INVESTIGATING THE PSYCHOLOGICAL FOUNDATIONS OF MORAL JUDGMENT

by

KATHARINE SAUNDERS

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Psychology

written under the direction of

Alan M. Leslie

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

JANUARY, 2014

ABSTRACT OF THE DISSERTATION

Investigating the Psychological Foundations of Moral Judgment

By KATHARINE SAUNDERS

Dissertation Director:

Alan M. Leslie

Is there an early developing neuro-cognitive structure that is specific to our moral sense? Recent research has begun to explore this question using a classic thought experiment known as the trolley problem. These "trolley studies" have uncovered what appears to be a universal pattern of moral intuitions in adults that some argue can only be explained by assuming implicit knowledge of complex moral principles. In this dissertation, I build on this work by testing preschoolers' and adults' tacit knowledge of the principle of double effect – a principle that has a long history within the fields of philosophy, religion, and law, and which has recently been proposed to underlie our moral intuitions in the trolley problem. I also investigate the role of perceived ingroup/outgroup structure in moral judgment – a factor which others have hypothesized to be a foundation of moral judgment.

Across three studies, preschoolers (studies 1 and 2) and adults (study 3) were tested on a series of dilemmas that were similar in structure to the traditional trolley problems, but involved property violations and assault (i.e. the apprehension of bodily

harm) rather than "personal" violations such as battery or homicide. In all three studies, participants showed a strong and stable pattern of intuitions consistent with the principle of double effect: dilemmas in which an individual was harmed as a foreseen side effect of saving five people were judged favorably, but dilemmas in which an individual was intentionally harmed as a means to saving five people were judged unfavorably. Four-year-olds and adults (but not three-year-olds) also disapproved of scenarios in which an agent knowingly allowed a preventable harm to occur. Manipulations of minimal ingroup/outgroup structure had little to no effect on either preschoolers' or adults' moral judgments in these dilemmas. Implications for the structure and development of moral judgment are discussed.

Acknowledgements and Dedication

I would like to thank my advisor, Alan M. Leslie, for his invaluable support, patience, and guidance throughout my graduate career. He has made me a better thinker, a better writer, and a better scientist, and he has always managed to make me laugh, even during the darker moments of my disser-pression. Special thanks to Lu Wang, Sydney Levine, Michelle Cheng, Melissa Kibbe, Deena Weisberg, Alex D'Esterre, Anton Scherbakov, Jennifer Jacobs, and all the members of the Cognitive Development Lab for their assistance in completing the research presented in this dissertation. I would also like to thank my committee members, Rochel Gelman and Gretchen Chapman, and John Mikhail for their help and support.

This dissertation is dedicated to my husband and cheerleader, Patrick DeSomma. Words cannot express my appreciation for his constant words of encouragement, his endless patience, and his unconditional love and support during this process (not to mention his expert culinary and computer skills, of which I shamelessly took advantage on many occasions). Thank you. I love you. You are my rock.

This research was supported by National Science Foundation grants BCS-0725169 and BCS-0922184.

Table of Contents

List of Tables

List of Figures

I. Introduction

The question of whether we are endowed with an innate intuitive moral sense has long been a subject of debate within the field of ethical philosophy. However, only recently has this question become a topic of empirical scientific inquiry for moral psychologists. Over the last few decades, a growing body of work has begun to explore the origins of our moral intuitions (i.e. the moral judgments that arise spontaneously and automatically, as distinct from judgments that are the result of conscious deliberative reasoning). Where do these intuitions come from? How (and when) do we begin to evaluate acts in terms of their deontic status (i.e. in terms of concepts such as *right*, *wrong*, *permissible*, *obligatory*, *forbidden*), and what are the cognitive processes involved in such judgments?

Much of this work has taken advantage of dilemmas developed by moral philosophers to investigate these questions. In particular, many studies have used a classic thought experiment known as the trolley problem (Foot, 1967) to test the moral intuitions of adults across a range of cultural backgrounds (e.g. Cushman, Young, & Hauser, 2006; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Hauser, 2006; Hauser, Cushman, Young, Jin, & Mikhail, 2007; Mikhail, 2002; O'Neill & Petrinovich, 1998; Petrinovich, O'Neill, & Jorgensen, 1993). These studies have uncovered what appears to be a universal pattern of moral intuitions that gives a critical role to the intentional and causal properties of action. In this dissertation, I build on this work by presenting a series of studies that suggest that preschoolers' moral judgments already bear at least some of the nuances that appear to be specific signatures of adult moral intuitions. Specifically, I show that both children and

adults show the same pattern of moral intuitions on a series of developmentally appropriate variations of the so-called "trolley problem."

In this Chapter (Chapter 1), I will begin by discussing the trolley problem, and why it is a particularly useful tool for investigating moral judgment in both adults and children. I will then discuss two theoretical approaches to studying moral intuition, and how each of these approaches has attempted to explain the pattern of intuitions observed in a class of trolley problems. In Chapter 2, I will turn to the developmental literature on children's moral cognition, and highlight some of the questions that the current literature leaves unanswered. In the remaining chapters, I will present a series of experiments I have conducted to address some of these questions. In Chapter 3, I ask whether preschoolers exhibit a pattern of intuitions consistent with the principle of double effect, a principle which has been hypothesized to underlie adult moral judgments. I also examine whether preschoolers consider it morally permissible to knowingly allow a preventable harm to occur. In Chapter 4, I ask whether the pattern of intuitions observed in Chapter 3 is susceptible to minimal ingroup/outgroup bias. Does it matter whether the recipient(s) of a harmful act belong to the child's ingroup or outgroup? In Chapter 5, I test adults on the same dilemmas that were presented to preschoolers in the previous chapter, as well as an additional set of dilemmas designed to test another potential source of ingroup bias in adult moral judgment: the identity of the moral agent. In Chapter 6, I summarize my findings and discuss how they inform our understanding of moral cognition and development.

1. The Trolley Problem

The trolley problem, a thought experiment first devised by Philippa Foot (1967), has recently become a popular experimental tool in moral psychology. Like most moral dilemmas, the trolley problem presents a fictional agent with a moral conflict: he must choose between taking an action that would result in both good and bad effects, and omission of that action. For example, in the "bystander" version of the trolley problem, a bystander sees that a runaway trolley is about to run over five people who are tied to the train track. If the bystander does nothing, the five people will be killed. Alternatively, the bystander can flip a switch that will divert the trolley onto a side track where only one person is standing. If he diverts the trolley onto the side track, the one person will die, but the five people will be saved.

This task provides a particularly powerful tool for studying moral judgment for several reasons. First, similar to judgment tasks in other areas of cognition (e.g. Chomsky's novel sentence tasks), it elicits a strong, spontaneous, and intuitive response to stimuli that are novel, contrived, and unfamiliar (Mikhail, 2002; Hauser et al., 2007a). Thus, unlike real-world dilemmas such as euthanasia or abortion, the trolley problem enables researchers to identify the kinds of stable intuitions that go beyond cultural norms, political and religious ideology, or personal preference. Second, because of its artificial nature, the trolley problem lends itself to careful and systematic manipulation. By varying individual parameters in this dilemma, we can begin to identify the universally salient factors that affect moral judgment, and the moral principles that may be operative in people's judgments. For example, consider the "footbridge" version of the trolley problem (Thomson, 1985): a bystander is standing on a footbridge overlooking a single train track. In order to stop the runaway trolley from killing the five people on

the train track, the bystander must push the large man standing next to him off the footbridge and in front of the oncoming train. The weight of the large man will stop the train, saving the five people but killing the large man.

Although the act of pushing the man in the footbridge problem results in the same outcome as flipping the switch in the bystander problem – five people are saved and one person dies – healthy adults across a wide range of demographic and socioeconomic groups find it morally permissible to throw the switch in the bystander problem, but not to push the man in the footbridge problem (Cushman et al., 2006; Greene et al., 2001; Greene et al., 2004; Hauser et al., 2007a). Furthermore, when asked to explain this pattern of intuitions (which will hitherto be referred to as the *double-effect effect* in this paper), participants are typically unable to do so (Cushman et al., 2006; Hauser et al., 2007a; O'Neill & Petrinovich, 1998; Petrinovich et al., 1993). In other words, although adults clearly see a moral distinction between these two dilemmas, they have difficulty identifying what that distinction might be.

These findings have several important implications. First, they suggest that at least some of our moral intuitions may reflect deep, widely shared moral principles that hold across a variety of populations. Second, the fact that adults are unable to explain their intuitions suggests that these principles are not open to conscious introspection. Third, while it is difficult to identify the specific moral principle or principles responsible for the double-effect effect (indeed, this is an ongoing subject of speculation among psychologists and philosophers), we can reject certain hypotheses outright. Deontological principles such as "killing is wrong," or utilitarian principles such as "maximize the overall good" fail to account for the moral distinction between the

bystander and footbridge problems, since the same number of people are killed and saved in each scenario. Conditional principles based on the action-descriptions present in the stimulus (e.g. If an act is of the type "pushing a person," then it is forbidden) are also unlikely candidates, as one can easily imagine a scenario in which pushing a person is morally permissible (e.g. pushing a man out of the way of an oncoming train), and scenarios in which throwing a switch is morally impermissible (e.g. throwing a switch to redirect a train away from one person towards five people) (Mikhail, 2011).

These observations lead to the conclusion, as articulated by John Mikhail (2011), that a model of moral judgment that adequately describes the double-effect effect "must be more elaborate and must involve complex, structure-dependent rules, whose basic operations are defined in relation to abstract categories that are only indirectly related to the stimulus" (p. 81). Specifically, Mikhail proposes that a number of complex, universal, and possibly innate moral principles may be systematically guiding our moral judgments in these dilemmas below the level of conscious awareness. These principles make up what Mikhail refers to as Universal Moral Grammar (UMG): a moral faculty of the mind/brain that enables us to determine the deontic status of acts and omissions according to their causal and intentional structure (analogous in many respects to Chomsky's Universal Grammar). This theory will inform much of the work I present in Chapters 3-5. In section 2, I will discuss this theory in greater detail, and outline the ways in which this theory accounts for the pattern of intuitions observed in a class of trolley problems. In section 3, I will discuss an alternative approach to understanding moral judgment that gives a privileged role to emotion.

2. Universal Moral Grammar

Mikhail's theory of UMG draws its inspiration primarily from the works of John Rawls and Noam Chomksy. In *A Theory of Justice* (1971), Rawls argued that moral philosophy can benefit from exploring the potential similarities between generative linguistics and morality. Following this "linguistic analogy," Mikhail has proposed a computational framework for studying moral judgment that draws on the concepts outlined in Chomksy's theory of Universal Grammar (UG) (Chomsky, N. 1965) (see also (Dwyer, 1999 and Hauser, 2006 for related theories). Under this framework, the moral psychologist is primarily concerned with uncovering the internally represented rules and operations that constitute our moral knowledge, and the mechanisms by which this knowledge is acquired.[1]

Like Chomsky's UG, UMG begins with a basic assumption that the mind is equipped with a universal mechanism that guides our learning about morality according to an innate set of abstract principles, rules and concepts. This mechanism endows us with "a natural readiness to compute mental representations of human acts and omissions in legally cognizable terms" (Mikhail, 2011, p. 101) by identifying which aspects of the environment are morally relevant. In many ways, this moral system or "moral faculty" is analogous to the human language faculty proposed by Chomsky (figure 1.1). According to Chomsky, the human language faculty provides the language learner with a set of core principles and basic assumptions about the universal structure of language. Although environmental input plays a significant role in language acquisition, an innate structure must be in place to guide acquisition of the language to which the child is exposed. This

[1] In some respects, this theory is also similar to the theories of Kant (1785/1964) and Kohlberg (1981, 1984), in that it emphasizes the role of dispassionate, principled reasoning in moral judgment. However, in contrast to Kant and Kohlberg, UMG assumes that such reasoning is intuitive (i.e. unconscious) rather than deliberative.

innate structure allows the child to acquire language rapidly, even when exposed to impoverished input, by constructing a *generative grammar*, defined by Chomsky as a finite set of recursive rules that allow a speaker to generate and understand an infinite number of grammatical sentences in her language.

Like the language faculty, the hypothesized moral faculty is also comprised of a set of universal and abstract principles that facilitate the acquisition of a *moral grammar*. In this case, the moral grammar, as defined by Mikhail, refers to "a complex and possibly domain-specific set of rules, concepts and principles that generates and relates mental representations of various types...[and] enables individuals to determine the deontic status of a potentially infinite number and variety of acts and omissions" (Mikhail, 2007, p. 238). Thus, just as Chomsky's generative grammar (i.e. competence) enables speakers to intuitively produce and understand an infinite number of original utterances in their native language (i.e. performance), the moral grammar allows us to form fluent, intuitive moral judgments about an infinite number of complex moral events, many of which we have never before encountered.

In this way, moral judgments (e.g. judgments of whether an act or omission is permissible or impermissible) are analogous to judgments of grammaticality. Chomsky and other linguists have demonstrated that a native speaker's ability to distinguish between grammatical and ungrammatical speech (e.g. *colorless green ideas sleep furiously*, versus *furiously sleep ideas green colorless* (Chomsky, 1957)) is rapid, automatic, and intuitive. In most cases, although we can easily identify an ungrammatical utterance, we are unable to articulate the principles underlying this knowledge, as they are often inaccessible to conscious thought (i.e. they are *operative,*

but not *express* principles (Chomsky, 1965). Similarly, according to UMG, our moral intuitions also rely on principles that operate below our conscious awareness.

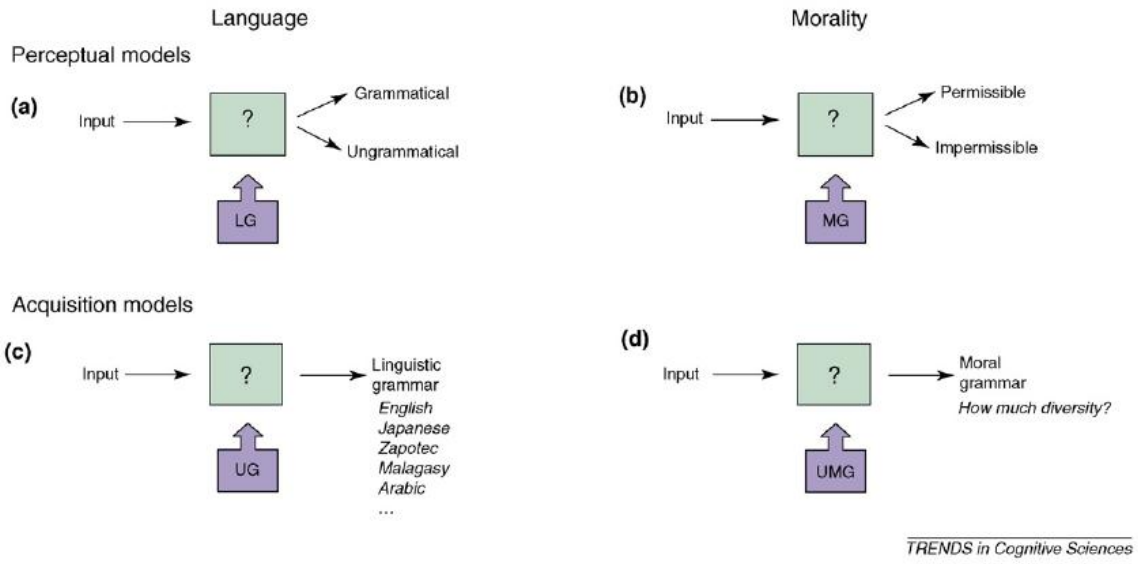

Figure 1.1. "Simple perceptual and acquisition models for language and morality." Reprinted from "Universal Moral Grammar: Theory, Evidence, and the Future" by J. Mikhail, 2007, *Trends in Cognitive Sciences, 11*(4), p. 145.

