4.2 Summary of verbal explanations in Experiment 2............................................................... 115
Table A.1 Summary of individual response patterns in Chapter 2................................................... 162
Table A.2 Supplementary summary of verbal explanations in Chapter 2........................................... 163
Table B.1 Scores before and after comprehension questions in Chapter 3 .................................... 170
List of Figures

Figure 2.1 Scores in Experiment 1 .................................................................................................................. 26
Figure 2.2 Scores in Experiment 2 .................................................................................................................. 34
Figure 3.1 Visual depiction of puppet show events ......................................................................................... 53
Figure 3.2 Scores in Experiment 1 .................................................................................................................. 59
Figure 3.3 Scores in Experiment 2A and 2B .................................................................................................... 64
Figure 3.4 Scores in Experiment 3 .................................................................................................................. 73
Figure 4.1 Image of Likert scale used in Experiment 1 and 2 ........................................................................... 93
Figure 4.2 Forced-choice scores in Experiment 1 ........................................................................................... 98
Figure 4.3 Likert scale ratings in Experiment 1 ............................................................................................... 100
Figure 4.4 Images from the prosocial and antisocial tower events1 ................................................................. 104
Figure 4.5 Forced-choice scores in Experiment 2 ............................................................................................. 106
Figure 4.6 Likert scale ratings in Experiment 2 ............................................................................................... 109
Acknowledgments

First and foremost I would like to thank my advisor, Dr. Kiley Hamlin, for her mentorship over the past six years. Beyond her insightful contributions to the studies contained in this dissertation, her immense knowledge and dedication to the field has inspired me in numerous ways. Kiley’s academic guidance and personal support have been invaluable. I would also like to thank the members of my dissertation committee, Dr. Andy Baron and Dr. Ara Norenzayan, for their insightful comments and valuable feedback on this work. I am incredibly grateful to all members (past and present) of the Centre for Infant Cognition, especially the lab managers and research assistants who spent hours helping to collect, transcribe, and code the data contained here. I extend a special thanks to the daycares and preschools who allowed me to conduct this research, as well as the families and children who participated in these studies. I am grateful for the financial support that funded this research: A grant from the Social Sciences and Humanities Research Council (SSHRC) of Canada awarded to J. K. Hamlin and a SSHRC Joseph-Armand Bombardier CGS Scholarship awarded to J. W. Van de Vondervoort.

Lastly, I would like to thank my family and friends for their support and encouragement, especially Margaret Winter (for countless hours spent on the phone supporting me from across the country and her detailed proofreading of this document) and Bixby the dog (for keeping me company while writing this document). Finally, I would like to thank Rob Irwin for his love and support; there is no way I could have completed this dissertation without you by my side.
Chapter 1: General Introduction

Humans routinely judge some people and actions as good, right, and deserving of praise and others as bad, wrong, and deserving of punishment. While the exact content of these judgments (e.g., which specific individuals and actions are condemned) can vary across individuals, time, and place, all typically-developing humans make moral judgments (Brown, 1991). Given the ubiquity of moral judgments and their significance to everyday social and moral functioning, it is perhaps unsurprising that psychologists have long studied the developmental trajectory of this evaluative moral sense.

Children’s Explicit Moral Judgments

Early psychological research characterized the emergence of morality as a slow process – young children were thought to be entirely insensitive to moral concerns, either because they completely lack a moral sense, or because their moral sense is rendered useless by egocentrism or cognitive limitations. In this view, a true moral sense results from extensive maturation and experience, and it is not until somewhat later in development that children make judgments that reflect moral principles (e.g., Freud, 1930/1961; Kohlberg, 1981; 1984; Piaget, 1932/1965).

One especially influential proponent of this view was Jean Piaget, who studied children’s explicit moral judgments through both direct interview and indirect observation. Piaget argued that children actively construct their understanding of morality throughout childhood by consciously reflecting upon their experiences. According to Piaget, very young children lack any kind of moral sense. Between ages 4 and 7, children learn to follow moral rules, at first viewing these rules as unchallengeable constraints dictated by authority figures. At this stage of moral development, children interpret rules literally and fail to consider various additional factors (e.g., whether a moral rule was broken intentionally or accidentally). To illustrate, Piaget famously
observed that children under 6 years of age considered a child who accidently broke 15 cups to be naughtier than a child who intentionally broke 1 cup, revealing an outcome focus when assigning blame; by age 6 children begin focusing on intent. By age 10, children’s experience of conflict and negotiations with peers leads to a more mature understanding of the complex and authority-independent nature of morality. This increased maturity in moral reasoning is accompanied by an appreciation of moral principles, such as justice. Piaget believed that initially children’s conception of justice is based in egalitarian principles (i.e., all individuals are treated the same), but that children develop an appreciation of equity, in which individuals’ specific needs are considered, around 12 years of age (Piaget, 1932/1965).

Lawrence Kohlberg extended Piaget’s findings by exploring shifts in moral reasoning across childhood and into adulthood. Kohlberg sought to determine when children’s moral judgments excluded egocentric, conventional, and legal concerns, instead focusing on universal moral principles. He famously investigated the development of moral judgment via the Moral Judgment Interview, in which children were presented with moral dilemmas (e.g., a man must decide whether to steal medicine to save his dying wife) and asked to explain what the protagonist should do and why. Kohlberg was most interested in the explicit reasoning behind children’s decisions, and observed that the content of these responses progressed through reliable stages across development: Before age 9 children focus on self-interest and external consequences (e.g., the man should not steal the drug because he might go to jail). In contrast, adolescents’ responses typically focus on societal values (e.g., the man should not steal the drug because it is against the law), while young adults increasingly describe moral principles that transcend social norms (e.g., he should not steal the drug because others need that same medicine). Like Piaget, Kohlberg believed that these changes resulted from interactions with
others that inspired active reflection on the appropriate justifications for moral decisions. Because Kohlberg did not observe reasoning consistent with universal moral principles until later in development, he concluded that mature moral judgments do not emerge before at least early adulthood (1981, 1984; Colby et al., 1983).

Kohlberg’s contemporaries explored how to more accurately and objectively quantify one’s stage of moral reasoning with the development of the Defining Issues Test, in which participants respond to moral dilemmas by ordering statements in terms of their importance to resolving the dilemma (Rest, 1979) and the Sociomoral Reflection Measure, in which participants are asked to explain the importance of certain moral propositions (Gibbs, Basinger, & Fuller, 1992). More recently, the Socio-Moral Reasoning Aptitude Level task depicts moral dilemmas via first-person perspective images, and participants are asked to provide an oral justification for their response to the dilemma (e.g., whether they would tell the teacher after witnessing a classmate cheating on a test; Dooley, Beauchamp, & Anderson, 2010).

Across these explicit-reasoning measurement tools, researchers have found that moral reasoning matures from childhood to adulthood among typically-developing individuals (Armon & Dawson, 1997; Colby et al., 1983; Chiasson, Vera-Estay, Lalonde, Dooley, & Beauchamp, 2017; Eisenberg, Cumberland, Guthrie, Murphy, & Shepard, 2005; Eisenberg, 1979; Gibbs et al., 1992). Advancements in moral reasoning are attributed to cognitive developments (e.g., executive functioning, including improved inhibition, attentional control, cognitive flexibility; Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Korkman, Kemp, & Kirk, 2001; Vera-Estay, Beauchamp, & Dooley, 2014), social-cognitive developments (e.g., empathy, theory of mind; Baird & Astington, 2004; Hoffman, 2000; Killen, Mulvev, Richardson, Jampol, & Woodward, 2011) and opportunities to engage in social perspective-taking (e.g., Gibbs, 2014).
That said, the idea that moral reasoning develops in a consistent manner has been debated. For example, there has been disagreement regarding whether moral reasoning develops similarly in males and females (see Gilligan, 1982 for the argument that men’s judgments reflects concerns regarding justice, while women’s judgments reflect emotions, empathy, and care for others), although evidence for gender difference in moral reasoning is inconsistent (Jaffee & Hyde, 2000; Walker, 1984). Furthermore, some differences in the development of explicit moral reasoning have been observed cross-culturally (e.g., differences in the extent to which adults produce judgments that focus solely on moral principles and exclude all social-conventional concerns; Gibbs, Basinger, Grime, & Snarey, 2007; Nisan & Kohlberg, 1982; Snarey, 1985).

Expanding the Study of Explicit Moral Judgments

The abovementioned work has been instrumental to the scientific study of explicit moral reasoning. That said, Piaget and Kohlberg’s conclusions regarding the precise developmental trajectory of children’s reasoning have not always been supported by later work. To illustrate, while researchers continue to observe that the tendency to incorporate intention into one’s moral judgments increases with age (e.g., Baird & Astington, 2004; Cushman, Sheketoff, Wharton, & Carey, 2013; see Margoni & Surian, 2016 for review), studies have shown that children under age 6 do not always focus on outcomes when assigning blame. For instance, 3- to 6-year-olds treat causing harm to others as more wrong that causing harm to oneself, even when both events produce the same outcome (Tisak, 1993). Further, when Piaget’s intention versus outcome tasks are simplified and processing demands are reduced (e.g., scenarios are accompanied by pictures, rather than being strictly verbal), children demonstrate sensitivity to intentions in their explicit moral judgments by age three (e.g., Nelson, 1980; Nobes, Panagiotaki, & Pawson, 2009; Yuill &
Perner, 1988; see Hilton & Kuhlmeier, 2019 for a discussion of how task features may influence young children’s production of intention-based moral judgments).

More recent studies have also explored the extent to which Kohlberg underestimated the moral reasoning abilities of young children. These studies revealed that children distinguish moral concerns (rules regarding justice, welfare, and rights) from social-conventional concerns (rules regarding social order) far earlier than Kohlberg proposed. In fact, studies eliciting verbal responses and explanations regarding morally relevant scenarios have shown that even young children appreciate the normative aspect of moral principles.

More specifically, an extensive body of research by prominent Social Domain Theorists has revealed that by 3 years of age children judge moral transgressions (e.g., hitting another child) as more likely to be wrong across contexts than social-conventional transgressions (e.g., wearing pyjamas to school, not saying “please”). By age 4, children also report that moral transgressions are more serious and less contingent upon explicit rules compared to social-conventional transgressions, and that moral transgressions are wrong even if not witnessed or condemned by an authority figure. This is the case both when children evaluate hypothetical (Smetana, 1981; Smetana & Braeges, 1990) and real-life (Smetana, Schlagman, & Adams, 1993) moral and conventional transgressions, and has been observed cross-culturally – for example, Chinese preschoolers in Hong Kong distinguish between moral and social-conventional transgressions by age 4 (Yau & Smetana, 2003), as do Colombian children by age 6 (Ardila-Rey & Killen, 2001). Overall, this more recent work demonstrates that even young children understand moral principles when age-appropriate methodologies are employed (see Nucci & Gingo, 2010; Smetana, 2006; Smetana, Jambon, & Ball, 2014; Turiel, 2006 for review). That said, these studies do continue to find age-related improvements in moral reasoning across the
childhood years (Smetana et al., 2012), especially when children are asked to reason about complex, non-prototypical moral dilemmas (Jambon & Smetana, 2014). Notably, such age-related improvements in explicit moral reasoning may be, at least in part, attributed to co-occurring improvements in language abilities (see Ball, Smetana, & Sturge-Apple, 2017; Smetana & Braeges, 1990 for evidence that language ability is positively associated with more mature moral reasoning in 2- and 3-year-olds).

What developmental processes account for this early understanding of moral principles? Echoing Piaget and Kohlberg’s proposals that moral development is driven by children’s evaluations of key experiences, it has been suggested that children construct distinctions between moral and non-moral concerns through interactions with the social environment (Turiel, 1983). For example, naturalistic observations of 2- and 3-year-olds’ interactions with their mothers and peers revealed that mothers display different responses to moral transgressions (e.g., focus on the victim’s rights or welfare) and conventional transgressions (e.g., focus on social order and regulation; Smetana, 1989, see also Nucci & Turiel, 1978). Likewise, mothers report intervening on their 2-year-olds’ moral transgressions with more inductive reasoning, physical restraint, and angry emotional reactions, compared to when their child performed non-moral transgressions related to practical concerns or potential harm to child themselves (Dahl & Campos, 2013; Dahl, Sherlock, Campos, & Theunissen, 2014). Reflecting upon these distinct responses may allow young children to distinguish between what is moral and what is conventional.

**Children’s Implicit Moral Evaluations**

Young children’s ability to identify moral concerns and reason about the nature of moral rules revealed that a moral sense is certainly operational by the preschool years. That said, the study of moral understanding via demonstrations of explicit reasoning does place some
restrictions upon moral theorizing. Specifically, a focus on explicit reasoning restricts the age at which one could find evidence for the emergence of an evaluative moral sense (i.e., beyond ages at which children can provide explicit, verbal responses).

To explore the presence of a moral sense before children are capable of explicit reasoning, researchers have developed methodologies that capitalize on infants’ and young children’s spontaneous reactions to morally relevant situations. Such methodologies have revealed that 2-year-olds spontaneously protest when they are the victims of moral transgressions (e.g., when a puppet attempts to throw away an object belonging to the child) but not following similar actions that do not involve violations (e.g., when a puppet attempts to throw away its own object; Rossano, Rakoczy, & Tomasello, 2011). Furthermore, observations of 2-year-olds’ social interactions reveal that children are more responsive to moral transgressions than conventional transgressions (e.g., more statements indicating pain or loss of property, emotional reactions such as crying, and physical retributions such as grabbing toys; Smetana, 1984), and that toddlers are more likely to comply and less likely to protest when caregivers intervene on their moral rather than conventional transgressions (Dahl, 2016). These studies demonstrate that 2-year-olds distinguish between moral and non-moral concerns when they are personally involved, even before they can provide explicit verbal explanations regarding moral reasoning.

Such methodologies have also been used to explore whether children recognize the universal nature of moral concerns. These studies reveal that 3-year-olds spontaneously protest both when they and others are the victim of moral transgressions (e.g., when a puppet attempts to throw away an object belonging to the child, or belonging to a third-party; Rossano et al., 2011; Vaish, Missana, & Tomasello, 2011). Three-year-olds also protest moral transgressions committed by both ingroup and outgroup members, but only protest conventional transgressions
performed by members of the ingroup (Schmidt, Rakoczy, & Tomasello, 2012). Taken together, these studies show that young children’s spontaneous responses are sensitive to the generalizability of moral principles, even when moral violations involve third parties.

**Investigating Infants’ Implicit Moral Evaluations**

Before children can generate active responses to moral events, researchers can explore infants’ reactions to observed third-party social and moral events. In these studies, infants often watch puppet show scenarios featuring a protagonist who repeatedly attempts to achieve a goal: To climb a steep hill, to open the lid on a box and reach an attractive toy, or to retrieve a dropped ball. Following a series of unsuccessful attempts, the protagonist’s goal is facilitated by a “helper” (who bumps the protagonist up the hill, helps him open the box, or returns the ball), or blocked by a “hinderer” (who bumps the protagonist down the hill, closes the lid on the box, or runs away holding the ball; Hamlin & Wynn, 2011; Hamlin et al., 2007; see also Buon et al., 2014; Scola, Holvoet, Arciszewski, & Picard, 2015).

When subsequently presented with the helper and hinderer, 3-month-olds selectively look toward helpers (Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2010), older infants selectively reach for helpers (Hamlin, 2015; Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007; see also Chae & Song, 2018; Scola et al., 2015; but see Salvadori et al., 2015 and Cowell & Decety, 2015), while older toddlers selectively help previously helpful individuals (Dahl, Schuck, & Campos, 2013) and selectively give resources to helpers (Hamlin, Wynn, Bloom, & Mahajan, 2011). Relatedly, infants and toddlers selectively approach and accept toys from those who distribute resources fairly rather than unfairly amongst third parties (Burns & Sommerville, 2014; Geraci & Surian, 2011; Lucca, Pospisil, & Sommerville, 2018). In all scenarios, selective
attention/approach is interpreted as a positive evaluation of that individual based on their actions in the helping/hindering or fairness display.

To explore the possibility that infants’ preferences are driven by low-level physical features of the prosocial and antisocial scenarios (e.g., particular types of movement in helping events; see Scarf, Imuta, Colombo, & Hayne, 2012), studies have included conditions in which “prosocial/antisocial” actors direct their efforts toward non-social targets instead of agents. In these conditions, “helpers/hinderers” or “fair/unfair” distributors push an inanimate circle up/down a hill, roll a ball toward/take a ball away from a mechanical claw, open/close a box with a claw, or place resources in front of objects. In all cases, infants show no preferences for either actor (Geraci & Surian, 2011; Hamlin et al., 2007; 2010; Hamlin & Wynn, 2011; see also Buon et al., 2014), suggestive that infants’ preferences in social versions of the scenarios are not solely due to perceptual aspects of scenarios.

While the evidence described thus far clearly demonstrates the social nature of infants’ evaluations, these studies alone do not strictly entail that infants’ prosocial preferences are rooted in a moral sense – that is, that infants have an impartial sense that helping is good and/or hindering is bad. However, a growing literature reveals that these preferences are sensitive to several factors that influence adults’ moral judgments. Specifically, infants’ evaluations privilege actors’ intentions over the outcomes they bring about (e.g., positively evaluating those who attempt to help others; Hamlin, 2013a; see also Woo, Steckler, Le, & Hamlin, 2017), are sensitive to actors’ knowledge states (i.e., whether the puppet could have known they were helping or hindering; Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013), are sensitive to whether the protagonist deserves reward or punishment (Hamlin, 2014a; Hamlin et al., 2011), and can privilege moral over egocentric concerns (i.e., in some situations, infants accept fewer
treats to avoid interacting with an unhelpful puppet; Tasimi & Wynn, 2016). Overall, while it remains an open question whether infants’ prosocial preferences are truly moral, this work does demonstrate that infants make sophisticated sociomoral evaluations.

Given that infants are (presumably) incapable of explicit reasoning about moral principles, their prosocial preferences are likely rooted in evaluative intuitions regarding morally relevant actions (i.e., the sense that certain people and actions are better than others and/or that some are good whereas others are bad). One possibility is that these intuitions are rooted in automatic emotional reactions that arise without conscious reasoning (e.g., Haidt & Joseph, 2008; but see Mikhail, 2007 for an account of non-emotional moral intuitions).

Why might moral intuitions emerge early in development? One possibility is that moral intuitions result from socialization children receive regarding which behaviors are appropriate and/or children’s active construction of moral knowledge based on relevant experiences. Through the internalization of socialized norms (e.g., Kochanska & Askan, 2006) or the active construction of moral concerns as distinct from social concerns (e.g., Dahl, Waltzer, & Gross, 2018), children may come to feel that certain actions are good or bad even before they can explicitly justify these intuitions. This possibility suggests that moral intuitions emerge along with early social (and moral) experiences: For instance, from the beginning of the second year of life caregivers encourage and positively reinforce infants’ prosocial behaviors (Dahl, 2015) and are especially responsive to younger toddlers’ moral transgressions (Smetana, 1984).

A non-mutually-exclusive possibility is that moral intuitions may have evolved to support humans’ unique tendency to engage in large-scale cooperation. While cooperation with others can be mutually beneficial, cooperative individuals can also be exploited by those who benefit from others’ contributions without ever reciprocating. Humans’ widespread engagement in
prosocial acts (i.e., costly actions intended to benefit others) and the successful functioning of large-scale groups suggests the evolution of mechanisms that maintain cooperation and reduce exploitation. These evolved mechanisms may include tendencies to evaluate others’ cooperative and uncooperative behaviors, to positively evaluate and selectively cooperate with those likely to reciprocate, to negatively evaluate and avoid cooperation with defectors, and to actively discourage (e.g., punish) exploitative behaviors that do occur (Axelrod & Hamilton, 1981; Boyd & Richerson, 1992; Cosmides & Tooby, 1992; Henrich & Henrich, 2007; Trivers, 1971). These evaluative tendencies may take the form of moral intuitions, which may in turn guide infants’ implicit preferences for prosocial others\(^1\) (see Hamlin, 2013b for discussion of an innate moral core) and contribute to the development of explicit morality. Notably, evolved intuitions may be evident extremely early in development, even before children experience explicit socialization or have the opportunity to reflect on their own social and morally-relevant experiences.

**The Current Dissertation**

As reviewed above, a growing literature suggests that infants may possess an implicit or intuition-based moral sense, while an established literature demonstrates that young children have an explicit understanding of morality. This dissertation connects these two bodies of work by examining young children’s explicit reactions to third-party helping and hindering displays similar to those created for use with infant populations. This work investigates whether adapting these paradigms reveals the relative maturity of preschoolers’ explicit moral judgments, and

---

\(^{1}\) In addition to positively evaluating/approaching prosocial others and negatively evaluating/avoiding antisocial others, very young children also selectively direct prosocial actions toward prosocial interaction partners (see Kuhlmeier, Dunfield, & O’Neill, 2014 for review). Notably, non-human primates also engage in selective prosocial acts; of note, this selective prosociality is thought to be rooted in affective or emotional processes rather than cognitive processes (Brosnan & de Waal, 2002; see Schino & Aureli, 2010 for discussion).
whether there is continuity between the implicit knowledge demonstrated by infants in previous work and young children’s explicit morality.

To examine the maturity of preschoolers’ explicit morality, the current studies investigate 3- to 5-year-olds’ sensitivity to several established features of moral judgments. The first feature to be considered is the impartiality of moral judgments; that is, moral judgments are inspired by interactions between unknown third parties and cannot be solely based on whether an action immediately benefits or harms the evaluator. Chapter 2 investigates 3- to 5-year-olds’ social preferences and moral judgments of those who help versus hinder an unfamiliar protagonist that unsuccessfully attempts to achieve an object-directed goal. Notably, these third-party interactions feature puppet characters that participants have not interacted with in the past, and have no reason to expect to interact with in the future. It is predicted that preschoolers will prefer helpers over hinderers, judge helpers to be “nicer” than hinderers, and selectively allocate punishment to hinderers, in alignment with infants’ selective reaching for helpers over hinderers in these puppet scenarios (Hamlin & Wynn, 2011).

The second feature of moral judgments considered here is their sensitivity to the intentions of the individual being evaluated. This feature reflects that, when deciding whether an individual is good or bad, a mature moral sense considers whether the individual planned to help or harm, regardless of whether that goal was achieved. Chapter 3 examines whether 3- and 4-year-olds’ social and moral judgments privilege helpers’ and hinderers’ intentions, even when these intentions conflict with the eventual outcome of their actions. Just as infants have been found to privilege intentions over outcomes (Hamlin, 2013a), it is predicted that preschoolers will positively evaluate those who intend to help a puppet protagonist, even if attempted helpers ultimately fail to bring about positive outcomes for the protagonist.
The final feature of moral judgments to be considered is their dependence upon the context in which actions are performed. This feature reflects that a mature moral sense does not follow simple behavioral rules, such as “facilitating goals is good and blocking goals is bad” or “goods should always be distributed equally.” Rather, there are situations in which prototypically positive actions like helping are negatively evaluated (e.g., when helping someone with a nefarious goal) and situations in which prototypically negative actions may be positively evaluated (e.g., when punishing deserving others). Chapter 4 explores whether 3- and 4-year-olds positively evaluate those who selectively help previously prosocial others and hinder previously antisocial others. It is predicted that preschoolers’ social and moral judgments will be sensitive to the context in which these actions occur, just as infants prefer those who “reward” and “punish” appropriately (Hamlin, 2014a; Hamlin et al., 2011). Lastly, Chapter 5 will discuss the main findings of this dissertation and elaborate on several remaining open questions. These open questions include distinguishing between social versus moral responding, the role of implicit versus explicit cognition in the production of moral judgments, and the role of experience, maturation, and culture on the development of moral judgments.

**Demographics of Study Participants**

All studies within this dissertation were conducted in Vancouver, British Columbia, Canada. The majority of child participants were recruited and tested in their community-based daycare centres (81%), with the remainder recruited through the Early Development Research Group at the University of British Columbia and tested in a campus research laboratory. Formal demographics were not collected at daycares. That said, English-language comprehension was assessed in each study via a warm-up task (Chapter 2 and 3) or scale familiarization task (Chapter 4). Further, children appeared to come from middle-class families representative of the
ethnic demographics of Vancouver. According to the 2016 Canadian Census (Statistics Canada, 2017), the median income of two parent families with children in Vancouver was $111,636 and $52,242 for single-parent families. Within Vancouver 73% of individuals between the ages of 25 and 64 had completed a post-secondary program (5% apprenticeship/trades certificates, 21% college/university diplomas, 30% undergraduate degrees, and 17% postgraduate degrees). The majority of Vancouver-based census respondents indicated that English was the language spoken most often in their home (72%), with the most common additional known languages being Cantonese (16%), Mandarin (11%), and French (10%). Finally, in terms of ethnic origins, most Vancouverites indicated their background was European (48%; most often British Isles and Western European origins) or Asian (50%; most often East and Southeast Asian origins).

As for children who participated on campus, the majority of their families provided demographic information (90%). Of those who completed the demographic questionnaire, 90% of parents had completed a post-secondary program (3.5% trade school diploma, 8% college certificates/diplomas, 15% professional degrees, 41% undergraduate degrees, 20.5% postgraduate degree, and 2% “other”). All children who participated on-campus heard at least 50% English on a daily basis (with the average being 90% English); the most frequent additional languages heard on a daily basis were French (20% of children), Cantonese (13% of children), and Spanish (10% of children). Finally, in terms of ethnic origins, most children were identified by their parents as European (44%), East and Southeast Asian (12%), or a combination of the two (19%). Chapter 5 of this dissertation will discuss the generalizability of the current studies and whether these findings are expected to hold beyond the population studied.
Chapter 2: Children’s Judgments of Third-Party Helpers and Hinderers

Introduction

As discussed in Chapter 1, human beings are remarkably judgmental. One example of this is our readiness to evaluate individuals based on their prosocial and antisocial actions: We condemn those that willfully harm and see them as deserving of punishment, and we celebrate those that selflessly help and see them as deserving of praise. Notably, humans make these moral judgments even in third-party situations, when the prosocial and antisocial actions being evaluated are of no consequence to the evaluator, and instead only effect unfamiliar others.

Even young children make moral judgments, suggestive that a moral sense exists (although is immature) in early childhood. By age 3-4 children distinguish moral versus non-moral concerns. These distinctions are evident in their explicit moral judgments (e.g., regarding the severity and generalizability of certain transgressions) and in their verbal justifications of these judgments: Whereas moral justifications tend to focus on issues of welfare and fairness, conventional justifications focus on concerns about authority, social order, and rules (Nucci & Weber 1995; see Smetana, 2006; Smetana, Jambon, & Ball, 2014 for reviews). Finally, by age 4 children treat moral versus immoral others differently (i.e., distributing more treats to prosocial versus antisocial others) and explicitly justify non-egalitarian treatment in terms of the individuals’ morally relevant behaviors (Kenward & Dahl, 2011).

Recent research (reviewed in Chapter 1) suggests that this moral sense may even be operational before children are capable of producing explicit moral judgments. By just 3 months of age, infants look longer at individuals who help a protagonist achieve its goal compared to individuals who thwart the protagonist’s efforts, and once infants can make visually guided reaches, they selectively reach for helpers over hinderers (see Hamlin, 2013b; Hamlin & Van de...
Vondervoort, 2018; Van de Vondervoort & Hamlin, 2018a for review). Critically, these prosocial preferences appear to reflect social, rather than merely perceptual, concerns: Infants show no preference for ‘helpers’ who direct similar physical behaviors towards inanimate objects (Hamlin & Wynn, 2011; Hamlin et al., 2007; 2010), or towards an animate character without a clear unfulfilled goal (Hamlin, 2015; see also Scarf, Imuta, Colombo, & Hayne, 2012).

These studies suggest that infants’ preferences reflect social concerns. However, as previously discussed in Chapter 1, the exact nature of these social preferences remains an open question. Indeed, infants’ prosocial preference may reflect non-moral social concerns, including analyses of who might belong to the infant’s social group, or who might benefit the infant in the future (e.g., Baillargeon et al., 2015; see also Tafreshi, Thompson, & Racine, 2014; but see responses by Hamlin, 2014b; Dahl, 2014). Potentially consistent with this social-but-not-moral interpretation, infants’ preferences for helpers versus hinderers are influenced by the protagonist’s similarity to infants themselves (Hamlin, Mahajan, Liberman, & Wynn, 2013): Infants prefer those who help similar others, but prefer those who hinder dissimilar others. Relatedly, from 4.5 months of age infants prefer those who help previously prosocial others, but prefer those who harm previously antisocial others (Hamlin, 2014a; Hamlin et al., 2011). Although these judgments could reflect a sense of who deserves reward and punishment (presumably a moral concern), they could instead reflect affiliative judgments, whereby infants think “the enemy of my enemy is my friend” (e.g., Heider, 1958).

On the other hand, the preferences described above may be based in moral concerns. Indeed, infants may select hinderers of hinderers because they feel that antisocial others should be punished (see also Meristo & Surian, 2013). Furthermore, certain conceptions of morality highlight the role of group-level concerns (e.g., Schweder, Much, Mahapatra, & Park, 1997;
Haidt & Joseph, 2008), and so the discovery that infants’ preferences are sensitive to target similarity does not necessarily indicate they are not of a moral nature. In addition, other studies have shown that infants’ evaluations are sensitive to factors that influence adults’ moral judgments; for example, whether an individual tried to help or to hinder even if they did not succeed (Hamlin, 2013a; see also Lee, Yun, Kim, & Song, 2015; Vaish et al., 2010), and whether an individual knew they were helping or hindering when they acted (Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013; see also Choi & Luo, 2015; Meristo & Surian, 2013; Sloane, Baillargeon, & Premack, 2012). Finally, in a recent study that directly pit self-interest against moral value, infants incurred costs (i.e., accepted fewer graham cracker treats) to avoid interacting with a hinderer, suggestive that infants can be motivated by more than simple self-interest (Tasimi & Wynn, 2016).

In summary, the body of work described above demonstrates that infants’ preferences in helping and hindering puppet scenarios are influenced by some of the same factors that influence adults’ social and moral judgments. Unfortunately, because preverbal infants can neither respond selectively following different prompts nor explain their responses, it is difficult to determine whether infants’ responses to helpers and hinderers are based on social or moral concerns. Conversely, it is possible to distinguish between social and moral responding in older children, who are more linguistically advanced than infants. Specifically, one can ask who children prefer (a positive social judgment), who they think is nice (a positive moral evaluation), and who they think should be punished (allocating a negative moral consequence). There is limited work exploring young children’s social and moral responses to third-party helping and hindering scenarios designed for infants (but see Kenward & Dahl, 2011 for preschoolers’ judgments following a show not previously tested on infants, and see Buon et al., 2014 for a comparison
between infants’ and toddlers’ responses to a prosocial and violent antisocial other); the current studies aimed to fill this gap.

**The Current Studies**

The current studies explored children’s explicit reasoning about characters involved in live helping and hindering puppet shows. Three- to 5-year-olds were presented with box and ball helping and hindering puppet shows originally used in Hamlin & Wynn (2011). These shows were chosen due to their use in many publications to date (e.g., Hamlin, 2013a, 2014a, Hamlin et. al., 2011; 2013; Salvadori et al., 2015; Steckler, Woo, & Hamlin, 2017; see also Scola et al., 2015), and because their use of simple props made them ideal for transporting to daycares, where most participants were tested. After being presented with the puppet scenarios, children were asked (1) who they “like”, (2) who is “nicer”, and (3) who “should get in trouble.” Following children’s identification of who should get in trouble, they were asked to explain this judgment. These questions were adapted from previous work in the moral development literature (e.g., Baird & Astington, 2004; Costanzo, Coie, Grumet, & Farnill, 1973; Cushman, Sheketoff, Wharton, & Carey, 2013; Leslie, Mallon, & DiCorcia, 2006; Zelazo, Helwig, & Lau, 1996), and were chosen because they reflect similar methods to those utilized with infants (e.g., asking “who do you like?” while presenting the helpful and unhelpful puppets, see Hamlin, 2013b).

It was predicted that at all ages, children’s social preferences and positive moral evaluations would favor helpers over hinderers, but that children would select hinderers over helpers as recipients of punishment. That said, there are two alternative possibilities: (1) That children would interpret the puppet shows as social-but-not-moral, and so would reliably prefer helpers over hinderers but not identify helpers as nicer or hinderers as deserving of punishment, and (2) That children would readily identify the moral value of helpers and hinderers, but that
their social judgments would reflect more idiosyncratic responding, resulting in chance-level responding on the liking question. Based on previous work showing that 3-year-olds can have difficulty producing interpretable responses to open-ended questions (e.g., Kenward & Dahl, 2011; Killen, Breton, Ferguson, & Handler, 1994), it was assumed that older children would provide more coherent verbal justifications than 3-year-olds. In addition, it was predicted that older children would be more likely than 3-year-olds to reference relevant social and/or moral considerations as the reason for their punishment allocations.

**Experiment 1**

**Method**

**Participants.** Twenty-five 3-year-olds ($M = 3;6$, range = 3;0-3;11, 13 girls), 24 four-year-olds ($M = 4;5$, range = 4;0-4;11, 11 girls), and 25 five-year-olds ($M = 5;4$, range = 5;0-5;9, 13 girls) participated. Before data collection began a pre-set stopping rule was established of 24 children per age; one extra 3-year-old and one extra 5-year-old were run due to scheduling issues. An additional 6 three-year-olds, 4 four-year-olds, and 2 five-year-olds were seen, but replaced due to unwillingness to watch the puppet show and/or to point to a puppet in response to questions (5), interference from parents or teachers (2), procedure errors (4), and failure to complete a warm-up task that required English comprehension (1).

**Procedure.** Children first completed a warm-up task with an image of a playground. Children were asked to point to a swing and slide, to verbally identify the color of a toy, and to identify their favorite outside activity. As noted above, one child was removed from the sample due to an inability to locate the swing or slide; children were retained in the sample if they pointed to these items (that is, verbal responses were not required).
Children then watched live puppet shows featuring a protagonist struggling to achieve his goal to open a box or retrieve a dropped ball in the presence of two other characters. All shows were identical to those used in previous infant studies (e.g., Hamlin & Wynn, 2011) except for two factors. First, in infant studies puppet shows are performed at one end of a long table, with a curtain lowering to occlude the “stage” between events; puppeteers are kept out of sight behind an additional curtain at the back of the stage. In contrast, in the current study shows were enacted either on the floor or a table directly in front of the child and with the puppeteer fully visible, as this was required to test in daycares. Second, a few non-valenced words were added to the shows for narration, as pilot testing revealed that this helped children remain attentive to the events. Each child watched both the box and ball scenarios in counterbalanced order.

**Box scenario.** Three characters performed the box puppet show. The experimenter enacted the protagonist saying “Hello!”, walking up to a clear box containing a purple whale toy, looking through the side of the box while saying “Look, a toy!”, and then unsuccessfully attempting to open the box five times. On the third attempt, the protagonist said, “Too heavy!”. On the fifth attempt, the helper assisted in protagonist in opening the box while saying “Open!” or the hinderer prevented the protagonist from opening the box by slamming the lid shut while saying “Close!” All narrations in both scenarios were spoken in a high-pitched, positive voice to indicate that it was the puppet speaking rather than the experimenter; speech was not modulated based on the valence of the behavior occurring. Children were shown two helping events and two hindering events in alternation for a total of four events.

**Ball scenario.** A different set of 3 characters performed the ball puppet show. The experimenter enacted the protagonist saying “Hello!”, walking up to a ball and saying “Look, a ball!”, before bouncing the ball three times. Following the third bounce, the protagonist dropped
the ball to one side while saying “Whoops!” The ball was either picked up and returned to the protagonist by the helper (who said, “Here!” while rolling the ball to the protagonist) or taken away by the hinderer (who said, “Goodbye!” while running away with the ball). Children saw two helping events and two hindering events in alternation for a total of four events.

**Social preference and moral judgments.** After watching each of the above scenarios, children were presented with the helper and hinderer and asked (in counterbalanced order) which puppet he or she preferred (i.e., “Which one of these guys do you like the most?”) and which puppet was nicer (i.e., “Which one of these guys was nicer?”). After children indicated a positive moral evaluation, they were asked whether the selected puppet was a “little bit nice or a lot nice” (order counterbalanced; we planned to examine niceness judgments on a 3-point scale from “not nice” to “a lot nice”, but because most children at each age responded that the selected puppet was “a lot” nice regardless of whether they selected the helper or hinderer, this question is not considered further). To reduce the likelihood that children would select the same puppet across the liking and nicer questions due to reaching perseveration, children were asked to point to each puppet in between these questions (e.g., “Point to the guy with a red shirt. Right! Point to the guy with the green shirt. Right!”). All children correctly identified the puppets’ shirt color; note that no reinforcement was given following responses to the liking and nicer questions). Children were then asked to allocate punishment to one of the puppets (i.e., “I think that one of these guys should get in trouble. Who should get in trouble?”) and then to justify this choice (i.e., “Why should he get in trouble?”). If the child did not provide a verbal response to the last question, they were prompted (e.g., “What do you think?”). Children were only asked to provide one verbal response to prevent interference between explanations about the same puppet show.
Two sets of three puppets were used to complete the two scenarios: a duck (protagonist) and two rabbits wearing red and green shirts (helper, hinderer; identity counterbalanced), and a wolf (protagonist) and two moose wearing blue and yellow shirts (helper, hinderer; identity counterbalanced). The duck and rabbits were utilized in the first round, and the wolf and moose were utilized in the second round. Because the order of scenarios was counterbalanced across children, which set of puppets enacted which scenario (box, ball) was also counterbalanced. Additional counterbalanced variables were the order of the events within each show (helping first, hindering first) and the side of the helper and hinderer puppets (right, left). For the question period the helper and hinderer remained on the same side as during the show.

**Transcription and coding procedure.** When permitted by the child’s parents and feasible within the daycare setting, participation in the study was audio and visually recorded. A research assistant transcribed children’s explanations regarding why the selected puppet should get in trouble from these recordings. When recording was not permitted (11 of 74 children), children’s explanations were transcribed during the study by the experimenter. Two additional independent research assistants who had not assisted with data collection or transcription then coded the content of children’s explanations according to the categories detailed below.

Based on an initial review of children’s explanations by the authors, it was determined that children’s responses were either informative, meaning the child referred to something about the puppet or puppet show, or uninformative, meaning they did not.

**Uninformative responses.** Uninformative responses were highly variable, and included cases in which the child did not provide a verbal response, provided an unintelligible verbal response, or provided a verbal response that did not include a justification for the child’s selection. These verbal responses included statements entirely unrelated to the characters or
events in the puppet show (e.g., “I don’t like skating”), repetitions of the question (e.g., “because he should get in trouble”), statements not providing a reason (e.g., “because”), and statements implying that the child was unsure (e.g., “I don’t know”).

**Informative responses.** Informative responses were those that related to the characters from the show and/or an action performed, and took one of the following forms:

*Relevant action.* The child appealed to the performance of the hindering action from the box or ball scenarios (e.g., “because he closed the box”, “because he didn’t give the ball back”)2.

*Irrelevant action.* The child appealed to the performance of a negative action that had not occurred in the box or ball scenarios (e.g., “because he hit somebody”). These responses might indicate that, when asked why a puppet should get in trouble, some children simply generated actions that generally lead someone to punishment.

*Relevant valence.* The child appealed to the positive or negative valence of the puppet or its actions; these statements were directly related to the puppet shows (e.g., “because he is mean [pointing to selected puppet]”, “because this guy is nicer” [pointing to unselected puppet]).

*Irrelevant valence.* The child appealed to the positive or negative valence of the puppet or its actions; these statements were not related to the puppet shows (e.g., “because he’s happy”).

*Non-social considerations.* The child did not reference positive or negative actions or valence. Responses in this category included physical descriptions of the puppet (e.g., “because he has long ears”), general disliking of the selected puppet (e.g., “he’s not the most favorite thing”), references to the previous puppet show (e.g., “because he closed it” following the ball show), and ambiguous statements about the puppets or shows (e.g., “it wasn’t saying anything”).

---

2 One 5-year-old in Experiment 1 pointed to the helper and appealed to the helping action from the box scenario when explaining why he should get in trouble (i.e., “because he opened it”).
Each child’s explanation was coded by two independent research assistants for the presence or absence of each response type; coders were blind to the referent (helper, hinderer) of the explanation but not whether the child was responding after the box or ball show. Informative response types were not mutually exclusive. Reliability across the six categories was strong (average Cohen’s kappa = .864, range = .701 to 1.000; see McHugh, 2012)\(^3\). Disagreements (which involved 14 of 148 responses) were then reviewed by the first author and the two coders. Eleven of these disagreements were immediately recognized as mistakes by one of the coders; typically, due to failure to consider whether the child was responding to a box or ball show (for example, “because he closed it” following the ball show). The other three statements were resolved by discussion among the two coders, in the presence of the first author.

Results

Social preference and moral judgments. Confirmatory analyses. For each test question (liking, niceness, trouble), children received a score of 1 if they responded in the direction of the hypothesis (i.e., liked the helper, judged the helper as nicer, assigned punishment to the hinderer) and 0 if they responded against the direction of the hypothesis. Children’s scores were summed across the two rounds resulting in three scores between 0 – 2 per child, one for each question type (see Figure 2.1). We predicted that, at all ages, children’s social preferences and positive moral evaluations would favor helpers over hinderers, but that children would select hinderers over helpers as recipients of punishment. To determine at which ages children did respond as predicted, a series of one-sample t-tests were conducted comparing children’s liking score, niceness score, and trouble score at each age to a chance score of 1.

---

\(^3\) Explanation coding was originally conducted by two different research assistants and resulted in substantially lower reliability (average Cohen’s kappa = .791, range = .340-1.000). The coding scheme was then clarified, a more extensive training procedure was created, and the data was re-coded by two different research assistants.
Contrary to our predictions, 3-year-olds did not reliably distinguish between the helper and hinderer when providing social or moral judgments. Three-year-olds’ liking score ($M = 1.08, SD = .64$) and niceness score ($M = 1.20, SD = .71$) did not differ from chance (one-sample t-tests comparing each score between $0 – 2$ to a chance score of $1$; liking: $t[24] = .625, p = .538, d = .125$; niceness: $t[24] = 1.414, p = .170, d = .283$). On the other hand, 3-year-olds reliably identified the hinderer as the appropriate recipient of punishment ($M = 1.32, SD = .75$; one-sample t-test, $t[24] = 2.138, p = .043, d = .428$). In contrast to 3-year-olds, both 4- and 5-year-olds selectively preferred the helper, judged the helper to be nicer, and allocated punishment to the hinderer. Four-year-olds’ liking score ($M = 1.42, SD = .83$), niceness score ($M = 1.54, SD = .72$), and trouble score ($M = 1.58, SD = .72$) all exceeded chance in the direction of the hypothesis (one-sample t-tests; liking: $t[23] = 2.460, p = .022, d = .502$; niceness: $t[23] = 3.680, p = .001, d = .751$; trouble: $t[23] = 3.984, p = .001, d = .813$). Likewise, 5-year-olds’ liking score ($M = 1.64, SD = .70$), niceness score ($M = 1.88, SD = .33$), and trouble score ($M = 1.88, SD = .33$) all exceeded chance in the predicted direction (one-sample t-tests; liking: $t[24] = 4.571, p < .001, d = .914$; niceness $t[24] = 13.266, p < .001, d = 2.653$; trouble: $t[24] = 13.266, p < .001, d = 2.653$). The dataset analyzed for all experiments in Chapter 2 is available on the Open Science Framework (osf.io/unm4a). See Table A.1 in Appendix A for individual patterns of responding within each round.
Figure 2.1. Mean liking, niceness, and trouble scores following the box and ball scenarios at each age in Experiment 1; *p < .05, **p < .01, ***p < .001, error bars reflect the SE.

*Exploratory analyses.* While differences in performance across age were not predicted, the analyses reported above reveal that whereas older children reliably formed both social and moral evaluations favoring the helper, 3-year-olds only reliably distributed punishment to the hinderer. To examine how age and/or question type influenced children’s tendency to respond in the direction of the hypothesis, we conducted a mixed-effect ANOVA with liking, niceness, and trouble scores as within-subjects variables (repeated-measure), and child’s age (3, 4, 5) and child’s gender (female, male) as between-subjects factors. This analysis revealed a main effect of question type ($F[1.860,126.478] = 5.170, p = .008, \eta_p^2 = .071$), a main effect of age ($F[2,68] = 7.309, p = .001, \eta_p^2 = .177$). There was no interaction between question type and age ($F[3.720,126.478] = .220, p = .916, \eta_p^2 = .006$), and no main effect or interactions involving gender (all $Fs < 2.420$, all $ps > .124$).

---

4 Mauchly’s Test of Sphericity indicated that the assumption of sphericity had been violated ($\chi^2 = 13.370, p = .001$). Since the estimated epsilon was greater than .750 (Greenhouse-Geisser: $\epsilon = .847$, Huynd-Feldt: $\epsilon = .930$), we applied the Huynd-Feldt correction (see Girden, 1992).
To probe the main effect of question type, a series of paired-samples t-tests using the Bonferroni corrected alpha value of .017 (.05/3) were used to compare scores on each question type across age. These tests revealed that children were more likely to respond in the direction of the hypothesis when asked which puppet should get in trouble (trouble scores: \( M = 1.60, SD = .66 \)) compared to which puppet they prefer (liking scores: \( M = 1.38, SD = .75 \); paired-samples t-test, \( t[73] = 2.704, p = .009, d = .314 \)). There was no difference between liking scores and niceness scores (\( M = 1.54, SD = .67 \); paired-samples t-test, \( t[73] = 2.250, p = .027, d = .262 \)) or between niceness scores and trouble scores (paired samples t-test, \( t[73] = 1.000, p = .321, d = .116 \)). To explore the main effect of age, a series of independent-samples t-tests using the Bonferroni corrected alpha value of .017 (.05/3) were used to compare overall scores (between 0 – 6) across the three question types at each age. These tests revealed that 5-year-olds (\( M = 5.40, SD = 1.26 \)) were more likely to respond in the direction of the hypothesis across test questions compared to 3-year-olds (\( M = 3.60, SD = 1.68 \); independent-samples t-test, \( t[48] = 4.282, p < .001, d = 1.236 \)). There was no difference between 3-year-olds’ overall scores and those of 4-year-olds (\( M = 4.54, SD = 2.00 \); independent-samples t-test, \( t[47] = 1.786, p = .081, d = .5121 \)) nor between 4-year-olds’ scores and 5-year-olds’ scores (independent-samples t-test, \( t[38.482] = 1.790, p = .081, d = .577 \)).

**Explanations regarding punishment. Confirmatory analyses.** To ensure that each child contributed equally to the dataset and avoid over-representing talkative children (who may have provided several types of informative responses), instances of each explanation type (uninformative, relevant action, irrelevant action, relevant valence, irrelevant valence, non-social considerations) are represented as proportions and averaged across the two rounds (see Table 2.1). We predicted that younger children would provide less interpretable explanations than older
children. To test this prediction, we used a factorial ANOVA to examine the effect of child’s age (3, 4, 5) and child’s gender (female, male) on the proportion of uninformative responses across rounds. This test revealed that the proportion of uninformative responses differed among the ages tested ($F[2,68] = 10.182, p < .001, \eta^2_p = .230$); there was no main effect of gender or interaction between age and gender (both $Fs < 1.363, all ps > .262$). To determine the nature of this difference, a series of independent samples t-tests compared the proportion of uninformative responses across ages. These analysis revealed that 3-year-olds ($M = .52, SD = .51$) were more likely to provide uninformative responses than were 4-year-olds ($M = .17, SD = .38$; independent samples t-test, $t[44.357] = 2.756, p = .008, d = .828$) or 5-year-olds ($M = .04, SD = .20$; independent samples t-test, $t[31.214] = 4.382, p < .001, d = 1.569$). This is consistent with previous work showing that it can be difficult to elicit coherent justifications from 3-year-olds (Kenward & Dahl, 2011; Killen et al., 1994). Four- and 5-year-olds were equally (un)likely to provide uninformative responses when explaining their allocations of punishment (independent samples t-test, $t[34.483] = 1.449, p = .156, d = .494$).

We also predicted that older children would provide more relevant references to the puppet’s actions or valence than younger children. To test this prediction, we combined the appeals to relevant actions and relevant valence into a single proportion of relevant responses. A factorial ANOVA examining the effect of the children’s age (3, 4, 5) and children’s gender (female, male) on the proportion of relevant responses revealed that the proportion of relevant responses differed among the ages tested ($F[2,68] = 11.075, p < .001, \eta^2_p = .246$); there was no main effect of gender or interaction between age and gender (both $Fs < 0.752, all ps > .475$). To determine the nature of this difference, a series of independent samples t-tests compared the proportion of relevant responses across ages. These analysis revealed that 3-year-olds ($M = .28,$
were less likely to appeal to relevant actions and valence than were 4-year-olds ($M = .73$, $SD = .44$; independent samples t-test, $t[47] = 3.491, p = .001, d = 1.018$) or 5-year-olds ($M = .82$, $SD = .38$; independent samples t-test, $t[48] = 4.542, p < .001, d = 1.311$). Four- and 5-year-olds were equally likely to reference relevant actions or valence when explaining their allocations of punishment (independent samples t-test, $t[47] = .774, p = .443, d = .226$). Older children were especially likely to reference the puppet’s hindering actions: 75% of 4-year-olds’ and 77% of 5-year-olds’ informative responses appealed the selected puppet’s unhelpful act as the reason he should get in trouble, compared to 50% of 3-year-olds’ informative responses. See Table A.2 in Appendix A for proportions of each response type following allocations of punishment to the hinderer, rather than the helper.

Discussion

Results from Experiment 1 suggest that 4- and 5-year-olds reliably form social preferences and moral judgments favoring helpers over hinderers in two scenarios previously used to evaluate infants’ sociomoral evaluations. At both ages children preferred the helper, judged the helper to be nicer, and allocated punishment to the hinderer. When asked to justify why the selected puppet should get in trouble, 4- and 5-year-olds most often referred to the puppet’s unhelpful action during the show. In contrast, 3-year-olds did not reliably distinguish between the helper and hinderer when asked who they prefer or who was nicer. Three-year-olds reliably identified hinderers as deserving of punishment, but approximately 50% of them did not provide informative explanations for this judgment.

Given that past research has demonstrated that preverbal infants reliably distinguish between helpers and hinderers in a forced choice paradigm (see Hamlin, 2013b) it is somewhat surprising that 3-year-olds failed to distinguish helpers from hinderers in 2 of our 3 question
types. That said, in the only other study of preschoolers’ explicit reactions to similar third-party helping and harming scenarios to date, Kenward and Dahl (2011) also observed that 3-year-olds failed to distinguish between helpers and harmers in a variety of explicit measures (but see Buon et al., 2014 for evidence that 29-month-olds can distinguish between prosocial and antisocial actors when responses to a range of explicit questions are considered together). Together, these results appear to indicate that the ability to successfully provide an implicit evaluation of helpers and hinderers in infancy is not sufficient for the provision of explicit judgments in early childhood. This pattern of performance has been observed with other types of social judgments, wherein infants show competence in implicit, non-verbal paradigms but young children initially fail to show competence in explicit verbal ones (e.g., tasks evaluating infants’ and children’s false-belief understanding; Baillargeon, Setoh, Sloane, Jin, & Bian, 2014; Leslie, German, & Polizzi, 2005; Wellman, Cross, & Watson, 2001).

On the other hand, it is possible that 3-year-olds’ inconsistent performance in Experiment 1 was not due to an inability to provide elicited social and moral judgments, but rather an inability to do so in our particular paradigm. For instance, 3-year-olds in Experiment 1 may have found it difficult to pay attention to the puppet shows, to process alternating events, or to remember the action performed by each puppet; indeed, Kenward and Dahl (2011) attribute 3-year-olds’ inability to distinguish between helpers and harmers when making explicit judgments to confusion regarding who did what to whom in their helping and harming scenarios. To explore this possibility, we ran an additional Experiment with 3-year-olds, in which children observed the very same scenarios as in Experiment 1, but with modifications designed to facilitate children’s comprehension and memory.
First, whereas in Experiment 1 helping and hindering scenarios were presented in alternation, in Experiment 2 acts were each repeated twice in a row. We reasoned this would facilitate children’s ability to remember who did what. In addition, after viewing 2 acts from each puppet but before receiving test questions, children were asked comprehension questions to confirm they remembered which puppet performed which physical action during the show. Comprehension questions required children to point to the character who opened/closed the box or gave/took the ball, and so were designed only to determine whether a child could identify each puppet’s action, rather than to encourage children to associate these actions with positive or negative valence. If children failed to correctly respond to comprehension questions, each puppet repeated its action once more, and the comprehension questions were repeated. To be included in the final sample, children had to pass comprehension questions by the end of the third round of puppet shows/comprehension questions (maximum of 4 action repetitions per puppet). We predicted that these modifications would improve 3-year-olds’ performance, such that 3-year-olds in Experiment 2 would report consistent social and moral judgments favoring the helper. We also predicted that 3-year-olds in Experiment 2 would provide more informative explanations that reference the puppet’s unhelpful action and/or negative valence as the reason he should get in trouble.

Experiment 2

Method

Participants. Twenty-four 3-year-olds ($M = 3;7$, range = 3;0-3;11, 10 girls) participated. One additional child was seen but replaced due to unwillingness to watch the puppet show.

Procedure. The only differences between Experiment 1 and Experiment 2 were that children in Experiment 2 watched two helping scenarios in a row and two hindering scenarios in
a row (helping/hindering order counterbalanced), and were asked comprehension questions before test questions. Following the box scenario, children were told, “One of these guys opened the box, and one of these guys closed the box.” Children were then asked to “point to the guy that opened the box” and to “point to the guy that closed the box” (order counterbalanced). Following the ball scenario, children were told, “One of these guys gave the ball back and one of these guys took the ball away.” Children were then asked to “point to the guy that gave the ball back” and to “point to the guy that took the ball away” (order counterbalanced).

If children responded incorrectly to comprehension questions following either scenario, they were shown the helping and hindering events in that scenario again (e.g., “I don’t think he opened the box. I’m going to show you those again”), and the comprehension questions were repeated. One child required an additional repetition of both the box and ball puppet shows, 2 children required an additional repetition of only the first puppet show (ball) and 1 child required two addition repetitions of the first puppet show (ball) before comprehension questions were answered correctly. Based on a pre-set criterion, 1 child’s responses following the box show were not included in the final analysis because the child responded incorrectly to comprehension questions in each of three rounds. After correctly responding to the comprehension questions, children were asked the same test questions as in Experiment 1.

Transcription and coding procedure. As in Experiment 1, two independent research assistants coded for the presence of each response type (uninformative, relevant action, irrelevant action, relevant valence, irrelevant valence, non-social considerations). Reliability across the six categories was moderate (average Cohen’s kappa = .765; see McHugh, 2012), and greatly influenced by 1 disagreement regarding whether a single statement reflected an irrelevant attribution. Without the irrelevant attribution category, reliability was almost perfect (average
Cohen’s kappa = .917, range = .760-1.000). Disagreements (which involved N = 4 explanations) were resolved as in Experiment 1. Two disagreements were immediately recognized as mistakes; one regarding the irrelevant attribution disagreement described above, and one due to failure to consider whether the child was responding to a box or ball show. The other two disagreements were resolved by discussion among the two coders in the presence of the first author.

Results

Social preference and moral judgments. Confirmatory analyses. As in Experiment 1, children received a score of 1 if they responded in the direction of the hypothesis for each test question (i.e., liked the helper, judged the helper as nicer, assigned punishment to the hinderer) and 0 if they responded against the direction of the hypothesis. Children’s scores were summed across the two rounds resulting in a score between 0 – 2 for each question type (see Figure 2.2). To determine whether children responded in the predicted direction at rates exceeding chance, a series of one-sample t-tests were conducted comparing 3-year-old’s liking score, niceness score, and trouble score to a chance score of 1.

In contrast to Experiment 1, 3-year-olds in Experiment 2 reliably distinguished between the helper and hinderer when providing both social and moral judgments. Three-year-olds’ liking score (M = 1.63, SD = .58), niceness score (M = 1.75, SD = .61), and trouble score (M = 1.67, SD = .64) all exceeded chance in the direction of the hypothesis (one-sample t-tests comparing each score between 0 – 2 to a chance score of 1; liking: t[23] = 5.318, p < .001, d = 1.085; niceness:

---

5 As in Experiment 1, the explanations in Experiment 2 were originally coded by two different research assistants and resulted in substantially lower reliability (average Cohen’s kappa = .733, range = .386-1.000). The coding scheme was then clarified, a more extensive training procedure was created, and the data was re-coded by two different research assistants.
$t[23] = 6.044, p < .001, d = 1.234$; trouble: $t[23] = 5.127, p < .001, d = 1.047$. See Table A.2 in Appendix A for a summary of individual patterns of responding within each round.

**Figure 2.2.** Mean liking, niceness, and trouble scores following the box and ball scenarios in Experiment 2; *** $p < .001$, error bars reflect the $SE$.

**Exploratory analyses.** The differential patterns of responding among 3-year-olds in Experiment 1 and 2 suggest that the addition of comprehension questions and/or presenting the helping and hindering scenarios twice in a row rather than in alternation improved young children’s performance (it is currently unclear whether one addition played a larger role than the other; we will return to this issue in the discussion). To formally examine how judgments differed across Experiments, a mixed-effect ANOVA on 3-year-olds’ responses with liking, niceness, and trouble scores as within-subjects variables (repeated-measure), and Experiment (1, 2) and child’s gender (female, male) as between-subjects factors was conducted. This analysis revealed a main effect of Experiment ($F[1,45] = 10.095, p = .003, \eta^2_p = .183$): 3-year-olds’ overall scores (between 0 – 6) summing across the three question types (liking, niceness, trouble) were higher in Experiment 2 ($M = 5.04, SD = 1.49$) compared to Experiment 1 ($M = 3.60, SD = \ldots$)
There was no main effects or interactions involving question type or gender (all $Fs < 1.230$, all $ps > .296$).

**Explanations regarding punishment.** As in Experiment 1, instances of each explanation type (uninformative, relevant action, irrelevant action, relevant valence, irrelevant valence, non-social considerations) were represented as proportions and averaged across the two rounds (see Table 2.1). We predicted that 3-year-olds in Experiment 2 would provide more informative explanations than did 3-year-olds in Experiment 1. A factorial ANOVA examined the effect of Experiment (1, 2) and child’s gender (female, male) on the proportion of uninformative responses. This analysis revealed a main effect Experiment ($F[1,45] = 4.615, p = .037, \eta_p^2 = .093$), such that 3-year-olds provided fewer uninformative explanation in Experiment 2 ($M = .23, SD = .39$) compared to Experiment 1 ($M = .52, SD = .51$); there was no main effect or interaction involving gender (both $Fs < .788$, both $ps > .379$).