2.1 The trolley problem according to UMG. As discussed in section 1, the trolley problems provide an ideal method for examining which principles or rules guide our judgments. Much like Chomsky's novel sentence examples, the trolley problems elicit spontaneous and intuitive responses to stimuli that are novel, contrived, and unfamiliar. Furthermore, these responses appear to be "stable stringent, and highly predictable" (Mikhail, 2011, p. 83), with little variation across differences in culture, nationality, ethnicity, race, gender, religion, age, or educational level (Hauser et al., 2007a; Mikhail, 2002; O'Neill & Petrinovich, 1998; Petrinovich et al., 1993). Furthermore, subtle manipulations of trolley problems that are otherwise identical produce drastically different intuitions, which participants often have difficulty explaining, suggesting that operative principles may be implicitly guiding participants' judgments in these scenarios (Cushman et al., 2006; Hauser et al., 2007a; Mikhail, 2002).

With this in mind, I return to the two trolley problems I introduced in section 1: the bystander problem and the footbridge problem. Why do participants make a moral distinction between these two very similar cases? According to Mikhail, participants' intuitions in these dilemmas can be explained by postulating two operative principles: the prohibition of intentional battery, and the principle of double effect (PDE). The prohibition of intentional battery is a common legal principle which "forbids purposefully or knowingly causing harmful or offensive contact with another individual or otherwise invading another individual's physical integrity without his or her consent" (Mikhail, 2007, p. 145). The principle of double effect is a long-established normative principle that was first formulated by Thomas Aquinas (1274/1988) in answer to the question of whether it is ever morally permissible to knowingly cause harm. In its most familiar form (and for the purposes of this dissertation), the PDE states:

> An otherwise prohibited action, such as battery, that has both good and bad side effects may be permissible if the prohibited act itself is not directly intended, the good but not the bad effects are directly intended, the good effects outweigh the bad effects, and no morally preferable alternative is available (Mikhail, 2007, p. 145. Also see Cushman et al., 2006; Fischer & Ravizza, 1992). [2]

If we apply these principles to the bystander and footbridge problems, as illustrated in figure 1.2, pushing the man in the footbridge problem is impermissible because the battery is intended as the *means* to saving the five people. In contrast, flipping the switch in the bystander problem is permissible because the battery of the man on the side track is merely an unavoidable *side effect* of preventing harm to five others.

---

[2] In his book *Elements of Moral Cognition* (2011), Mikhail expands on the formal definitions of certain concepts in each of these principles, and postulates the following additional principles as part of the moral grammar: the principle of naturally liberty, the prohibition of intentional homicide, the self-preservation principle, a moral calculus of risk, and the rescue principle. Although I will not be discussing these principles here, I will return to the last principle (the rescue principle) in Chapters 3-5.

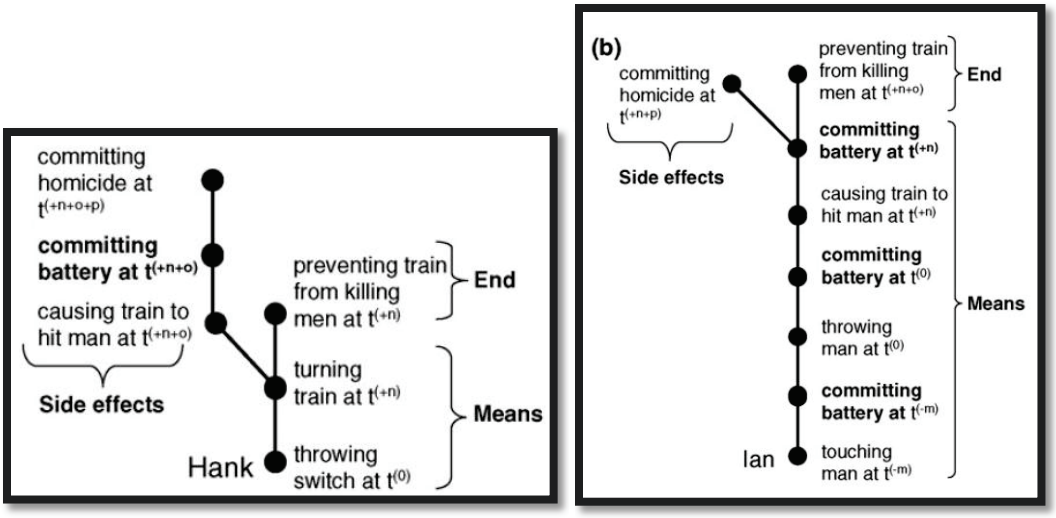

Figure 1.2. Structural Descriptions of Action Plans in Bystander and Footbridge Problems. Adapted from "Universal Moral Grammar: Theory, Evidence, and the Future" by J. Mikhail, 2007, *Trends in Cognitive Sciences, 11*(4), p. 150.

Mikhail proposes that a primary feature of the moral faculty is the ability to compute structural descriptions of acts/omissions (like the structural descriptions represented in figure 1.2) in terms of these causal and intentional properties (i.e. means, ends, side effects). But how does the mind/brain accomplish this? After all, these structural descriptions are not inherent in the stimuli themselves; they are mental representations – "a pattern of organization that is imposed on the stimuli by the mind itself" (Mikhail, 2007, p. 145). For example, the intentions of the bystander are never explicitly stated in the trolley problem; we must *infer* from the stimuli (in this case, written text) what effects of his action were intended effects (and which of the intended effects was his end or goal versus means to that end), and what effects were foreseen but unintended side effects. How does the mind recover properties like ends, means, and side effects from the stimulus?

According to Mikhail's model, this is achieved via a sequence of operations or "conversion rules," the details of which he fleshes out in meticulous detail in his book *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment* (2011). These conversion rules include:

> (i) identifying the various action descriptions in the stimulus, (ii) placing them in an appropriate temporal order, (iii) decomposing them into their underlying causative and semantic structures, (iv) applying certain moral and logical principles to these underlying structures to generate representations of good and bad effects, (v) computing the intentional structure of the relevant acts and omissions by inferring (in the absence of conflicting evidence) that agents intend good effects and avoid bad ones, and (vi) deriving representations of morally salient acts like battery and situating them in the correct location of one's act tree. (Mikhail, 2007, p. 146).

Finally, Mikhail assumes that the behavioral outputs of our moral system, permissibility judgments, reflect our tacit knowledge of "the basic principles of deontic logic" (Mikhail, 2011, 124). The fact that all natural languages appear to contain words or phrases to express deontic concepts like *obligatory*, *permissible*, and *forbidden* (figure 1.3a) supports this view (Bybee & Fleischman, 1995; Mikhail, 2011). Furthermore, the logical relations between these concepts can be represented in a "square of opposition and equipollence" (figure 1.3b), which suggests that these concepts may be reduced to only one "deontic primitive."[3] This primitive, together with the concepts of act ('A') and omission ('not-A'), constitute what Mikhail refers to as a formalizable deontic logic – the deontic component of moral competence (Mikhail, 2007).

[3] In his model, Mikhail assumes the deontic primitive is the concept *forbidden*.

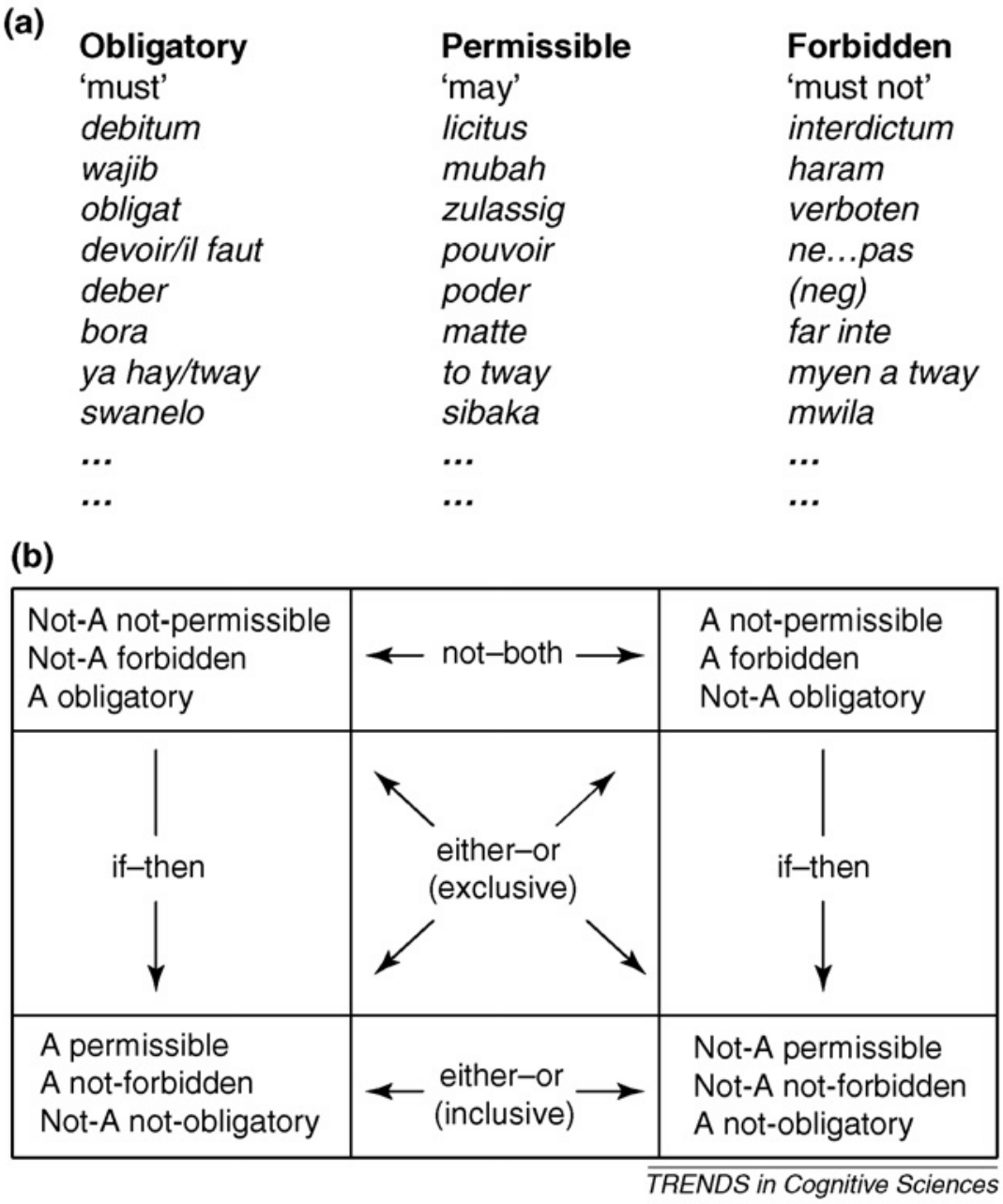

Figure 1.3. "Three deontic concepts (a) and square of opposition and equipollence (b)." Adapted from "Universal Moral Grammar: Theory, Evidence, and the Future" by J. Mikhail, 2007, *Trends in Cognitive Sciences, 11*(4), p. 144.

In sum, according to Mikhail's moral grammar hypothesis, our mature moral knowledge consists of conversion rules that allow us to compute structural descriptions of a given stimulus, the deontic principles that operate over those structural descriptions, and a formalized deontic logic that enables us to assign a deontic status to a given

act/omission in the stimulus. In section 2.2 I discuss some of the evidence that supports this theory.

2.2 Evidence for UMG. Do the operative principles Mikhail hypothesizes to be part of the moral grammar (the prohibition of intentional battery, and the PDE) accurately predict participants' moral judgments? To test this hypothesis, starting in the 1990s, Mikhail and colleagues conducted a series of studies investigating participants' responses to a set of 12 carefully controlled trolley problems (see Appendix A for the complete text of these scenarios) (Mikhail, 2002; Mikhail, Sorrentino & Spelke, 1998). They found that both the prohibition of intentional battery and the PDE correctly predicted the moral intuitions of a diverse sample of participants on all 12 trolley problems. Participants judged acts as permissible only if all four conditions of the PDE were met:

1) the prohibited act itself (in this case, battery) is not directly intended
2) the good but not the bad effects are directly intended
3) the good effects outweigh the bad effects
4) no morally preferable alternative is available

And participants judged forceful contact as impermissible only if it violated the prohibition of intentional battery (i.e. if it was non-consensual). For example, participants judged the implied consent scenario, in which the bystander pushes the man out of the way of the train, as permissible.

Furthermore, Mikhail and colleagues found that the act trees they outlined for each trolley problem not only accurately predicted and explained participants' moral intuitions in all 12 trolley problems, they also explained the *variance* of permissibility judgments in some of the trolley problems. For example, participants' permissibility

judgments across six of the trolley problems increased linearly as a function of how many counts of battery the moral agent committed prior to as a means to his goal. Mikhail et al. also found that when they asked participants to justify their moral judgments of these dilemmas, few participants were capable of providing "logically adequate" justifications for their intuitions, even when "logically adequate" justifications were leniently defined to include any justifications that "state[d] a reason, rule, or principle, or...otherwise identif[ied] at least one feature of the given scenario – even one that was obviously immaterial, irrelevant, arbitrary, or ad hoc – that could in principle generate the corresponding judgment" (Mikhail, 2002, p. 22). In other words, their results supported the hypothesis that these principles are non-introspectable.

In 2006, Cushman, Young, and Hauser conducted a study with similar aims in mind. The study consisting of 18 pairs of carefully controlled trolley problems designed to test participants' conscious versus intuitive knowledge of three principles:

- *The action principle*: Harm caused by action is morally worse than equivalent harm caused by omission.
- *The intention principle*[4]: Harm intended as the means to a goal is morally worse than equivalent harm foreseen as the side effect of a goal.
- *The contact principle*[5]: Using physical contact to cause harm to a victim is morally worse than causing equivalent (Cushman et al., 2006, p. 1083)

They found that while each of the three principles independently and reliably predicted participants' judgments, only the action principle and the contact principle were accessible to conscious introspection; less than a third of participants were able to explain

---

[4] Cushman et al.'s (2006) definition of the intention principle is similar to Mikhail's formulation of the PDE, but focuses exclusively on the first two of its four conditions: the prohibited act itself is not directly intended, and the good but not the bad effects are directly intended.

[5] It should be noted that unlike Mikhail's definition of battery, Cushman et al.'s (2006) definition of physical contact was restricted to direct physical contact between agent and patient, and did not include the use of an object or instrument to commit battery. They also did not include any cases where contact was granted or implied.

their pattern of responses when their judgments differed according to the intention principle (i.e. the PDE). Furthermore, although the majority of participants *were* able to explain their judgments in the case of the action/omission principle and the contact principle, many participants expressed doubt that physical contact *should* be a morally relevant factor. In other words, despite the fact that their intuitions reflected a moral distinction between using physical contact to cause harm and causing equivalent harm without physical contact, upon reflection, participants often felt that this distinction was morally invalid. Thus, the authors tentatively suggest that while the contact principle was implicitly guiding participants' moral judgments during the task, they only became aware of it through the process of post-hoc reasoning, at which point they rejected it as morally invalid. These findings suggest that while both the contact principle and the act/omission principle are expressed principles (i.e. accessible after conscious introspection), the PDE is an operative principle – one that guides our moral intuitions beneath the level of conscious awareness.

A subsequent study by Hauser, Cushman, Young, Jin, and Mikhail (2007) which focused exclusively on the PDE provided additional support for this hypothesis. Using a web-based technology, they collected judgments and justifications from 5,000 participants covering 120 different countries – the largest and most diverse sample in any trolley study to date. Participants were asked to evaluate four trolley problems of critical interest to the researchers: the bystander problem, the footbridge problem, and two problems that were designed to differ only in terms of whether the battery to one person was intended as the means to saving five people (the loop track problem), or was merely a foreseen side effect (the man-in-front problem) (see Appendix A for a description of

these dilemmas). Thus, the latter pair of dilemmas (borrowed from Mikhail's original 12) attempted to isolate the PDE from other potentially relevant factors in the other two dilemmas, such as the action-description of the basic act, the degree of physical contact between the bystander and the victim, the temporal order of the good and bad effects, whether the act was personal or impersonal (as defined by Greene et al., 2001), and whether a new threat was introduced or an existing threat was redirected. The researchers found that the PDE consistently predicted participants' pattern of judgments across all four dilemmas, regardless of gender, age, ethnicity, religion, national affiliation, education level, and exposure to moral philosophy. Furthermore, participants were again unable to provide sufficient justifications for their judgments, providing further support for the moral grammar hypothesis.