We also predicted that 3-year-olds in Experiment 2 would provide more relevant references to the puppets’ actions or valence than in Experiment 1. To test this prediction, we combined the appeals to relevant actions and relevant valence into a single proportion of relevant responses. A factorial ANOVA examined the effect of Experiment (1, 2) and child’s gender (female, male) on the proportion of relevant responses and found no difference across Experiments ($F[1,45] = 1.647, p = .206, \eta_p^2 = .035$); 3-year-olds were equally likely to reference the puppet’s relevant action or valence when explaining their allocations of punishment in Experiment 1 ($M = .28, SD = .46$) and Experiment 2 ($M = .46, SD = .49$); there was no main effect or interaction involving gender (both $Fs < .068$, both $ps > .796$). That said, references to the relevant hindering action from the puppet show were the most common type of informative
explanations (53%) in Experiment 2. See Table A.2 in Appendix A for proportions of each response type following allocations of punishment to the unhelpful rather than helpful puppet.

**Discussion**

Overall, after correctly identifying which puppet performed which action (e.g., opened the box versus closed the box), 3-year-olds demonstrated social preferences and moral judgments that favored the helper. Specifically, across two helping and hindering scenarios children preferred the helper, indicated that the helper was nicer, and selected the hinderer to get in trouble. These findings suggest that 3-year-olds’ failure to reliably distinguish between the helper and hinderer in Experiment 1 was not due to an inability to interpret the helpful and unhelpful actions as socially and/or morally relevant. Rather, these results suggest that younger children in Experiment 1 had difficulty processing or remembering the helpful and unhelpful actions as presented in the first experiment, which was improved by the methodological modifications made in Experiment 2. These methodological modifications also resulted in fewer uninformative explanations when 3-year-olds in Experiment 2 justified their allocation of punishment to one puppet. To the extent that 3-year-olds provided informative explanations, they most frequently appealed to the puppet’s unhelpful action during the puppet show as the reason he should get in trouble. However, there was no difference in the proportion of appeals to relevant social considerations (actions and valence) among 3-year-olds in Experiment 1 and 2.

**General Discussion**

Across two experiments, 3-5-year-olds observed third-party helping and hindering scenarios and made both social preference judgments and moral judgments. When presented with a character that helped a protagonist achieve his goal and a character that thwarted the protagonist’s goal, 4- and 5-year olds in Experiment 1 indicated that they preferred the helper,
evaluated the helper as “nicer” than the hinderer, and selectively allocated punishment to the hinderer. When asked to justify their allocation of punishment, children typically appealed to social considerations—primarily the selected puppet’s performance of an unhelpful action. In contrast, 3-year-olds in Experiment 1 did not provide consistent social preferences or moral judgments, but instead only reliably allocated punishment to the hinderer. Despite allocating punishment to the hinderer, 3-year-olds seldomly referred to relevant social considerations when explaining their allocation. In Experiment 2, 3-year-olds participated in a simplified procedure in which puppet’s actions were presented sequentially rather than in alternation, and answered comprehension questions to assess their understanding of which puppets performed which actions. If children failed the comprehension questions, they were shown the puppet shows again. Following the successful answering of these comprehension questions, 3-year-olds also indicated that they preferred the helper, evaluated the helper as “nicer” than the hinderer, and selectively allocated punishment to the hinderer. When providing informative explanations regarding why the selected puppet should get in trouble, 3-year-olds most commonly appealed to social and/or moral considerations, such as the puppet’s hindering action.

While the current studies utilized two of the helping and hindering puppet shows designed to explore infants’ social evaluations (see Hamlin & Wynn, 2011), there were some modifications to the presentation of the scenarios. Notably, the addition of a few non-valenced words throughout the puppet show, which were added to capture preschoolers’ attention and highlight the protagonist’s goal and struggles (e.g., “Look, a toy. Too heavy!” in the box scenario), the action performed by the helper (e.g., “Open!”) and the action performed by the hinderer (e.g., “Close!”). A potential concern is that these words unduly guided children’s interpretation of the helping and hindering scenarios. We do not think this is the case for two
reasons; first, the added words were not positive or negative (i.e., opening and closing are not inherently good or bad actions), and second, the added words were consistently said in a positive tone of voice to avoid suggesting that one act was more positive than another. That said, it is an empirical question how children interpret the helping and hindering scenarios without narration.

Given the similarity between the procedures previously used to examine infants’ evaluations and those currently used to examine children’s judgments, and infants’ success in forming implicit evaluations following the helping and hindering scenarios (see Hamlin, 2013b), it may be surprising that 3-year-olds in Experiment 1 did not prefer the helper over the hinderer or judge the helper to be nicer than the hinderer. However, it should be noted that there are several potentially important differences between the current procedures and those typically utilized with infants (beyond the addition of narration, as discussed above). Specifically, infants are exposed to more instances of the helping and hindering scenarios (a minimum of 6 total events, as opposed to the 4 total events in Experiment 1). Infants also observe the scenarios from the end of a long table, with a curtain blocking the experimenter from view. In the current studies, the scenarios were enacted directly in front of the child with the experimenter fully visible. Thus, it is possible that 3-year-olds in Experiment 1 would have provided consistent social and moral judgments if they observed more instances of the scenarios and/or were tested without the potential distractions of the nearby puppets and/or experimenter.

Further modifications to the helping and hindering puppet show paradigm were introduced in Experiment 2. These modifications significantly improved 3-year-olds’ performance, resulting in consistent social and moral judgments and more interpretable explanations regarding their punishment allocations (cf. Kenward & Dahl, 2011). It is currently unclear which of these modifications is/are responsible for their improvement: The sequential
rather than alternating presentation of helping and hindering scenarios, the answering of comprehension questions regarding each puppet’s action, or the repetition of the helping and hindering scenarios after incorrect responses to comprehension questions. While not definitive, it seems unlikely that 3-year-olds’ performance improved due to increased exposure to the helping and hindering scenarios. Few children answered comprehension questions incorrectly on the first round and thus, on average, children saw 4.26 acts in Experiment 2 versus 4 acts (2 helping, 2 hindering) in Experiment 1. In addition, we also feel it is unlikely that 3-year-olds’ improved performance in Experiment 2 was due to answering comprehension questions. A recent study in our lab also presented young children with puppet show events presented sequentially rather than in alternation, and compared children’s social and moral judgments before and after answering comprehension questions. Even before answering comprehension questions, 3-year-olds acted like 3-year-olds in the current Experiment 2, providing reliable moral judgments regarding helpers and hinders (Chapter 3 of this dissertation: Van de Vondervoort & Hamlin, 2018b). Overall, then, it seems that the sequential rather than alternating presentation of helping and hindering scenarios in Experiment 2 is likely responsible for the difference in 3-year-olds’ responding across Experiments 1 and 2, presumably because sequential presentation facilitates processing of which puppet performed which action. That said, more work is necessary to determine how young children process and interpret these helping and hindering scenarios.

These studies contribute to a growing literature regarding children’s responses to morally valenced actions. Previous work has demonstrated infants’ preferences for prosocial rather than antisocial others during the first year of life (e.g., Hamlin & Wynn, 2011; Hamlin et al., 2007). By ages 3-4 children clearly distinguish between moral transgressions and conventional or personal transgressions (e.g., Smetana, 2006 for review) and spontaneously protest the
performance of antisocial behaviors (e.g., Rossano et al., 2011). The current studies suggest that 3- to 5-year-olds evaluate goal facilitation and goal blocking as morally relevant behaviors, and form a variety of moral judgments (i.e., positive moral evaluations, allocation of punishment) based on these actions. Furthermore, when asked to justify a moral judgment (the allocation of punishment), children often appeal to the performance of unhelpful actions within the hindering scenario. While it is possible that appeals to the hindering action from the puppet show reflect non-social concerns (e.g., reports regarding the physical action performed) or social-but-not-moral concerns (e.g., reports regarding the puppet’s role in blocking the protagonist’s goal without any sense that this action was inappropriate), it is also possible that these explanations reflect moral concerns. Specifically, appeals to the puppet’s hindering action when explaining why he should get in trouble may suggest that children’s moral judgments have a normative basis – that the puppet ought not have blocked the protagonist’s goal and so should be punished – rather than a non-social or idiosyncratic, social basis.

In the current studies, we asked children to only explain their moral judgments to prevent explanations regarding moral judgments from influencing explanations regarding social preferences and vice versa. To further explore the cognitions underlying preferences for helpful versus unhelpful others, future work should explore children’s justifications for their social preferences. To illustrate, it would be informative to determine the extent to which children appeal to non-social considerations when explaining social preferences, such as idiosyncratic preferences based on the puppet’s appearance or neutral aspects of the puppet show display (e.g., non-helping or hindering actions performed during the show). If children are more likely to appeal to social considerations when explaining moral judgments and to appeal to non-social
considerations when explaining social preferences, this would suggest that social and moral responses to the helping and hindering scenarios are (or can be) distinct.

In addition, although the methods utilized here were chosen based on a host of previous work in the moral development literature (e.g., Baird & Astington, 2004; Costanzo et al., 1973; Cushman et al., 2013; Leslie et al., 2006; Zelazo et al., 1996), it would also be fruitful to explore children’s reactions to helping and hindering using other common moral development methodologies. For example, our methods could be modified to examine children’s spontaneous protests to helpful versus unhelpful acts (for instance, if puppets were to warn about what they were going to do before they do it, children might step in and instruct them not to; e.g., Rossano et al., 2011; Vaish et al., 2011), or to examine children’s assessments of whether it was “OK” versus “not OK” for a puppet to behave as it did, given the existence of authority sanctions, unique contexts, etc. (for review see Smetana, 2006). Of note, a judgment regarding whether a single puppet’s actions are “OK” or “not OK” is distinct from the forced-choice judgments preschoolers provided in the current studies. While categorical judgments that the helper’s actions were acceptable and the hinderer’s actions were unacceptable would be consistent with the findings of the current studies, it is an open question whether children provide categorical moral judgments following helping and hindering scenarios. These methods would further help to elucidate the social versus moral nature of children’s reactions to helpers and hinderers.

A final open question is whether infants’ implicit preferences for helpers over hinderers are related to children’s explicit social and moral judgments. A fruitful area for future research would be exploring the continuity of moral evaluation and behavior across the development of individual children, beginning with puppet preferences in infancy. Are those children who were more likely to prefer prosocial puppets in infancy also more likely to identify prosocial puppets
as nicer and allocate punishment to hinderers in early childhood? Do these children perform better on other measures of social and/or moral functioning than peers who were less likely to prefer prosocial puppets in infancy? Establishing whether there is within-individual continuity from infancy through childhood would greatly add to the debate regarding whether infants’ responses should be considered precursors to a moral sense. In sum, the current studies shed light on preschoolers’ explicit evaluations regarding third-party helpers and hinderers, and suggest multiple avenues for future research.
### Table 2.1

Proportion of explanations containing each response type in Experiment 1 and Experiment 2

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Age</th>
<th>Uninformative</th>
<th>Informative</th>
<th>Relevant</th>
<th>Irrelevant</th>
<th>Relevant</th>
<th>Irrelevant</th>
<th>Non-social</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>action</td>
<td>action</td>
<td>valence</td>
<td>valence</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>.52</td>
<td></td>
<td>.24</td>
<td>.00</td>
<td>.04</td>
<td>.08</td>
<td>.12</td>
</tr>
<tr>
<td>4</td>
<td>.17</td>
<td></td>
<td>.63</td>
<td>.04</td>
<td>.10</td>
<td>.00</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.04</td>
<td></td>
<td>.74</td>
<td>.04</td>
<td>.08</td>
<td>.04</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>.23</td>
<td>.41</td>
<td>.02</td>
<td>.05</td>
<td>.04</td>
<td>.25</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The informative response categories were not mutually exclusive and proportions sum to 1.
Chapter 3: Children’s Consideration of Others’ Intentions

Introduction

When considering whether an action is good or praiseworthy versus bad or blameworthy, adults are sensitive to both an agent’s mental states (intentions, beliefs, desires) and the outcomes they bring about. While in some cases adults do condemn those that unintentionally cause harm (e.g., Cushman, Dreber, Wang, & Costa, 2009; Cushman & Greene, 2012; Gino, Moore, & Bazerman, 2008), adults typically privilege intentions over outcomes when making moral judgments (e.g., Cushman, 2008; Malle, 1999; Young, Cushman, Hauser, & Saxe, 2007). The ability to incorporate mental state information into moral judgments, rather than focus strictly on the outcomes of morally relevant actions, has long been considered a hallmark of moral maturity (Kohlberg, 1981; 1984; Piaget, 1932/1965).

Beginning with the work of Jean Piaget, researchers have explored when this feature of the mature moral sense becomes operational, and have documented a developmental transition whereby children’s moral judgments initially focus on outcomes and only later shift to focusing on intentions. For example, as described in Chapter 1, Piaget found that younger children tended to judge a child who accidently broke 15 cups as naughtier than a disobedient child who broke one, and it was not until age 8 – 10 that children focused on others’ intentions by more positively evaluating the child who accidentally caused a large negative consequence (1932/1965).

Subsequent studies revealed that Piaget’s methodology led him to underestimate the age at which children can use mental states to inform their moral judgments, suggestive that the centrality of intentions in moral judgments does not require many years of maturation, teaching, and/or relevant experiences to emerge. For example, young children incorporate intentions into their moral judgments when intentions are explicitly stated or otherwise made salient, when
intentions are deconfounded from outcomes (e.g., consequences are held constant while intentions vary), and when a larger variety of test questions are used (e.g., asking about the agent rather than the acceptability of the act; e.g., Armsby, 1971; Bearison & Isaacs, 1975; Berg-Cross, 1975; Buchanan & Thompson 1973; Chandler, Greenspan, & Barenboim, 1973; Constanzo, Coie, Grumet, & Farnhill 1973; Cushman, Sheketoff, Wharton, & Carey, 2013; Farnill, 1974; Karniol, 1978; Nelson-Le Gall, 1985; Nobes, Panagiotaki, & Bartholomew, 2016; see Hilton & Kuhlmeier, 2019 for a discussion of how task procedures may encourage or discourage young children’s production of intention-based moral judgments). Under these circumstances, even 3-year-olds’ judgments show sensitivity to intentions (Nelson, 1980; Nobes, Panagiotaki, & Pawson, 2009; Yuill & Perner, 1988). That said, a host of studies have repeatedly demonstrated that young children initially privilege outcome over intention when the two are in conflict, and increasingly consider intention as they age (e.g., Armsby, 1971; Baird & Astington, 2004; Costanzo et al., 1973; Helwig, Zelazo, & Wilson, 2001; Imamoglu, 1975; Killen, Mulvey, Richardson, Jampol, & Woodward, 2011; Li & Tomasello, 2018; Margoni & Surian, 2017; Moran & O’Brien, 1983; Zelazo, Helwig, & Lau, 1996; see Margoni & Surian, 2016 for review).

Notably, children’s ability to incorporate intentions into moral judgments has typically been tested using vignette-based tasks, in which experimenters narrate illustrated stories and then probe children’s explicit judgments (but see Farnill, 1974, Chandler et al., 1973, and Li & Tomasello, 2018 for use of videotaped scenes). These judgments include both verbal and Likert scale ratings of action acceptability (e.g., “Is it okay for [her] to [perform that act]? How good is it for [her] to [perform that act]? Is it really, really good, or just a little good, or just okay?”, Zelazo et al., 1996) and/or the moral worth of a character (e.g., “Is [he] a good boy or a bad boy?”, Costanzo et al., 1973).
Clearly, the tasks described above require that children can process a verbally-presented story and respond to explicit questioning. These requirements may underrepresent the abilities of children who are unwilling or unable to engage in explicit questioning. Thus, researchers have recently developed tasks measuring more implicit forms of evaluation. Rather than asking for responses to specific test questions, some researchers have explored early sensitivity to others’ prosocial and antisocial intentions using age-appropriate behavioral tasks. Such studies provide further evidence that young children are sensitive to others’ intentions. For example, in one study 3-year-olds were less likely to help an adult who attempted but failed to harm a third party compared to a neutral adult (Vaish, Carpenter, & Tomasello, 2010), while in another study 3- and 4-year-olds were more likely to spontaneously correct punishments imposed on others who accidentally rather than intentionally caused the same harmful outcome (Chernyak & Sobel, 2016). In a final example, 5- and 6-year-olds used informants’ past intentions and outcomes when determining who to trust when searching for a prize (Liu, Vanderbilt, & Heyman, 2013).

Critically, more implicit forms of evaluation, in which neither the story presentations nor the response measures require verbal abilities, also allow for the study of preverbal children who are less than 3 years of age. To illustrate, in one study 5- and 9-month-old infants watched a live-action puppet show featuring a protagonist puppet who repeatedly tried but failed to open a box containing an attractive toy. In alternation, a helper puppet assisted the protagonist in opening the box so that he could access the toy, and a hinderer puppet slammed the box shut, preventing the protagonist from achieving his goal. When subsequently presented with a choice between the helper and hinderer, both 5- and 9-month-olds preferentially reached for the helper rather than the hinderer puppet, suggestive that infants differentially evaluated prosocial versus antisocial others (Hamlin & Wynn, 2011; for replications and related findings see Buon et al., 2014;
Hamlin, 2015; Hamlin, Wynn, & Bloom, 2007; 2010; Scola, Holvoet, Arciszewski, & Picard, 2015; Steckler, Woo, & Hamlin, 2017; for failure to replicate see Salvadori et al., 2015; see also Scarf, Imuta, Colombo, & Hayne, 2012).

These implicit paradigms have also been utilized to explore infants’ sensitivity to third-party scenarios in which intentions and outcomes conflict. In one such task, 5- and 8-month-olds watched puppet shows in which successful and unsuccessful helpers and hinderers intervened following a protagonist’s repeated failure to open a box (Hamlin, 2013a). The successful helper and hinderer achieved their respective goals to either assist or thwart the protagonist’s goal (as in Hamlin & Wynn, 2011). Conversely, the failed helper and hinderer brought about an outcome that conflicted with their intention: the failed helper tried but failed to open the box, while the failed hinderer tried but failed to prevent the protagonist from opening the box. When presented with different combinations of successful and failed helpers and hinderers, 8-month-olds preferentially reached for puppets with helpful intentions, regardless of the outcome that occurred (i.e., successful helpers over failed hinderers, failed helpers over successful hinderers, and failed helpers over failed hinderers). In contrast, when presented with two puppets who had demonstrated the same intention (i.e., successful helper and failed helper, failed hinderer and successful hinderer), 8-month-olds showed no preference for either puppet, suggestive that they did not evaluate characters based on the outcomes they were associated with. Unlike 8-month-olds, 5-month-olds preferentially reached for successful helpers over successful hinderers, but showed no preferences when presented with any failed puppet (Hamlin, 2013a). Thus, infants’ sociomoral evaluations appear to privilege intentions over outcomes by 8 months of age, but not at 5 months (for related evidence with accidental help and harm see Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013; Woo, Steckler, Le, & Hamlin, 2017).
In another task measuring infants’ expectations about characters involved in failed attempts to help and harm, 12- and 16-month-olds watched a video featuring a protagonist unsuccessfully attempting to climb a hill. Two characters alternately intervened: A successful hinderer who pushed the protagonist down the hill, and either a successful helper or unsuccessful helper (Lee, Yun, Kim, & Song, 2015). Subsequently, looking times suggested that 16-month-olds expected the protagonist to approach the character who had intended to help, even if he failed to do so and the protagonist’s outcome was negative. In contrast, 12-month-olds expected the protagonist to approach the successful helper rather than the hinderer, but only to approach the failed helper over the hinderer when outcome information was removed from the video (Lee et al., 2015). Together, these studies demonstrate that although a salient outcome may disrupt this sensitivity, infants are sensitive to others’ intentions to help or hinder – even when intentions and outcomes conflict. Indeed, unlike much work with young children (e.g., Armsby, 1971; Baird & Astington, 2004; Costanzo et al., 1973; Helwig et al., 2001; Killen et al., 2011; Margoni & Surian, 2017; Moran & O’Brien, 1983; Yuill & Perner, 1988; Zelazo et al., 1996), to our knowledge no work to date have provided evidence that infants’ third-party social evaluations and expectations either rely solely on outcome or initially privilege outcome over intention.

What accounts for this apparent developmental discontinuity, whereby infants seem to privilege intentions but young children privilege outcomes? One possibility is that presentation of the social interactions via live puppet shows or videos, rather than illustrated vignettes, might facilitate understanding, in that a fully acted-out scenario provides richer information than does a narrated short vignette (see Farnill, 1974 and Chandler et al., 1973 for evidence that children are sensitive to the intentions of characters in videotaped scenes from age 6). If so, then presenting preschoolers with live puppet shows may facilitate relatively more mature moral reasoning – that
is, positive evaluations of those with positive intentions and negative evaluations of those with negative intentions, irrespective of the outcomes characters bring about.

**The Current Studies**

The current studies explore whether preschoolers’ social and moral judgments privilege intentions, even when agents’ intentions conflict with the outcome of their actions. Scenarios were enacted via a live puppet show and based on shows previously used to explore infants’ sociomoral evaluations of characters with varying intention who are associated with varied outcomes (Hamlin, 2013a). Children viewed events in which a protagonist unsuccessfully attempted to open a box to reach a toy inside (as in Hamlin & Wynn, 2011). Two additional puppets intervened: *helpers* demonstrated a positive intention to assist the protagonist, while *hinderers* demonstrated a negative intention to prevent the protagonist from achieving his goal. The helper and hinderer puppets were either *successful* in bringing about their objective, or *failed* to assist or thwart the protagonist’s goal. Thus, across studies, the protagonist interacted with four puppets: (1) successful helpers who try and help the protagonist achieve his goal, resulting in a positive outcome for the protagonist, (2) successful hinderers who try and block the protagonist’s goal, resulting in a negative outcome for the protagonist, (3) failed helpers who unsuccessfully try to help the protagonist achieve his goal, resulting in a negative outcome for the protagonist, and (4) failed hinderers who unsuccessfully try to block the protagonist’s goal, resulting in a positive outcome for the protagonist.

Each child was presented with two distinct events (e.g., failed helping and successful hindering), and then were asked three test questions: (1) which of the two puppets they “like”, (2) which was “nicer”, and (3) which “should get in trouble”. After children identified who should get in trouble, they were asked to explain this judgment. While these forced-choice
questions do not allow conclusions regarding whether children (for example) think either puppet is “nice” (rather than “nicer”), these questions have been used to examine 3- to 5-year-olds’ social and moral judgments following helping and hindering puppet shows in which intentions and outcomes were not in conflict (Van de Vondervoort & Hamlin, 2017), are consistent with the forced-choice nature of infants’ evaluations in past work, and are consistent with questions previously used to explore young children’s explicit moral judgments (e.g., Baird & Astington, 2004; Costanzo et al., 1973; Cushman et al., 2013; Zelazo et al., 1996). Following the first round of liking/niceness/punishment test questions, children answered comprehension questions regarding the puppets’ actions and the outcome of each event and then answered the same test questions again. Comprehension questions ensured that children attended to both the failed/successful helper/hinderer’s actions and the outcome for the protagonist.

Experiment 1 explored whether 3- and 4-year-olds utilize actors’ mental states to inform their social and moral judgments when outcomes are equivalent. Children observed a live puppet show featuring a protagonist who failed to achieve his goal to open a box. In the “positive outcome” condition, a successful helper and failed hinderer intervened, resulting in a positive outcome for the protagonist (i.e., the box was opened and the toy was reached). In the “negative outcome” condition, a failed helper and successful hinderer intervened, always resulting in a negative outcome for the protagonist (i.e., the box was closed and the toy was not reached). Experiment 2 then examined whether 3- and 4-year-olds’ judgments privilege actors’ mental states or the outcome of their actions when these intentions and outcomes conflict. Children observed live puppet show events featuring the same protagonist; a failed helper intervened to bring about a negative outcome for the protagonist, while a failed hinderer was associated with a positive outcome. Finally, Experiment 3 investigated whether 3- and 4-year-olds’ judgments
were sensitive to outcomes when actors’ mental states were equivalent. In one condition a successful helper and a failed helper intervened in the protagonist’s struggle, while in the second condition a successful hinderer and a failed hinderer intervened; critically, both puppets in each condition had the same intention but brought about opposite outcomes.

Based on work showing that young children are sensitive to intentions when outcomes are equivalent across scenarios in vignette tasks (e.g., Cushman et al. [2013] found that children evaluated a character who attempted but failed to cause harm more negatively than a character who successfully brought about an intended positive outcome by age 4; see also Chernyak & Sobel, 2016) and that infants privilege agents’ intentions following similar scenarios (Hamlin, 2013a), it was predicted that 3- and 4-year-olds in Experiment 1 would report liking the character with the positive intention, judge the character with the positive intention as nicer, and allocate punishment to the character with the negative intention, even though the characters were not distinguishable based on the valence of their associated outcomes. Further, based on work showing that 3-year-olds can have difficulty producing interpretable responses to open-ended questions (e.g., Kenward & Dahl, 2011; Van de Vondervoort & Hamlin, 2017), it was predicted that in this and all further experiments, 4-year-olds would provide more informative verbal justifications than 3-year-olds. It was also predicted that 4-year-olds would be more likely than 3-year-olds to reference sociomoral considerations as the reason for their punishment allocations, including references to the characters’ successful or unsuccessful attempts to block the protagonist’s goal. Children’s gender was not expected to influence responding, although this possibility was explored similarly in this and all further experiments, as this is common in developmental work (e.g., Helwig et al., 2001; Nobes et al., 2009).
Experiment 1

Method

Participants. Twenty-four 3-year-olds ($M = 3;6$, range = 3;2-3;11, 13 girls) and 24 4-year-olds ($M = 4;6$, range = 4;0-4;11, 16 girls) participated in the positive outcome condition, while 26 3-year-olds ($M = 3;6$, range = 3;0-3;11, 15 girls) and 24 4-year-olds ($M = 4;4$, range = 4;0-4;10, 12 girls) participated in the negative outcome condition. Before data collection began a pre-set stopping rule was established of 24 children per age per condition; two extra 3-year-olds were run due to scheduling issues. An additional 26 3-year-olds were tested but replaced due to failure to complete an English language warm-up (2), procedure error (1), unwillingness to participate (1), and a color and/or side preference that resulted in pointing to the same puppet across all test questions in one or both rounds (22). An additional eight 4-year-olds were tested but replaced due to color/side preferences. The decision to remove children that displayed a color/side preference in one or both rounds of test questions was pre-set following a pilot study, as children who judged that the same puppet is “liked”, “nicer” and “should get in trouble” appeared unmotivated and/or that they did not understand the test questions. Appendix B includes key analyses including children with color/side preferences; results are essentially identical in all experiments and do not influence the interpretations reported here.

Procedure. Warm-up. Children were shown a picture of a playground and asked to find the swing and slide, and to name the color of a toy and their favorite outside activity. Before data collection began it was decided that children would be replaced in the sample if they were unable/unwilling to locate the swing or slide via pointing; verbal responses were not required.

Puppet show. Children participated in either the positive outcome condition or the negative outcome condition. All children watched a live puppet show featuring a protagonist
struggling to achieve his goal to open a box and reach an attractive toy; a second and third puppet then intervened (successful helper and failed hinderer or failed helper and successful hinderer; see Figure 3.1). Puppet events were based on previous infant studies (Hamlin, 2013a; see also Hamlin & Wynn, 2011), with two notable differences (as in Van de Vondervoort & Hamlin, 2017). First, for infants, the puppet events were enacted at the end of a long table and a curtain was lowered between events to hide the puppets; experimenters were hidden behind a curtain at the back of the table. Events in the current experiments were enacted on the floor or a table directly in front of the child and with the experimenter visible. Second, a few non-valenced words were added to the events for narration. All narrations were produced in a high-pitched, positive voice to indicate that the puppet was speaking rather than the experimenter; speech was not modulated based on the valence of the puppets’ intention or the eventual outcome.

**Figure 3.1.** Visual depiction of the puppet show events.
Children watched four puppet events; two successful helper and two failed hinderer events in the positive outcome condition, two failed helper events and two successful hinderer events in negative outcome condition. At the start of each event, the successful/failed helper/hinderer puppets were seated on either side and back from a clear box containing a purple whale toy. The experimenter enacted the protagonist walking up to the box, looking through the side of the box while saying “Look, a toy!”, and unsuccessfully attempting to open the box five times. During the third to fourth attempt, the protagonist said, “Too heavy!”. On the fifth attempt, the successful/failed helper/hinderer intervened:

**Successful helper.** In successful helper events the puppet ran forward, joined the protagonist’s struggle, and aided in opening the box while saying, “Open!” The puppet then ran away and the protagonist laid facedown, grasping the toy inside the box while saying “Toy!”

**Failed helper.** In failed helper events the puppet ran forward and joined the protagonist’s attempts to open the box three more times; during the first attempt the puppet said “Open!” The puppet then ran away and the protagonist laid facedown beside the box while saying “No toy!”

**Successful hinderer.** In successful hinderer events the puppet ran forward and jumped on the box, slamming it closed while saying “Close!” The puppet then ran away and the protagonist laid facedown beside the box while saying “No toy!”

**Failed hinderer.** In failed hinderer events the puppet ran forward and jumped on the box, slamming it closed while saying “Close!” The protagonist then struggled to open the box while the puppet jumped on the box twice more before running away. After another struggle, the protagonist successfully opened the box and laid facedown, grasping the toy while saying “Toy!”

---

6 In the failed hinderer events in Hamlin (2013) the puppet jumped on the box once; here the number of failed attempts (three) was equated across the failed helper and failed hinderer events. See Experiment 2 in the current paper for children’s judgements of the failed hinderer following one versus three attempts to close the box.
The narration during each event was designed to highlight the intervening puppets’ intention and the eventual outcome. Children were shown each event twice in a row, for a total of four events. Three puppets were used: A duck (protagonist) and two rabbits wearing a red and a green shirt (failed/successful helper/hinderer, identity counterbalanced). Additional counterbalanced variables were event order (red first, green first) and side of the puppets (red right, red left). For the question period puppets remained on the same side as during the show.

**Test questions.** Following the puppet events, children were presented with the successful/failed helper/hinderer and asked (in counterbalanced order) which puppet they preferred (i.e., “Which one of these guys do you like the most?”) and which was nicer (i.e., “Which one of these guys was nicer?”)⁷. To reduce response perseveration, children were asked to point to each puppet in between the liking and niceness questions (e.g., “Point to the guy with a red/green shirt. Right!”). Children were then asked which puppet deserved punishment and to explain this choice (i.e., “I think that one of these guys should get in trouble. Who should get in trouble? Why should he get in trouble?”). Children were prompted if they did not explain their punishment allocation (e.g., “What do you think?”). Children then answered comprehension questions and were asked the same test questions again. For each test question, children received a score of 1 if they responded in the direction of the hypothesis and 0 if not, resulting in a total of six scores (three test questions, two rounds of questioning) between 0 – 1 per child. One 4-year-old in the positive outcome condition responded that both puppets were liked in round one and after children identified the nicer puppet, they were asked whether that puppet was a “little bit nice or a lot nice” (order counterbalanced). The initial plan was to examine niceness judgments on a 3-point scale from “not nice” to “a lot nice”, but because children were not trained on this scale prior to testing and because most children at each age responded that the selected puppet was “a lot” nice regardless of which puppet they indicated was nicer, this question is not considered further.
one 3-year-old in the negative outcome condition responded, “I don’t know” when asked which puppet should be punished in round two; these responses were scored as against hypothesis.

**Comprehension questions.** Following the first round of test questions, children were shown each event type and asked one comprehension question about the intervening puppet’s action (e.g., for successful puppets, “Did he open the box or close the box?” and for failed puppets, “Did he try to open the box or try to close the box?”) and one comprehension question about the outcome for the protagonist (i.e., “Did the duck get the toy?”). If answered incorrectly, children were shown the event again and the question was repeated (e.g., “I don’t think he opened the box. Let me show you that one again”). If a comprehension question was answered incorrectly twice, children were corrected (e.g., “He opened the box. This bunny opened the box.”). Across experiments, 73% of children answered all four comprehension questions correctly the first time; only 3% of children were corrected before they answered test questions.

**Transcription and coding.** When permitted by caregivers and possible within the preschool, participation was audio and visually recorded. A research assistant transcribed children’s verbal explanations from these recordings. When recording was not permitted (53 of 295 children across experiments), explanations were transcribed during the study by the experimenter. Two additional research assistants who were not involved in data collection or transcription coded children’s explanations according to the following categories:

**Uninformative responses.** Uninformative responses included those in which children provided no verbal response, unintelligible responses, or verbal responses that did not include a justification for the punishment allocation. These verbal responses included statements unrelated to the puppet events (e.g., “there’s a big storm”), statements without a justification (e.g., “because in trouble”), and statements that the child was unsure (e.g., “I don’t know”).
**Informative responses.** Informative responses were related to the shows and included:

**Protagonist’s goal.** References to the protagonist’s goal to open the box and/or reach the toy inside the box (e.g., “the duck [protagonist] was trying to open it”).

**Relevant action.** References to the puppet’s attempted or completed helping or hindering action (e.g., “he was trying to close the box”, “he closed the box”, “because she didn’t open it”).

**Irrelevant Action.** References to positively or negatively valenced actions not from the shows (e.g., “because he punched this one”). While inaccurate, these responses may reflect children’s beliefs about actions that typically lead to punishment.

**Relevant skill valence.** References to the positive or negative nature of the puppet’s ability to open or close the box (e.g., “not strong”, “he wasn’t doing it very good”)

**Relevant general valence.** References to the positive or negative valence of the puppet or its actions that were related to the shows, but not related to the puppet’s ability to open or close the box (e.g., “he [selected puppet] was mean”, “this one [unselected puppet] was nicer”).

**Irrelevant valence.** References to the positive or negative valence of the puppet or its actions that were not directly related to the shows (e.g., “he’s mad”).

**Non-social considerations.** Responses that did not include sociomoral content, such as physical descriptions of the puppet (e.g., “he’s soft”), general disliking of the puppet (e.g., “I like the green shirt one”), descriptions of neutral acts (e.g., “he’s playing”), and ambiguous statements (e.g., “he does a lot of things”).

---

8 A need for this category had not been identified when the positive intention condition of Experiment 3 was initially coded. This category was added once it was clear that some children were using skill explanations. The data was then entirely recoded without discussing any statements within that condition. Due to the order in which the data was coded, this category was not included in the coding scheme for Experiment 2A; inspection of the transcriptions by the first author revealed it was not necessary to recode as no children used skill explanations. The full coding scheme, including skill explanations, was utilized for all other experiments in Chapter 3.
Each explanation was coded by two independent research assistants for the presence or absence of each response type; coders were blind to the referent (failed/successful helper/hinderer) of the explanation. Informative response types were not mutually exclusive. To avoid over-representing talkative children, whose explanations may have contained several types of informative responses, instances of each explanation type were represented as proportions and averaged across the two rounds. Reliability across the eight categories was strong (average Cohen’s kappa = .812; see McHugh, 2012). Disagreements were resolved by discussion among the two coders and the first author.

**Results**

**Test questions.** To explore whether responses differed before and after comprehension questions, a series of mixed-effect ANOVAs were conducted with round one scores and round two scores as within-subjects variables, and age (3, 4) and gender (female, male) as between-subjects factors. When compared to a Bonferroni-corrected alpha value of .017 (.05/3), there were no main effects of round or interactions involving round of questioning within the positive outcome condition (all $F$s < 6.039, $p$s > .017, $\eta^2_p$s < .122) or the negative outcome condition (all $F$s < 3.437, $p$s > .069, $\eta^2_p$s < .069). Thus, children’s scores were summed across rounds resulting in three scores between 0 – 2 per child (liking, niceness, trouble scores). Appendix B shows scores in each round for all experiments; in all experiments, results from the first round are similar to those reported here and do not influence the interpretations presented in the main text. The dataset generated and analyzed for all experiments in Chapter 3 can be found on the Open Science Framework (osf.io/mgzq7).

**Confirmatory analyses.** To determine at what age(s) liking, niceness, and trouble scores differed from chance, a series of one-sample t-tests compared scores at each age to a chance
score of 1. Three-year-olds in the positive outcome condition did not distinguish between the puppets when reporting who they liked ($p = .137$), while 4-year-olds liked the successful helper ($p = .015$). Both ages judged the successful helper to be nicer ($ps < .001$) and allocated punishment to the failed hinderer ($ps < .001$). In the negative outcome condition, both ages liked the failed helper (3-year-olds: $p = .047$; 4-year-olds: $p = .005$), judged the failed helper as nicer ($ps < .001$), and allocated punishment to the successful hinderer ($ps < .001$; see Figure 3.2 and Table 3.1 for descriptive and test statistics).

**Figure 3.2.** Mean liking, niceness, and trouble scores at each age in Experiment 1. Each score ranges between 0 – 2 with higher values indicating higher rates of with-hypothesis responding across two rounds of questioning; *$p < .05$, **$p < .01$, ***$p < .001$, error bars reflect the $SE$.**

**Exploratory analyses.** To examine whether age, gender, and/or question type influenced children’s tendency to respond in the direction of the hypothesis, two mixed-effect ANOVAs were conducted with question type (liking, niceness, and trouble scores) as a within-subjects variable (repeated-measure), and age (3, 4) and gender (female, male) as between-subjects factors. In the positive outcome condition, there was only a main effect of question type ($F[1.463,64.364] = 14.794, p < .001, \eta^2_p = .252$; all other $Fs < 1.542, ps > .220, \eta^2_p's < .035$). To
explore the main effect of question type, a series of paired-samples t-tests using the Bonferroni corrected alpha value of .017 (.05/3) was used to compare scores on each question type across age. In the positive outcome condition, children were less likely to respond in the direction of the hypothesis when asked which puppet they liked ($M = 1.33, SD = .78$) compared to which puppet was nicer ($M = 1.85, SD = .41$; $t[47] = 4.518, p < .001, d = .652$) and which puppet should get in trouble ($M = 1.79, SD = .54$; $t[47] = 4.276, p < .001, d = .617$); there was no difference between niceness and trouble scores ($t[47] = 1.000, p = .322, d = .144$).

In the negative outcome condition, there was a main effect of question type ($F[1.293,59.466] = 6.363, p = .009, \eta^2_p = .122$) and an interaction between age and gender ($F[1,46] = 4.483, p = .040, \eta^2_p = .089$; all other $Fs < .846, ps > .362, \eta^2_p's < .019$). To explore the main effect of question type, a series of paired-samples t-tests using the Bonferroni corrected alpha value of .017 (.05/3) was used to compare scores on each question type across age. Children were again less likely to respond in the direction of the hypothesis when asked which puppet they liked ($M = 1.42, SD = .81$) compared to which puppet was nicer ($M = 1.74, SD = .56$; $t[49] = 2.947, p = .005, d = .417$) and which puppet should get in trouble ($M = 1.66, SD = .66$; $t[49] = 2.585, p = .013, d = .366$); again there was no difference between niceness and trouble scores ($t[49] = 1.661, p = .103, d = .235$). To explore the interaction between age and gender, two independent-samples t-tests using the Bonferroni corrected alpha value of .025 (.05/2) were used to compare overall scores (summing liking, niceness, and trouble scores, resulting in a score between 0 – 6 for each child) across the three question types. Among 3-year-olds, males were more likely to respond in the direction of the hypothesis ($M = 5.55, SD = .69$) than were females ($M = 4.07, SD = 2.05$; $t[18.024] = 2.600, p = .018, d = .945$); there was no
difference between 4-year-olds males’ \((M = 4.67, SD = 2.15)\) and females’ scores \((M = 5.25, SD = 1.36; t[22] = .796, p = .435, d = .339)\).

**Punishment explanations.** Children most frequently appealed to relevant (un)successful helping or hindering actions when explaining their punishment allocations: 44% of 3-year-olds and 65% of 4-year-olds in the positive outcome condition, and 52% of 3-year-olds and 59% of 4-year-olds in the negative outcome condition did so (see Table 3.2). In their statements, nearly all children referenced the puppet’s attempted or completed hindering action (e.g., “he was closing it”, “he closed the lid”), although one 4-year-old in the negative outcome condition referenced a helping action (i.e., “he’s trying to open that” in round one, and “because he was opening the box” in round two when explaining why the failed helper should be punished).

To test whether younger children provide less interpretable explanations, two factorial ANOVAs examined the effect of age (3, 4) and gender (female, male) on the proportion of uninformative responses across rounds. While the proportion of uninformative responses was greater among 3-year-olds compared to 4-year-olds in the positive outcome condition \((F[1,44] = 4.117, p = .049, \eta^2_p = .086; \text{all other } Fs < 2.694, ps > .107, \eta^2_p s < .059)\), there was no difference in the proportion of uninformative responses across age in the negative outcome condition \((F[1,46] = 1.665, p = .203, \eta^2_p = .035; \text{all other } Fs < 2.055, ps > .158, \eta^2_p s < .044)\).

Finally, to test whether 4-year-olds would be more likely than 3-year-olds to reference relevant sociomoral content when explaining their punishment allocations, appeals to the protagonist’s goal, relevant actions, relevant general valence, and relevant skill valence were combined into a single “relevant responses” category. A factorial ANOVA examining the effect of age (3, 4) and gender (female, male) on the proportion of relevant responses revealed that both ages provided equally relevant responses in both the positive outcome condition \((F[1,44] = \)
2.867, \( p = .097, \eta_p^2 = .061 \); all other Fs < .187, ps > .667, \( \eta_p^2 \)s < .005) and the negative outcome condition (\( F[1,46] = .180, p = .674, \eta_p^2 = .004 \); all other Fs < .551, ps > .461, \( \eta_p^2 \)s < .013).

**Discussion**

Overall, Experiment 1 demonstrated that preschoolers distinguish between characters with opposing intentions when outcomes are uninformative. When presented with a successful helper and failed hinderer who both brought about a positive outcome, 3-year-olds showed no preference for either character, while 4-year-olds preferred the successful helper. Both 3- and 4-year-olds judged the successful helper to be nicer and allocated punishment to the failed hinderer. When presented with a failed helper and successful hinderer who both brought about a negative outcome, both 3- and 4-year-olds preferred the failed helper, judged the failed helper as nicer, and allocated punishment to the successful hinderer. Across conditions, children were more likely to respond in the direction of the hypothesis regarding moral (niceness/punishment) rather than social questions (liking), and most often referenced the character’s attempted or completed hindering action when explaining which character should get in trouble.

Results from Experiment 1 are consistent with past work in which young children demonstrate sensitivity to others’ intentions when intentions do not conflict with the outcomes brought about (Buchanan & Thompson, 1973; Costanzo et al., 1973; Farnill, 1974; Nelson, 1980). Experiment 2 sought to determine whether children still privilege intentions when they do conflict with outcomes. Children observed a puppet show featuring a protagonist who unsuccessfully attempted to open a box. A failed helper and failed hinderer intervened; both characters brought about outcomes that conflicted with their intention. Based on past work showing that older preschoolers can incorporate intention information into their vignette-based judgments (e.g., Cushman et al. [2013] found that children evaluate accidental harm more
positively than attempted harm by age 5), younger preschoolers’ sensitivity to intentions following puppet show events in Experiment 1, and past work showing that infants privilege agents’ intentions following these puppet events (Hamlin, 2013a), it was predicted that children would report liking the character with the positive intention, judge the character with the positive intention as nicer, and allocate punishment to the character with the negative intention.

### Experiment 2A

**Method**

**Participants.** Twenty-four 3-year-olds ($M = 3;5$, range = 3;0-3;11, 13 girls) and 24 four-year-olds ($M = 4;6$, range = 4;0-4;11, 9 girls) participated. An additional 15 3-year-olds were replaced due to unwillingness to participate (1), caregiver interference (1), failure to accept correction during comprehension questions (2), and color/side preferences (11). An additional 3 4-year-olds were replaced due to unwillingness to participate (1) and color/side preferences (2).

**Procedure.** The warm-up task, test questions, comprehension questions (i.e., "Did he try to open the box or try to close the box? Did the duck get the toy?"), transcriptions and coding procedures were identical to Experiment 1.

**Puppet show.** Children watched a live puppet show featuring a protagonist struggling to open a box; a second and third puppet intervened (failed helper in two events, failed hinderer in two events). All details were identical to those in Experiment 1, except for the failed hinderer:

**Failed hinderer.** In failed hinderer events, the puppet ran forward and jumped on the box, slamming it closed while saying “Close!” The puppet then ran away. After another struggle, the protagonist successfully opened the box and laid facedown, grasping the toy while saying “Toy!” Note that unlike in the positive outcome condition of Experiment 1, the failed hinderer jumped on the box once rather than three times; this mirrors the failed hinderer events in Hamlin (2013).
Results

Test questions. A series of mixed-effect ANOVAs explored whether responses differed before and after comprehension questions; this revealed no main effects of round or interactions involving round of questioning on liking, niceness, or trouble scores (all Fs < 5.686, ps > .020, $\eta^2_p$s < .115). Children’s scores were summed across rounds resulting in three scores between 0 – 2 per child (liking, niceness, trouble).

Confirmatory analyses. A series of one-sample t-tests comparing liking, niceness, and trouble scores at each age to a chance score of one revealed that younger children did not distinguish between the puppets: 3-year-olds’ liking ($p = .357$), niceness ($p = .203$), and trouble ($p = .417$) scores did not differ from chance. In contrast, 4-year-olds liked the failed helper ($p = .002$), judged the failed helper as nicer ($p < .001$), and allocated punishment to the failed hinderer ($p < .001$; see Figure 3.3 and Table 3.1).

Figure 3.3. Mean liking, niceness, and trouble scores at each age in Experiment 2A and 2B. Each score ranges between 0 – 2 with higher values indicating higher rates of with-hypothesis responding across two rounds of questioning; *$p < .05$, **$p < .01$, ***$p < .001$, error bars reflect the SE.
Exploratory analyses. A mixed-effect ANOVA was used to examine whether age, gender, and/or question type influenced children’s tendency to respond in the direction of the hypothesis. This revealed only a main effect of age ($F[1,44] = 7.214, p = .010, \eta^2_p = .141$; all other $F$s < 2.449, $ps > .114, \eta^2_p$s < .054), such that 4-year-olds’ overall scores across the three questions types ($M = 5.13, SD = 1.78$) were higher than 3-year-olds’ ($M = 3.50, SD = 2.06$).

Punishment explanations. The most frequent response among 3-year-olds were uninformative (40%), while their most informative responses were appeals to relevant (attempted) helping or hindering actions (29%); nearly all these appeals referenced a hindering action (e.g., “this one tried to close the box”, “because he closed it”), although one 3-year-old referenced a helping action in the second round (i.e., “because he’s trying to open” when explaining why the failed helper should be punished). The most frequent response among 4-year-olds were appeals to relevant (attempted) hindering actions (42%; see Table 3.2). A factorial ANOVA examined the effect of age and gender on the proportion of uninformative responses across rounds. While there was no main effect of age or gender (all $F$s < 2.396, $ps > .128, \eta^2_p$s < .053), there was an interaction between these factors ($F[1,44] = 6.649, p = .013, \eta^2_p = .131$), such that 3-year-olds males ($M = .55, SD = .47$) provided more uninformative explanations than 4-year-old males ($M = .07, SD = .26; t(14.377) = 3.047, p = .008, d = 1.373$), while female 3-year-olds ($M = .27, SD = .39$) and 4-year-olds ($M = .39, SD = .49$) provided the same proportion of uninformative responses ($t[20] = .642, p = .528, d = .292$). Finally, a factorial ANOVA examining the effect of age and gender on the proportion of relevant sociomoral responses (protagonist’s goal, relevant actions, relevant general valence, and relevant skill valence) revealed that 4-year-olds provided more relevant responses than 3-year-olds ($F[1,44] = 4.594, p = .038, \eta^2_p = .095$; all other $F$s < 1.695, $ps > .199, \eta^2_p$s < .038).
Discussion

Overall, Experiment 2A reveals that 4-year-olds, but not 3-year-olds, privilege intentions when making social and moral judgments. When presented with a failed helper and failed hinderer, 4-year-olds preferred the failed helper, judged the failed helper to nicer, and allocated punishment to the failed hinderer. Four-year-olds’ explanations of their punishment allocations most frequently referenced the puppet’s (attempted) hindering action. In contrast, 3-year-olds failed to distinguish between the puppets when asked which puppet was liked, nicer, and should be punished. Given their chance responding to test questions, it is unsurprising that 3-year-olds’ explanations regarding punishment allocations were largely uninformative.

Given young children’s documented struggle to privilege intentions when intentions and outcomes conflict (e.g., Costanzo et al., 1973; Cushman et al., 2013), one possibility is that 3-year-olds simply do not use intentions to inform their sociomoral judgments when individuals can instead be distinguished by outcomes. Although 3-year-olds did not reliably distinguish characters by either intention or outcome, it is possible that they are in a transitional stage. An alternative possibility is that 3-year-olds can privilege intentions, but that the puppet shows in Experiment 2A did not adequately convey this mental state information to them. Specifically, the failed hinderer demonstrated his intention to close the box only once before the protagonist successfully opened it (in contrast, the failed helper demonstrated its intent three times); this may have made the strength of the failed hinderer’s negative intent somewhat ambiguous, rendering the distinction between the characters unclear.

Experiment 2B explored whether children privilege intentions when the failed hinderer’s intentions were made more salient. Children observed a failed helper and failed hinderer intervene in the protagonist’s struggle to open a box. As in Experiment 2A, the failed helper
attempted to open the box three times. Unlike in Experiment 2A, the failed hinderer also
demonstrated his negative intention three times, by repeatedly slamming the box closed. It was
predicted that both 3- and 4-year-olds would report liking the failed helper, judge the failed
helper as nicer, and allocate punishment to the failed hinderer.

**Experiment 2B**

**Method**

**Participants.** Twenty-four 3-year-olds ($M = 3;6$, range = 3;0-3;11, 11 girls) and 25 four-
year-olds ($M = 4;6$, range = 4;0-4;11, 12 girls) participated. The pre-set sample size was 24
children per age per condition; one extra 4-year-old was run due to scheduling issues. An
additional 13 3-year-olds were replaced due to procedure errors (3) and color/side preferences
(10). An additional 5 4-year-olds were replaced due to procedure error (1) and color/side
preferences (4).

**Procedure.** The warm-up task, test questions, comprehension questions (i.e., "Did he try
to open the box or try to close the box? Did the duck get the toy?") transcriptions and coding
procedures were identical to previous experiments.

**Puppet show.** Children watched a live puppet show featuring a protagonist struggling to
open a box; a second and third puppet intervened (failed helper in two events, failed hinderer in
two events). All puppet show details were identical to Experiment 1.

**Results**

**Test questions.** A series of mixed-effect ANOVAs explored whether responses differed
before and after comprehension questions; this revealed no main effect of round on niceness or
trouble scores, and no interactions involving round for liking, niceness, or trouble scores
(Bonferroni-corrected alpha value of .017 [0.050/3]; all $F$s < 4.511, $p$s > .038, $\eta^2$s < .092).
However, liking scores were higher after comprehension questions ($M = .69$, $SD = .47$) versus beforehand ($M = .53$, $SD = .50$; $F[1,45] = 7.420$, $p = .009$, $\eta_p^2 = .142$). Because round of questioning had no effect on liking scores in other experiments and consistently had no effect on niceness or trouble scores, children’s scores were summed across rounds resulting in three scores between 0 – 2 per child (liking, niceness, trouble).

**Confirmatory analyses.** A series of one-sample t-tests comparing liking, niceness, and trouble scores at each age to a chance score of one revealed that children did not prefer either the failed helper or hinderer: 3-year-olds’ liking scores ($p = .479$) and 4-year-olds’ liking scores ($p = .073$) did not differ from chance. However, both ages judged the failed helper to be nicer ($ps < .001$) and allocated punishment to the failed hinderer ($ps < .001$; see Figure 3.3 and Table 3.1).

**Exploratory analyses.** A mixed-effect ANOVA examining whether age, gender, and/or question type influenced children’s tendency to respond in the direction of the hypothesis revealed a main effect of question type ($F[1.208,54.356] = 14.550$, $p < .001$, $\eta_p^2 = .244$; all other $Fs < 3.821$, $ps > .056$, $\eta_p^2s < .079$). To explore this main effect, a series of paired-samples t-tests using the Bonferroni corrected alpha value of .017 (.050/3) were used to compare scores on each question type across age. Children were less likely to respond in the direction of the hypothesis when asked which puppet they liked ($M = 1.22$, $SD = .85$) compared to which puppet was nicer ($M = 1.73$, $SD = .53$; $t[48] = 3.900$, $p < .001$, $d = .557$) and which puppet should get in trouble ($M = 1.73$, $SD = .57$; $t[48] = 4.228$, $p < .001$, $d = .604$); there was no difference between niceness and trouble scores ($t[48] = .000$, $p = 1.000$, $d = .000$).

**Punishment explanations.** When asked to explain why the selected puppet should get in trouble, responses most frequently included appeals to relevant (attempted) helping or hindering actions: 39% of 3-year-olds and 79% of 4-year-olds (see Table 3.2). While these appeals
typically referenced a hindering action (e.g., “because he tried to close the box”), one 4-year-old referenced the failed helper’s action in both rounds (i.e., “because he said open” when explaining why the failed helper should be punished). A factorial ANOVA examined the effect of age and gender on the proportion of uninformative responses across rounds and found that the proportion of uninformative responses was greater among 3-year-olds compared to 4-year-olds ($F[1,45] = 7.438, p = .009, \eta_p^2 = .142$; all other $Fs < 2.129, ps > .151, \eta_p^2s < .046$). Finally, a factorial ANOVA examining the effect of age and gender on the proportion of relevant sociomoral responses (protagonist’s goal, relevant actions, relevant general valence, and relevant skill valence) revealed that 4-year-olds provided more relevant responses than 3-year-olds ($F[1,45] = 11.502, p = .001, \eta_p^2 = .204$; all other $Fs < 3.518, ps > .066, \eta_p^2s < .073$).

**Discussion**

Overall, Experiment 2B demonstrated that 3-year-olds can privilege intention over outcomes when making moral judgments when intentions are clarified (i.e., by having the failed hinderer demonstrate his intention to close the box three times rather than once). While 3-year-olds in Experiment 2A and 2B showed no preference for either the failed helper or failed hinderer, 3-year-olds in Experiment 2B judged the failed helper to be nicer and the failed hinderer to be more deserving of punishment. Four-year-olds in Experiment 2B also showed no preference for the failed helper or failed hinderer (c.f. Experiment 2A), but judged the failed helper as nicer and allocated punishment to the failed hinderer. This pattern suggests that, like in previous experiments, both 3- and 4-year-olds’ moral judgments (niceness, trouble) favor the failed helper more robustly than their social judgments (liking). The most frequent explanation for the allocation of punishment were references to the failed hinderer’s hindering action.
Experiment 3 explored whether children utilize outcomes to make social and moral judgments when characters’ intentions are identical. When intentions are uninformative, sociomoral judgments may favor individuals associated with positive versus negative outcomes, as these individuals may be associated with positive versus negative outcomes again in the future. Alternatively, judgments may favor individuals who are successful in bringing about their intended outcome, whatever it may be. Indeed, previous work has shown a relationship between judgments of competence and judgments of prosociality in children of this age (Landrum, Pflaum, & Mills, 2016; Brosseau-Liard & Birch, 2010; Stipek & Daniels, 1990; but see Fusaro, Corriveau, & Harris, 2011). That said, in a previous study infants tested with similar conditions did not distinguish characters who differed only on outcome (Hamlin, 2013a). 

In Experiment 3, children observed a puppet show featuring a protagonist unsuccessfully attempting to open a box. In the “positive intention” condition, a successful helper and failed helper intervened; both characters had a positive intention but the successful character brought about a positive outcome for the protagonist and the failed character brought about a negative outcome. In the “negative intention” condition, a successful hinderer and a failed hinderer intervened; both puppets had a negative intention but the successful character brought about a negative outcome for the protagonist and the failed character was associated with a positive outcome. Given previous work showing children’s bias towards outcomes when making sociomoral judgments, it was predicted that both 3- and 4-year-olds would prefer characters who caused or were associated with positive outcomes. Thus, in the positive intention condition it was predicted that children would prefer the successful helper, judge the successful helper as nicer, and allocate punishment to the failed helper, and in the negative intention condition it was
predicted children at both ages would prefer the failed hinderer, judge the failed hinderer as nicer, and allocate punishment to the successful hinderer.

**Experiment 3**

**Method**

**Participants.** Twenty-four 3-year-olds ($M = 3;6$, range = 3;0-3;10, 13 girls) and 26 4-year-olds ($M = 4;5$, range = 4;0-4;11, 14 girls) participated in the positive intention condition, while 23 3-year-olds ($M = 3;6$, range = 3;2-3;11, 13 girls) and 27 4-year-olds ($M = 4;6$, range = 4;0-4;11, 14 girls) participated in the negative intention condition. The pre-set sample size was 24 children per age per condition; five additional 4-year-olds were run due to scheduling issues and one child was initially recruited and tested as a 3-year-old but it was later learned that the child was 2-years-old at the time of testing. An additional 32 3-year-olds were tested but replaced due to unwillingness to participate (5), failure to complete an English language warm-up (1), and color/side preferences (26). An additional 17 4-year-olds were tested but replaced due to unwillingness to participate (2), failure to complete an English language warm-up (2), refusal to accept corrections following comprehension questions (1), and color/side preferences (12).

**Procedure.** The warm-up task, test questions, comprehension questions (for successful puppets, “Did he open the box or close the box? Did the duck get the toy?” and for unsuccessful puppets, "Did he try to open the box or try to close the box? Did the duck get the toy?") transcriptions and coding procedures were identical to previous experiments. One 4-year-old in the positive intention condition indicated that neither puppet should get in trouble in round two, while one 3-year-old in the negative intention condition indicated neither puppet was liked or nicer in both rounds and three 4-year-olds indicated neither puppet was nicer in one or both
rounds. These responses were scored as against the hypothesis that children would respond based on outcome.