Taken together, these findings suggest that adults universally possess unconscious knowledge of at least two operative principles: the prohibition of intentional battery, and the PDE. However, this interpretation of the data is controversial in that it is distinctly rationalist. In Mikhail's own words:

> The critical issue…is not whether moral intuitions are linked to emotions – clearly they are – but how to characterize the appraisal system those intuitions presuppose, and in particular whether that system incorporates elements of a sophisticated jurisprudence (Mikhail, 2011, p. 39).

In the following section (Section 3), I outline alternative theories that give a critical causal role to emotion in moral judgment. I will discuss some of the limitations of these theories, and why Mikhail's UMG theory is a useful framework from which to proceed when investigating moral development. In Chapter 2, I will address some of the developmental questions raised by Mikhail's UMG theory, and the questions that remain unanswered in the literature.

3. Emotion theories

Emotion theories draw their inspiration from eighteenth century philosopher David Hume (1739-1740/1978), who argued that moral judgments were the result of our "passions," rather than of reason. For Hume and many contemporary emotion theorists, we arrive at our moral judgments without any kind of principled reasoning. Instead, "moral emotions," defined by Haidt (2003) as "those emotions that are linked to the interests or welfare either of society as a whole or at least of persons other than the judge or agent" (p. 853), are the primary source of our moral intuitions (e.g. Damasio, 1994; Greene & Haidt, 2002; Haidt, 2001, 2003; Moll & de Oliveira-Souza, 2007; Nichols, 2004; Prinz, 2004, 2007).

3.1 Behavioral and neurobiological evidence for moral emotions. Much of the behavioral evidence in support of this claim comes from emotional priming studies showing a relationship between moral judgment and feelings of disgust. For example, a series of studies by Schnall, Haidt, Clore, and Jordan (2008) found that moral judgments were more severe when participants were primed with disgust-inducing stimuli, such as a bad smell, a dirty room, a memory of a physically disgusting experience, or a disgusting film clip (Schnall et al., 2008). Yet another experiment by Wheatley and Haidt (2005) found that when participants were hypnotically primed to feel disgust when they encountered a neutral target word, they tended to judge moral transgressions more severely when the target word was embedded in the moral vignette.

Many emotion theorists explain participants' particular sensitivity to disgust by appealing to an evolutionary theory of "moral disgust" (e.g. Haidt, McCauley, & Rozin, 1994; Haidt, Rozin, McCauley, & Imada, 1997; Lerner, Small, & Loewenstein, 2004;

Miller, 1997; Moll et al., 2005; Rozin, 1997; Rozin & Fallon, 1987; Rozin, Haidt, & McCauley, 2008; Rozin, Lowery, Imada, & Haidt, 1999; Wheatley & Haidt, 2005). They argue that what originated as an instinctive oral distaste for harmful/poisonous food evolved into a "core disgust" for offensive sensory properties of natural substances such as bodily fluids and excretions, and eventually, into higher-order forms of disgust such as a "moral disgust" for sex-related acts, violations of the body, death, degrading or polluting influences, and contact with morally contaminated objects, people, or social groups, etc.

Additional evidence for this claim comes from neurobiological data on brain regions associated with both emotional and moral stimuli. For example, Jorge Moll and colleagues found that overlapping brain regions – namely the lateral and medial orbitofrontal cortex – were recruited for two different domains of disgust: pure disgust ("disgust devoid of moral context") and moral disgust/indignation (Moll et al., 2005). Another study by Moll and colleagues found that participants who were scanned while viewing pictures of morally-relevant stimuli and pictures of non-moral but emotionally salient stimuli showed activation of the anterior insula, amygdala, and subcortical structures for both types of stimuli (Moll et al., 2002).

The behavioral and neurobiological evidence described above suggests that the brain regions responsible for emotional and moral processing are intimately linked. These findings on their own, however, provide insufficient evidence for the claim that that emotions *cause* moral judgments. As Huebner, Dwyer, and Hauser (2009) note, it is possible that emotional priming may simply produce an additive effect, such that it enhances the severity of a transgression that has already been judged as morally wrong.

Negative emotions such as disgust might also simply serve to draw our attention to morally salient features of the environment, such as ends, means, and side effects. On the other hand, emotional priming might interfere or distort judgments generated by the moral system (Mikhail, 2011). Either way, these data are not sufficient for assigning a causal role to emotion in moral judgment. While the neurobiological data clearly indicate that emotions like disgust are related to moral judgment, this relationship may be correlational, not causal.[6]

3.2 Neuropsychological evidence for moral emotions. Another source of evidence for emotion theories comes from individuals who are impaired in their ability to process certain emotions, such as psychopaths and VMPFC patients.

3.2.1 Psychopaths. James Blair has pioneered neuropsychological research on psychopaths, a population prone to antisocial and immoral behavior, and appearing to lack social emotions like remorse, guilt, and empathy (Hare & Quinn, 1971). A study examining emotion attribution in healthy and psychopathic individuals found that although psychopaths were capable of attributing happiness, sadness, and embarrassment to others, they differed significantly from controls in their attributions of guilt; While controls typically attributed the target emotion of guilt to protagonists in guilt stories, psychopaths tend to attribute feelings of happiness or indifference to the protagonists (Blair, Jones, Clark, & Smith, 1995). A brain imaging study also found that when responding to emotionally valenced words, psychopaths show reduced activation in emotion-related brain regions when compared to controls (Kiehl et al., 2001).

---

[6] This point is further supported by the fact that although children show signs of "core" disgust in response to certain stimuli such as feces or foods around age 2 or 3 (Rozin, Hammer, Oster, Horowitz, & Marmora, 1986), they do not show a disgust response to more abstract stimuli such as contamination until later childhood (e.g. Fallon, Rozin, & Pliner, 1984; Rozin & Fallon, 1985), and yet children are capable making a variety of nuanced moral distinctions before this age (see Chapter 2 for a review).

In addition to these emotional deficits, psychopaths show certain deficiencies in moral understanding as well. Specifically, unlike their non-psychopathic counterparts (Nucci & Nucci, 1982; Nucci & Turiel, 1978; Smetana, 1981), psychopaths are unable to distinguish between moral rules – those which are generalizable, unalterable, and independent of authority dictates (ex: do not kill) – and conventional rules – those which are changeable and relative to the social context (ex: do not wear pajamas to a wedding) (Blair, 1995). That is, psychopaths are more likely to view all violations as conventional transgressions - forbidden only as long as an authority figure enforced the rule. Furthermore, Blair (1997) found that although children who scored high on the Psychopathy Screening Device *were* capable of making this moral/conventional distinction, it was less pronounced relative to controls. This finding is particularly interesting, considering that typically developing children as young as three years old (Smetana, 1981) and even children with autism (Blair, 1996; Leslie, Mallon, & DiCorcia, 2006) make this distinction.

Blair and others have argued that adult psychopaths fail to distinguish between moral and conventional transgressions precisely because they have trouble experiencing emotions and attributing emotions to others. In particular, Blair argues that psychopaths lack the ability to empathize with the emotional distress of others, which he claims is essential for healthy moral development (see Chapter 2 for a more detailed discussion of Blair's theory; see also Hoffman, 2000). However, this empathy account does not clearly account for why children with psychopathic tendencies out-perform adult psychopaths on this task. Blair's account is also insufficient to explain subsequent data collected by Cima, Tonnaer, and Hauser (2010), which showed that even though adult psychopaths

show diminished emotional processing relative to controls, they do not differ from controls in their pattern of judgments on footbridge and bystander trolley problems – a pattern of judgments which emotion theorists attribute to a heightened emotional response in the footbridge problem (Greene et al., 2001, 2004; Moll & de Oliveira-Souza, 2007). Cima et al. also found no correlation between psychopaths' moral judgments and their PCL-R scores (Psychopathic Checklist-Revised) or the nature of their criminal convictions.

In light of these finding, Cima et al. (2010) propose that rather than impairing moral *judgment*, psychopaths' emotional deficiencies contribute to their immoral *behavior* (see also Hauser, 2006; Huebner et al., 2009). According to this view, "although psychopaths clearly have an emotional deficit, their failure to distinguish between moral and social conventions may result from a failure to bind emotions with a theory about which actions are right or wrong" (Hauser, 2006, p. 237). In other words, psychopaths know the difference between right and wrong, but simply do not care (i.e. they do not experience the emotions typically associated with committing a moral transgression).

3.2.2 VMPFC Patients. Individuals with brain lesions are another population that exhibit distinct emotional deficits that appear to be associated with moral judgment. Specifically, work by neuroscientist Antonio Damasio has focused on individuals with adult-onset damage to the ventromedial prefrontal cortex (VMPFC), a brain region believed to be linked to social emotions such as compassion, shame, guilt, and contempt (Anderson, Barrash, Bechara, Tranel, 2006; Damasio, 1994; Damasio, Tranal, & Damasio, 1990; Koenigs & Tranel, 2007).

To a large extent, the behavior of these patients mirrors that of the 19th century patient Phineas Gage, who became famous after a tragic head injury left him with an altered personality and impaired reasoning abilities (Damasio, 1994). Damasio refers to this pattern of behavior as "acquired sociopathy," because it is often characterized by irreverence and apparent disregard for social norms, indifference to the feelings of others, and generally flat affect - all characteristics that are also typical of sociopaths. Similar to psychopaths, patients with VMPFC damage also show little or no arousal response to stimuli that typically induce high arousal responses in healthy participants (e.g. pictures of body mutilation, death, etc.) (Damasio et al., 1990).

Interestingly, although these individuals possess intact IQ, memory, language, attention, and problem-solving abilities, and can even demonstrate explicit social and moral knowledge (Bechara, Tranel, Damasio, & Damasio, 1996; Saver & Damasio, 1991), their moral judgments systematically deviate from those of normal subjects in "high-conflict" trolley problems (Koenigs et al., 2007) – a subset of "personal"/emotionally salient dilemmas such as the footbridge problem, in which reaction times are slower and judgment variance is higher (for further discussion of "personal" dilemmas (Greene et al., 2001, 2004) refer to section 3.3.1). Whereas healthy adults discriminate between bystander and footbridge versions of the trolley dilemma, VMPFC patients treat these two cases similarly, "producing an abnormally 'utilitarian' pattern of judgments" on personal, high conflict moral dilemmas (Koenigs et al., 2007, p. 908).

Emotion theorists such as Moll and de Oliveira-Souza (2007) explain these results by suggesting that the VMPFC is primarily responsible for "prosocial"/empathetic moral

emotions such as guilt, compassion, and interpersonal attachment. They argue that whereas normal participants reject the utilitarian outcome in the footbridge dilemma because the emotional salience of pushing the fat man triggers these prosocial moral emotions, patients with damage to the VMPFC judge this dilemma as acceptable because of a deficit in prosocial moral emotions.[7]

These lesion studies perhaps provide the strongest evidence for emotion theories so far. However, the degree to which emotions causally contribute to moral judgment is still inconclusive. Given the emotional deficiencies of VMPFC patients, it is somewhat surprising that these patients' moral judgments so closely align with those of healthy adults on many other tasks. For example, most patients attain the second level of moral development on Kohlberg's (Colby & Kohlberg, 1987) widely used paradigm of the Standard Moral Interview – a level that is characteristic of most healthy adults (Saver & Damasio, 1991). Furthermore, their responses to trolley problems are virtually identical to those of controls on impersonal and low-conflict personal dilemmas such as the bystander problem. While emotion theories provide one explanation for their deviant judgments in high-conflict personal dilemmas, others such as Huebner et al. (2009) have argued that VMPFC patients respond differently from controls in these dilemmas because they "fail to treat the morally salient features of high-conflict dilemmas as morally salient" (p. 4). For example, they may focus on the consequences of an action, but ignore the means by which those consequences occurred. Under this view, "deviant outputs [are

[7] Although Koenigs et al. (2007) agree that "for a selective set of moral dilemmas,....[our] findings support a necessary role for emotion in the generation of [judgments of right and wrong]," they align themselves with the dual process theory discussed below, arguing that social emotions do not play a role in resolving *all* moral dilemmas.

therefore] a result of deviant inputs, rather than a result of a deficit in moral processing per se" (p. 4).

Perhaps the biggest limitation of the emotion theories discussed so far is that none of these theories are specific enough to model. At best, they are incomplete, as none provide a satisfactory account for what specific properties of the stimulus are necessary to trigger an emotional response, let alone how the mind identifies when those properties are present in the stimulus. What is it that distinguishes disgust-eliciting scenarios from empathy-eliciting scenarios, for example? And why do we respond with stronger feelings of disgust to some moral stimuli than to others? An adequate theory must provide an explanation of the evaluative processes that trigger these emotions. However, simplistic perceptual models quickly fall apart when we try to explain moral violations such as deceit, violations of trust, and broken promises. For each perceptual model, one can easily think of counterexamples in which the hypothesized perceptual trigger is present but the corresponding emotion or judgment is not elicited, and vice versa. For example, we often experience empathetic distress when witnessing natural disasters or no-fault car accidents, and yet we do not consider these to be moral transgressions. Likewise, many of the behaviors we view as morally wrong are non-violent, and often do not elicit visible or audible signs of distress in the victim, or at least not in close temporal proximity to the violation itself (e.g. stealing and lying are wrong, even when the victim is unaware of the theft or deceit).

These emotion theories also do a particularly poor job of predicting or explaining why healthy participants morally distinguish between similar and highly abstract stimuli such as the trolley dilemmas. Emotion theorists argue that it is the degree of emotional

salience, as well as the type of emotion/brain region being activated by the scenario, that drives these differential judgments. And yet they give no account of what properties of the stimuli make them more or less emotionally salient, how the stimuli are mentally represented, or why personal dilemmas elicit empathetic emotions, for example, but impersonal dilemmas do not. In the following section, I describe an additional emotion-based theory that has attempted to address some of the limitations of the theories just described.

3.3 Dual Process Theory. Another group of emotion theories generally referred to as "dual process theories" (Kahneman & Frederick, 2002) have been proposed to explain the source of our moral judgments. While these theories are quite similar in many respects to the emotion theories described above, dual process theories view moral judgments as products of two distinct systems: a social-emotional system (i.e. System I), and an abstract reasoning system (i.e. System II). System I is rapid, intuitive, and automatic, while system II is slow, effortful, conscious, and deliberate. Dual process theories propose that both these systems are causal factors in our moral judgments. For example, according to Joshua Greene's dual-process model (e.g. Greene et al., 2001), reason and emotion systems both generate moral intuitions, which are occasionally in conflict with each other. The extent to which a particular moral scenario engages each of these systems determines the moral judgment. Thus, whereas emotion theorists like Haidt argue that System I is primarily responsible for moral judgments (although he allows that System II may play a role in other domains of cognition, he rejects it as a source for moral judgments), Greene allows for instances in which System II may override System I.

3.3.1 The Trolley problem according to Greene's model. In 2001, Greene et al. conducted an fMRI study to test the following hypothesis:

> the crucial difference between the [bystander] dilemma and the footbridge dilemma lies in the latter's tendency to engage people's emotions in a way that the former does not. The thought of pushing someone to his death is, we propose, more emotionally salient than the thought of hitting a switch that will cause a trolley to produce similar consequences, and it is this emotional response that accounts for people's tendency to treat these cases differently (Greene et al., 2001, p. 2016).

In particular, Greene et al. (2001) predicted that scenarios like the footbridge problem would engage the emotional system of the brain to a greater extent because those dilemmas involved actions that were "up close and personal." They defined a moral violation as "personal" if it "(a) could reasonably be expected to lead to serious bodily harm (b) to a particular person or a member or members of a particular group of people (c) where this harm is not the result of defecting an existing threat onto a different party" (p. 2107). Greene et al. (2004) later expanded on this definition with the following explanation:

> One can think of these three criteria in terms of 'ME HURT YOU.' The "HURT" criterion picks out the most primitive kinds of harmful violations (e.g., Summary assault rather than insider trading) while the "YOU" criterion ensures that the victim be vividly represented as an individual. Finally, the "ME" condition captures a notion of "agency," requiring that the action spring in a direct way from the agent's will, that it be "authored" rather than merely "edited" by the agent. (Greene et al., 2004, p. 389).

Thus, according to Greene et al. (2001, 2004) when participants contemplated trolley dilemmas that met these 'personal' criteria, such as the footbridge problem, they would be more likely to give deontological responses (i.e. judge them as 'inappropriate') because their emotional system would be driving the judgment. In contrast, when participants contemplated impersonal dilemmas (those that failed to meet the personal

criteria above) such as the bystander problem, or non-moral scenarios, they would be more likely to give utilitarian responses (i.e. judge them as 'appropriate') because the reasoning system was determining the judgment.

In support of this hypothesis, Greene et al. (2001) found that brain areas associated with emotion (i.e. the medial prefrontal gyrus, posterior cingulate gyrus, and angular gyrus) were more active when participants responded to personal dilemmas such as the footbridge trolley problem, whereas that brain areas associated with abstract reasoning and problem solving (i.e. the DLPFC) were more active when subjects responded to "impersonal" dilemmas such as the bystander problem, as well as to non-moral scenarios. Furthermore, they found that on trials in which participants gave "emotionally incongruent" responses to personal dilemmas (e.g. indicated that pushing the fat man was 'appropriate'), their responses showed longer reaction times than on trials in which they gave congruent responses; however, no differences in reaction time were observed on impersonal trials. Greene et al. (2001) argued that these findings provided support for the hypothesis that emotionally incongruent (i.e. utilitarian) responses to personal dilemmas were more difficult (and therefore require longer RTs) due to the increased cognitive control required to overcome "emotional interference."

Two follow-up studies provided additional support for this claim. Another neuroimaging study by Greene et al. (2004) showed that brain regions previously associated with cognitive conflict (i.e. the anterior cingulated cortex (Botvinick, Braver, Barch, Carter, & Cohen, 2001), and cognitive control/abstract reasoning (i.e. the DLPFC (Miller & Cohen, 2001)) exhibited increased activation on high-RT trials and trials in which participants gave emotionally incongruent (i.e. utilitarian) responses to personal

dilemmas. Finally, a third study by Greene, Morelli, Lowenberg, Nystrom, & Cohen (2008) asked participants to evaluate particularly difficult "high conflict" (Koenigs et al., 2007) personal scenarios while simultaneously performing a digital search task. Greene (2008) predicted that because utilitarian judgments in these scenarios required cognitive control, increasing the cognitive load by adding a simultaneous digital search task would reduce the frequency of such judgments, and increase reaction time for utilitarian responses. Although the prediction that participants would give fewer utilitarian responses on cognitive load trials was not supported, findings did support the prediction that cognitive load would selectively increase the average reaction time for utilitarian judgments, but not for deontological judgments.

3.3.2 Limitations of Greene's theory. Taken together, these studies provide strong evidence that trolley problems recruit both emotion and reasoning centers of the brain, and increased activation of these regions varies systematically depending on which version of the trolley problem is presented to the participant. Furthermore, these studies support the claim that participants morally discriminate between different kinds of trolley problems, such as between the footbridge and bystander problems, based on mutually exclusive criteria. Any theory of moral intuition is therefore descriptively inadequate unless it provides an explanation for why our responses to these dilemmas differ. Greene and colleagues should be commended for attempting to identify the "psychologically essential features" of the stimuli that guide participants' judgments (Greene et al., 2001, p. 2017), a question that is entirely neglected by most emotion theorists. Nevertheless, as Greene has since acknowledged, the (largely perceptual) criteria he and his collaborators proposed in 2001 are descriptively inadequate (Greene, 2009; Mikhail, 2007, 2011), as

they fail to explain or predict people's moral judgments on many of the original 12 trolley problems designed by Mikhail.

For example, consider the Implied Consent scenario (Mikhail, 2011) in which a man walking across the train tracks is about to be killed by the oncoming train. In order to save him, the bystander must throw him out of the way. To do so, however, would "likely cause serious bodily harm" to the man, and this harm would not result from the deflection of an existing threat. Therefore, according to Greene's criteria, this scenario would qualify as a personal scenario, and we should expect participants to disapprove of throwing the man out of the way. Nevertheless, not surprisingly, 93% of Mikhail's participants judged this action as permissible. (Other counterexamples include the Expensive Equipment, Intentional Homicide, Loop Track, Better Alternative, Disproportional Death, and Drop Man scenarios. See Appendix A for a description of each).

Greene et al. (2001, 2004) have also been criticized for confounds in their stimuli such as a tendency to use more "colorful language" and more frequent references to family and close acquaintances when describing 'personal' moral dilemmas. They were also criticized for their failure to control for other factors that differed between personal and impersonal dilemmas, such as whether the dilemma involved physical contact with the victim, harm as a means to an end, and violence such as murder and rape (Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006; McGuire, Langdon, Coltheart, & Mackenzie, 2009; Mikhail, 2008). Furthermore, the fact that Greene et al. presented dilemmas in the second person is problematic (Mikhail, 2008), particularly in light of evidence suggesting that participants process behavioral prediction questions (e.g.

questions about what the participant would actually do) differently from questions about what is wrong to do (Borg et al. 2006; Royzman & Baron, 2002). Many have also questioned whether asking participants to judge a particular action as "appropriate" or "inappropriate" was the best way to measure moral judgment, as an action may be "inappropriate" for a variety of reasons that have nothing to do with morality (Borg et al., 2006; Mikhail, 2008).

In order to resolve some of these questions, McGuire et al. (2009) conducted an item analysis of Greene et al.'s (2001) data. Their analysis revealed that the interaction in RT between dilemma type ('personal' and 'impersonal') and response ('appropriate' or 'inappropriate') was driven by a small number of the 60 dilemmas presented to subjects and was therefore "not generalizable to other putative populations of moral dilemmas" (p. 579). Furthermore, when McGuire et al. excluded dilemmas from the analysis that less than 5% of the subjects judged to be 'appropriate' (because participants responded remarkably quickly to these items), this interaction became non-significant, indicating that participants responded more slowly to 'personal' dilemmas in general, regardless of whether they judged them to be 'appropriate' or 'inappropriate.' In other words, the difference in RT between utilitarian and deontological judgments of high-conflict 'personal' dilemmas was no longer significant; the interaction between dilemma type and response was simply due to consistently rapid responses of 'inappropriate' to a subset of 'personal' dilemmas, rather than to slower responses of 'appropriate' to 'personal' dilemmas. This reanalysis of Greene et al.'s findings suggests that the personal/impersonal distinction is not supported by behavioral evidence, and severely weakens the dual process theory as a model for how we form moral judgments.

Greene (2009) has responded to these criticisms by arguing that the primary hypothesis of the dual process theory - that deontological responses are driven by the emotional/automatic system, while utilitarian judgments are driven by controlled cognitive processes – still stands, even without support for the personal/impersonal distinction. According to Greene (2009), "we need not know how, exactly, the footbridge and [bystander] dilemmas differ in order to know that they engage dissociable processing systems" (p. 583). However, we *do* need to know how these dilemmas differ if we are to construct a descriptively adequate model of moral judgment (Mikhail, 2011). A descriptively adequate model should be capable of predicting participants' responses to any given stimulus. Greene's prediction that participants will respond deontologically when their emotional system is engaged is insufficient without a description of *what engages the emotional system*. Greene's theory fails to answer two fundamental questions: what are the "psychologically essential criteria" for triggering the emotional system, and how does the mind identify whether a given stimulus meets these criteria? When considering stimuli that are as complex and abstract as the trolley problems, it becomes apparent that the mind must perform some kind of analysis and interpretation of these dilemmas. It is difficult to imagine how the mind might go about extracting the "psychologically essential criteria" from these stimuli without forming some kind of structural representation of these scenarios.[8]

[8] More recent work by Greene et al. (2009) has begun to address these issues. Their current model proposes that an interaction between "personal force…when the force that directly impacts the victim is generated by the agent's muscles" and the agent's intention underlies participant's emotional responses to problems like the footbridge dilemma (p. 1). However, in contrast to Mikhail, they argue for an "embodied" goal representation involving personal force, in which "our sense of an action's moral wrongness is tethered to its more basic motor properties, and specifically that the intention factor is intimately bound up with our sensitivity to personal force" (p. 8).

Finally, many of the criticisms I directed at the emotion theories above also apply to Greene's dual process theory. Namely, the extent to which emotions and conscious reasoning play *causal* roles in moral judgment is still unclear. Although the footbridge and bystander dilemmas systematically elicit increased activity in brain regions associated with emotion and reasoning respectively, these data are correlational, and thus cannot speak to any causal relationships between emotions, reasoning, and moral judgment. Emotion and reasoning processes are clearly involved when participants evaluate these moral dilemmas, but we cannot say *at what point* these processes become involved, as temporal resolution for fMRI data is limited. The behavioral evidence from Greene et al. (2008) showing the effect of cognitive load on RT for utilitarian responses is more promising; however, these data fail to rule out alternative explanations. For example, if, as computational theorists such as John Mikhail claim, trolley dilemmas "elicit mental representations that differ in their structural properties" (Mikhail, 2011, p. 356), then the length and complexity of the computations required for a given dilemma might also be associated with selective differences in RT.

4. Conclusions

When compared to emotion and dual process theories, Mikhail's computational UMG theory fills in many of the gaps that the other theories seem to neglect. Not only does the theory of UMG account for much of the data the other theories cannot, it does more than provide correlations between particular neural networks and moral judgments. To uncover the source of our moral judgments, we must first recognize the problems we are faced with: What aspects of the environment does the mind take as input to arrive at a moral judgment? How does the brain distinguish morally relevant features from those

that are irrelevant to morality? What is the nature of our mature moral knowledge, and how do we acquire such knowledge? Mikhail's computational theory provides a strong framework for answering these kinds of questions. Furthermore, it opens doors to previously unstudied aspects of moral cognition, drawing attention to the competence-performance distinction in the moral domain and the potential factors that constrain the range of the world's moral cultures.

5. Overview of the dissertation research

In the current Chapter, I discussed Mikhail's UMG theory, which hypothesizes that we posses unconscious knowledge of the PDE and the prohibition of intentional battery (i.e that these principles are part of our moral grammar), and that this knowledge explains the stable pattern of intuitions observed in adult moral judgments of the trolley problems. If this is true, the computations involved in generating these moral intuitions are quite complex and sophisticated; the PDE relies not only on our ability to identify universally prohibited "prima facie" wrongs (such as battery), but also on our ability to mentally represent acts/omissions in terms of their causal and intentional properties (i.e. in terms of proximal causes, ends, means, side effects), to weigh both the probabilities and the magnitudes of that action's good and bad effects, and to compute and compare alternative actions and their respective causal and intentional properties (Mikhail, 2002).

This hypothesis raises interesting questions regarding the development of moral judgment. At what age do these capacities emerge? At what age do children generate deontic judgments in accordance with the PDE? And how do they acquire such a principle, given that adults are typically unaware that the PDE is operative in their judgments? Following Chomsky's poverty of the stimulus argument (Chomsky, 1986;

Dwyer, 1999), Mikhail proposes that the PDE may be part of our innate biological endowment:

> On reflection, it seems doubtful that children are affirmatively taught to generate the specific representations presupposed by this principle to any significant extent. We thus seem faced with the possibility that certain moral principles emerge and become operative in the exercise of moral judgment that are neither explicitly taught, nor derivable in any obvious way from the data of sensory experience. In short, we appear confronted with an example of what Chomsky calls the phenomenon of the "poverty of the stimulus" in the moral domain (Mikhail, 2002, p. 15).

In Chapter 2, I discuss the argument for the poverty of the stimulus further, and some potential evidence in support of this claim. While some evidence indicates that children's moral judgments are sensitive to the causal and intentional properties of action from an early age, virtually no studies have investigated whether children younger than 8-years-old generate deontic judgments in accordance with the PDE. I take up this question in the subsequent chapters of this dissertation.

In Chapters 3-5, I present a series of studies in which preschoolers (Chapters 3 and 4) and adults (Chapter 5) were asked to evaluate double-effect dilemmas designed to be similar in structure to the bystander and footbridge problems described above. However, unlike the classic trolley problems, the dilemmas in the current studies involved violations of personal property and assault (i.e. the apprehension of battery), rather than battery and homicide. I also investigated whether participants considered it morally permissible to refrain from preventing a foreseeable harmful outcome in these scenarios. Thus, not only are these studies the first to explicitly investigate whether the PDE is operative in children as young as three years of age, they are also the first to use trolley problems that do not involve bodily harm (in either the adult literature or the developmental literature), and the first to investigate children's knowledge of the duty of

rescue (i.e. knowingly allowing a preventable harm to occur is impermissible, unless preventing harm requires unjustified costs (Mikhail, 2011)).

In Chapters 4 and 5, I also investigate another aspect of moral cognition that has received little attention in the developmental literature: the role of ingroup/outgroup structure on moral judgment. The study of children's representation of social categories is a new topic in social cognition that has attracted significant attention in recent years. However, few studies have explored the intersection between this aspect of social cognition and moral judgment in preschoolers. In Chapter 4, I ask whether group structure (i.e. whether the moral patient(s) belong to the child's ingroup or outgroup) affects children's moral judgments of double-effect scenarios. In Chapter 5, I ask whether group structure affects adults' moral judgments of the same dilemmas that were presented to preschoolers in Chapter 4, as well as an additional set of dilemmas in which the identity of the moral agent also varied. Although there is some evidence that group structure can affect adult's moral intuitions of trolley problems involving "real" groups, and there is substantial evidence that the mere presence of "minimal" groups can induce adult ingroup bias on a variety non-moral tasks, this is the first study to investigate the effect of minimal group structure on adults' intuitions in double-effect scenarios, and the first to manipulate the identity of both the moral agent and the moral patient in such scenarios.

II. Moral acquisition and development

How and when does the capacity for moral judgment first emerge? Current evidence suggests that children's ability to evaluate others based on their social interactions occurs quite early in development. In their first year of life, infants already distinguish between proscial and negative interactions (Premack & Premack, 1997), prefer agents who help other agents achieve their goals over agents who thwart the goals of others (Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007) and even prefer agents who reward helpers and punish hinderers (Hamlin, Wynn, Bloom, & Mahajan, 2011). By 12 months of age, infants not only prefer prosocial agents themselves, they also expect others to show a preference for prosocial agents over antisocial agents (Kuhlmeier, Wynn, & Bloom, 2003), and between 12 and 18 months of age, infants spontaneously engage in helping behavior themselves when they see an agent in need (e.g. Liszkowski, Carpenter, Striano, & Tomasello, 2006; Warneken & Tomasello, 2007), unless that agent has harmful intentions (Vaish, Carpenter, & Tomasello, 2010).

These findings are intriguing, particularly when one considers that 3-month-olds – the youngest age group to show a preference for prosocial agents so far (Hamlin, Wynn, & Bloom, 2010) – have been exposed to very little environmental input that might explain how they could have learned this preference for helpers over hinderers. Furthermore, this evidence is consistent with other findings that suggest infants are sensitive to morally relevant properties of action such as an agent's goal and action plan (e.g., Behne, Carpenter, Call, & Tomasello, 2005; Carpenter, Call, & Tomasello, 2005; Carpenter, Akhtar, & Tomasello, 1998; Csibra, Gergely, Bíró, Koos, & Brockbank, 1999;

Gergely, Bekkering, & Kiraly, 2002; Gergely, Nádasdy, Csibra, & Bíró, 1995; Johnson, Booth, & O'Hearn, 2001; Luo & Baillargeon, 2005; Meltzoff, 1995; Premack & Premack, 1997; Woodward, 1998; Woodward & Sommerville, 2000). However, the fact that infants prefer "nice" agents to "mean" agents is insufficient for attributing moral judgment to infants (i.e. knowledge of deontic concepts such as obligatory, permissible, and forbidden). Preferring prosocial agents to antisocial agents is not the same thing as assigning moral value to their actions. If we are to understand how and when moral acquisition occurs, we must ask several important questions: First, at what point in development do children demonstrate social evaluations that clearly go beyond simple judgments of preference or approval? Second, what is the nature of children's moral knowledge at this age, and how does it compare to that of adults? What factors do children take to be morally relevant? Finally, what is the nature of the moral input children receive, and is it sufficient to explain their moral knowledge at this age? In other words, does Chomsky's poverty of the stimulus argument apply to moral acquisition?