**Puppet show.** Children watched a live puppet show featuring a protagonist struggling to open a box; a second and third puppet intervened (two successful helper events and two failed helper events in the positive intention condition, two successful hinderer events and two failed helper events in the negative intention condition). All puppet show details were identical to those in Experiment 1.

**Results**

**Test questions.** A series of mixed-effect ANOVAs explored whether responses differed before and after comprehension questions; this revealed no main effect of round or interactions involving round on liking, niceness, or trouble scores within the positive intention (Bonferroni-corrected alpha value of .017 [.05/3]; all $F$s < 2.896, $p$s > .095, $\eta^2$s < .060) or the negative intention condition (all $F$s < 5.220, $p$s > .026, $\eta^2$s < .103). Children’s scores were again summed across rounds resulting in three scores between 0 – 2 per child (liking, niceness, trouble).

**Confirmatory analyses.** A series of one-sample t-tests comparing liking, niceness, and trouble scores at each age to a chance score of one revealed that 3-year-olds in the positive intention condition did not distinguish between the puppets when reporting who they liked ($p = 1.000$) or when allocating punishment ($p = .357$), though they did judge the successful helper to be nicer than the failed helper ($p = .026$). In contrast, 4-year-olds in the positive intention condition liked the successful helper ($p = .005$), judged the successful helper to be nicer ($p = .005$), and allocated punishment to the failed helper ($p = .026$). Children did not differentiate between the puppets for any test questions in the negative intention condition: 3- and 4-year-olds’ liking (3-year-olds: $p = .604$; 4-year-olds: $p = .070$), niceness (3-year-olds: $p = .575$; 4-
year-olds: \( p = .814 \), and trouble scores (3-year-olds: \( p = .604 \); 4-year-olds: \( p = .814, d = .046 \)) did not differ from chance (see Figure 3.4 and Table 3.1).

**Figure 3.4.** Mean liking, niceness, and trouble scores at each age in Experiment 3. Each score ranges between 0 – 2 with higher values indicating higher rates of with-hypothesis responding across two rounds of questioning; *\( p < .05 \), **\( p < .01 \), error bars reflect the SE.

**Exploratory analyses.** Two mixed-effect ANOVAs revealed no effect of age, gender, or question type on children’s tendency to respond in the direction of the hypothesis in the positive or negative intention condition (all \( F_s < 2.967, ps > .071, \eta^2_p s < .062 \)).

**Punishment explanations.** In the positive intention condition, 3-year-olds’ explanations regarding punishment allocation were mostly uninformative (48%). Four-year-olds also provided many uninformative responses (40%), and although neither puppet intended to close the box, 4-year-olds’ appeals to relevant actions most often referenced the puppet’s failure to help (e.g., “because she didn’t open it”, “he didn’t help the duck open the box”; 42%). In the negative intention condition, 3-year-olds’ responses were largely uninformative (37%) or appeals to relevant hindering actions (36%); 4-year-olds most often appealed to relevant hindering actions (34%), but also provided a number of uninformative responses (26%; see Table 3.2). Two
factorial ANOVAs found no effect of age or gender on the proportion of uninformative responses across rounds in either condition (all $F$s < 1.359, $ps > .249$, $\eta_p^2$s < .030) and two factorial ANOVAs found no effect of age or gender on the proportion of relevant responses (protagonist’s goal, relevant actions, relevant general valence, and relevant skill valence) in either condition (all $F$s < 2.761, $ps > .102$, $\eta_p^2$s < .058).

**Discussion**

Overall, Experiment 3 reveals that children’s social and moral judgments are not uniformly based on outcomes when intentions are identical. When positively intentioned characters brought about distinct outcomes, 3-year-olds judged the successful helper as nicer, but did not prefer or allocate punishment to either the failed or successful helper. Given their chance responding when asked which puppet should get in trouble, it is unsurprising that 3-year-olds’ explanations for this judgment were often uninformative. In contrast, 4-year-olds consistently utilized outcomes to inform their sociomoral judgments of positively intentioned characters (i.e., 4-year-olds liked the successful helper, judged the successful helper as nicer, and allocated punishment to the failed helper). Among 4-year-olds’ informative responses, explanations for punishment allocations most often referenced the puppet as having failed to help the protagonist, although neither puppet intended to thwart the protagonist’s goal. In contrast to the positive-intention condition, when negatively-intentioned characters brought about distinct outcomes, both age groups responded at chance levels when asked which puppet was liked, nicer, and should get in trouble. When asked to explain their allocation of punishment, 3-year-olds’ responses were most frequently uninformative, while 4-year-olds most often appealed to the puppet’s attempted or completed hindering action.
General Discussion

Experiments 1–3 provide evidence that 3- and 4-year-olds readily produce sociomoral judgments based on character’s intentions, rather than strictly on the outcomes these characters achieve. After observing live-action puppet shows in which characters’ intentions are fully acted out and the consequences of their actions can be directly observed, preschoolers were asked to provide a social judgment (i.e., which of two puppets is liked) and moral judgments (i.e., which of two puppets was nicer and which should be punished). Children were also asked to verbally justify their allocations of punishment. When characters could only be distinguished based on their intentions because outcomes were uninformative (Experiment 1), both 3- and 4-year-olds’ moral judgments revealed an intention focus, while children’s social judgments were less consistent: 4-year-olds, but not 3-year-olds, liked the successful helper over the failed hinderer, and both ages preferred the failed helper over the successful hinderer. When both intentions and outcomes conflicted in Experiment 2A, 4-year-olds’ social and moral judgments showed an intention focus, while 3-year-olds did not distinguish between the puppets. When the failed hinderer’s intention was further highlighted in Experiment 2B (i.e., the failed hinderer attempted to block the protagonist’s goal three times instead of once, the same number of attempts as the failed helper), children’s moral judgments showed a consistent focus on intention over outcome, though neither 3- nor 4-year-olds consistently preferred one character over the other. Finally, when characters had identical intentions but brought about opposing outcomes in Experiment 3, 4-year-olds’ social and moral judgments showed an outcome focus when comparing two characters with positive intentions, while 3-year-olds judged the successful helper to nicer than the failed helper and responded at chance when judging which character was preferred and which
should receive punishment. Both ages responded at chance in all comparisons involving two negatively-intentioned puppets.

Across all experiments, children’s most frequent informative justifications for their punishment allocation were appeals to the character’s hindering action. This was the case regardless of whether the action was successful or unsuccessful (e.g., children explained that a failed hinderer should get in trouble because he [tried to] block the protagonist’s goal), and whether the character had intended to bring about a negative outcome (e.g., when comparing a failed and successful helper, children explained that the failed helper should get in trouble because he did not allow the protagonist’s goal to be achieved). While it was predicted that 3-year-olds would provide more uninformative responses than 4-year-olds (see Kenward & Dahl, 2011; Van de Vondervoort & Hamlin, 2017) and that 4-year-olds would be more likely than 3-year-olds to provide more relevant sociomoral considerations in their explanations, these predictions were largely unsupported.

These results provide evidence that young children can privilege intention over outcome when making moral judgments. Contrary to evidence suggesting that a focus on intentions develops after the early preschool years (Piaget, 1932/1965; see also Armsby, 1971; Baird & Astington, 2004; Costanzo et al., 1973; Helwig et al., 2001; Killen et al., 2011; Margoni & Surian, 2017; Moran & O’Brien, 1983; Yuill & Perner, 1988; Zelazo et al., 1996), both 3- and 4-year-olds’ forced-choice judgments regarding niceness and the allocation of punishment were based on which character displayed a positive versus negative intention to help or hinder a third party, regardless of the outcome achieved (except for 3-year-olds in Experiment 2A, in which the negatively-intentioned character’s intention may have been unclear). This was the case when the characters being evaluated had opposing intentions but brought about the same outcome, and
when both characters’ intentions and outcomes conflicted. Further, the consistency between children’s responding in the current study and infants’ responses to similar scenarios (Hamlin, 2013a) suggests that sensitivity to others’ intentions develops earlier than previously thought and aligns with claims of continuity in moral competence from infancy (i.e., infants’ spontaneous preferences for well-intentioned individuals) to early childhood (i.e., preschoolers’ elicited judgments; see Margoni & Surian, 2016 for discussion).

Surprisingly, when outcomes were the only way to distinguish between characters, children did not consistently show an outcome bias: Although 4-year-olds liked a successful helper over a failed helper, judged the successful helper to be nicer than the failed helper, and allocated punishment to the failed helper, they did not distinguish between a successful and failed hinderer on any test questions. Three-year-olds fared even worse, and responded above chance in the outcome conditions only when asked whether the successful versus failed helper was nicer. These results are surprising, both given past work suggestive of an outcome bias in this age group and because children could have alternatively distinguished between the characters based on which character successfully brought about their intended outcome (see Landrum et al., 2016; Brosseau-Liard & Birch, 2010; Stipek & Daniels, 1990 for evidence that young children’s judgments of competence and prosociality are related); however (aside from some positive evaluation of the successful versus failed helper), children showed no consistent evidence of either strategy.

One potential concern is that 3-year-olds’ inability to distinguish between the failed and successful helpers when allocating punishment and both ages’ consistent failure to distinguish between the failed and successful hinderers may be due to the moral judgment questions asked. For instance, it is potentially unclear how to respond when asked which of two characters with
positive intentions should get in trouble, or which of two characters with negative intentions is nicer. This ambiguity may have resulted in the observed chance-level responses in Experiment 3. However, it is important to note that these questions are only unclear if children are evaluating the puppets based on their intentions. That is, if children were evaluating characters in terms of outcomes brought about, there would have been a clear answer to which of the two positively-intentioned puppets was nicer (i.e., the successful helper) and to which of the two negatively-intentioned puppets should get in trouble (i.e., the successful hinderer). Thus, it seems clear that children were not uniformly utilizing an outcome bias to answer these moral judgment questions, even when this was the only way that they might have distinguished between the characters.

Another potential concern is that the chance-level responding among 3-year-olds allocating punishment in the positive intention condition and among both 3- and 4-year-olds in the negative outcome condition of Experiment 3 is due to differences in how the same intentions were displayed across characters. Specifically, in the positive intention condition, the successful helper enacted his positive intention once (at which point he successfully aids the protagonist in opening the box) while the failed helper enacted his positive intention three times (i.e., by repeatedly struggling with the protagonist to try and open the box). Positive evaluations of both the successful helper’s positive outcome and the failed helper’s repeated well-intentioned efforts may have resulted chance-level responding among 3-year-olds; 4-year-olds’ judgments favored the successful helper despite this concern. In the negative intention condition, the successful hinderer enacted his negative intention once before the protagonist’s goal is thwarted while the failed hinderer enacted his negative intention three times (i.e., by repeatedly slamming the box shut before the protagonist was eventually able to achieve his goal). Negative evaluations of both the outcome of the successful hinderer’s action and the failed hinderer’s repeated negative-
intentioned efforts may have resulted in chance-level responding among 3- and 4-year-olds when comparing these two characters. It is possible that equating the intention displays (i.e., the successful helper tries to open the box three times before being successful, the successful hinderer slams the box closed three times before the protagonist fails to achieve his goal) would result in a consistent outcome bias when the characters’ intentions are equivalent. This possibility should be explored in future studies utilizing live-action puppet shows.

Regardless of children’s judgments when intentions are equivalent, the current studies show that 3- and 4-year-olds can use intentions to form moral judgements when outcomes are equivalent (Experiment 1) and can privilege intentions over outcomes when the two conflict (Experiment 2B). What accounts for children’s ability to privilege intentions following a live puppet show, compared to previous studies utilizing illustrated stories, in which young children initially fail to privilege intentions, especially when intentions and outcomes conflict (e.g., Cushman et al., 2013; Killen et al., 2011)? One possibility is that puppet shows allow characters’ intentions to be fully acted out, making intentions more salient than when explained during a vignette (even when intentions are explicitly stated). Likewise, processing demands may be reduced when children can observe the events unfold, rather than needing to infer what happened between images illustrating a vignette-based task. Finally, the pragmatic demands of puppet show-based tasks versus vignette-based tasks may account for differences in young children’s responding. For example, forced-choice comparisons between two puppets (e.g., asking which of two puppets is nicer) may allow children to distinguish between actors in a way that cannot be observed when children are instead asked to evaluate each character independently (e.g., asking whether each puppet is nice; see Hilton & Kuhlmeier, 2019 for further discussion of how differences in stimuli presentation and test questions can alter the saliency of intention versus
outcome information and influence young children’s production of intention-based judgments). Further, children’s pragmatic reasoning about the experimenter’s own intentions may lead them to focus on outcomes following vignettes if caregivers are more likely to use stories rather than pretend play to explain norms of behavior (e.g., stories depicting punishment for harms caused, regardless of the character’s intentions; see Westra & Carruthers, 2017 for a discussion of how children’s pragmatic reasoning may influence their performance on false-belief tasks). Future studies should probe these possibilities by directly comparing children’s judgments following vignette and puppet show versions of the same scenarios.

While the current studies provide evidence that children’s moral judgments are intention-based, 3- and 4-year-olds’ social judgments less consistently showed an intention-bias. Specifically, exploratory analyses revealed an effect of question type in the positive and negative outcome condition of Experiment 1 and in Experiment 2B, such that children were less likely to respond in the direction of the hypothesis (i.e., that children would favor the character with positive intentions over the character with negative intentions) when asked which character they liked, as opposed to when making moral judgments about niceness and punishment. These analyses suggest that the puppet show events were more consistently viewed as morally relevant as opposed to socially relevant, and that idiosyncratic preferences (e.g., preferences based on the puppet’s appearance) may have influenced children’s social judgments more than their moral judgments. Children in the current studies were only asked to explain one judgment to prevent contamination between explanations regarding the allocation of punishment and explanations regarding social preferences. That said, future studies should explore whether children’s social preferences are justified by appeals to the puppets’ helpful intention or by appeals to other aspects of the puppets or the events within the puppet show.
Finally, there are several remaining open questions regarding the developmental trajectory of sensitivity to others’ intentions. First, it is currently unknown whether infants’ implicit preferences for characters with positive intentions over characters with negative intentions, regardless of outcome (Hamlin, 2013a), are related to preschoolers’ explicit sociomoral judgments following similar puppet show displays. While it is possible that infants’ implicit preferences and young children’s explicit judgments are distinct, it may be that sociomoral functioning in infancy is related to explicit moral development later in life. Relatedly, more work is needed to accurately characterize the use of intentions in moral judgments across the lifespan. This could be accomplished by utilizing the same stimuli to examine intention-based judgments in infants, preschoolers, older children, and adults. While the current studies adapted live puppet show stimuli previously shown to infants, practical concerns restricted our sample to the preschool years, rather than the broader age range necessary to make strong conclusions regarding the continuity of intention sensitivity across the lifespan. Lastly, it is also an open question whether an early focus on intentions is universal, and if so, how this develops into adult-like moral responses across a variety of cultures. Given variability in the extent to which adults from small-scale, non-Western societies incorporate intentionality in moral judgments (Barrett et al., 2016), it is possible that early moral judgments differ along important dimensions, or that infants and young children in both Western and non-Western societies share an early sensitivity to intentions that is refined according to their culture. Exploring the development of implicit evaluations and explicit judgments within and across diverse individuals over time would greatly contribute to our understanding of how intention and outcome information becomes integrated in mature moral judgments.
Table 3.1

Descriptive and test statistics for confirmatory analyses t-tests

<table>
<thead>
<tr>
<th>Experiment Condition</th>
<th>1</th>
<th>2A</th>
<th>2B</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive outcome</td>
<td>Negative outcome</td>
<td>Positive intention</td>
<td>Negative intention</td>
</tr>
<tr>
<td>df</td>
<td>23</td>
<td>25</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>M</td>
<td>1.25</td>
<td>1.35</td>
<td>1.17</td>
<td>1.79</td>
</tr>
<tr>
<td>SD</td>
<td>0.79</td>
<td>0.85</td>
<td>0.87</td>
<td>0.59</td>
</tr>
<tr>
<td>t</td>
<td>1.543</td>
<td>2.632</td>
<td>0.941</td>
<td>0.00</td>
</tr>
<tr>
<td>d</td>
<td>0.315</td>
<td>0.409</td>
<td>0.641</td>
<td>0.00</td>
</tr>
<tr>
<td>df</td>
<td>23</td>
<td>25</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>M</td>
<td>1.75</td>
<td>1.73</td>
<td>1.21</td>
<td>1.58</td>
</tr>
<tr>
<td>SD</td>
<td>0.53</td>
<td>0.53</td>
<td>0.78</td>
<td>0.58</td>
</tr>
<tr>
<td>t</td>
<td>6.912</td>
<td>6.171</td>
<td>1.310</td>
<td>0.00</td>
</tr>
<tr>
<td>d</td>
<td>1.411</td>
<td>1.210</td>
<td>0.627</td>
<td>0.00</td>
</tr>
<tr>
<td>df</td>
<td>23</td>
<td>25</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>M</td>
<td>1.71</td>
<td>1.62</td>
<td>1.25</td>
<td>1.58</td>
</tr>
<tr>
<td>SD</td>
<td>0.62</td>
<td>0.70</td>
<td>0.74</td>
<td>0.65</td>
</tr>
<tr>
<td>t</td>
<td>5.560</td>
<td>4.500</td>
<td>0.827</td>
<td>0.00</td>
</tr>
<tr>
<td>d</td>
<td>1.135</td>
<td>0.883</td>
<td>1.135</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. Mean scores range between 0 – 2 with higher values indicating higher rates of with-hypothesis responding across two rounds of questioning.
Table 3.2

Proportions of explanations containing each response type in Experiments 1 – 3

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Age</th>
<th>Uninformative Responses</th>
<th>Informative Responses</th>
<th>Relevant Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Protagonist’s goal</td>
<td>Relevant action</td>
</tr>
<tr>
<td>1 (positive outcome condition)</td>
<td>3</td>
<td>.31</td>
<td>.03</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.13</td>
<td>.02</td>
<td>.65</td>
</tr>
<tr>
<td>1 (negative outcome condition)</td>
<td>3</td>
<td>.08</td>
<td>0</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.19</td>
<td>.02</td>
<td>.59</td>
</tr>
<tr>
<td>2A (opposing outcomes/intentions)</td>
<td>3</td>
<td>.40</td>
<td>0</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.19</td>
<td>.03</td>
<td>.42</td>
</tr>
<tr>
<td>2B (opposing outcomes/intentions)</td>
<td>3</td>
<td>.27</td>
<td>0</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.04</td>
<td>.05</td>
<td>.79</td>
</tr>
<tr>
<td>3 (positive intention condition)</td>
<td>3</td>
<td>.48</td>
<td>0</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.40</td>
<td>0</td>
<td>.42</td>
</tr>
<tr>
<td>3 (negative intention condition)</td>
<td>3</td>
<td>.37</td>
<td>.02</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.26</td>
<td>.02</td>
<td>.34</td>
</tr>
</tbody>
</table>

Note. Proportion. The informative response categories were not mutually exclusive and proportions sum to 1. Relevant responses include appeals to the protagonist’s goal, relevant actions, relevant general valence, and relevant skill valence.
Chapter 4: Children’s Consideration of Contextual Information

Introduction

While intentionally helping others is typically considered prosocial and intentionally hindering others is typically considered antisocial, this is not always the case. Instead, there are contexts in which helpful actions are evaluated negatively (e.g., helping someone achieve a harmful goal, rewarding someone’s wrongdoing) and unhelpful actions are evaluated positively (e.g., blocking someone’s antisocial goal, appropriately punishing someone’s wrongdoing). It is necessary for mature moral thinkers to consider the context in which helpful and unhelpful actions occur, and use this contextual information to form their moral judgments.

Chapter 4 of this dissertation explores young children’s judgments regarding helpful versus unhelpful actions directed towards previously prosocial individuals (who presumably deserve reward) versus antisocial individuals (who presumably deserve punishment). We have chosen to explore context-dependent judgments regarding the treatment of previously prosocial versus antisocial individuals following arguments that this ability was critical to the evolution and continued functioning of human’s cooperative groups, and mature moral thinkers’ (i.e., adults’) production of such context-dependent judgments.

More specifically, direct reciprocity involves individuals selectivity interacting with those who have helped them and avoiding those who have harmed them. These selective responses to previously prosocial versus antisocial others can encourage future cooperative acts and discourage uncooperative ones (Trivers, 1971; Axelrod & Hamilton, 1981). That said, cooperation amongst large groups individuals cannot likely be explained by direct reciprocity alone (Boyd & Richerson, 1998). Instead, large-scale cooperative groups are made possible by individuals who are also concerned with how others are treated. These individuals maintain
cooperative groups via indirect reciprocity (i.e., selectively rewarding those who have benefited third parties without direct material benefits; Alexander, 1987; Nowak & Sigmund, 1998; 2005) and altruistic or third-party punishment (i.e., punishing those who have wronged third parties, even when doing so is costly and when the initial transgression did not affect the punisher; Fehr & Fischbacher, 2004; Fehr & Gächter, 2002; Yamagishi, 1986). Adults find engagement in such altruistic punishment personally rewarding, as evidenced by the activation of neurological reward systems when allocating punishment to individuals who fail to cooperate in an economic exchange (de Quervain et al., 2004; see also Singer et al., 2006).

In addition to engagement in direct reciprocity and altruistic punishment, adults positively evaluate others who act in accordance with these principles. For example, in the context of laboratory-based economic games, adults support those who engage in indirect reciprocity (that is, reward those who reward cooperators; Kiyonari & Barclay, 2008). While in this study adults did not approve of altruistic punishers, other work has demonstrated that adults judge those who punish non-cooperators as more trustworthy and worthy of respect than non-punishers (Barclay, 2006; see also Nelissen, 2008 for evidence that the trustworthiness of altruistic punishers increases with the costliness of the punishment). Likewise, a more recent vignette-based study found that third-party punishers were judged as more likeable compared to those who failed to intervene following the initial transgression (Gordon, Madden, & Lea, 2014). Finally, there is even evidence that adults punish individuals who observe third-party transgressions and fail to enact punishment themselves (Martin, Jordan, Rand, & Cushman, 2019), suggestive that engagement in altruistic punishment is not only preferred but considered normative (i.e., an action that one ought to perform).
Broadly consistent with claims that indirect reciprocity and altruistic punishment are central to the evolution of human cooperation, are findings that even human infants detect and evaluate social interactions featuring these principles. For example, in one study 10-month-olds looked longer toward an event in which an individual rewarded someone who previously distributed resources unfairly rather than fairly, suggestive that infants expected indirect reciprocity in this situation and were surprised when the individual rewarded the unfair distributor. Further, these expectations depended on the individual having witnessed the initial distribution, and thus knowing who had behaved fairly/unfairly (Meristo & Surian, 2013).

Related studies show that 10-month-olds look longer when a character who observed fair versus unfair distributions among third parties later approaches the fair rather than unfair distributor (Geraci & Surian, 2011) and look longer when antisocial actions (hitting, taking) are directed toward previously unfair versus fair distributors (Meristo & Surian, 2014). Of note, infants looking times in these latter studies do not suggest they expected fair distributors to be treated well and unfair distributors to be treated poorly, since infants looked longer toward events consistent with indirect reciprocity and third-party punishment rather than events that violated these principles. That said, these studies do suggest that infants distinguish/detect violations of indirect reciprocity and third-party punishment across a variety of situations.

Finally, a series of studies has demonstrated that infants preferences for helpful rather than unhelpful others depends on the context in which helping and hindering actions occur. In these studies, infants observed a live puppet show featuring a character who unsuccessfully attempted to open the lid on a box to reach an attractive toy. A prosocial puppet helped the character open the box and an antisocial puppet slammed the lid closed. Infants then saw a second scenario in which either the previously prosocial or antisocial puppet played with and
dropped its ball. Across two events, the ball was returned by a helper puppet and taken away by a hinderer puppet. Infants were then presented with the helper and hinderer and their preference was recorded as the first puppet contacted via a visually-guided reach. While 5-month-olds preferred helpers across contexts, 8- and 19-month-olds selectively reached for puppets that helped prosocial others and puppets that hindered antisocial others (Hamlin et al., 2011). When younger infants were habituated to the initial prosocial/antisocial puppet events (rather than seeing each event only once), 4.5-month-olds also demonstrated context-dependent preferences (Hamlin, 2014a), suggestive that younger infants’ indiscriminate preferences for helpers in past work was due to limited memory and processing capacities (e.g., forgetting whether the target puppet was previously prosocial or antisocial). Together, these results are suggestive that infants and toddlers positively evaluate those who engage in indirect reciprocity and those who engage in appropriate third-party punishment.

While the studies described above show that infants are capable of context-dependent evaluations, it is unclear whether these evaluations are based on moral concerns (i.e., an impartial sense that others should be appropriately rewarded or punished) or non-moral social concerns. For example, one social explanation of these findings is that infants prefer those who share their (negative) evaluation of the antisocial individual who slammed the lid of the box closed. This shared evaluation might breed mutual liking and affiliation between the infant and the “punisher” (see Heider, 1958), resulting in their preference for the puppet who took the ball away. While the complexity of such a social judgment is impressive, such a preference does not necessarily speak to the development of mature moral judgments.

Unfortunately, infants’ linguistic limitations make it difficult to further probe their intuitions about these situations: Infants lack the ability to understand and respond to different
types of questions (e.g., questions about who is liked/preferred versus who performed a morally acceptable actions) and cannot verbally explain their preferences in behavioral tasks. In contrast, it is possible to determine whether linguistically-competent children form social and/or moral judgments about helpers and hinderers based on the context in which prosocial/antisocial actions are performed. This makes the study of young children’s judgments regarding such scenarios an important area of study for those interested in the development of mature moral judgments.

That said, just one published paper has addressed this question. Li and Tomasello (2018) presented 3- and 5-year-olds with video-recorded puppet show events in which individuals successfully or unsuccessfully attempted to help or harm a previously prosocial or antisocial character. More specifically, children first saw either a prosocial monkey puppet who shared a carrot with a third party or an antisocial monkey puppet who hit a third party and stole its ball. Children then saw the previously prosocial or antisocial monkey attempt to open the lid on a box of toys. Across four events, the prosocial/antisocial monkey interacted with a successful helper (successfully enacted positive intention to open the box, resulting in a positive outcome for monkey), failed helper (unsuccessfully attempted to enact positive intention, resulting in negative outcome for monkey), successful hinderer (successfully enacted negative intention to keep lid closed, resulting in negative outcome for monkey), and failed hinderer (unsuccessfully attempted to enact negative intention, resulting in positive outcome for monkey). In all four failed/successful helper/hinderer events, a bystander puppet observed the interactions.

After each event children were asked to make a social judgment (“who do you like more [failed/successful helper/hinderer or bystander]?”), a moral judgment (rating the acceptability of the failed/successful helper/hinderer on a 5-point Likert scale from big frown [completely right] to big smile [completely right]), and to distribute stickers between the failed/successful
helper/hinderer and bystander. It was found that the helper/hinderer’s intentions influenced children’s social judgments, moral judgments, and sticker distributions; 5-year-olds were more influenced by intention information than 3-year-olds. Notably, the monkey puppet’s previous prosocial versus antisocial behavior only influenced 5-year-olds’ social judgments (i.e., their liking of helper/hinderers compared to the bystander puppet); this factor had no influence on children’s moral judgments or sticker distributions (Li & Tomasello, 2018).

The literature reviewed thus far suggests that infants are influenced by an individual’s past prosocial/antisocial actions when producing implicit evaluations (Hamlin, 2014a; Hamlin et al., 2011), 3-year-olds are insensitive to this information, and 5-year-olds produce context-dependent social (but not moral) judgments in an explicit response task (Li & Tomasello, 2018). That said, the memory demands of the task utilized by Li and Tomasello (2018) may have contributed to children’s limited use of context in their social and judgments. Specifically, in this study children viewed the prosocial/antisocial monkey’s actions and each of the failed/successful helper/hinderer’s actions only once, children evaluated four different characters three different ways across the task, and no comprehension questions or memory checks were included. Just as previous work found that the presentation of multiple puppet scenarios taxed younger infants’ memory and processing capabilities (Hamlin, 2014a; Hamlin et al., 2011), it is possible that the number of test questions and the within-subjects nature of this task may have taxed children’s memory and processing capacities. Consequently, 3-year-olds may have simply forgotten whether the monkey puppet was previously prosocial/antisocial, and 5-year-olds may have only been able to incorporate this information into a simple forced-choice questions (i.e., point to which of two puppets is liked) but not more complex tasks such as rating the acceptability of an action or allocating resources. Overall, it remains unclear whether preschoolers can consider
context (i.e., previously prosocial versus antisocial actions) when making moral judgments, and whether context can influence the social preferences of children under 5 years.

The Current Studies

The current studies explore whether 3- and 4-year-olds’ social and moral judgments of helpful and unhelpful characters differ depending on whether these actions are directed towards a previously prosocial or antisocial individual. Scenarios were animated and presented on a laptop computer. In both studies, children were asked to make three forced-choice judgments: (1) which of the two characters (helper, hinderer) they “like”, (2) which they would want to be “friends with”, and (3) which character “did the right thing.” After identifying which character did the right thing, children were asked to explain this judgment and explain why the unselected character behaved as it did. Finally, the children were asked to rate the acceptability of the helper and hinderer’s action on a 5-point Likert scale ranging from “really bad” to “really good”.