In this chapter, I will review some of the moral developmental literature with these questions in mind. While evidence in the field of moral cognition is limited in this regard, there is some evidence to suggest that children are indeed equipped with quite sophisticated moral knowledge that is unlikely to be acquired solely through cultural transmission (Dwyer, 1999; Mikhail, 2002).

1. The Moral/Conventional distinction

Research over the last two decades has shown that young children are able to acquire permissibility rules over a remarkably short period of time. Children as young as

26 months can reason appropriately about the permissibility of committing moral transgressions such as hitting another child and conventional transgressions such as wearing pajamas to school (Smetana & Braeges, 1990). By three years of age, children not only correctly identify these transgressions as impermissible, they judge moral transgressions as more serious and deserving of greater punishment than conventional transgressions. They also recognize that while moral transgressions are universally wrong, and wrong independent of rules or authority, conventional transgressions are rule-contingent, flexible, and dependent on authority (Turiel, 1978, 1983; Nucci & Turiel, 1978; Nucci, 2001; Nucci & Turiel, 1993; Nucci, Turiel, & Encarnacion-Gawrych, 1983; Smetana, 1981, 1983, 1985; Smetana & Braeges, 1990). Furthermore, they recognize that in the personal domain, what is "good" or "bad" is a matter of personal choice, but moral and conventional rules about good and bad apply to everyone, and are not a matter of personal choice (Nucci & Weber, 1995).

How do children acquire this knowledge? Although children do receive moral feedback and instruction from their parents, it is unclear how this input allows children to make the abstract distinction between moral and conventional rules. When parents do teach their children permissibility rules, it is often in the form of post-hoc admonishments about specific acts (e.g. "Don't hit your sister!", "Don't play with your food!", "Say 'please'!"). While these corrections are potentially informative regarding specific acts in specific contexts, they are insufficient for extracting more general rules about moral vs. conventional violations (Dwyer, 1999, 2006).

Another possibility is that children learn the moral conventional distinction based on the degree of arousal their parent exhibits and the severity of the punishment the child

receives for a given transgression. However, this hypothesis is also unlikely (see Dwyer, 1999, 2006). Consider the child who has colored all over the family's new couch, versus the child who has pinched her brother. Although the former transgression is a conventional one, it is more likely to elicit a strong emotional reaction from the child's parent and receive a comparable, if not more severe punishment. Furthermore, children often get away with bad behavior without being caught or punished, and are often punished for things they didn't do (especially if they have siblings). Likewise, children's good behavior often goes unnoticed or unrewarded. If children learn permissibility rules and the moral/conventional distinction based on the degree of punishment each transgression receives, children would have to commit or observe a large number of moral and conventional transgressions before the age of three, and each (or most) of these transgressions would have to result in a level of punishment that was proportionate to the type of transgression committed. However, given the previous examples, it is very possible that children receive conflicting and inconsistent feedback for moral and conventional transgressions.

A third but related possibility is that children learn to distinguish between moral and conventional violations based on domain-general emotional processes (in the case of moral violations) (Blair, 1995; Haidt, 2001; Nichols, 2004). For example, according to Blair's (1995) developmental model, children learn to inhibit actions that will result in harm to others by way of a violence inhibition mechanism (VIM) similar to inhibition mechanisms that have been proposed to control aggression in some animal species (Eibl-Eibesfeldt, 1970; Lorenz, 1966). Blair argues that when typically developing children provoke emotional distress in others, distress cues such as "a sad facial expression or the

sight and sound of tears" trigger a withdrawal response, accompanied by negative feelings such as guilt, remorse, sympathy, and empathy for the victim (Blair, 1995, p. 3). These emotions serve as negative reinforcement for actions that generate distress cues in others, and eventually children learn to inhibit such actions. Once a child associates negative feelings with these kinds of actions, he is capable of recognizing immoral acts even in the absence of distress cues. Blair argues that children are capable of the moral/conventional distinction because only moral violations become associated with signs of distress; "Since conventional transgressions, by definition, do not result in victims, they are therefore never paired with distress cues and will not therefore become stimuli for the activation of VIM" (Blair, 1995, p. 7).

However, Blair's empathy account fails to adequately explain our moral competence for several reasons. For one thing, we often experience empathetic distress when witnessing natural disasters or car accidents, for example, and yet we do not consider these to be moral transgressions (unless the driver was driving recklessly, etc.). Likewise, many of the behaviors we view as moral transgressions are non-violent, and often do not elicit visible or audible signs of distress, or at least not in close temporal proximity to the violation itself (e.g. stealing and lying are wrong, even when the victim is currently unaware of the theft or deceit). Furthermore, several studies appear to contradict Blair's hypothesis that witnessing emotional distress is a necessary prerequisite for young children to view an act as immoral. For example, a study by Vaish, Carpenter, and Tomasello (2009) found that 18- to 25-month-old infants showed more concern and subsequent prosocial behavior towards adult victims of property violations (versus controls) even when the victims showed no signs of emotional distress.

Findings from another study by Leslie, Mallon, and DiCorcia (2006b) also directly contradict Blair's explanation of the moral/conventional distinction. Leslie et al. showed that not only did children with autism, a population known to also show difficulties in attributing mental and emotional states to others (e.g. Baron-Cohen, Leslie, & Frith, 1985; Hobson, 1993), match the performance of typically developing children on the moral/conventional task (to be fair, a replication of Blair's (1996) findings), both typically developing preschoolers and autistic children also passed a "cry baby" version of the task, in which a character showed signs of distress even though he was not the victim of any moral violation (in this case, even though he already had his own cookie, he burst into tears when another agent refused to give him her cookie). In other words, both typically developing children and children with autism recognized when a character's emotional distress was unreasonable or unjustified, and did not morally condemn actions that resulted in such "cry baby" displays of distress. These findings suggest that children do not make the moral/conventional distinction by merely responding to whether a "victim" shows signs of distress.

In the following section, I explore the evidence for a fourth hypothesis: that children's moral judgments rely on domain-specific moral computations that are specifically attuned to a set of underlying causal and intentional properties – properties which most conventional violations lack (Huebner, Lee, & Hauser, 2010). In particular, current evidence suggests that when evaluating a given action, children represent the intentional structure of the act (including the intentions, goals and beliefs of both the moral agent and the moral patient), as well as the causal structure of the act (i.e. in terms of concepts such as "agent" "patient" "cause" and "effect").

2. The role of intention in moral development

Despite Piaget's (1932/1965) earlier claims that children younger than 7 years old fail to attend to morally relevant factors such as an agent's motives, or whether a given act was intentional or accidental, current evidence suggests that infants in the first year of life already represent the actions of other agents as rational and goal-directed (e.g., Csibra et al., 1999; Csibra, Biro, Koos, & Gergely, 2003; Premack & Premack, 1997; Woodward, 1998), and can distinguish intentional actions from accidental or unsuccessful actions (e.g., Behne et al., 2005; Carpenter et al., 1998; Woodward, 1999) and ends from means (i.e. the agent's plan of action for achieving her end/goal) (e.g., Carpenter et al., 2005; Gergely et al., 1995; Gergely et al., 2002).

More importantly, children's moral judgments show an early sensitivity to the agent's goal/desire, knowledge, and intention. For example, when an agent's motives are explicit, salient, and available, children as young as three years old assign more blame to agents with bad motives than to agents with good motives (Nelson, 1980). Furthermore, when evaluating identical actions/outcomes, three-year-olds judge foreseeable outcomes as more intentional than unforeseeable outcomes (Nelson-Le-Gall, 1985), and assign more blame to agents who knowingly/intentionally cause harm than to agents who unknowingly/accidentally cause harm (Nelson-Le-Gall, 1985; Nunez & Harris, 1998); Wellman et al., 1979). Likewise, they judge agents who knowingly bring about positive outcomes as more morally praiseworthy than agents who unknowingly do so (Nelson-Le Gall, 1985). A study by Siegal and Peterson (1998) showed that three-year-olds are even sensitive to the subtle distinction between an intentional lie, an innocent mistake (e.g.

uttering a falsehood based on a false belief), and a careless/negligent mistake (one in which the agent fails to use the knowledge available to him to avoid harm).

These distinctions, seem difficult to explain given the input children receive. First, these distinctions rely on inferences about an agent's state of knowledge, beliefs, goals, and intentions, which are not directly recoverable from the stimulus[9]. Particularly in the case of a moral violation involving intentional/antisocial deception, it is difficult to construct a theory that relies on purely low-level perceptual cues to explain children's evaluations. Second, it is unlikely that children are explicitly taught to make a moral distinction between innocent and negligent mistakes, good and bad motives, foreseen and unforeseen outcomes, and intentional and accidental acts, especially because the actions/outcomes must be held constant for each of these factors to become salient. In fact, when younger children are asked to integrate *conflicting* information regarding actions/outcomes and an agent's mental state, younger children tend to default to an outcome-oriented pattern of moral judgments (Killen, Mulvey, Richardson, Jampol, & Woodward, 2011; Yuill, 1984; Yuill & Perner, 1988; Zelazo, Helwig, & Lau, 1996). Given that children can distinguish between good and bad motives when the outcomes are held constant (and given that young children have difficulty integrating conflicting information in general (e.g., Diamond, Kirkham, & Amso, 2002)), these findings suggest that younger children default to outcome-heavy judgments when intentions and outcomes conflict because outcome information is more salient, most likely because it is directly observable. How then, do children learn to make the subtle moral distinctions described

[9] A growing body of research on theory of mind suggests that children make these inferences by way of an innate domain-specific theory of mind mechanism (Leslie, 1987; 1994) which allows them to extract information about an agent's mental state from observable behavior (e.g. Onishi & Baillargeon, 2005; Onishi et al., 2007; Southgate et al., 2007).

above, when those distinctions are so often obscured by other more salient factors in their environment?

Findings by Leslie, Knobe, and Cohen (2006a) show yet another aspect of children's moral knowledge that is difficult to explain without postulating some kind of appraisal mechanism. In a preschool version of the side-effect effect task (Knobe, 2003), Leslie et al. showed that the moral valence of a foreseen but disavowed side-effect (one which the agent foresees but does not care about) influenced preschoolers' judgments of whether an agent brought about the side effect intentionally. Previous studies have found that adults judge a negative foreseen side effect as intentional, and a positive foreseen side effect as unintentional (Knobe, 2003). To test this "side-effect effect" in preschoolers, Leslie et al. presented children with stories in which a character's main effect (ex: bringing a pet frog to a friend's house) resulted in either a positive side effect (ex: making the friend happy), or a negative side effect (ex: making the friend sad). Children were told in both conditions that the character knew but did not care about the side effect. Children were then asked if the character brought about the side effect on purpose. Leslie et al. (2006a) found that children who understood the concept of "not caring" exhibited the side-effect effect: bad side effects were judged as intentional, and good side effects were judged as unintentional. However, three-year-olds who failed the caring question (does character X care that character Y will get happy/upset?) did not show the side-effect effect (i.e. they defaulted to a "yes" bias in both good side-effect and bad side-effect stories).

As the authors point out, preschoolers are unlikely to have heard anyone articulate the rule that bad side effects are brought about on purpose and good side effects are not.

Furthermore, the fact that children only showed the side-effect effect once they were able to process that the agent did not care about the foreseen side-effect suggests that the side-effect effect "is not acquired gradually, but emerges immediately following its prerequisite" (p. 426). In this case, the prerequisite is the concept "not caring." The fact that young children have difficulty with this concept suggests that children begin with a default "presumption of caring," perhaps related to Mikhail's (2011) proposed "presumption of good intentions," whereby we assume that an agent "is a person of good will, who pursues good and avoids evil" unless we are given sufficient evidence to the contrary (Mikhail, 2011, p. 173).[10]

These findings also suggest that not only do judgments of intentionality influence moral judgments, but moral judgments may also influence judgments of intentionality. Leslie et al.(2006a) propose two domain-specific hypotheses to explain this phenomenon. According to both hypotheses, the side-effect effect is the product of two different systems at work: a theory of mind mechanism that generates judgments of intentionality, and a moral system that generates moral judgments. However, under the first hypothesis, the theory of mind mechanism "may have a parameter for moral valence of outcomes...The value of this parameter would influence judgments of purpose, but would be obtained from processes external to theory of mind, such as moral judgment" (p. 426). Alternatively, under the second hypothesis, the moral system 'could take in information

[10] This hypothesis is further supported by a recent study by Laguttata and Sayfan (2013) that showed age-related increases among children ages 4-10 in the ability to take into account an agent's past harmful behavior when making future behavioral predictions. It is also consistent with previous studies showing that children expect positive behavior even when they have been given neutral or negative intention information (Grant & Mills, 2011), disregard negative behavioral information, or require more evidence of negative behaviors than positive behaviors to make character attributions (Boseovski, 2010; Boseovski & Lee, 2006, 2008), and expect negative behaviors to improve over time (Heyman & Giles, 2004; Lockhart, Chang, & Story, 2002).

about the situation and the agent's mental states. Then it could use this information to determine whether or not the behavior was morally bad and, on that basis, produce as output a judgment of whether or not it was performed intentionally" (p. 426).

3. Unanswered questions

The studies described in this Chapter begin to outline the problem of descriptive adequacy inherent in any perceptual/domain-general developmental model. The input children receive is unlikely to account for their early sophisticated moral knowledge. However, if we are to accept Mikhail's theory of UMG, stronger evidence of the emergence of complex operative principles early in moral judgment is required. In Chapter 1, I discussed two principles that Mikhail proposes are likely to be part of the universal moral grammar: the prohibition of intentional battery, and the principle of double effect. Although some of the evidence presented in the current chapter hints that children already posses the prerequisite concepts for the prohibition of intentional battery such as *intentional* or *purposeful*, and other work suggests that children also possess the concept *non-consensual* – for example, children as young as three years of age already understand the privileges of ownership, such as the right to grant or exclude others from using one's property (Neary, Friedman, & Burnstein, 2009; Rossano, Rakoczy, & Tomasello, 2011) – virtually no studies have examined young children's knowledge of complex principles such as the PDE[11].

Furthermore, Mikhail makes strong claims about the impartial nature of the moral faculty that have yet to be tested empirically. Like Rawls (1971), Mikhail assumes that there is a principled distinction between "considered judgments" – judgments in which

[11] But see my discussion of Pellizzoni, Siegal, and Surian (2010) in Chapter 3. See also my discussion of Mikhail's (2002) trolley study in Chapter 3, which only tested children young as 8 years old.

our moral capacities are most likely to be displayed without distortion" (Rawls, 1971, p. 47) – and prejudices. Under this view, although we may behave in ways that are prejudiced or discriminatory, such behavior is not an accurate reflection of our underlying moral competence. In other words, the moral grammar underlying our judgments is impartial, but factors exogenous to the moral system, such as the feelings, attitudes, and stereotypes we form toward certain groups or individuals, may (either consciously or unconsciously) bias or distort our moral judgments and actions. This hypothesis raises interesting theoretical predictions – namely, that when making moral evaluations, young children should privilege morally-relevant information, such as causal and intentional information, over morally-irrelevant information, such as information about an agent's and/or patient's social group. However, until recently, the interaction between moral cognition and group cognition has been largely neglected in the developmental literature.

In the following chapters I present a series of studies that begin to address these questions.

III. Investigating the principle of double effect

When, if ever, is it morally permissible to knowingly cause harm? In his *Summa Theologiae*, Thomas (1274/1988) was the first to introduce the principle of double effect (PDE) in answer to this question. He observed that a single action can often generate multiple foreseen outcomes or effects, some of which are intended effects, and some of which are unintended side effects. He argued that the justification for an act such as self defense, which produces both the good effect of saving one's life, as well as the bad effect of killing the attacker, must rest in the distinction between intending the good effect vs. intending the bad effect (under the condition that the good effect is proportionate to or outweighs the bad effect). Aquinas's principle has since been reformulated in the following manner (Fischer & Ravizza, 1992; Mikhail, 2000): an act that results in harmful effects is permissible if and only if the intended effects of the act are good, the harmful effects are only unavoidable side effects, and not means, of achieving the good effects, the good effects outweigh the bad effects, and no morally preferable alternative is available. In this chapter, I ask whether the moral judgments of preschoolers accord with this principle.