Of note, the forced-choice questions in the current studies differ from those in Chapter 2 and 3 of this dissertation, in which children were asked which of two characters was liked, which character was “nicer”, and which character “should get in trouble” (Van de Vondervoort & Hamlin, 2017; 2018b). Given the readiness with which preschoolers identify niceness with prosocial intentions/actions and trouble with antisocial actions/intentions (see Chapter 2 and 3 of this dissertation), we reasoned that children in the current studies might identify the most recently observed helper as nicer and allocate punishment to the most recent hinderer, regardless of whether these actions targeted a previously prosocial or antisocial individual. Thus, in an attempt to explore children’s more global evaluations of the social interactions, based on the combination of all events displayed, we decided to ask an additional social judgment question (i.e., who do you want to be “friends with”) and a different moral judgment question (i.e., who
“did the right thing”). We reasoned that these questions may be less likely to direct children to answer based only on the helping/hindering just observed, and instead potentially answer based on their evaluations of the helper and hinderer in context. Further, previous work has shown that adults view altruistic punishers as most trustworthy and worthy of respect than non-punishers, but that altruistic punishers are not seen as nicer (Barclay, 2006). Finally, we also asked children to evaluate the helper and hinderer’s actions on a 5-point Likert scale in the current studies. We reasoned that scale-based ratings might be more sensitive to small differences in judgments across prosocial and antisocial contexts (e.g., if children evaluate those who help prosocial others as very good, while those who help antisocial others are a little bit good).

In Experiment 1, preschoolers observed either a prosocial individual who helped a third party complete its block tower, or an antisocial individual who broke the tower. This prosocial/antisocial individual then played with and dropped a ball, which was returned by a helper and taken away by a hinderer. It was predicted that 3- and 4-year-olds’ social and moral judgments would favor the helper over the hinderer when these characters interacted with a previously prosocial individual (i.e., children would selectively like the helper, want to be friends with the helper, report that the helper did the right thing, rate the helper’s action positively and rate the hinderer’s action negatively). In contrast, it was predicted that children would be less likely to favor the helper over the hinderer when these characters interacted with a previously antisocial individual. These predictions were based on previous work showing that even infants make implicit context-dependent evaluations of helpers and hinderers in similarly-structured interactions (Hamlin, 2014a; Hamlin et al., 2011) and based on young children’s sensitivity to other types of contextual information when making explicit social and moral judgments (e.g., whether individuals intended to help versus harm; see Chapter 3 of this dissertation). That said,
recent work has also suggested that children under 5 years may not consider this type of contextual information (see Li & Tomasello, 2018 for evidence that 3-year-olds’ judgments do not depend on the previous prosocial/antisocial actions of the individual being helped/hindered).

**Experiment 1**

**Method**

**Participants.** Before data collection began, we established and pre-registered a pre-set stopping rule of 24 children per age per condition (see osf.io/yfquz). Twenty-four 3-year-olds (M age = 3;7, range = 3;1-3;11, 13 girls) and 24 4-year-olds (M age = 4;6, range = 4;0-4;10, 14 girls) participated in the prosocial target condition, while 24 3-year-olds (M age = 3;7, range = 3;2-3;11, 11 girls) and 24 4-year-olds (M age = 4;7, range = 4;1-4;11, 16 girls) participated in the antisocial target condition. Seven additional 3-year-olds and 5 4-year-olds were seen and replaced in the final sample; 8 children were unable/unwilling to complete the scale familiarization task, 2 children refused to participate/answer questions, and there were 2 procedural errors.

**Procedure. Scale familiarization task.** Children were first familiarized with a printed five-point Likert scale used to rate the acceptability of actions (see Figure 4.1; adapted from Smetana, Ball, Jambon, & Yoo, 2018). The scale items ranged from “really bad” (large red thumbs down) to “really good” (large green thumbs up), with mid-sized thumbs up/down representing intermediate values and a blank image representing “just okay”. Children practiced using the scale to rate the acceptability of different actions and were then required to correctly identify each point on the scale; children who were unwilling or unable to complete the scale familiarization task were replaced in the final sample.
Figure 4.1. Likert scale used to indicate the acceptability of an action.

**Animated social interactions.** Children were randomly assigned to either the prosocial target condition or the antisocial target condition. In both conditions children watched animated social interactions presented on a laptop computer; all interactions featured characters with eyes that engaged in self-propelled motion. Depending on condition children watched either the prosocial tower event, in which an orange star helped a green circle complete a block tower, or the antisocial tower event, in which the orange star knocked the circle’s tower down. Following the prosocial/antisocial event, children saw the orange star and were asked two comprehension questions (i.e., “Did the orange star help build the tower or break the tower? Was that a good thing to do or a bad thing to do?). If children answered either question incorrectly, they watched the event again. If children answered incorrectly a second time, they were corrected (e.g., “Actually he broke the tower. That orange star knocked the tower down.”). To encourage recall, the experimenter agreed with children’s correct response to the comprehension questions before the prosocial/antisocial event was repeated (e.g., “That’s right! He did break the tower and that was a bad thing to do. Let’s see that bad thing one more time.”).

All children then watched the helpful and unhelpful ball events, in which the orange star character (previously prosocial or antisocial, depending on condition) bounced a soccer ball three times before accidentally dropping it. In the helpful event the ball was returned by a square character, while in the unhelpful event the ball was taken away by a square character (blue
square and yellow square; identity and side counterbalanced). Children watched the helpful and unhelpful events once each (order counterbalanced), and at the start of each event they were asked to recall the previous action of the orange star (e.g., “Let’s see what happens next. But first, did that orange star help build or break the tower?”). If children answered incorrectly, they were corrected before the animation continued. Following the helpful and unhelpful events, children saw the blue square and yellow square and were asked two comprehension questions (i.e., “Who gave the ball back? Who took the ball away?”). As above, if children answered either question incorrectly the animation and relevant question(s) were repeated, and then the child was corrected if necessary. See the Open Science Framework (OSF) registration component for examples of the animated social interactions for both experiments (osf.io/4ucgd).

**Test questions.** Following the helpful/unhelpful event comprehension questions, children continued to see the blue and yellow square characters (ball event helper and hinderer) and made three forced-choice judgments: *Which one of these guys do you like the most? Which one would you be friends with? Which one of these guys did the right thing?* The order of the first two questions (liking, friendship) was counterbalanced and to reduce response perseveration, children were asked to point to each character before the final forced-choice question (e.g., “Point to the blue one. Point to the yellow one.”). After the forced-choice questions, children were asked to explain why the selected character did the right thing and why the unselected character did not (e.g., if the child selects the helper, “Why did he do the right thing? Why did he give the ball back? What about this one? Why did he do that? Why did he take the ball away?”). Finally, children watched the helpful and unhelpful events once more; at the start of each event they were asked to recall the previous actions of the orange square and after each event they were asked to rate the acceptability of the helper/hinderer’s action on the Likert-scale (e.g., “What do you
think about what the blue square did? Point to the picture”). Across both ages, 53% of children answered all eight comprehension questions correctly, while 38% needed at least one question repeated; only 9% of children were ever corrected on a comprehension question after repetition.

Transcription and coding procedure. When permitted by children’s caregivers and feasible within the daycare setting (e.g., when in a private room without non-participating children in the background of the video), participation in the study was audio and visually recorded. A research assistant then transcribed children’s explanations regarding why the selected puppet did the right thing and why the unselected puppet behaved as it did. When recording did not occur, children’s explanations were transcribed during the study by the experimenter. Two independent research assistants then coded the content of children’s explanations according to the categories detailed below:

Uninformative responses. Uninformative responses included refusals to provide an explanation, unintelligible responses, responses unrelated to the animated social interactions (e.g., “I was at Tim Hortons”), and statements that the child was unsure (e.g., “I don’t know”).

Informative responses. Informative responses were related to the animated social interactions and took one of the following forms:

Moral actions and motivations. Child explained why the character did (not) do the right thing via appeals to morally-relevant actions or attributions. Morally-relevant actions were the giving/taking acts from the ball events (e.g., “he gave the ball back”, “he took it away”), while morally-relevant attributions were indications that character was prosocial/cooperative (e.g., nice, good, friendly) or antisocial/uncooperative (e.g., mean, bad, naughty, bully).

---

9 For our purposes, “being friendly” was coded as a moral rather than social motivation. We felt this was most appropriate given children’s and teacher’s use of “being a friend” to describe helpful, cooperative, appropriate behavior in the classroom and our goal of capturing when children explain the character’s actions via appeals to its prosocial/cooperative behavior.
Social actions and motivations. Child explained why the character did (not) do the right thing without reference to the giving/taking acts from the ball events (e.g., “he’s not playing”, “because he wanted to”, “he liked the ball”) and/or by appealing to a social-but-not-moral motivation (e.g., happy, silly, funny, mad, angry). For each statement in this category, coders also considered whether the responses could potentially indicate that the helper/hinderer’s action was motivated by a Moral Emotion; that is, motivated by positive or negative reactions to the prosocial/antisocial character’s actions in the tower event (e.g., happiness in the prosocial target condition; anger or sadness in the antisocial target condition). Of note, since participants rarely specified the basis for a character’s emotions, it is purely speculative that any of these statements reflect the attribution of moral emotions. Further, anytime a child’s explanation was inconsistent with the attribution of an appropriate moral emotion (e.g., saying the helper in prosocial target condition was angry or giving a non-moral reason for the emotion, such as “he was mad he didn't get ball first”) this supplementary code was not applied.

Target’s prosocial/antisocial action. Child explained why the character did (not) do the right thing by referencing the previously prosocial/antisocial actions performed by the target of helping/hindering (e.g., “because the star was nice”, “the star breaked it”).

Each child’s explanations were coded by two independent research assistants for the presence or absence of each response type. Informative response types were not mutually exclusive. Reliability across the five categories was strong (average Cohen’s kappa = .829, range = .603 to 1.000; see McHugh, 2012). Disagreements were resolved by discussion among the two coders, in the presence of the first author.
Results

Forced-choice judgments. Confirmatory analyses. A series of binomial tests were used to determine at what age(s) liking, friendship, and rightness scores differed from chance (see Figure 4.2). We predicted that children’s judgments would favor the helper in the prosocial target but not necessarily in the antisocial target condition; see OSF registration for documented predictions and analysis plan (osf.io/yfquz). Contrary to these predictions, children showed the same pattern of responding across conditions. In both conditions, 3-year-olds selected the helper/hinderer at chance levels when asked about liking (prosocial target: 17/24 selected helper, $p = .064$; antisocial target: 11/24 selected helper, $p = .839$) and friendship (prosocial target: 16/24 selected helper, $p = .152$; antisocial target: 13/24 selected helper, $p = .839$), but indicated the helper did the right thing at rates exceeding chance (prosocial target: 20/24 selected helper, $p = .002$; antisocial target: 18/24 selected helper, $p = .023$). Conversely, in both conditions, 4-year-olds selected the helper/hinderer at chance levels when asked about friendship (prosocial target: 15/24 selected helper, $p = .307$; antisocial target: 16/24 selected helper, $p = .152$), but liked the helper (prosocial target: 20/24 selected helper, $p = .002$; antisocial target: 18/24 selected helper, $p = .023$) and identified the helper as having done the right thing (prosocial target: 22/24 selected helper, $p < 0.001$; antisocial target: 23/24 selected helper, $p < 0.001$). The dataset analyzed for all experiments in Chapter 4 is available on the OSF (osf.io/jm85z).

---

10 For both Experiment 1 and 2, the OSF registration document states that a series of one-samples t-tests will be used to determine at what ages children’s liking, friendship, and rightness scores differ from chance responding. However, for both experiments, a series of binomial tests is reported. This is the appropriate choice given that the data is binary and not continuous.
Figure 4.2. Percent of children who selected the helper over the hinderer when asked about liking, friendship, and rightness at each age in Experiment 1; *p < .05, **p < .01, ***p < .001.

To determine whether judgments in each age group (3-year-olds, 4-year-olds) differed depending on condition (prosocial target, antisocial target), children were scored 1 if they selected the helper and 0 if they selected the hinderer for each forced-choice questions (liking, friendship, rightness); these scores were then summed, resulting in a score between 0 – 3 for each child. Contrary to our prediction that scores would be higher in the prosocial versus antisocial condition, neither 3-year-olds’ scores (prosocial target: $M = 2.21, SD = .66$; antisocial target: $M = 1.75, SD = 1.03$; independent-samples t-test, $t(39.046) = .074, p = .693, d = .541$), nor 4-year-olds’ scores differed across condition (prosocial target: $M = 2.38, SD = .82$; antisocial target: $M = 2.38, SD = .875$; independent-samples t-test, $t(46) = .000, p = 1.000, d = .006$).

**Exploratory analyses.** To explore whether age, gender, and/or question type influenced children’s responses, two mixed-effect ANOVAs were conducted with question type (liking, friendship, rightness) as a within-subjects variable and child’s age (3, 4) and child’s gender (female, male) as between-subjects factors. These analyses revealed a main effect of question type in both the prosocial target condition ($F[2,88] = 3.340, p = .040, \eta^2 = .071$) and the
antisocial target condition ($F[2,88] = 7.861, p = .001, \eta_p^2 = .152$); there were no other significant main effects or interactions in the prosocial target (all $Fs < .833, all ps > .437$) or antisocial target condition (all $Fs < 3.107, all ps > .084$).

To explore the main effect of question type, three paired-samples t-tests using the Bonferroni corrected alpha value of .017 (.050/3) were used to compare each question type across age within each condition separately. In the prosocial target condition, children were more likely to select the helper when asked about rightness ($M = .88, SD = .334$) compared to friendship ($M = .65, SD = .48$; paired-samples t-test, $t(47) = 3.081, p = .003, d = .445$). Children were equally likely to select the helper when asked about liking ($M = .77, SD = .43$) and friendship (paired-samples t-test, $t(47) = 1.288, p = .204, d = .186$) and when asked about liking and rightness (paired-samples t-test, $t(47) = 1.300, p = .200, d = .188$). In the antisocial target condition, children were more likely to select the helper when asked about rightness ($M = .85, SD = .36$) compared to either social judgment question; that is, when compared to liking ($M = .60, SD = .49$; paired-samples t-test, $t[47] = 3.580, p = .001, d = .518$) or compared to friendship ($M = .60, SD = .49$; paired-samples t-test, $t[47] = 3.293, p = .002, d = .475$). Children were equally likely to select the helper when asked about liking and friendship (paired-samples t-test, $t[47] = .000, p = 1.000, d = .000$).

**Action acceptability ratings.** Children’s Likert scale ratings of the helper/hinderer’s actions were scored from 1 (really bad) to 5 (really good; see Figure 4.3). As predicted, in the prosocial target condition, children rated the helper’s action (3-year-olds: $M = 4.50, SD = .978$; 4-year-olds: $M = 4.71, SD = .81$) more positively than the hinderer’s action (3-year-olds: $M = 1.42, SD = .58$; 4-year-olds: $M = 1.38, SD = 1.01$; Wilcoxon signed-rank tests: 3-year-olds: $Z = 4.316, p < .001, r = 0.623$; 4-year-olds: $Z = 4.423, p = .001; r = 0.638$). Contrary to our
prediction that children would rate the helper and hinderer’s actions similarly in the antisocial
target condition, children again rated the helper’s action (3-year-olds: $M = 3.88$, $SD = 1.60$; 4-
year-olds: $M = 4.67$, $SD = .76$) more positively than the hinderer’s action (3-year-olds: $M = 1.88$,
$SD = 1.19$; 4-year-olds: $M = 1.50$, $SD = .89$; Wilcoxon signed-rank tests: 3-year-olds: $Z = 3.306$,
$p = .001$, $r = 0.477$; 4-year-olds: $Z = 4.296$, $p < .001$; $r = 0.620$).

Finally, while we predicted that ratings would differ between the prosocial target and
antisocial target condition, children at both ages rated the helper’s actions similarly across
conditions (Mann-Whitney $U$ tests; 3-year-olds: $Z = 1.059$, $p = .290$, $r = .153$; 4-year-olds: $Z =
.636$, $p = .525$, $r = .092$) and the hinderer’s actions similarly across conditions (Mann-Whitney $U$
tests; 3-year-olds: $Z = 1.312$, $p = .190$, $r = .189$; 4-year-olds: $Z = 1.420$, $p = .156$, $r = .205$).

**Figure 4.3.** Mean action acceptability ratings in the prosocial target condition and antisocial
target condition at each age in Experiment 1; error bars reflect the standard error of the mean.

**Verbal explanations.** For each response type (uninformative, moral actions/motivations,
social actions/motivations, moral emotions, target’s prosocial/antisocial action), children
received a score of 1 or 0 based on the presence or absence of that response type in their verbal
explanation (see Table 4.1). When explaining why the selected character did the right thing,
children most often appealed to moral actions/motivations (e.g., “because he gave the ball back”, “because he was nice”): 54% of 3-year-olds and 66% of 4-year-olds in the prosocial target condition and 75% of 3-year-olds and 92% of 4-year-olds in the antisocial target condition. Likewise, when explaining why the unselected character behaved as it did, children were again most likely to appeal to moral actions/motivations (e.g., “that guy took the ball” or “because he's very mean”): 46% of 3-year-olds and 67% of 4-year-olds in the prosocial target condition and 50% of 3-year-olds and 63% of 4-year-olds in the antisocial target condition.

Overall, in Experiment 1 children’s social and moral judgments of helpers versus hinderers were not influenced by whether the target of these actions had been prosocial or antisocial. Both 3- and 4-year-olds rated helping positively and hindering negatively, regardless of whether the helper/hinderer was interacting with a prosocial or antisocial individual. On one hand, these results were unexpected given that infants display context-dependent preferences for helpers and hinderers in similarly-structured situations (Hamlin, 2014a; Hamlin et al., 2011). On the other hand, a related study also found that 3-year-olds’ judgments of helpers and hinderers were not influenced by whether the target was prosocial or antisocial (Li & Tomasello, 2018).

Experiment 2 explored whether preschoolers’ failure to incorporate the target’s previous behavior into their social and more judgments reflects their belief that helping is positive and hindering is negative regardless of the context, or whether this failure was an artifact of the current methodology. These alternatives were explored by altering three aspects of the animated social interactions from Experiment 1 to potentially increase children’s production of context-dependent judgments. Firstly, in Experiment 2, the prosocial/antisocial individual displayed its cooperative versus uncooperative intentions repeatedly, rather than once (i.e., the prosocial character helped a third party complete two visually distinct block towers, while the antisocial
character broke both towers). It was predicted that the repeated destruction of the towers would be interpreted as a more severe transgression, which might heighten context-dependent judgments given that third-party punishment is typically viewed as more acceptable when norm violations are more severe (see Eriksson, Andersson, & Strimling, 2017 for evidence that increasing the severity of the norm violation increases the perceived acceptability of punishments enacted by peers). Secondly, in Experiment 2, the initial prosocial/antisocial actions were directly observed by the characters who would later help and hinderer the prosocial/antisocial individual; thus, it was no longer ambiguous whether those enacting “rewards” or “punishments” were aware of the target’s deservingness. It was predicted that this would increase context-dependent judgments given past work demonstrating that infants only expect others to engage in indirect reciprocity when these individuals witnessed the initial events (Meristo & Surian, 2013) and work showing that infants only prefer helpers over hinderers when these helpers could have known they were acting to allow another character to achieve his unfulfilled goal (Hamlin et al., 2013). Finally, in Experiment 2, the helper and hinderer directly matched the actions of the target individual, by either helping the target complete its own block tower or breaking the target’s tower. It was predicted that this would promote context-dependent judgments by suggesting that the hindering action was directly related to the initial transgression.

Experiment 2

Method

Participants. Before data collection began, we established and pre-registered a pre-set stopping rule of 24 children per age per condition (see osf.io/g6r9t). Twenty-four 3-year-olds (M age = 3;7, range = 3;0-3;11, 13 girls) and 24 4-year-olds (M age = 4;5, range = 4;0-4;11, 13 girls) participated in the prosocial target condition, while 24 3-year-olds (M age = 3;6, range = 3;1-
3;11, 14 girls) and 24 4-year-olds (M age = 4;6, range = 4;0-4;11, 13 girls) participated in the antisocial target condition. Five additional 3-year-olds and 5 4-year-olds were seen and replaced in the final sample; 9 children were unable/unwilling to complete the scale familiarization task, 1 child refused to participate/answer questions.

**Procedure. Scale familiarization task.** As in Experiment 1, children were familiarized with a five-point Likert scale used to rate the acceptability of actions.

**Animated social interactions.** Children were randomly assigned to either the prosocial target condition or the antisocial target condition. Depending on condition children either watched the prosocial tower events, in which an orange star helped a purple circle complete two different block towers, or the antisocial tower events, in which the orange star knocked two of the circle’s towers down. In both conditions, the tower events were observed by a yellow square and blue square character, who would subsequently help/hinder the prosocial/antisocial star character (described below; see Figure 4.4). Following each prosocial/antisocial event, children were shown an image of the star and then an image of the squares behind the short wall while asked comprehension questions regarding the star’s action and the squares’ knowledge of that action (i.e., “Did the orange star help build the tower or break the tower? Was that a good thing to do or a bad thing to do? Did the squares see him do that?”). If children answered any comprehension question incorrectly, the animation and question was repeated. If the answer answered incorrectly a second time, then they were corrected. The experimenter agreed with children’s correct response to each comprehension question to encourage recall.
All children then watched the helpful and unhelpful events, in which the orange star character (previously prosocial or antisocial, depending on condition) began to build a block tower. In the helpful event a square character completed the star’s block tower, while in the unhelpful event a square character knocked down the tower (blue square and yellow square; identity and side counterbalanced). Children watched the helpful and unhelpful events once each (order counterbalanced). At the start of each event they were asked to recall the previous actions of the orange star and the squares’ observation of those actions (e.g., “Was he breaking the circle’s towers before? Did the squares see him break those towers?”). If children answered either question incorrectly, they were corrected before the animation continued. Following the helpful and unhelpful events, children saw the blue square and yellow square and were asked two comprehension questions (i.e., “Who helped the star build his tower? Who broke the star’s tower?”). As above, if children answered either question incorrectly the animation and relevant question(s) were repeated, and then the child was corrected (if necessary).

**Test questions.** The test question procedure was identical to that of Experiment 1, except that in Experiment 2 children were asked to recall both the previous actions of the orange square and the squares’ observation of that action before watching the helpful/unhelpful events. Recall that each animation is repeated before children are asked to rate the acceptability of the helper and hinderer’s action on the Likert-scale. Across both age groups, 50% of children answered all
sixteen comprehension questions correctly, while 36% needed at least one question repeated; 14% of children were corrected on a comprehension question after repetition.

**Transcription and coding procedure.** Transcription and coding followed the same procedure as in Experiment 1 except that *Moral Actions* now refer to the building/breaking acts from the helpful/unhelpful tower events (e.g., “because he build it”, "he broke the tower”) and the addition of the following informative response category:

*Helper/hinderer’s knowledge.* Child explained why the character did (not) do the right thing by referencing that the square characters observed the star’s previous prosocial/antisocial action (e.g., “the yellow square were looking”, “because he saw it do the right thing and then he wanted to do it”).

As in Experiment 1, each child’s explanations were coded by two independent research assistants for the presence or absence of each response type (uninformative, moral actions/motivations, social actions/motivations, moral emotions, target’s prosocial/antisocial action, helper/hinderer’s knowledge). Reliability across the six categories was strong (average Cohen’s kappa = .858, range = .740 to 1.000; see McHugh, 2012).

**Results**

**Forced-choice judgments. Confirmatory analyses.** A series of binomial tests were used to determine at what age(s) liking, friendship, and rightness scores differed from chance (see Figure 4.5). We predicted that children’s judgments would favor the helper in the prosocial target condition but not necessarily in the antisocial target condition; see OSF registration for documented predictions and analysis plan (osf.io/g6r9t). Contrary to these predictions, and as in Experiment 1, children at both ages showed the same pattern of responding across conditions. In both conditions of Experiment 2, 3-year-olds selected the helper/hinderer at chance levels when
asked about friendship (prosocial target: 12/23 selected helper, $p = 1.000$; antisocial target: 16/23 selected helper, $p = 0.093$), but liked the helper (prosocial target, antisocial target: 19/24 selected helper, $p = 0.007$) and indicated the helper did the right thing (prosocial target: 23/24 selected helper, $p < .001$; antisocial target: 20/24 selected helper, $p = 0.002$). In contrast, in both conditions, 4-year-olds liked the helper (prosocial target: 22/24 selected helper, $p < .001$; antisocial target: 18/24 selected helper, $p = 0.023$), selected the helper to be friends with (prosocial target: 22/24 selected helper, $p < .001$; antisocial target: 20/24 selected helper, $p = 0.002$), and judged the helper as having done the right thing (prosocial target, antisocial target: 22/24 selected helper, $p < .001$).

Figure 4.5. Percent of children who selected the helper over the hinderer when asked about liking, friendship, and rightness at each age in Experiment 2; *$p < .05$, **$p < .01$, ***$p < .001$.

To determine whether judgments in each age group (3-year-olds, 4-year-olds) differed depending on condition (prosocial target, antisocial target), children were scored 1 if they

---

11 One 3-year-old in the prosocial target condition and one 3-year-old in the antisocial target condition refused to select a character when asked whether they would be friends with the helper or hinderer.
selected the helper and 0 if not for each forced-choice questions (liking, friendship, rightness); these scores were then summed, resulting in a score between 0 – 3 for each child. Contrary to our prediction that scores would be higher in the prosocial versus antisocial condition (and as in Experiment 1), neither 3-year-olds’ scores (prosocial target: $M = 2.25$, $SD = .74$; antisocial target: $M = 2.29$, $SD = .75$; independent-samples t-test, $t(46) = .194$, $p = .847$, $d = .058$), nor 4-year-olds’ scores differed across condition (prosocial target: $M = 2.75$, $SD = .44$; antisocial target: $M = 2.50$, $SD = .66$; independent-samples t-test, $t(40.214) = 1.543$, $p = .131$, $d = .455$).

**Exploratory analyses.** To explore whether age, gender, and/or question type influenced children’s responses, two mixed-effect ANOVAs were conducted with question type (liking, friendship, rightness) as a within-subjects variable and child’s age (3, 4) and child’s gender (female, male) as between-subjects factors. Unlike Experiment 1, which showed a main effect of question type in each condition, the present analyses revealed no main effects nor interactions in the antisocial target condition (all $Fs < 1.738$, all $ps > .193$). In contrast, in the prosocial target condition, there was a main effect of question type ($F[1.747,75.115] = 4.909$, $p = .013$, $\eta^2 = .102$), a main effect of age ($F[1,43] = 6.401$, $p = .015$, $\eta^2 = .130$), and an interaction between the two ($F[.747,75.115] = 5.174$, $p = .011$, $\eta^2 = .107$); all other $Fs < .944$, all other $ps > .382$. To explore these main effects and interaction, a series of t-tests were conducted using the Bonferroni corrected alpha value of .007 (.050/7). Three paired-samples t-tests compared responses to each question type across age, and showed no significant differences when compared to the adjusted alpha value (liking: $M = .87$, $SD = .34$; friendship: $M = .72$, $SD = .45$; rightness: $M = .94$, $SD = .25$; all $ps > .010$). An independent-samples t-test comparing younger versus older children’s scores across question type (summed liking, friendship, rightness scores between 0-3) revealed that 4-year-olds ($M = 2.75$, $SD = .44$) were more likely to select the helper
than 3-year-olds ($M = 2.25, SD = .74; t[37.660] = 2.849, p = .007, d = .840$). Finally, to probe the interaction between question type and age, three independent-samples t-tests compared 3-year-olds’ versus 4-year-olds’ responses to each question type: 4-year-olds were more likely to select the helper as a friend ($M = .92, SD = .28$) compared to 3-year-olds ($M = .52, SD = .51; t[33.983] = 3.261, p = .003, d = .984$); there was no difference when asked about liking (3-year-olds: $M = .79, SD = .42$; 4-year-olds: $M = .92, SD = .28; t[40.542] = 1.220, p = .229, d = .360$) or rightness (3-year-olds: $M = .96, SD = .20$; 4-year-olds, $M = .92, SD = .28; t[46] = .586, p = .561, d = .118$).

**Action acceptability ratings.** Children’s Likert scale ratings of the helper/hinderer’s actions were again scored from 1 (really bad) to 5 (really good; see Figure 4.6). As predicted, in the prosocial target condition, children rated the helper’s action (3-year-olds: $M = 4.67, SD = .76$; 4-year-olds: $M = 4.92, SD = 0.28$) more positively than the hinderer’s action (3-year-olds: $M = 1.25, SD = .53$; 4-year-olds: $M = 1.13, SD = .34$; Wilcoxon signed-rank tests: 3-year-olds: $Z = 4.455, p < .001, r = 0.643$; 4-year-olds: $Z = 4.611, p < .001; r = 0.666$). Contrary to our prediction that children would rate the helper and hinderer’s actions similarly in the antisocial target condition, children again rated the helper’s action (as in Experiment 1; 3-year-olds: $M = 4.46, SD = 1.06$; 4-year-olds: $M = 4.71, SD = .91$) more positively than the hinderer’s action (3-year-olds: $M = 1.28, SD = .88$; 4-year-olds: $M = 1.21, SD = .51$; Wilcoxon signed-rank tests: 3-year-olds: $Z = 4.203, p < .001, r = 0.607$; 4-year-olds: $Z = 4.455, p < .001; r = 0.643$).