The Principle of Double Effect

As discussed in Chapter 1, The PDE has often been invoked to explain our divergent intuitions in the bystander and footbridge versions of the Trolley Problem (e.g. Fischer & Ravizza, 1992; Mikhail, 2000). According to the PDE, flipping the switch to divert the trolley to the side track in the bystander problem is permissible because the bystander intends only the good effects of his action (i.e. saving the five people), while the death of the one person is an unavoidable *side effect* of flipping the switch; however,

the act of pushing the man in the footbridge problem is impermissible because the bad effect (the death of the large man) is intended as the *means* of achieving that good end.

A growing body of cross-cultural research suggests that adults intuitively understand this principle. Adults across a wide range of demographics reliably show a pattern of responses consistent with the PDE on multiple versions of the trolley problem (Cushman et al., 2006; Greene et al., 2001, 2004; Hauser, 2006; Hauser et al., 2007a; Mikhail, 2002; O'Neill & Petrinovich, 1998; Petrinovich et al., 1993), even when other potentially relevant factors are carefully controlled, such as the degree of physical contact between the bystander and the victim, the temporal order of the good and bad effects, whether the act is personal (i.e. emotionally salient) or impersonal, and whether a new threat is introduced or an existing threat is redirected (Mikhail, 2002; Cushman et al., 2006; Hauser, Cushman, Young, Jin,& Mikhail, 2007). Nevertheless, when asked to justify their responses, adults are typically unable to articulate this principle, suggesting that while the PDE is operative in their judgments, it is not accessible to conscious reasoning (Mikhail, 2002; Cushman et al., 2006; Hauser et al., 2007a).

This raises an interesting developmental question – namely, how and when do we acquire this principle? Given that adults are generally unaware that such a principle is guiding their judgments, it is unlikely that the PDE is explicitly taught to children. Nevertheless, some have hypothesized that children may acquire this principle at an early age, as part of an innate appraisal system that mentally represents the causal and intentional properties of acts and omissions (Mikhail, 2002; Hauser, 2006). If this is the case, children should show an early sensitivity to the causal and intentional properties of actions, even when evaluating morally complex acts such as those in the trolley problem

(i.e. acts with multiple outcomes and intentions (Mikhail, 2002)). However, to date, there have only been two studies examining preschool children's intuitive reasoning about acts that produce "double effects."

The first study, by Leslie, Knobe, and Cohen (2006), presented children with stories in which a character's intended action (ex: bringing a pet frog to a friend's house) resulted in either a positive side effect (ex: making the friend happy), or a negative side effect (ex: making the friend sad). Children were told in both conditions that the character knew but did not care about the side effect. Children were then asked if the character brought about the side effect on purpose. Leslie et al. found that children who understood the concept of "not caring" exhibited the same "side-effect effect" that had been previously found in adults (Knobe, 2003): bad side effects were judged as intentional, whereas good side effects were judged as unintentional.[12] Thus, even when presented with morally complex acts, preschool children appear to grasp the distinction between main effects (intended outcomes) and side effects. But do they make a moral distinction between harm that is an intended means and harm that is a foreseen but unavoidable side effect?

The second study, by Pellizzoni, Siegal, and Surian (2010), used modified versions of the footbridge and bystander problems to test preschoolers' understanding of the *contact principle*, which states that harmful actions involving physical contact with the victim are morally worse than equally harmful actions involving no physical contact (eg. Greene et al., 2001; Cushman et al., 2006; Hauser et al., 2007a). At the beginning of

---

[12] A study by Pellizzoni, Siegal, & Surian (2009) replicated and extended these findings by showing that many children continued to show the side-effect effect even when the agent was described as lacking foreknowledge of the outcome, and even in cases where the agent possessed foreknowledge but their state of caring is unspecified.

each dilemma, a ball was shown rolling towards five Lego play-people on a wooden track. In the footbridge dilemma, a smaller Lego person had to push a bigger Lego person in front the rolling ball in order to save the five Lego people. In the bystander dilemma, the bystander had to pull a string, which redirected the ball away from the five Lego people and onto another track, on which only one Lego person was standing. Children were asked to decide what the bystander should do (e.g. push the person or not push the person), and (in a second experiment) what was the *right* thing to do. Pellizzoni et al., found that like adults, preschoolers showed the double-effect effect; that is, they advocated action in the bystander dilemma but did not advocate action in the footbridge dilemma. Furthermore, in response to a third dilemma in which pulling the chord to save one person resulted in harming five others, children did not advocate action. Thus, in all three dilemmas, children's judgments mirrored those of adults; rather than following a simple "something must be done" principle, children only advocated action when it prevented harm to the greatest number of people, and only when the action did not involve harmful physical contact with another person as a means of saving the larger number of people.

The authors concluded from this finding that preschoolers are indeed sensitive to the contact principle. However, given Pellizzoni et al.'s stimuli, it is unclear whether children's moral judgments in these dilemmas were guided by the contact principle alone, since the footbridge dilemma differed from the bystander dilemma not only in terms of harmful *contact*, but also in terms of whether battery was committed as a *means* or as a *side effect*. Thus, it is possible that both the contact principle and the PDE were guiding

children's judgments in these dilemmas. Alternatively, it is possible that unlike adults, children possess unconscious knowledge of only one of these principles.

Only one study with older children, conducted by John Mikhail (2002), has directly investigated whether children possess tacit knowledge of the PDE. Thirty children ages 8-12 years old were asked to read either an "intentional battery" trolley problem, (in which a doctor must remove the organs of one healthy patient as a means of saving five other people in need of an organ transplant), or a "forseeable battery" trolley problem (similar to the bystander problem, except that the bystander is driving the runaway train), and then indicate whether the proposed action was "wrong" and provide a written explanation for their response. Mikhail et al. found that participants were more likely to say the proposed action was wrong in the case of intentional battery than in the case of foreseeable battery. Furthermore, the majority of children's justifications were logically inadequate, although only marginally less adequate than those of adults. Mikhail tentatively concluded that, like adults, children ages 8-12 unconsciously rely on principles such as the PDE and the prohibition of intentional battery (which forbids purposefully or knowingly causing non-consensual bodily contact with another person) when evaluating a morally complex act. However, given the older age range of the children in the study, and the small sample size, these findings provide only weak support for the hypothesis that the PDE is innately specified.

The Current Study

The primary aim of the current study was to investigate whether preschool children possess tacit knowledge of the PDE; that is, I asked whether preschoolers, like adults, morally distinguish between harming one person as a means to helping others (i.e.

harm as a main effect), versus harming one person as a foreseen side effect of helping others. Like Mikhail and Pellizzoni et al., I asked children to evaluate novel scenarios with the same logical structure and abstract moral content as the traditional trolley problems. However, unlike in previous studies, I used scenarios in which the agents made no physical contact with the victims; instead, the current scenarios involved a more abstract moral violation: trespass to a victims' personal property. I chose to use dilemmas involving property violations for several reasons: First, we wanted to rule out harmful physical contact as an explanation of impermissibility. Second, I wanted to use scenarios that appealed to and could be understood by three-year-olds. Current evidence suggests that the concept of death is difficult for preschoolers to grasp (Carey, 1988), but they nevertheless appear to have quite sophisticated reasoning abilities when it comes to ownership and property rights violations (e.g. Beggan (1992), Berti, Bombi, & Lis (1982), Friedman & Neary, 2008; Neary, Friedman & Burnstein, 2009; Rossano, Rakoczy, & Tomasello, 2011). Third, I was interested in whether the prohibition of intentional harm in the PDE extends to acts of harm other than battery. To date, all previously tested trolley problems (in both adults and children) involve only two kinds of moral prohibition: homicide and/or battery (i.e. non-consensual bodily contact). Nevertheless, it is possible that children recognize other forms of legal trespass (e.g. violations of personal property, assault, etc.) as moral prohibitions. If this is the case, children should show the double-effect effect (the pattern of judgments found in the bystander and footbridge problems) even when evaluating scenarios that do not involve homicide or battery.

In addition to testing children's knowledge of the PDE and the prohibition of trespass to personal property, I investigated another aspect of moral reasoning that has received little attention in the developmental literature: the duty of care/rescue. In tort law, a duty of care is a legal obligation to avoid acts or omissions that are likely to cause harm to others. For example, under US common law, parents have a legal duty to care for their children, doctors have a duty to care for their patients, drivers have a duty to care for others on the road, employers have a duty to ensure the safety of their employees at work, etc.). Similarly, a duty of rescue (though not legally recognized in the US) is the obligation to come to the aid of another, if doing so does not put the rescuer (or other individuals) in unreasonable danger. Researchers such as John Mikhail (2011), have argued that such legal obligations reflect an underlying universal moral grammar that includes not only negative duties to avoid causing harm under certain conditions (such as the prohibition of intentional battery and the PDE), but also positive duties to actively prevent harm in certain cases. In the current study, I made a first attempt at testing preschoolers' knowledge of the duty to rescue by including a third scenario in which the agent chooses to do nothing, thereby letting the five people be harmed.

Method

Participants

Participants were 52 preschoolers divided into two age groups: 26 three-year-olds (16 girls) between the ages of 36 and 47 months (M = 42.5 months, SD = 3.7 months), and 26 four- and five-year-olds (15 girls) between the ages of 48 and 76 months (M =

55.9 months, *SD* = 7.1 months). An additional 13 children were eliminated from the study, (5) for failing to cooperate, and (8) for failing to pass the training phase.

Materials and Procedure

Training phase. Children were tested individually in a quiet location at their preschool. All testing sessions were videotaped for future scoring. Prior to testing, children were introduced to the "Pink Scale" (figure 3.1).

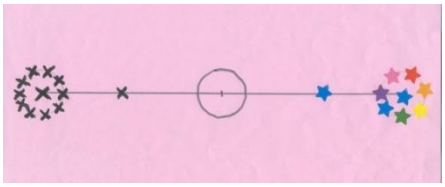

Figure 3.1. The Pink Scale. Children were told to point to the stars when something was good, the X's when something was bad, and the circle when something was "just ok." They were told that one star/X meant "a little good/bad," and lots of stars/X's meant "really good/bad."

Children were told to use the Pink Scale to show when something was good, bad, or "just ok" by pointing to the stars on the right when something was good, pointing to the X's on the left when something was bad, and pointing to the circle in the middle when something was just ok. They were told that lots of stars meant that something was "really good" and lots of X's meant that something was "really bad." A single star meant that something was "a little good" and a single X meant that something was "a little bad". After the scale was explained to them, children were asked to rate various items such as ice cream, eating bugs, and water, by pointing to the appropriate point on the scale.

Once participants were comfortable using all five points on the scale, they were given two training stories (one boy story, and one girl story), each involving two actions (one harming action, and one helping action). The first picture in each story introduced two characters of the same gender, A and B. The following two pictures were presented simultaneously; one picture showed character A harming character B (e.g. hitting him), and the other picture showed A helping B (e.g. giving him a cookie). For each picture,

children were asked whether A *should* have done what he/she did (*normative question*), and then to rate A's actions using the Pink Scale (*ratings question*). Children who passed at least one of these measures (i.e. responded appropriately to at least 3 out 4 normative questions, or at least 3 out of 4 ratings questions) continued to the test phase.

Test phase. Children were presented with a series of three computer-animated dilemmas on a laptop monitor. Each dilemma began with following introduction (figure 3.2):

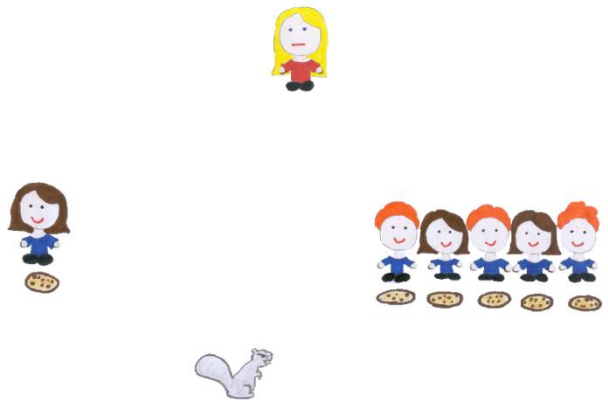

Figure 3.2. Introduction. This is Jane. Jane is in the park today. And there are some other people in the park too. There are lots of people over here, and there is one person over here. And look! They all have a snack. These people over here have a snack, and this person over here also has a snack. What does it look like they are eating? That's right! They all have cookies. But Uh oh! What is that? That's right. That is a sneaky squirrel. And do you know what he likes to do? He likes to eat other people's food! And he sees all those yummy cookies over there, so he is going over there to eat all of those cookies! That will make these people very sad. Well, Jane sees that sneaky squirrel, and Jane knows that the sneaky squirrel is going to eat those cookies and make those people sad. Let's see what Jane does.

Participants then watched Jane take one of three actions.

In the side effect dilemma (figure 3.3), participants watched Jane place a wall between the squirrel and the five cookies, thus redirecting the squirrel towards the cookie on the left.

Figure 3.3. Side effect dilemma. Well, Jane puts up a wall. Jane knows that if she puts up a wall next to these people, the squirrel will go over here and eat this person's cookie instead. But now this person sad. Let's watch that again.

In the omission dilemma (figure 3.4), the participants were told that Jane chose to do nothing, so the squirrel ate all five cookies on the right.[13]

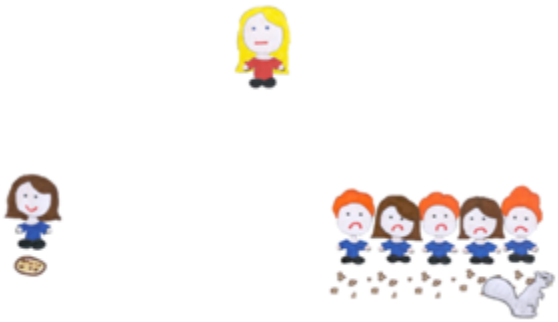

Figure 3.4. Omission dilemma. Well, Jane doesn't do anything. She just stands there and does nothing. Jane knows that if she just does nothing, the squirrel will go over here and eat all these people's cookies. And now they are all sad. Let's watch that again.

In the main effect dilemma (figure 3.5), participants watched Jane take the cookie on the left and feed it to the squirrel in order to prevent the squirrel from eating the cookies on the right.

[13] Although we refer to this dilemma as the "omission" dilemma, it is technically an "inaction" dilemma, as omission generally implies abstaining from a particular act (e.g. *not* "flipping the switch"), whereas in our omission story, it is not entirely clear *what* act is being omitted.

Figure 3.5. Main effect dilemma. Well, Jane takes this person's cookie and gives it to the squirrel. Jane knows that if she takes this person's cookie and gives it to the squirrel, the squirrel will eat this person's cookie instead. But now this person is sad. Let's watch that again.

The order in which the three dilemmas were presented was counterbalanced such that the "omission" dilemma was always presented as the second dilemma in the series, and the other two alternated as the first and last dilemmas presented. Before each new dilemma, children were told that they were going to see the same story again, but this time Jane would do something a little different. For all three dilemmas, participants were asked control questions before Jane acted, to ensure that they understood what the squirrel was about to do, and how the five children would feel if their cookies were eaten. Participants then watched each cartoon twice. During the second run-through, children were questioned to make sure they understood what Jane had done, what the squirrel had done as a result of her action (or inaction), and how the victim(s) felt when their cookie(s) were eaten (See Appendix B for a complete script).[14] If a child did not answer a control question correctly, he or she was corrected by the experimenter, and the relevant portion of the story was retold; the child's comprehension was then checked again before proceeding.

Participants were then asked two test questions: 1) "Should Jane have done that?" (*normative question*) 2) "Can you show me on the Pink Scale? Was that a good thing to do, a bad thing to do, or an ok thing to do? Was it a little good/bad or really good/bad? (*ratings question*). Pink Scale ratings were scored as -2 for "really bad," -1 for "a little bad," 0 for "just ok," +1 for "a little good," and +2 for "a really good." Responses to all control and test questions were coded by at least two independent observers. If discrepancies occurred between the observations, a third coder was used.

Results

Results are organized into two sections, beginning with the comparison between side effect and main effect dilemmas, and ending with the omission results. Within each

[14] No "knowledge" control question was asked because a prior pilot study revealed that children understood that Jane foresees the consequences of her actions in all three dilemmas. Previous studies also suggest that children often default to an assumption of shared knowledge (Roth & Leslie, 1998; Wimmer, Hogrefe, & Perner, 1988). In fact, 3-year-olds sometimes have difficulty understanding that an actor does NOT know something.

section, results for each of the two measures (normative question, ratings) are presented in turn. Preliminary analyses indicated that there were no significant gender differences for either measure, so gender was dropped from further analyses.

1. Did preschoolers show the double-effect effect?

Normative question. Overall, 81% of children approved of Jane's action in the side effect dilemma (Binomial test, $N = 52$, $p < .001$, two tailed), while only 25% of children approved of Jane's action in the main effect dilemma (Binomial test, $N = 52$, $p < .001$, two-tailed). Of the 52 participants, 33 children (63%) showed the double-effect effect (i.e. responded that Jane have done what she did in the side effect dilemma but should not have done what she did in the main effect dilemma). Only 4 participants (8%) showed the reverse pattern of judgment, (McNemar test, $\chi^2(1, N = 52) = 21.189$, $p < .001$, two-tailed). Nine children gave a pattern of judgments consistent with utilitarianism (i.e. advocated action to save the greatest number of people in both dilemmas), and six children responded that Jane should not have acted in either dilemma. No age effects were found (Fisher's exact test, all *p*s > .05) (see figure 3.6, left panel).

To eliminate the possibility of order effects, first-trial side effect and main effect responses were analyzed, with comparisons made between subjects. Of the 26 children who saw the side effect dilemma first, 23 (88%) responded 'yes' to the normative question (Binomial test, $N = 26$, $p < .001$, two-tailed), whereas only 8 out of 26 children (31%) who saw the main effect dilemma first did so (Binomial test, $N = 26$, $p = .076$, two-tailed). Responses to the two dilemmas differed significantly, Fisher's exact test, $\chi^2(1, N = 26)$, $p < .001$, two-tailed. No age differences were found (*ps* > .05) (see figure

3.6, right panel). These results indicate that children's responses to the normative question were consistent with the PDE.

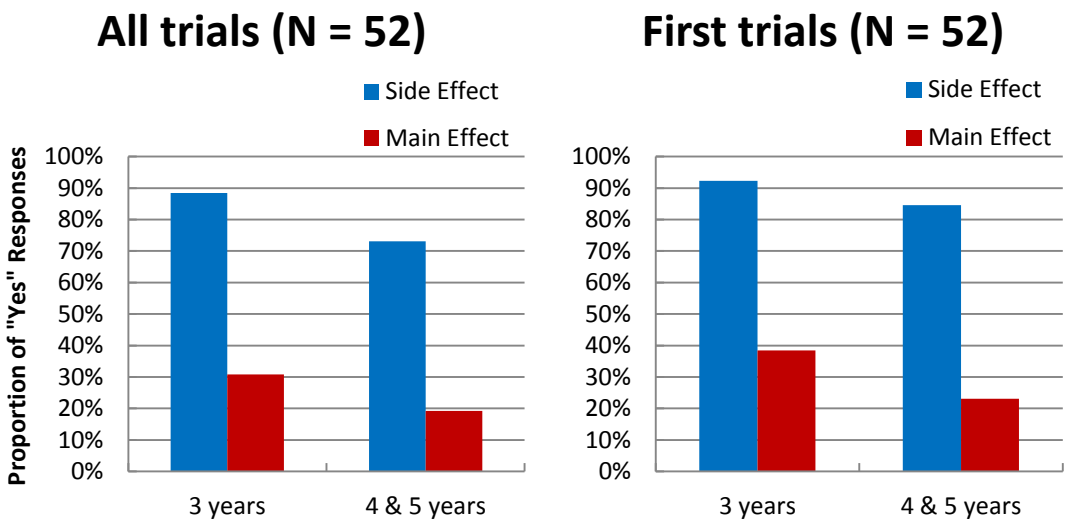

Figure 3.6. Children's normative judgments as a function of age and dilemma. The percentages of children in each age group who responded "yes" to the normative question for all side effect and main effect dilemmas (left panel), and for dilemmas that were presented first (right panel).

Ratings question. Ratings for the side effect and main effect dilemmas were analyzed using a mixed-design ANOVA with dilemma (2: side-effect, main effect) as a within-subjects factor, and age (2: 3-year-olds, 4- and 5-year-olds) and order (2: side-effect first, main effect first) as between-subjects factors. A large effect of dilemma was found, $F(1, 48) = 44.977$, $p < .001$, $\eta_p^2 = .484$, but no main effect of age, $F(1, 48) = .005$, $p = .943$, $\eta_p^2 < .001$, or order, $F(1, 48) = .879$, $p = .353$, $\eta_p^2 = .018$ was found. As predicted by the PDE, the effect of dilemma reflected a tendency to judge main effect dilemmas ($M = -1.40$, $SE = 0.15$) as more severe than side-effect dilemmas ($M = 0.19$, $SE = 0.21$). This result was confirmed non-parametrically, Wilcoxon signed-ranks test, $N = 52$, $Z = -4.674$, $p < .001$, $r = .65$ (See Appendix C for the distribution of children's ratings in each dilemma).

However, the effect of dilemma was qualified by a small but significant dilemma x age x order interaction, $F(1, 48) = 9.930$, $p = .003$, $\eta_p^2 = .171$. Analyses of simple effects revealed that three-year-olds rated the side effect dilemma significantly higher than the main effect dilemma when the main effect dilemma was presented first, $F(1, 48) = 19.038$, $p < .001$, $\eta_p^2 = .284$, but their ratings did not differ significantly when the side effect dilemma as presented first, $F(1, 48) = 3.160$, $p = .082$, $\eta_p^2 = .062$. Conversely, four- and five-year-olds' showed the double-effect effect only when the side effect dilemma was presented first, $F(1, 48) = 30.189$, $p < .001$, $\eta_p^2 = .386$, but their ratings did not differ significantly when the main effect dilemma was presented first, $F(1, 48) = 3.160$, $p = .082$, $\eta_p^2 = .062$ (see figure 3.7).

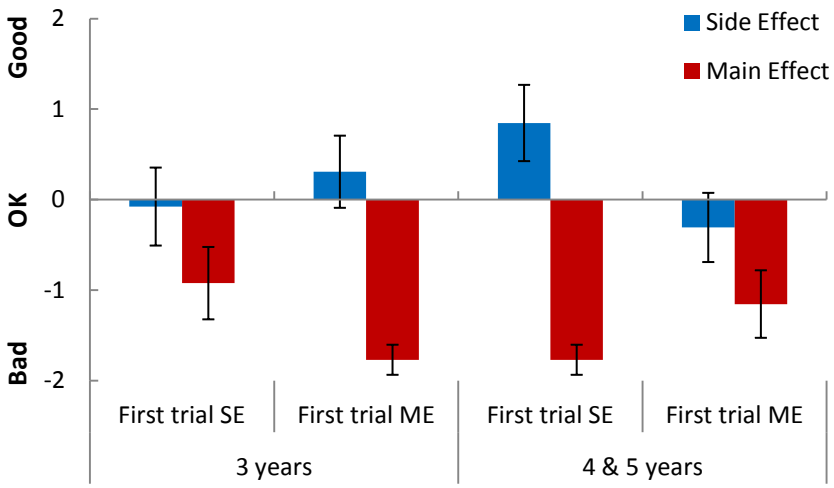

Figure 3.7. Children's ratings as a function of age, order, and dilemma. This figure shows children's ratings for side effect and main effect dilemmas as a function of their age, and whether the side effect dilemma was presented as the first dilemma (first trial SE) or the last dilemma (first trial ME). Error bars show standard error of the mean.

To eliminate these order effects, a two-way ANOVA was used to compare first-trial side effect responses (i.e. ratings in the side effect dilemma only for children who saw the side-effect dilemma first) to first-trial main effect responses (i.e. ratings in the main effect dilemma only for children who saw the main effect dilemma first), with dilemma and age as between-subjects factors. As predicted, first-trial side effect ratings

(M = 0.38, SE = 0.30) were significantly higher than first-trial main effect ratings (M = -1.46, SE = 0.21), $F(1, 48) = 25.743$, $p < .001$, $\eta_p^2 = .349$. This result was confirmed non-parametrically (Mann-Whitney U = 131, $p < .001$, r = .56). Three-year-olds were also found to give lower ratings on average (M = -0.92, SE = 0.30) than older children (M = -0.15, SE = 0.40), $F(1, 48) = 4.469$, $p = .040$, $\eta_p^2 = .085$.

2. How did preschoolers judge omission?

Normative question. Overall, 60% of participants disapproved of Jane's decision to do nothing in the omission dilemma. However, this proportion did not differ significantly from chance, (Binomial test, N = 52, p = .212). A binary logistic regression analysis was used to test the main and interactive effects of age (2: three-year-olds, 4- and 5-year-olds) and order (2: side effect first, main effect first) on normative responses in the omission dilemma. Using a forward stepwise procedure, the best-fitting model included a main effect of age, Wald $\chi^2(1, N = 52) = 8.850$, $p = .003$, $\beta = 1.905$, odds ratio (OR) = 6.720. No main or interactive effects of order were included in the model (ps > .05). Inspection of the data (figure 3.8) indicated that a significant proportion of four- and five-year-olds judged the omission dilemma as impermissible, whereas three-year-olds responded at chance: 21 out of 26 four- and five-year-olds (81%) responded "no" to the normative question in the omission dilemma (Binomial test, N = 26, p = .003, two-tailed). By contrast, only 10 out of 26 three-year-olds (38%) did so (Binomial test, N = 26, p = .33, two-tailed). This model had an overall correct prediction rate of 71.2%. A test of this model against a constant-only model was statistically significant, $p < .001$.

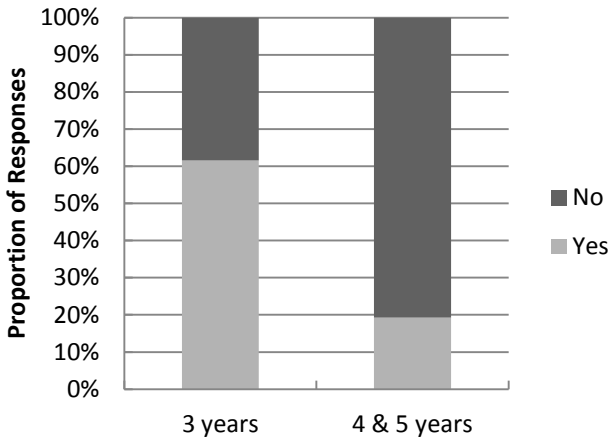

Figure 3.8. Omission: Children's normative judgments by age. This figure shows the distribution of children's responses to the question, "Should she have done that?" in the omission dilemma by age group.

As illustrated in figure 3.9, when the proportion of children's "yes" responses in the omission dilemma was compared to "yes " responses in the other two dilemmas, children were significantly more likely to say "yes" in the side effect dilemma than in the omission dilemma, McNemar test, $\chi^2(1, N = 52) = 13.793$, $p < .001$, two-tailed, but the proportion of "yes" responses in the omission and main effect dilemmas were equally low, McNemar test, $\chi^2(1, N = 52) = 3.063$, $p = .08$, two-tailed.

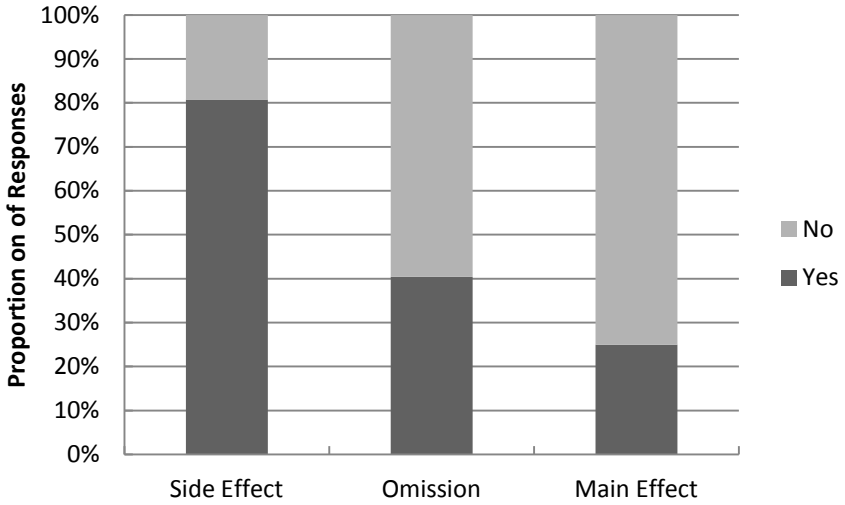

Figure 3.9. Children's normative judgments by dilemma. This figure shows the distribution of children's responses to the question, "Should she have done that?" for each of the three dilemmas.

Ratings. Preschoolers across both age groups rated the omission dilemma negatively, with an average rating of -0.83 ($SE = 0.18$), which was significantly different

from chance, $t(51) = 4.554$, $p < .001$, two-tailed. A 2 x 2 ANOVA showed no effects of age or order on omission ratings ($ps > .05$).

Ratings for the omission dilemma were compared to side effect and main effect dilemmas using a 3 (dilemma: side effect, omission, main effect) x 2 (age: 3-year-olds, 4- and 5-year-olds) x order (side effect first, main effect first) repeated measures ANOVA. As expected, mean ratings differed significantly between dilemmas, $F(2, 96) = 23.798$, $p < .001$, $\eta_p^2 = .331$. Post-hoc tests using the Bonferroni correction revealed that the omission dilemma was rated significantly lower than the side effect dilemma, $p < .001$, and significantly higher than the main effect dilemma, $p = .049$ (figure 3.10).

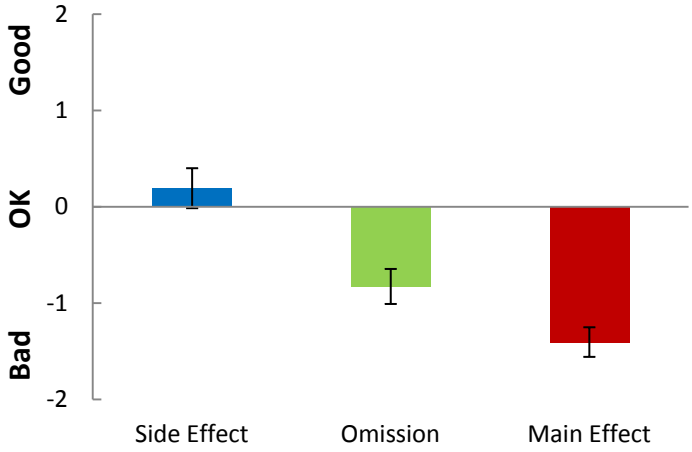

Figure 3.10. Children's average ratings for each of the three dilemmas. Error bars show standard error of the mean.

However, the effect of dilemma was qualified by a small but significant interaction between dilemma, age, and order, $F(2, 96) = 6.202$, $p = .003$, $\eta_p^2 = .114$. Simple effects indicated that three-year-olds' mean ratings differed significantly among the dilemmas only when the main effect dilemma was presented first, $F(1, 47) = 9.839$, $p < .001$, $\eta_p^2 = .295$, but not when the side effect dilemma was presented first, $F(1, 47) = 1.667$, $p = .200$, $\eta_p^2 = .006$, whereas four- and five-year-olds' mean ratings differed significantly among the dilemmas only when the side effect dilemma was presented first,

$F(1, 47) = 15.885$, $p < .001$, $\eta_p^2 = .403$, but not when the main effect dilemma was presented first, $F(1, 47) = 1.550$, $p = .223$, $\eta_p^2 = .062$. Follow-up tests indicated that three-year-olds rated the side effect dilemma significantly higher than the omission dilemma, $p = .009$ (and significantly higher than the main effect dilemma, $p < .001$) only when the main effect dilemma was presented first. Four- and five-year-olds rated the side effect dilemma significantly higher than the omission dilemma, $p < .001$ (and significantly higher than the main effect dilemma, $p < .001$) only when the side effect dilemma was presented first. (see figure 3.11).