Finally, while we predicted that ratings would differ between the prosocial and antisocial target condition, as in Experiment 1, children rated the helper’s actions similarly across conditions (Mann-Whitney $U$ tests; 3-year-olds: $Z = .693, p = .488, r = .100$; 4-year-olds: $Z = .545, p = .586, r = .079$) and the hinderer’s actions similarly across conditions (Mann-Whitney $U$ tests; 3-year-olds: $Z = .338, p = .735, r = .049$; 4-year-olds: $Z = .455, p = .649, r = .066$).
Verbal Explanations. For response type (uninformative, moral actions/motivations, social actions/motivations, moral emotions, target’s prosocial/antisocial action, helper/hinderer’s knowledge), children received a score of 1 or 0 based on the presence or absence of that response type in their verbal explanation (see Table 4.2). When explaining why the selected character did the right thing, children most often appealed to moral actions/motivations (e.g., “because he helped build the tower.”, “he did something good”): 54% of 3-year-olds and 79% of 4-year-olds in the prosocial target condition, 75% of 4-year-olds in the antisocial target condition. The exception being that 3-year-olds in the antisocial target condition were equally likely to provide uninformative responses as appeals to moral actions/motivations (46%). As in Experiment 1, when explaining why the unselected character behaved as it did, children were most likely to appeal to moral actions/motivations (e.g., “because he's broken it” or “because he wasn't a very nice guy”): 50% of 3-year-olds and 71% of 4-year-olds in the prosocial target condition and 75% of 3-year-olds and 63% of 4-year-olds in the antisocial target condition.

Overall, in Experiment 2 children’s social and moral judgments continued to be unaffected by whether the target of helping/hindering was previously prosocial or antisocial. As
in Experiment 1, 3- and 4-year-olds rated helping positively and hindering negatively, regardless of whether the target of these actions was a prosocial or antisocial individual. The general discussion will consider the extent to which the current studies suggest that preschoolers simply do not consider a target’s previous behavior relevant to the evaluation of helpers and hinderers, and how these findings are limited to the specific sets of stimuli utilized in Experiment 1 and 2.

**General Discussion**

Across two experiments, 3- and 4-year-olds evaluated characters who helped versus hindered a previously prosocial or antisocial individual. In both experiments children’s social judgments (liking, friendship) and moral judgments (rightness) were unaffected by whether the helper and hinderer were interacting with a prosocial or antisocial target. Children in both experiments also rated helping positively and hindering negatively, regardless of the context in which these actions were performed. This was the case even though children’s recall of the target’s prosocial/antisocial action was supported throughout Experiment 1 and 2 via comprehension questions and memory checks regarding the target’s previous actions directly before this character was helped or hindered; thus, it is unlikely that children’s failure to consider contextual information was due to memory demands (i.e., forgetting the target’s previous actions). Furthermore, children’s judgments were unaffected by context even when the prosocial and antisocial individuals displayed their positive/negative actions multiple times, when the helper and hinderer directly observed the prosocial/antisocial acts, and when the helper and hinderer’s actions mirrored the prosocial/antisocial character’s previous actions (Experiment 2).

On one hand, related work has previously documented children’s insensitivity to context in similar situations. In one study, 3-year-olds’ social judgments (liking) and moral judgments (rightness/wrongness ratings) were unaffected by whether helpers and hinderers interacted with
previously prosocial or antisocial others; 5-year-olds’ moral judgments were likewise unaffected by this information (Li & Tomasello, 2018). That said, there are several reasons it is surprising that preschoolers’ judgments were insensitive to the context in which helping/hindering actions were performed. First of all, even infants form context-dependent judgments of helpers and hinderers in similarly-structured situations (Hamlin, 2014a; Hamlin et al., 2011). Further, Chapter 2 and 3 of this dissertation demonstrated significant overlap between infants’ implicit preferences and preschoolers’ explicit social and moral judgments. Thus, it was unexpected that preschoolers’ explicit judgments failed to align with infants’ preferences in the current studies.

Second, preschoolers act in accordance with the principle of indirect reciprocity. For example, when given the opportunity to distribute an unequal number of resources between a helpful puppet who assisted a third party and a harmful puppet who violently prevented the third party from achieving its goal, 4.5-year-olds gave more resources to the helper (Kenward & Dahl, 2011; see also Li & Tomasello, 2018). Relatedly, when asked to help a puppet distribute an unequal number of resources, 3.5-year-olds directed the puppet to give more resources to characters who previously shared with a third party compared to characters that did not (Olson & Spelke, 2008). Given 3- and 4-year-olds own engagement in indirect reciprocity, it is surprising that they did not differentially evaluate others who did versus did not do the same.

Finally, young children also act in accordance with the principle of third-party punishment: Toddlers selectively take treats from puppets who have previously hindered a third party rather than helped (Hamlin et al., 2011), 3- to 6-year-olds will enact third-party punishment even when doing so is personally costly (i.e., by closing an attractive slide and thus preventing the child themselves and the antisocial other from playing on the slide in the future; Yudkin, Van Bavel, Rhodes, 2019), and 5-year-olds will punish previously antisocial adults (by allocating
them unpleasant resources) when this can be done anonymously rather than in person (Kenward & Östh, 2015). Preschoolers will also correct situations in which punishment is allocated inappropriately: 3- and 4-year-olds spontaneously rectify unfair situations imposed by adults who punish accidental transgressors, even when doing so is personally costly (Chernyak & Sobel, 2016), and when asked to retell stories in which victims rather than perpetrators of transgressions are punished, 4-year-olds will enact punishment of the wrongdoer instead (Kenward & Östh, 2012). Overall, given young children’s own engagement in third-party punishment and their sensitivity to inappropriate third-party punishment, it is surprising that preschoolers do not positively evaluate those who punish antisocial others in the current studies.

Why might preschoolers have failed to consider the target’s previous prosocial versus antisocial actions in the current studies? One possibility is that the ability to use this information when forming moral judgments emerges after the preschool years. This development may be tied to greater experience with context-dependent rules in daily interactions, explicit teaching of moral norms regarding reward and punishment, and/or more mature social-cognitive skills (e.g., theory-of-mind, abstract reasoning, etc.). If this is the case, then it may be the infants’ preferences in similar situations (Hamlin, 2014a; Hamlin et al., 2011) reflect social rather than moral concerns. For example, if infants’ preferences for those who hinder antisocial others is based on a social preference (e.g., affiliation based on shared dislike of hinderer) rather than an evaluation akin to moral approval of third-party punishment, then we would not necessarily expect young children’s explicit judgments to align with these implicit social preferences.

Another possibility is that preschoolers are capable of producing context-dependent judgments, but that the stimuli used in studies to date has not appropriately depicted scenarios in which third-party punishment is appropriate. For example, in Experiment 2, the severity of the
initial transgression was equivalent to the punishment (i.e., the antisocial star character knocked down a third party’s towers, the hinderer knocked down the star’s tower). This mirroring of the antisocial star’s actions was intended to show that the “punishment” was related to and directly proportionate to the initial transgression. However, it is possible that hindering would be evaluated more positively if the initial transgression was more severe than the punishment (e.g., if the hinderer simply takes blocks away from an antisocial character who destroyed a third parties’ block tower). Additionally, the social status of the helper and hinderer compared to the target character is ambiguous; it is unclear whether the hinderer character has the authority to allocate punishment for the target's wrongdoing. Perhaps hindering of antisocial others would be evaluated more positively when enacted by an authority figure or appropriate institution (see Cushman, 2015; for example, if the hinderer is a teacher who takes the blocks away from an antisocial character who destroyed a third parties’ block tower in a classroom setting). Overall, future studies are needed to determine when young children begin to reliably produce explicit social and moral judgments favoring those who act in accordance with the principles of indirect reciprocity and altruistic punishment.
Table 4.1

Mean number of explanations containing each response type in Experiment 1

<table>
<thead>
<tr>
<th>Condition/Response</th>
<th>Age</th>
<th>Uninformative Responses</th>
<th>Informative Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Moral</td>
</tr>
<tr>
<td>Prosocial target – selected character</td>
<td>3</td>
<td>.29</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.08</td>
<td>.66</td>
</tr>
<tr>
<td>Prosocial target – unselected character</td>
<td>3</td>
<td>.25</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>.67</td>
</tr>
<tr>
<td>Antisocial target – selected character</td>
<td>3</td>
<td>.08</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.04</td>
<td>.92</td>
</tr>
<tr>
<td>Antisocial target – unselected character</td>
<td>3</td>
<td>.08</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.04</td>
<td>.63</td>
</tr>
</tbody>
</table>

Note. The informative response categories were not mutually exclusive, and the moral emotions category only applies when children have used positive moral emotions in the prosocial target condition or negative moral emotions in the antisocial target condition.
Table 4.2

**Mean number of explanations containing each response type in Experiment 2**

<table>
<thead>
<tr>
<th>Condition/Response</th>
<th>Age</th>
<th>Uninformative Responses</th>
<th>Informative Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Moral</td>
</tr>
<tr>
<td>Prosocial target – selected character</td>
<td>3</td>
<td>.29</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.21</td>
<td>.79</td>
</tr>
<tr>
<td>Prosocial target – unselected character</td>
<td>3</td>
<td>.21</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.17</td>
<td>.71</td>
</tr>
<tr>
<td>Antisocial target – selected character</td>
<td>3</td>
<td>.46</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.13</td>
<td>.75</td>
</tr>
<tr>
<td>Antisocial target – unselected character</td>
<td>3</td>
<td>.17</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.13</td>
<td>.63</td>
</tr>
</tbody>
</table>

**Note.** The informative response categories were not mutually exclusive, and the moral emotions category only applies when children have used positive moral emotions in the prosocial target condition or negative moral emotions in the antisocial target condition.
Chapter 5: General Conclusion

Summary of Main Findings

This dissertation presented three sets of studies exploring preschoolers’ explicit social and moral judgments of third-party helpers and hinderers. Each empirical chapter explored sensitivity to a different feature of moral judgments (i.e., concern for third parties, sensitivity to intentions, and sensitivity to context); this was accomplished by adapting paradigms used to demonstrate infants’ implicit preferences for prosocial others. It was predicted that the use of these paradigms would reveal the relative maturity of preschoolers’ explicit moral judgments.

Chapter 2 examined young children’s impartial moral judgments following the observation of third-party puppet interactions. In Experiment 1, 3- to 5-year-olds (N=74) observed helping and hindering scenarios previously utilized to explore sociomoral evaluation in preverbal infants. Whereas 3-year-olds in Experiment 1 did not reliably distinguish between the helper and hinderer when reporting social preferences or moral judgments, both 4- and 5-year-olds preferred the helper, judged the helper to be “nicer” than the hinderer, selectively allocated punishment to the hinderer, and were able to justify their punishment allocations. A simplified procedure and the addition of comprehension questions in Experiment 2 (N = 24) improved 3-year-olds’ performance, suggestive that their performance in Experiment 1 was likely due to processing or memory difficulties rather than an inability to engage in explicit social and moral evaluation. This set of studies revealed that young children readily interpret helping and hindering scenarios as socially and morally relevant.

Chapter 3 explored preschoolers’ sensitivity to others' intentions to help versus hinder. A substantial literature has suggested that preschoolers’ moral judgments initially privilege outcomes over intention (e.g., Armsby, 1971; Baird & Astington, 2004; Costanzo et al., 1973;
Helwig et al., 2001; Imamoglu, 1975; Killen et al., 2011; Li & Tomasello, 2018; Margoni & Surian, 2017; Moran & O’Brien, 1983; Zelazo et al., 1996; see Margoni & Surian 2016 for review). The four studies presented in Chapter 3 reveal that, in contrast, even younger preschoolers can privilege intentions when evaluating characters who successfully or unsuccessfully help or hinder a third party. Following a live-action puppet show, children were asked the same series of questions as in Chapter 2: which puppet was liked, which puppet was nicer, which puppet should be punished (and why). In Experiment 1, 2A, and 2B (N = 195), 3- and 4-year-olds evaluated characters with distinct intentions to help or to hinder, who were associated with either positive or negative outcomes. Both ages judged characters with more positive intentions as nicer, and allocated punishment to characters with more negative intentions; neither of these tendencies depended on the outcomes the characters were associated with. Three-year-olds’ responses were somewhat less consistent than were 4-year-olds’, in that 3-year-olds’ judgments were disrupted by ambiguous harmful intent. Notably, children’s social judgments (i.e., who do you like) were less consistent than their moral judgments (i.e., who was nicer, who should get in trouble). In Experiment 3 (N = 100), children evaluated characters with the same intention but who were associated with different outcomes. Children showed inconsistent responding across age and outcome valence, but only 4-year-olds evaluating two characters with positive intentions reliably responded based on outcome. When providing informative responses in all three studies, children most frequently explained their punishment allocations by appealing to the puppet’s (attempted) hindering action or failure to help. These findings raise questions as to what underlies different response patterns across studies in the literature, and suggest that observing moral scenarios fully acted out (e.g., not vignette-based tasks) may facilitate young children’s intention-based moral judgments.
Finally, Chapter 4 explored preschoolers’ sensitivity to the context in which helping and hindering actions are performed; that is, whether preschoolers evaluate helpful and unhelpful actions differently depending upon whether the actions are directed towards prosocial others (who deserve rewards) or antisocial others (who deserve punishment). In two experiments 3- and 4-year-olds (N = 192) observed scenarios in which a prosocial or antisocial individual helped or hindered a third party. The prosocial/antisocial individual was then helped by one character and hindered by another. Preschoolers then judged whether they liked the helper or hinderer, which character they would be friends with, and which character did the right thing (and why). Children also rated the acceptability of the helper and hinderer’s actions on a 5-point Likert scale. In both Experiment 1 and 2, children’s judgments were unaffected by whether the helper and hinderer were interacting with a prosocial or antisocial individual. For example, 3- and 4-year-olds judged helping positively and hindering negatively regardless of who these actions were directed toward. Future studies should further examine the extent to which preschoolers positively evaluate those who act in accordance with the principles of indirect reciprocity and altruistic punishment. Altogether, this dissertation documents important areas of continuity and discontinuity between infants’ implicit preferences and preschoolers’ explicit judgments.

**Implications for the Study of Infants’ Moral Sense**

Infants’ preferences for those who help versus hinder third parties (see Hamlin & Wynn, 2011) aligned with young children’s explicit social and moral judgments (Chapter 2). Likewise, infants’ reliance on intention information when forming preferences regarding failed/successful helpers/hinderers (Hamlin, 2013) aligned with preschoolers’ explicit moral (but not necessarily social) judgments of similar characters (Chapter 3: Experiment 1, 2B, and the negative intention condition of Experiment 3). Such continuity is at least suggestive that infants’ preferences may
be an early-emerging component of our mature moral sense. That said, the cross-sectional nature of the comparisons being made here prevent any strong conclusions regarding the relation between infants’ evaluations and children’s judgments. More direct evidence is needed, such as exploring the continuity between judgments in infancy and childhood within the same child (see Tan, Mikami, Hamlin, 2018 for initial evidence that sociomoral competency in infancy predicts sociomoral functioning in preschool). As for areas of discontinuity, it was found that preschoolers can use outcomes to inform their social and moral judgments when intentions are uninformative (i.e., 4-year-olds use of outcome information to judge characters with equivalent positive intentions in Experiment 3 of Chapter 3), while infants do not (Hamlin, 2013). Further, while infants prefer those who help prosocial others and hinderer antisocial others (Hamlin, 2014; Hamlin et al., 2011), it remains unclear whether young children use individuals’ past behavior when evaluating acts of helping and hindering (Chapter 4; see also Li & Tomasello, 2018). These discontinuities highlight the need to further explore the link between infants’ implicit evaluations and young children’s explicit judgments. For example, do infants’ preferences reflect social or moral concerns regarding how prosocial versus antisocial others are treated? Do young children show implicit preferences for those who help prosocial others and hinder antisocial others? At what age and under what conditions do children report positive evaluations of those who reward/punish appropriately?

**Implications for the Study of Children’s Moral Judgments**

Beyond implications for the interpretation of infants’ sociomoral preferences, this dissertation makes significant contributions to the understanding of young children’s moral judgments. First, Chapter 2 is one of a relatively small number of studies exploring children’s explicit judgments and verbal reasoning regarding helping and hindering puppet show displays
(compared to the more common vignette-based presentation of more prototypical moral violations, such as physical harm; see Nucci & Gingo, 2010; Smetana, 2006; Smetana, Jambon, & Ball, 2014; Turiel, 2006 for review). Prior to this work, Kenward and Dahl (2011) had presented 3- and 4-year-olds with live puppet shows featuring a protagonist that struggled to achieve its goal; the protagonist was assisted by a helper puppet and thwarted by a violent hinderer puppet. In these studies, 4.5-year-olds but not 3-year-olds reported liking the helper over the violent hinderer and provided a reasonable explanation for their puppet preferences (see supplementary materials for Kenward and Dahl, 2011); note that 3-year-olds in these studies struggled to answer comprehension questions regarding the puppet’s actions during the show. The studies in Chapter 2 of this dissertation reveal that 3-year-olds also prefer helpers over hinderers, and both ages form explicit moral judgments favouring helpers over hinderers. Overall, this work demonstrates that children view helping/hindering a third parties’ goal as both socially and morally relevant, and provides convergent evidence that from 3 years children can reason about moral norms and transgressions.

Second, Chapter 3 of this dissertation documents a relatively mature consideration of intentions among 3-year-olds. While 3-year-olds’ have demonstrated sensitivity to intentions in previous work (i.e., when confounding factors are removed and intentions are made explicit or otherwise salient; Armsby, 1971; Nelson, 1980; Nobes et al., 2009; Yuill & Perner, 1988; see Hilton & Kuhlmeier, 2019 for discussion), these studies are limited in number, especially compared to the number of studies showing that the ability to privilege intentions over outcomes develops after age 3 or 4 (e.g., Costanzo et al., 1973; Cushman et al., 2013; King, 1971; Margoni & Surian, 2017). For example, even recent studies using video-recorded puppet shows continue to find that 3-year-olds show limited (if any) use of intention information when forming social
and moral judgment (e.g., Li and Tomasello, 2018). In contrast, Chapter 3 demonstrates that both 3- and 4-year-olds can privilege intentions over outcomes when making moral judgments.

Finally, Chapter 4 is one of only two studies exploring young children’s explicit judgments of those who help/hinder prosocial/antisocial others (see also Li and Tomasello, 2018). While these studies provide additional evidence that preschoolers prefer helpers over hinderers (Experiment 2) and form moral judgments that favor helpers (Experiment 1 and 2), children’s judgments were not influenced by the context in which helping/hindering occurred. It is still very much an open question whether/when preschoolers positively evaluate those who act in accordance with the principles of indirect reciprocity and altruistic punishment; future studies should continue to explore these questions.

**Limitations and Unanswered Questions**

All that said, there are limitations to the current studies and open questions that cannot be resolved with the current data. Several of these will be discussed in the remainder of this chapter, including the appropriateness of the sample sizes utilized, distinguishing social versus moral responding in the current studies, whether children’s responses can be explained via affective tagging, the role of implicit versus explicit processes in the production of moral judgments, and the role of maturation, experience, and culture in shaping moral judgment.

**Power to Detect Small Effects**

One limitation of the current studies is the sample size utilized (i.e., 24 children per age per condition). This sample size was based on standard practices in the field, rather than an a priori power analysis of the sample needed to detect each expected effect within a reasonable degree of confidence. Based on the effect sizes observed, post-hoc power analyses suggest that the studies in Chapter 2 were appropriately powered to detect non-chance responding when
asked whether the helper or hinderer puppet was liked, nicer, and should get in trouble
(Experiment 1 and 2: average \( d = 1.041, N = 24, \alpha = .05, \text{power} = 0.998 \)). A similar analysis also
suggests that the studies in Chapter 3 were appropriately powered to detect children’s use of
intentions when forming social and moral judgments in conditions where intentions were
informative (Experiment 1, 2A, and 2B: average \( d = 1.090, N = 24, \alpha = .05, \text{power} = 0.999 \)).

Unfortunately, the use of this sample size when investigating whether preschoolers form
context-dependent judgments was likely problematic. Post-hoc power analyses based on the
observed effect sizes in Chapter 4 suggest that Experiment 1 and 2 were woefully underpowered
to detect a difference in responding between the prosocial target condition and the antisocial
target condition, both when comparing summed forced-choice responses (liking, friendship,
rightness: average \( d = 0.265, N = 24, \alpha = .050, \text{power} = 0.146 \)) and when comparing action
acceptability ratings (average \( d = 0.259, N = 24, \alpha = .050, \text{power} = 0.138 \)) across conditions.
While the observed effect sizes may not be accurate representations of the true effect in the
population (given the sampling error associated with small samples), these post hoc power
analyses certainly show that \( N = 24 \) was an inappropriate sample size to use when exploring the
presence of small effects. Future studies investigating the presence of context-dependent social
and moral judgments using a similar paradigm would be appropriately powered at a sample size
of approximately \( N = 239 \) per condition (average \( d = 0.261, \alpha = .050, \text{power} = 0.800 \); all power
analyses conducted in G*Power from Faul, Erdfelder, Lang, & Buchner, 2007).

Distinguishing Social Versus Moral Responding

Despite the sample size issues in Chapter 4, this dissertation has provided evidence that
preschoolers’ explicit judgments are influenced by several factors that determine adults’ moral
judgments, including concern about the treatment of third parties (Chapters 2 – 4) and sensitivity
to others’ intentions rather than the outcomes achieved (Chapter 3). Throughout this dissertation, I have discussed children’s judgments as social (i.e., who do you like, who do you want to be friends with) versus moral (i.e., who was nicer, who should get in trouble, who did the right thing). This categorization reflects that social concerns and affiliations can be based on idiosyncratic preferences (e.g., friendships based on liking the same music), while moral concerns reflect relatively more meaningful attributions regarding whether actions are nice/mean, right/wrong, and deserving of praise/blame or reward/punishment.

However, one concern is that children were not distinguishing test questions in the same manner as the experimenter. That is, children may not have been differentially responding based on the content of the test questions (i.e., not making strictly social judgments when asked about liking/friendship and/or strictly moral judgments when asked about niceness/rightness/allocation of punishment). Instead, children may have been responding to all questions based on social concerns. For instance, children’s judgments favoring intentionally helpful individuals and opposing intentionally unhelpful others may be based on an analysis of who is likely to (at least attempt to) provide help to the child in the future. If children were indeed responding to all test questions based on egocentric social concerns, then this dissertation speaks to children’s social (but not moral) development, since moral judgments are based on concerns about how others ought to be treated, regardless of the implications for the child themselves.

Other researchers have distinguished social versus moral responding by asking children to make “criterion judgments” that reflect defining aspects of the moral domain (e.g., evaluating actions based on obligatoriness, impersonality, alterability, universality, social consensus, rule contingency, authority jurisdiction; Turiel, 1983). As reviewed in Chapter 1, by age 3 or 4 children distinguish moral transgressions (e.g., causing physical harm) from social-conventional
transgressions (e.g., violating dress code) and judge moral transgression as more serious, more likely to be wrong across contexts, less contingent on explicit rules, and wrong regardless of whether the transgression is witnessed or condemned by an authority figure (see Nucci & Gingo, 2010; Smetana, 2006; Smetana, Jambon, & Ball, 2014; Turiel, 2006 for review). Relatedly, researchers have also distinguished social versus moral responding by comparing children’s judgments about characters who violated moral versus social norms. For example, in a recent study 2- to 4-year-olds watched pairs of live puppet shows in which one puppet committed a moral transgression (i.e., hitting, teasing) while the other committed a social-conventional transgression (i.e., putting toys away in the wrong place, not sitting down during story time). Children then evaluated the puppets (i.e., which was more wrong, which deserved more punishment) and reported their relative preferences between puppets (i.e., which puppet they liked, which they wanted to be friends with). While younger children selected between the transgressors at chance when providing evaluations and preferences, older children preferred conventional transgressors and evaluated moral transgressors as more wrong and deserving of punishment (Smetana, Ball, Jambon, & Yoo, 2018). These results are consistent with the work reviewed above suggestive that domain distinctions are in place by 3 or 4 years of age.

While the current studies did not seek to distinguish social versus moral responding by having children make criterion judgments or compare helpers/hinderers to those who upheld/violated social conventions, there is evidence that children in Chapters 2 – 4 were differentially responding based on the content of the various test questions. First, there is some preliminary evidence that preschoolers’ moral judgments were more consistent than their social ones. That is, whenever children were asked to compare helpers and hinderers, social judgments (i.e., who do you like, who do you want to be friends with) favoring the helper were almost
always numerically lower than moral judgments favoring the helper (i.e., who should get in trouble, who did the right thing)\textsuperscript{12}. Second, whenever there was a significant effect of question type on children’s judgments, follow-up analyses revealed that children were less likely to respond in the direction of the hypotheses on social rather than moral judgments\textsuperscript{13}. Thirdly, in Chapters 2 – 4 children most frequently explained their moral judgments (i.e., allocations of punishment in Chapter 2 and 3, justifications of rightness in Chapter 4) via explicit references to morally relevant actions (i.e., the helping/hindering actions from the puppet show/animated events) or moral attributions/motivations (e.g., “he was mean”, “because he’s good”). While this certainly does not guarantee that children’s internal reasoning was based on moral concerns, appeals to morally-relevant content are more consistent with moral understanding than if children’s explanations were often based on non-moral, social concerns (e.g., “he’s very silly”, “because he wanted to”) or non-social concerns (e.g., “he’s soft”, “I like the green shirt one”). Overall, while the present data cannot conclusively determine whether children produced distinctly social versus moral judgments, there are reasons to think that children understood the social (i.e., liking, friendship) versus moral (i.e., trouble, rightness) test questions as intended. That all said, the relationship between social concerns and moral concerns is very much an area of active debate. As discussed above, one possibility is that moral concerns are entirely

\textsuperscript{12} Social preferences for the helper were numerically lower than moral judgments favoring the helper. This was the case for all ages in Chapter 2 Experiment 1 and 2; for all ages/conditions in Chapter 3 Experiment 1, all ages in Experiment 2B, and 4-year-olds in Experiment 2A (3-year-olds responded at chance for all test questions); and for all ages/condition in Chapter 4 Experiment 1 and 2, with the exception of 4-year-olds in the prosocial target condition of Experiment 2, who were near ceiling in their social and moral judgments favoring the helper.

\textsuperscript{13} When there was a significant effect of question type, children were less likely to respond in the direction of the hypothesis when making social versus moral judgments. In Chapter 2 Experiment 1 children were less likely to select the helper when asked about liking versus trouble; in Chapter 3 Experiment 1 (both conditions) and 2A children were less likely to respond as predicted when asked about liking versus niceness or trouble; in Chapter 4 Experiment 1 and 2 children were less likely to select the helper when asked about friendship versus rightness in the prosocial target condition and when asked about liking or friendship compared to rightness in the antisocial target condition of Experiment 1.
distinct from social concerns. This distinction may align with the criterion judgments used to define moral versus social-conventional norms (e.g., evaluating actions based on obligatoriness, impersonality, alterability, universality, social consensus, rule contingency, authority jurisdiction) and is supported by evidence that even young children distinguish moral norms from social-conventional norms along several dimensions (see Smetana, 2006; Turiel, 1983). An alternative possibility is that moral norms are a subset of social norms. For example, it has been proposed that moral norms represent an important class of “norms with feeling”, and that affective responses to moral transgressions (and disgusting social-conventional transgressions) determine which actions are judged to be impermissible and authority independent (Nichols, 2002; see also Kelly, Stich, Haley, Eng, & Fessler, 2007). This view suggests that moral norms are not completely distinct from social-conventional concerns, as this affective response to moral concerns arises in other situations as well. The overlap between social and moral norms is further supported by neurological evidence demonstrating that the processing of moral content recruits several brain areas involved in social understanding, rather than specifically moral processing (see Young & Dungan, 2012 for discussion; see also Greene & Haidt, 2002); though recent evidence suggests that a sub-region of the posterior superior temporal sulcus may selectively process information about the prosocial/antisocial nature of social interactions (i.e., whether an agent is being helped or hindered; Isik, Koldewyn, Beeler, & Kanwisher, 2017).

**Indiscriminate Responses Based on Affective Tagging**

In addition to responding based solely on egocentric social concerns, another potential concern is that children responded to all test questions based on the affective valence of the characters’ interactions (i.e., indiscriminately attributing positive [negative] descriptors to characters associated with positive [negative] intentions and/or outcomes), without considering
the moral content of the characters’ behavior. This may occur via affective tagging, in which the affect/valence associated with a positive or negative event becomes associated with the individual experiencing that event (see Olson, Dunham, Dweck, Spelke, & Banaji, 2008). To illustrate how young children’s moral judgments may be based on the affective valence of non-moral content, consider that 3- to 5-year-olds judge lucky individuals who happen to experience positive events to be nicer than unlucky individuals (Olson et al., 2008), while 5-year-olds predict that individuals who label objects accurately will be more prosocial than inaccurate labelers (Brosseau-Liard & Birch, 2010; see also Stipek & Daniels, 1990; but see Fusaro, Corriveau, & Harris, 2011).

If children were indeed responding indiscriminately to all test questions based on affective tagging, then this dissertation cannot make strong conclusions regarding children’s evaluations of moral content. However, there are several reasons to think that preschoolers in the current studies were not responding based on affective tagging. First, there is evidence (described in the previous section) that children responded differentially to social versus moral test questions. This response pattern counters the affective tagging account, which suggests that children’s social and moral judgments would be identical (i.e., based only on the positive/negative valence of the interaction and the test question). Second, infants’ reactions to helpers and hinderers in similar situations cannot be explained via affective tagging. To illustrate, when presented with helping and hindering scenarios featuring previously prosocial or antisocial protagonists, 4.5- and 8-month-olds preferred helpers of prosocial others and hinderers of antisocial others (Hamlin, 2014a; Hamlin et al., 2011); the affective tagging account would instead predict positive evaluations/preferences for helpers, regardless of the context in which helping is performed. The affective tagging account would likewise predict that positive
(negative) actions would be directed towards those associated with positive (negative) events. In contrast to this account, while toddlers selectively give treats to prosocial others and take treats from antisocial others, they also give treats to characters who have been the victims of antisocial behaviour (Hamlin, 2014a). Together, these findings are broadly suggestive that preschoolers’ responses to similar helping and hindering scenarios also cannot be reduced to a simple analysis of the affect/valence associated with an interaction.

**Distinguishing Implicit Versus Explicit Responding**

In addition to disentangling social versus moral responding, another potential concern is the ability to distinguish between implicit versus explicit responding. Throughout this dissertation I have referred to children’s moral judgments as “explicit” in nature. This label is intended to reflect that linguistically-competent preschoolers are capable of conscious, deliberate thought and thus could possess (at least some) explicit knowledge of moral principles. It is plausible that children’s moral judgments in Chapters 2 – 4 were rooted in such explicit knowledge: Judgments were elicited by the experimenter’s verbal questioning, there is preliminary evidence that children responded differentially based on question type, and preschoolers could provide reasonable explanations for their moral judgments (see above for details). This explicit responding is contrasted with the “implicit” sociomoral evaluations demonstrated by infants in previous work (see Hamlin, 2013b; Hamlin & Van de Vondervoort, 2018; Van de Vondervoort & Hamlin, 2018a for review). As discussed in Chapter 1, infants’ behavioral responses when presented with helpers and hinderers presumably reflect solely automatic, unconscious responses rooted in intuitions about moral content. This is assumed to be the case, given that preverbal infants are (presumably) incapable of explicit reasoning about
moral principles and unaffected by the content of the language included in these tasks (e.g., when the experimenter asks, “Who do you like?”).

That said, the current data alone cannot determine the relative contribution of explicit versus implicit processes to children’s moral judgments in Chapters 2–4. For example, it is possible that children responded to the experimenter’s verbal questions based on their implicit moral intuitions rather explicit moral reasoning. In this view, children’s verbal explanations are generated after-the-fact to support their forced-choice judgments. This possibility is consistent with arguments that even adults’ moral judgments are rooted in intuitions rather than reason-based processes (e.g., Haidt, 2001); that is, individuals are faced with morally-relevant events, evaluative intuitions arise quickly and automatically in response to these events, and moral judgments reflect the content of these intuitions. Moral reasoning may follow when arguments are needed to support an already-made judgment.

It has been proposed that these moral intuitions may be rooted in automatic emotional reactions to moral concerns (Haidt & Joseph, 2008). This account is consistent with demonstrations that certain emotions can unconsciously influence adults’ responses to moral violations. For example, negative emotions (e.g., disgust, anger) can increase the severity of moral judgments, leading to greater condemnation or punishment of moral transgressions (Inbar,

14 In fact, the characterization of cognitive processes as either “implicit” or “explicit” may in itself be overly simplistic. For example, Karmiloff-Smith (1992) distinguishes between implicit knowledge (level I) and multiple levels of explicit knowledge (levels E1-3); the Representational Redescription (RR) model posits that implicit representations are translated into increasingly higher formats with learning and development. Importantly, this view distinguishes between unconscious explicit knowledge (level E1) and explicit knowledge that can be consciously reflected upon and verbally reported (level E2-3). Of note, it remains to be seen whether preschoolers’ moral judgments reflect the cognitive flexibility that characterizes levels E2-3 (e.g., whether 3- and 4-year-olds produce context-dependent judgments of helpers and hinderers; see Chapter 4).

15 Alternatively, moral intuitions may be non-emotional in nature, and instead arise from a complex set of rules, concepts, and principles that allow individuals to determine which acts are right versus wrong. Just as an innate “universal grammar” may underlie the rapid learning of language (Chomsky, 1995), a non-emotional universal moral grammar may underlie unconscious moral intuitions (Hauser, 2006; Mikhail, 2007).
Pizarro, Knobe & Bloom, 2009; Lerner, Goldberg, & Tetlock, 1998; Rottman, Kelemen, & Young, 2014; Schnall, Benton, & Harvey, 2008; Schnall, Haidt, Clore & Jordan, 2008; Seidel & Prinz, 2013a; Wheatley & Haidt, 2005; but see Johnson, Cheung & Donnellan, 2014 and Landy & Goodwin, 2015). In contrast, positive emotions can reduce the condemnation of hypothetical moral violations (Seidel & Prinz, 2013b; Strohminger, Lewis, & Meyer, 2011) and increase endorsement of principle-based responses to moral dilemmas (Valdesolo & DeSteno, 2006).

The role of emotions in moral judgments is also evident when examining the relative activation of emotion-related versus reason-related brain areas. For instance, when adults are presented with moral dilemmas that pit principle-based moral rules against emotionally-charged actions (e.g., would you kill your child to save five strangers), the tendency to reject harmful actions that align with utilitarian principles is associated with greater recruitment of emotion-based brain areas versus reason-based brain areas (Greene, Sommerville, Nystrom, Darley & Cohen, 2001; see also Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). When moral dilemmas are less emotionally charged and/or when emotional processing is not engaged, adults are instead more likely to endorse utilitarian actions (e.g., acting to save the greatest number of lives, even if that means killing a family member). This logical application of moral principles is associated with greater relative activation of reason-related brain areas (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Shenhav & Greene, 2014; see Greene, 2013 and Greene & Haidt, 2002 for review).

The causal role of emotion in moral judgment is evident when examining the judgments of adults with damage to emotion-related brain areas, who are more likely to endorse utilitarian moral principles compared to healthy adults (Koenigs et al., 2007; Mendez, Anderson, & Shapira, 2005). In fact, the functioning of these brain areas early in development is critical to the
typical development of moral judgments: Whereas individuals who acquire damage to emotion-related brain areas during childhood demonstrate impaired moral reasoning as adults, individuals who acquire similar lesions during adulthood demonstrate intact moral reasoning (Anderson, Bechara, Damasio, Tranel & Damasio, 1999; Taber-Thomas et al., 2014). Further suggestive that emotional processing is especially critical to moral judgment early in life is evidence that observing moral violations is associated with stronger activation of emotion-related brain areas in younger versus older children (Decety, Michalska & Kinzler, 2012).

Future studies could directly explore whether preschoolers’ responses in the current paradigms are the result of emotion-based intuitions or explicit reasoning. This could be accomplished by asking children to make forced-choice questions about helpers and hinderers under time pressure or cognitive load (which would presumably disrupt deliberate but not intuition-based reasoning), or after children’s emotions have been experimentally manipulated (which would disrupt emotion-based intuitions but not deliberate reasoning). If, for example, preschoolers’ positive/negative evaluations of those who intend to help/hinder remain intact despite the inducement of negative/positive emotions, then this would provide some evidence that these judgments are not solely based on emotion-based intuitions. This contrasts predictions about infants’ prosocial preferences, which I expect to be driven by emotional reactions to the observation of helping and hindering. This latter prediction is based on the proposal that infants’ implicit moral sense is rooted in evolved emotion-based intuitions regarding moral content (see Chapter 1 for further discussion), and consistent with evidence that infants and toddlers display more positive affect during the observation of helping versus hindering puppet shows (Steckler, Liberman, Van de Vondervoort, Slevinsky, & Hamlin, 2018). Overall, it may be that humans are born with emotion-based, implicit intuitions regarding moral content and that (at least some of)
these intuitions become explicit by early childhood; while this does not entail that preschoolers’ moral judgments are produced independent of emotion, it does place an increased importance on the role of deliberate, conscious reasoning with age.

**The Influence of Maturation and Experience**

The development of moral judgments from intuition-based (in infancy) to a combination of intuition and reason (in childhood and adulthood) allows humans to engage in more complex moral reasoning with age. Many factors likely contribute to this transition from implicit to explicit moral knowledge, including the acquisition of language, the maturation of social and cognitive abilities, and the accumulation of relevant experiences. Together, these factors allow individuals to engage in more complex moral reasoning than supported by intuition alone; this explicit moral reasoning undergoes tremendous development after the preschool years.

Notably, early-emerging moral intuitions alone may prove insensitive to some factors that influence mature moral judgments, such as differences in how advantaged versus disadvantaged others ought to be treated. To illustrate, infants expect individuals to distribute resources equally among recipients (Schmidt & Sommerville, 2011, see also Meristo, Strid, & Surian, 2016; Sommerville, Schmidt, Yun, & Burns, 2013) and prefer equal distributors over unequal distributors (Burns & Sommerville, 2014; Geraci & Surian, 2011; Lucca, Pospisil, & Sommerville, 2018). These expectations are not inflexible, as evidenced by work showing that toddlers only expect fair distributions when recipients are equally meritorious, but not when only one individual works to complete an assigned task (Sloane, Baillargeon, & Premack, 2012). However, early expectations regarding resource distribution do not always align with more mature moral judgments. For example, recent studies have shown that toddlers expect resource rich individuals (i.e., those with relatively more items) to receive more items than individuals
who are resource poor (Sitch, Ake, & Sommerville, 2016). This potentially suggests that moral intuitions regarding resource distributions do not (always) align with concerns related to equity (i.e., positive evaluations of preferential treatment for disadvantaged others). Additional evidence demonstrates that explicit reasoning about resource allocations in the face of inequality develops across childhood. For example, when asked to evaluate distributions of resources between one individual with lots of resources and one individual with no resources, 3- and 4-year-olds judged equal distributions to be fair, 5- and 6-year-olds judged both equal and equitable distributions as fair, while 7- and 8-year-olds judged equal allocations less positively (Rizzo & Killen, 2016). Altogether, these studies show important age-related differences between infants’ implicit expectations and older children’s explicit judgments regarding morally-relevant content (i.e., how resources ought to be distributed amongst advantaged versus disadvantaged individuals).

The age-related development of children’s explicit judgments is also evident in their reasoning about moral dilemmas, in which multiple moral concerns conflict. To illustrate, consider scenarios in which a prototypical moral transgression is committed to prevent a greater harm (e.g., hitting/calling someone a bad name to prevent them from causing severe physical harm to a third party). While moral intuitions regarding the avoidance of harm may lead to negative evaluations of the transgressor, children offer increasingly forgiving evaluations of such transgressors between the ages of 5 and 11 years of age (Jambon & Smetana, 2014). These results suggest that, with age, children become increasingly able to integrate competing evaluations (e.g., negative evaluations of those who cause harm, positive evaluations of those who prevent harm) into their explicit moral judgments.

Such age-related changes in the ability to integrate conflicting evaluations into a single moral judgment may, in fact, account for preschoolers’ failure to report context-dependent
judgments of helpers and hinders in Chapter 4 of this dissertation. That is, for moral judgments of the hinderer to be more positive in the antisocial versus prosocial target conditions of Experiment 1 and 2, children needed to integrate their positive evaluation of the helper, their negative evaluation of the hinderer, and their (presumed) positive evaluation of those who punish appropriately. If explicit reasoning plays a role in these judgments, then it is plausible that preschoolers struggled to endorse hindering in the face of recently-made-explicit moral knowledge that helping is typically good and hindering is typically bad. With age, children may be more likely to report judgments favoring the hinderer of antisocial others (see Li & Tomasello, 2018 for evidence that 5-year-olds distinguish between prosocial and antisocial contexts when making social judgments); these changes may be due to domain-general increases in reasoning abilities, rather than changes in the moral evaluations elicited by these scenarios.

Note that infants are not likely to experience a similar conflict due to competing intuitions when faced with similar scenarios: If their implicit preference for hinderers of antisocial others simply arise based on moral intuitions, then no conscious conflict occurs to disrupt their expression of behavioral responses (i.e., reaching preferences) based on these intuitions.

**The Influence of Culture**

In addition to age-related changes due to maturation and experience, the culture in which children are raised plays a significant role in the development of moral judgment. One intriguing possibility is that cultural influences build upon the moral intuitions discussed above. In this view, moral intuitions are an innate, universal component of typically-developing human cognition (see Hamlin, 2013b). Consistent with this hypothesis is growing evidence that various populations of infants possess early-emerging preferences for prosocial (i.e., helpful, fair) individuals over antisocial (i.e., unhelpful, unfair) ones. Such preferences have been documented
in Canada (e.g., Steckler et al., 2017; see also Woo et al., 2017), France (Buon et al., 2014; Scola et al., 2015), Italy (Geraci & Surian, 2011), Japan (Shimizu, Senzaki, & Uleman, 2018; see also Kanakogi et al., 2017 for evidence that infants prefer those who intervene to protect third parties from harm), Korea (Chae & Song, 2018), and the United States (e.g., Burns & Sommerville, 2014; Hamlin et al., 2007; 2010; Hamlin & Wynn, 2011); but see Salvadori et al. (2015) for a failed replication in Hungary.

Importantly though, even if humans possess innate intuitions about certain moral concerns, these intuitions will be shaped by cultural norms. The influence of culture on moral development is evident in children’s moral behaviors (e.g., cross-cultural differences in the performance of costly prosocial actions and acts of distributive justice; House et al., 2011; Schäfer, Haun, & Tomasello, 2015) and moral judgments (e.g., cross-cultural differences in the development of perceptions regarding fair resource distributions; Blake et al., 2015). Such differences in moral development across cultures raise questions regarding the generalizability of the results described in this dissertation. For example, adults’ consideration of intentions when forming moral judgments differs cross-culturally (see Barrett et al., 2016). To illustrate, when presented with vignettes featuring accidents, failed attempts to harm, and intentional moral violations, Indigenous iTaukei Fijians from Yasawa Island (who normatively avoid inferences about others’ private mental states) placed more emphasis on outcomes than Indo-Fijians (who do not avoid mental state inferences) and North Americans (who tend to interpret behavior in terms of mental states) when judging permissibility and punish-worthiness (McNamara, Willard, Norenzayan, & Henrich, 2019). These differences in the use of intention information may likewise influence even young children’s social and moral judgments regarding the failed versus successful helpers and hinderers presented in Chapter 3 of this dissertation (but see McNamara,
2016 for evidence that both North American and Yasawan children’s preferences for failed/successful helpers and hinderers focus on intentions from middle-childhood).

Relevant to Chapter 4 of this dissertation is evidence that punishment behaviors vary across cultures. This includes differences in the punishment of antisocial others: For example, one study exploring the punishment behavior of adults in 15 diverse societies found that while at least some adults in all populations paid to punish unfair economic distributions between third parties, this willingness to engage in costly third-party punishment varied substantially across the cultures studied (Henrich et al., 2006). There is also substantial variation in responses to prosocial others, including the existence of antisocial punishment (i.e., the punishment of prosocial others) in some cultures. For instance, while university students around the world punished individuals who made minimal contributions to the public good in one economic experiment, participants in some cultures also punished high contributors. Suggestive that punishment behaviors reflect social norms, higher rates of antisocial punishment were associated with weak norms of civic cooperation and weak rule of law (i.e., low trust in law enforcement intuitions; Herrmann, Thöni, & Gächter, 2008). Such cross-cultural differences suggest that children’s context-dependent judgments of helpers and hinderers may vary between Western and non-Western populations. That is, if future studies reveal that Canadian preschoolers do in fact positively evaluate those who selectively help prosocial others and punish antisocial others (c.f. Chapter 4), then it is possible that such tendencies would develop/be expressed differently in different cultures. Future work needs to explore these possibilities by presenting young children in a variety of contexts with similar (though culturally-appropriate) tasks.
Concluding Remarks

In conclusion, this dissertation makes a meaningful contribution to our understanding of young children’s social and moral judgments of those who help versus hinder third parties. These studies also document important continuities and potential discontinuities between infant’s implicit evaluations of prosocial versus antisocial others and young children’s explicit judgments. Future studies should continue to explore the nature of children’s moral judgments, including the relative contribution of implicit and explicit processes, as well as the role of maturation, experience, and culture in the development of moral evaluations.
References


Gordon, D. S., Madden, J. R., & Lea, S. E. (2014). Both loved and feared: Third party punishers are viewed as formidable and likeable, but these reputational benefits may only be open to dominant individuals. *PloS ONE, 9*(10), e110045.


Appendix A

Social Preference and Moral Judgments

To visualize individual patterns of responding in Chapter 2, children’s judgments within each round were characterized as evidence for moral responding, social responding, moral and social responding, or another pattern of judgments. Moral responding refers to children who provided consistent moral judgments (judged the helper as nicer, and allocated punishment to the hinderer) but preferred the hinderer. Social responding refers to children who preferred the helper but did not provide consistent moral judgments (i.e., judging the helper or hinderer as both nicer and deserving of punishment; judging the hinderer as nicer and helper as deserving of punishment). Moral and social responding refers to children who preferred the helper, judged the helper as nicer, and allocated punishment to the hinderer. Finally, “other” patterns of responses included those who preferred the hinderer and provided inconsistent moral judgments, or those who selected the hinderer across the three test questions. See Table A.1 for the number of children providing each response pattern in Experiment 1 and 2.

Explanations Regarding Punishment

To visualize how children justified appropriate allocations of punishment in Chapter 1, we examined the explanations provided by children who responded that the hinderer, rather than the helper, should get in trouble. Table A.2 shows the proportion of each explanation type (uninformative, relevant action, irrelevant action, relevant valence, irrelevant valence, non-social considerations) within each round among only those children who allocated punishment to the hinderer.
Table A.1

*Number of children providing social and/or moral judgments in Experiment 1 and 2*

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Age</th>
<th>Round 1</th>
<th></th>
<th>Round 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Moral</td>
<td>Social</td>
<td>Other</td>
<td>Moral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and</td>
<td></td>
<td></td>
<td>and</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>11</td>
<td>2</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>17</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Table A.2

Proportion of explanations containing each response type in Experiment 1 and Experiment 2 among children who allocated punishment to the hinderer

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Age</th>
<th>Round 1</th>
<th>Round 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Uninformative</td>
<td>Informative</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relevant action</td>
<td>Irrelevant action</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>.59</td>
<td>.29</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>.11</td>
<td>.68</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>.04</td>
<td>.76</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>.16</td>
<td>.53</td>
</tr>
</tbody>
</table>

*Note.* The informative response categories were not mutually exclusive and proportions sum to 1.
Appendix B

Excluded Participants

Following a pilot study, we decided that in all experiments in Chapter 3 we would remove and replace children that indicated the same puppet was “liked”, “nicer”, and “should get in trouble” in one or both rounds of test questions. Indicating the same puppet across positively and negatively valenced test questions suggested that children were unmotivated and/or did not understand the test questions. Further, since neither the color of the puppets’ shirts (red, green) nor their location (right, left) was varied across test questions within the same child, these children’s responses may have reflected a side and/or color preference. Confirmatory analyses, including these children’s social and moral judgments, are reported here. Overall, response patterns are very similar when children with a side/color preference are included in the analyses.

Experiment 1

Methods

Participants. The full sample included 72 3-year-olds ($M = 3;6$, range = 3;0-3;11, 39 girls) and 56 4-year-olds ($M = 4;5$, range = 4;0-4;11, 32 girls).

Results

To explore whether responses differed before and after comprehension questions, we conducted a series of mixed-effect ANOVAs with round one and two scores as within-subjects variables, and age (3, 4) and gender (female, male) as between-subjects factors. When compared to a Bonferroni-corrected alpha value of .017 (.050/3), there was no main effects of round or interactions involving round of questioning on liking, niceness, or trouble scores in the negative outcome condition (all $Fs < 4.853$, all $ps > .030$, all $\eta^2 ps < .073$). However, in the positive
outcome condition, niceness scores were higher after comprehension questions ($M = .92, SD = .28$) versus beforehand ($M = .79, SD = .41; F[1,57] = 6.636, p = .013, \eta_p^2 = .104$; all other $Fs < 1.446$, all $ps > .233$, all $\eta_p^2$s < .026). Because round of questioning had no effect on niceness scores in other experiments and consistently had no effect on liking or trouble scores, children’s scores were summed across the two rounds resulting in three scores between 0 – 2 per child.

A series of one-sample $t$-tests comparing children’s liking, niceness, and trouble scores at each age to a chance score of one revealed the same pattern of results as in the main text. Three-year-olds in the positive outcome condition did not distinguish between the puppets when reporting who they liked ($M = 1.03, SD = .82; t[31] = .215, p = .831, d = .038$), while 4-year-olds liked the successful helper ($M = 1.38, SD = .78; t[28] = 2.635, p = .014, d = .489$). Both ages judged the successful helper to be nicer (3-year-olds: $M = 1.56, SD = .67; t[31] = 4.756, p < .001, d = .841$; 4-year-olds: $M = 1.86, SD = .44; t[28] = 10.524, p < .001, d = 1.954$) and allocated punishment to the failed hinderer (3-year-olds: $M = 1.66, SD = .60; t[31] = 6.171, p < .001, d = 1.091$; 4-year-olds: $M = 1.69, SD = .66; t[28] = 5.625, p < .001, d = 1.045$). Both ages in the negative outcome condition liked the failed helper (3-year-olds: $M = 1.38, SD = .84; t[39] = 2.831, p = .007, d = .448$; 4-year-olds: $M = 1.52, SD = .75; t[26] = 3.578, p = .001, d = .689$), judged the failed helper as nicer (3-year-olds: $M = 1.68, SD = .62; t[39] = 6.936, p < .001, d = 1.097$; 4-year-olds: $M = 1.67, SD = 62; t[26] = 5.586, p < .001, d = 1.075$), and allocated punishment to the successful hinderer (3-year-olds: $M = 1.30, SD = .82; t[39] = 2.306, p = .027, d = .365$; 4-year-olds: $M = 1.63, SD = .63; t[26] = 5.199, p < .001, d = 1.001$).
Experiment 2A

Methods

Participants. The full sample included 35 3-year-olds (M= 3;5, range = 3;0-3;11, 16 girls) and 26 4-year-olds (M= 4;6, range = 4;0-4;11, 10 girls).

Results

A series of mixed-effect ANOVAs explored whether responses differed before and after comprehension questions; this revealed no main effects of round or interactions involving round of questioning on liking, niceness, or trouble scores (Bonferroni-corrected alpha value of .017 [.050/3]; all Fs < 2.414, all ps > .125, all ηp²s < .042). Children’s scores were summed across the two rounds resulting in three scores between 0 – 2 per child (liking, niceness, trouble).

A series of one-sample t-tests comparing children’s liking, niceness, and trouble scores to a chance score of 1 revealed the same pattern of results as in the main text. Younger children did not distinguish between the puppets: 3-year-olds’ liking (M = 1.17, SD = .82; t[34] = 1.234, p = .226, d = .209), niceness (M = 1.17, SD = .75; t[34] = 1.358, p = .183, d = .230), and trouble (M = 1.09, SD = .78; t[34] = .649, p = .521, d = .110) scores did not differ from chance. In contrast, 4-year-olds liked the failed helper (M = 1.50, SD = .81; t[25] = 3.138, p = .004, d = .615), judged the failed helper as nicer (M = 1.73, SD = .67; t[25] = 5.588, p < .001, d = 1.096), and allocated punishment to the failed hinderer (M = 1.73, SD = .60; t[25] = 6.171, p < .001, d = 1.210).

Experiment 2B

Methods

Participants. The full sample included 34 3-year-olds (M= 3;6, range = 3;0-3;11, 16 girls) and 29 4-year-olds (M= 4;5, range = 4;0-4;11, 13 girls).
Results

A series of mixed-effect ANOVAs explored whether responses differed before and after comprehension questions; this revealed no main effects of round or interactions involving round of questioning on liking, niceness, or trouble scores (Bonferroni-corrected alpha value of .017 [.050/3]; all $F$s < 5.441, all $p$s > .022, all $\eta^2_p$s < .085). Children’s scores were summed across the two rounds resulting in three scores between 0 – 2 per child (liking, niceness, trouble).

A series of one-sample t-tests comparing children’s liking, niceness, and trouble scores at each age to a chance score of one revealed a similar pattern of results as in the main text. As in the main text, 3-year-olds’s liking scores ($M = 1.26, SD = .82$) do not differ from chance ($t[33] = 1.864, p = .071, d = .320$). However, while in the main text 4-year-olds also showed no preference for either puppet, we now see that 4-year-olds prefer the failed helper ($M = 1.31, SD = .81; t[28] = 2.073, p = .048, d = .385$). As in the main text, both ages reliably judged the failed helper to be nicer (3-year-olds: $M = 1.53, SD = .61; t[33] = 5.022, p < .001, d = .861$; 4-year-olds: $M = 1.79, SD = .56; t[28] = 7.636, p < .001, d = 1.418$) and allocated punishment to the failed hinderer (3-year-olds: $M = 1.38, SD = .74; t[33] = 3.016, p = .005, d = .517$; year-olds: $M = 1.83, SD = .47; t[28] = 9.519, p < .001, d = 1.768$).

Experiment 3

Participants

The full sample included 73 3-year-olds ($M= 3;6, \text{ range } = 3;0-3;11, 36 \text{ girls}$) and 65 4-year-olds ($M= 4;5, \text{ range } = 4;0-4;11, 32 \text{ girls}$).
Results

A series of mixed-effect ANOVAs explored whether responses differed before and after comprehension questions; this revealed no main effects of round or interactions involving round of questioning on liking, niceness, or trouble scores in the positive (Bonferroni-corrected alpha value of .017 [.050/3]; all $F$s < 2.666, all $p$s > .106, all $\eta^2$s < .040) or negative intention condition (all $F$s < 2.900, all $p$s > .093, all $\eta^2$s < .044).

A series of one-sample t-tests comparing children’s liking, niceness, and trouble scores at each age to a chance score of one revealed a similar pattern of results to those reported in the main text. While in the positive intention condition of the main text, 3-year-olds distinguished between the successful helper and the failed helper for one test question and judged the successful helper to be nicer, they did not distinguish between the puppets for any test question in the full sample: 3-year-olds responded at chance when asked which puppet was liked ($M = .92, SD = .83; t[36] = .595, p = .556, d = .098$), nicer ($M = 1.24, SD = .76; t[36] = 1.946, p = .059, d = .320$), and should receive punishment ($M = 1.11, SD = .81; t[36] = .813, p = .422, d = .134$). As in the main text, 4-year-olds in the positive intention condition liked the successful helper ($M = 1.42, SD = .79; t[32] = 3.078, p = .004, d = .536$), judged the successful helper to be nicer ($M = 1.39, SD = .75; t[32] = 3.028, p = .005, d = .527$), and allocated punishment to the failed helper ($M = 1.36, SD = .70; t[32] = 2.988, p = .005, d = .520$).

As in the main text, children did not differentiate between the puppets for any test questions in the negative intention condition: 3- and 4-year-olds’ liking (3-year-olds: $M = 1.08, SD = .84; t[35] = .595, p = .556, d = .099$; 4-year-olds: $M = .781, SD = .75; t[31] = 1.648, p = .109, d = .291$), niceness (3-year-olds: $M = 1.08, SD = .77; t[35] = .649, p = .520, d = .108$; 4-
year-olds: $M = 1.00, SD = .80; t[31] = .000, p = 1.000, d = .000$), and trouble scores (3-year-olds: $M = .94, SD = .71; t[35] = .466, p = .644, d = .078$; 4-year-olds: $M = 1.00, SD = .80; t[31] = .000, p = 1.000, d = .000$) did not differ from chance.

Results by Round

Table B.1 presents the number of children at each age in each condition answering in the direction of the hypothesis in round one (before answering comprehension questions) and in round two (after answering comprehension questions). These children did not display a color/side preference. To determine whether responses deviated from chance, a binomial p value was calculated for each question; significant values are noted below.
Table B.1

Number of with hypothesis responses before and after comprehension questions

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Condition</th>
<th>Age</th>
<th>N</th>
<th>Round 1</th>
<th>Round 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Liking Question</td>
<td>Niceness Question</td>
</tr>
<tr>
<td>1</td>
<td>Positive outcome</td>
<td>3</td>
<td>24</td>
<td>15</td>
<td>19**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>24</td>
<td>16</td>
<td>23***</td>
</tr>
<tr>
<td></td>
<td>Negative outcome</td>
<td>3</td>
<td>26</td>
<td>16</td>
<td>22**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>24</td>
<td>17</td>
<td>20**</td>
</tr>
<tr>
<td>2A</td>
<td></td>
<td>3</td>
<td>24</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>24</td>
<td>18*</td>
<td>22***</td>
</tr>
<tr>
<td>2B</td>
<td></td>
<td>3</td>
<td>24</td>
<td>10</td>
<td>19**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>25</td>
<td>16</td>
<td>23***</td>
</tr>
<tr>
<td>3</td>
<td>Positive intention</td>
<td>3</td>
<td>24</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>26</td>
<td>20*</td>
<td>19*</td>
</tr>
<tr>
<td></td>
<td>Negative intention</td>
<td>3</td>
<td>23</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>27</td>
<td>11</td>
<td>14</td>
</tr>
</tbody>
</table>

Note. Binomial $p^* < .05$; $** < .01$; $*** < .001$. 