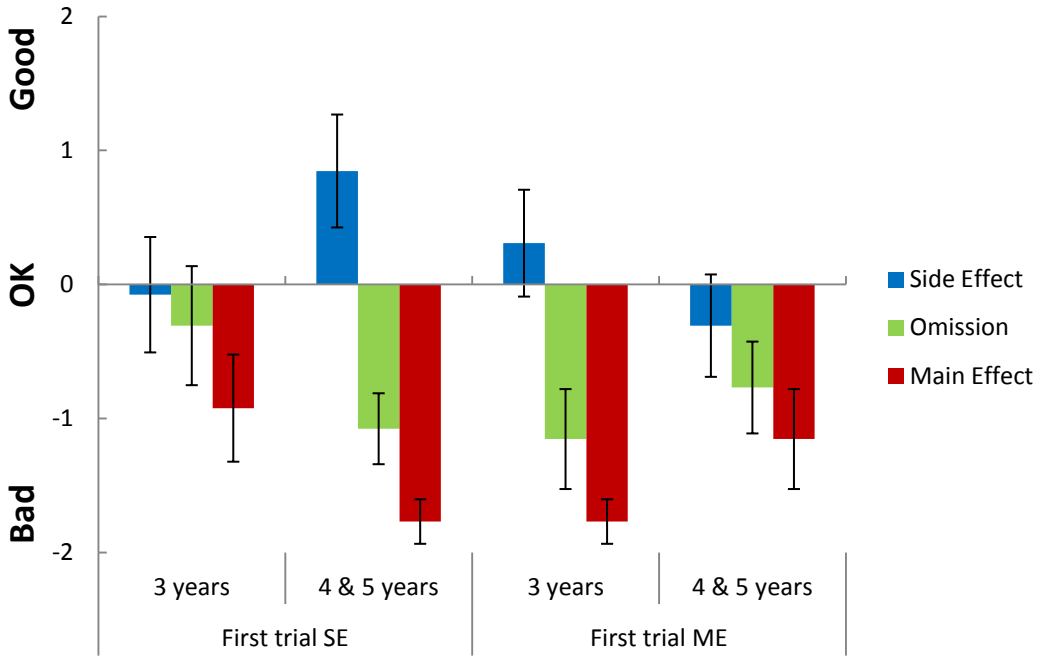

Figure 3.11. Children's ratings as a function of order, age, and dilemma. This figure shows children's average ratings for each of the three dilemmas as a function of age, and whether the side effect dilemma was presented as the first dilemma (first trial SE) or the last dilemma (first trial ME). Error bars show standard error of the mean.

Discussion

The current study used moral dilemmas similar in structure to the so-called trolley problem to investigate whether preschoolers' moral judgments are consistent with

Aquinas's principle of double effect, which states that an act that results in a harmful effect is morally permissible if and only if the harmful effect is a foreseen but unintended side effect of bringing about a greater good, but not if the harmful effect is the means to bringing about the greater good. I found that children as young as three years old showed an adult-like pattern in their normative judgments consistent with Aquinas's principle: children advocated saving five people at the cost of harming one person when the harm was a foreseen *side effect* of saving the five people, but not when the harm was the means (i.e. *main effect*) to saving the five people. I also found a similar pattern in children's ratings, although this pattern appeared to be qualified by different anchoring effects in each age group: Three-year-olds' rated the agent's action in the side effect dilemma significantly higher than the agent's action in the main effect dilemma only when the main effect dilemma was presented first, whereas four- and five-year-olds did so only when the side effect dilemma was presented first. Because the number of participants in each age x order condition was fairly small, it is premature to draw any meaningful conclusions from these order effects. However, these findings are consistent with evidence of order effects in the adult literature (Lombrozo, 2009; Mikhail, 2002; Petrinovich & O'Neill, 1996; Schwitzgebel & Cushman, 2012; Wiegmann, Okan, & Nagel, 2012). Nevertheless, when only first trial responses were compared, both age groups showed a dominant pattern of ratings consistent with the PDE: first-trial side effect dilemmas were rated significantly higher than first-trial main effect dilemmas.

These results conceptually replicate the results of Pellizzoni, Siegal, and Surian (2010), but also extend their findings in important ways. Pellizzoni et al. found that preschoolers advocated intervening to save five people at the cost of harming one person

only when such intervention did not involve physical contact with the victim. The trolley problems in the current study, though structurally similar to those of Pellizzoni et al., involved no physical contact with the victims, or even any physical harm (i.e. battery) to the victims. Instead, the harmful acts involved violations of personal property, and their negative effects were purely psychological (provoking sadness in the victim).[15] Nevertheless, children continued to show the double-effect effect, suggesting that the same abstract pattern of intuitive reasoning holds across different kinds of acts and moral violations.

Together with the findings of Pellizzoni et al., the current results suggest that children's moral intuitions in these scenarios are not consequence-driven. Children do not simply follow a utilitarian principle such as "acts that maximize happiness are permissible", nor do they follow a simple "do no harm" heuristic (i.e. "acts that result in harm are impermissible"). Moreover, children's moral judgments do not appear to be based solely on simple perceptual features such as bodily contact. Instead, the current findings point to a more nuanced moral reasoning in preschoolers that gives a critical role to the causal and intentional structure of actions. In particular, these results imply that preschoolers, like adults, make a moral distinction between intended means and foreseen side effects.

However, this conclusion is speculative, as the current findings are also consistent with alternative explanations. For example, it is possible that children in the current study were simply following the rule "don't take things that aren't yours." Although

---

[15] Our measures also differed slightly from those of Pellizzoni et al. While their questions required children to make a forced choice about what the character should do next (X or not X?), our measures required children to make post-hoc judgments (Should Jane have done X?) and rate the agent's action on a permissibility scale.

further research is needed to definitively rule out this explanation, I suspect that such a hypothesis is unlikely to adequately describe preschoolers' intuitions for many of the same reasons that the rule "don't push people" is inadequate to describe adult and children's intuitions in the trolley problem. For one thing, it generates the rather unlikely prediction that children would disapprove of taking another person's property even in cases where such an act is clearly permissible. For example, imagine a scenario in which an agent takes another person's cookie in order to protect it from an imminent threat (such as a hungry squirrel), or a scenario in which an agent knows that the cookie has been contaminated, and is attempting to prevent the owner from eating it and getting sick. These cases both involve the action description "taking another person's property," but are permissible because they involve cases of implied consent; that is, taking the other person's property is done with the intention of furthering the other person's (inferred) goals (Mikhail, 2011). It would be surprising if children did not recognize the distinction between "taking another person's property" in the case of implied consent versus in the main effect dilemma. After all, children distinguish between other kinds of acts on the basis of their intentional structure quite early in development. For example, infants as young as six-months-old distinguish between helpful pushing and harmful pushing (i.e. battery) (Hamlin et al., 2007), and three-year-olds distinguish between intentionally telling a falsehood (i.e. lying), versus telling a falsehood due to an honest mistake, versus telling a falsehood due to carelessness or negligence (Siegal & Peterson, 1998). Given that children make these kinds of subtle distinctions, among others, it is arguably more parsimonious to attribute a limited number of abstract, structure-dependent principles to

children, than to postulate a potentially infinite series of stimulus-specific, case-by-case prohibitions in order to fully account for children's moral knowledge.

Perhaps a more plausible explanation that has been proposed as an alternative to the means/side effect distinction is the distinction between introducing a new threat versus redirecting an existing threat. Although some evidence suggests that this distinction cannot account for adults' moral judgments on a subset of trolley problems (Hauser et al., 2007a Mikhail, 2002)[16], this distinction could nevertheless account for children's divergent moral judgments in the current study.

Yet another relevant distinction that has been explored in the adult literature points to the temporal order of the good and bad effects in the trolley problem: In the bystander problem, the five people are saved before the one person is harmed, whereas in the footbridge problem, the one person is harmed before the five people are saved. Again, this distinction does not appear to be descriptively adequate for explaining adults' moral intuitions (Hauser et al., 2007a; Mikhail, 2002; Sinnott-Armstrong, 2008), but it may nevertheless be adequate to describe the moral judgments of preschoolers in the current study.[17]

Further research is needed to explore these alternative explanations. Indeed, the explanation for why adults make a moral distinction between the bystander and footbridge problems is an ongoing subject of debate among psychologists and philosophers. Still, it is likely that at least some of the factors that contribute to this

[16] See the discussion of the Loop Track and Man-In-Front problems in Chapter 1

[17] However, recall that the children in Pellizzoni et al.'s (2010) study disapproved of scenarios in which pulling the chord to save one person resulted in harming five others. In this case, children disapproved of an action in which the good effects occurred before the bad effects. Therefore, even if such a temporal-based principle is operative in children's judgments, it must require not only the condition that the good effects must come before (or concur with) the bad effects, but also that the good effects must outweigh the bad effects – conditions which are also part of the PDE.

pattern of intuitions in both adults and children involve abstract structural representations of properties and relations that are not directly observable in the stimulus.

In addition to testing preschoolers' knowledge of the PDE, this study is the first to ask children to morally evaluate a scenario in which an agent refrains from preventing a foreseeable harmful outcome. On the normative measure, four- and five-year-olds generally responded that the agent should *not* have "done nothing" (i.e. should have done something) in the omission dilemma. However, three-year-olds responded at chance to the normative question (I will return to this finding later). On the ratings measure, children in both age groups tended to judge that "doing nothing" in this dilemma was bad, and tended to prefer action (vs. inaction) when one person was harmed as a foreseen side effect of saving the five people. However, children did not prefer action when one person was harmed as a means of preventing harm to five people. In fact, despite the fact that five people were harmed in the omission dilemma and only one person was harmed in the main effect dilemma, children tended to rate the omission dilemma more positively than the main effect dilemma. This finding reinforces the claim that children do not simply follow a "something must be done" heuristic (Pellizzoni et al., 2010), nor are children's judgments purely outcome-driven. Instead, children's moral judgments appear to be sensitive to the manner and intentions with which harm occurs. In particular, Isuggest that this pattern of judgments may reflect underlying knowledge of the duty to rescue – the moral obligation to prevent a harmful outcome, unless doing so requires unjustified costs. If so, this finding supports John Mikhail's speculation that "an adequate rescue principle must occupy a subordinate position in a "lexically ordered"

scheme of principles, in which at least some negative duties to avoid harm are ranked higher than at least some positive duties to prevent harm" (Mikhail, 2011, p. 145).

However, the results in the current study provide only limited support for this hypothesis. It is important to note that the duty to rescue is fundamentally comparative, as it involves comparing an act or omission with its least harmful alternative (Mikhail, 2011). However, it is not clear what comparison children were making in the current study when they judged that "doing nothing" in the omission dilemma was bad. For one thing, the least harmful alternative to "doing nothing" was not explicitly stated, nor is it obvious what that alternative would be. Furthermore, it is not clear that "doing nothing" was actually perceived as an omission in the current study. Indeed, the "omission" dilemma would perhaps be more accurately termed an "inaction" dilemma, as omission generally implies abstaining from a particular act (e.g. *not* "flipping the switch"), whereas "doing nothing" in the current study potentially entailed omitting an unknown number of acts, including (but not limited to) the act in the previous dilemma[18].

This brings us to the question of why three-year-olds performed at chance on the normative measure in this dilemma. There are a few possible explanations for this result. First, younger children may simply have had difficulty parsing the (admittedly awkward) question "Should she have done nothing?" Second, given the comparative nature of the rescue principle, and the ambiguities of the omission dilemma discussed above, three-year-olds may have found the demands of a binary (impermissible/permissible) response particularly difficult. In the omission dilemma, the agent neither caused harm, nor did she prevent it. According to the rescue principle, in order to judge "doing nothing" in the

[18] This may also pose potential problems for meeting the "no better alternative" condition of the PDE in the other two dilemmas. I return to this issue in Chapter 6.

omission dilemma as impermissible, the child must infer that the agent *could* have intervened in a permissible/low-risk manner, but chose not to. In other words, the child must keep in mind two acts (and their effects) simultaneously: the current act (knowingly harmful omission) and its least harmful alternative. This kind of counterfactual reasoning is cognitively demanding, particularly for younger children (most likely due to working memory constraints) (Robinson & Beck, 2000)[19]. Furthermore, as already mentioned, the least harmful alternative to doing nothing was not explicitly stated in the omission dilemma; children therefore had to generate their own "least harmful alternative," increasing the demands of the task even further.

More research is needed to explore preschoolers' understanding of the duty of rescue, and the conditions under which children consider it morally obligatory to intervene on another's behalf. The development of children's understanding of choice in these scenarios is also in need of further study. Often, the choice an agent did not make but could have made (or an action the agent knowingly could have taken but did not take) is relevant to moral judgment. However, it is not clear how we generate the agent's "least harmful alternative" when a potentially limitless number of actions could be taken. How are these possibilities constrained, and how do these constraints affect moral judgment?

Future research should also investigate a principle that, although closely related to the principles investigated in this study, was not directly tested in the current study: the action/omission principle (Cushman et al., 2006), otherwise known as the omission bias (Baron & Ritov, 2004; Spranca, Minsk, & Baron, 1991), or the Doctrine of Doing and

[19] For four-year-olds, the ability to generate a least-harmful alternative may have been facilitated by first observing a permissible alternative to doing nothing (i.e. the side effect dilemma), as indicated by the fact that none of the older children who saw the side effect dilemma first judged the omission dilemma as permissible, whereas 5 out of the 13 children who saw the main effect dilemma first judged the omission dilemma as permissible.

Allowing (Quinn, 1989; Rachels, 1975). Research with adults has shown that we tend to judge harm by *commission* as morally worse than equivalent harm by *omission* (Baron & Ritov, 2004; Cushman et al., 2006; Kamm, 1998; Spranca et al., 1991), but few developmental studies have investigated whether children show this bias. In the current study, I did not test children's knowledge of this principle directly, since I did not ask children to judge a control dilemma in which five people were harmed as a result of the agent's *action*. However, the fact that children tended to judge the omission dilemma as more permissible than the main effect dilemma, even though more people were harmed in the omission dilemma, suggests that preschoolers may be indeed be sensitive to the distinction between actively causing harm and allowing harm to occur.

IV. Investigating the role of group structure in moral judgment

In Chapter 3, I showed that like adults, children as young as three years old show an asymmetry in their moral judgments of "trolley-like" dilemmas, approving of an agent's choice to save the greatest number of people only in cases where harm to another person was not intended as the means of doing so. I also showed that preschoolers disapprove of an agent's choice not to act when she could have prevented harm to others. In Chapters 4 and 5, I investigate whether this pattern of intuitions is influenced by perceived ingroup/outgroup structure – a factor which some have hypothesized to be a foundation of moral judgment (e.g. Haidt & Graham, 2007).

The Question of Moral Impartiality

In rationalist moral theory, impartiality is considered a fundamental component of morality. In *A Theory of Justice*, John Rawls argues that a truly rational moral judgment must be made from the "original position":

> A purely hypothetical situation…[in which] no one knows his place in society, his class position or social status, nor does anyone know his fortune in the distribution of natural assets and abilities, his intelligence, strength and the like….The principles of justice are chosen behind a veil of ignorance" (Rawls, 1971).

At least in the Western world, we generally uphold this view of rational impartial justice as well. For example, in the United States, our legal system is based on the fundamental principle that all humans should be treated as having equal moral status under the law. And yet no one can deny that our history is rife with examples in which one social group applied a supposedly "universal" moral imperative (e.g. "though shalt not kill", "all men are created equal", etc.) only to members of their ingroup. From slavery to genocide,

humans have committed countless atrocities against certain social groups while simultaneously claiming to uphold fair and impartial moral principles. How then do we reconcile this pattern of exclusion, prejudice, and violence against the "other" with a rational and impartial sense of justice?

One theory, as advocated by Mikhail (2011) and Hauser (2006), assumes that there is a principled distinction between "considered judgments" – judgments in which our moral capacities are most likely to be displayed without distortion" (Rawls, 1971, p. 47) – and prejudices. Under this view, although we may behave in ways that are prejudiced or discriminatory, such behavior is not an accurate reflection of our underlying moral competence. In other words, the moral grammar underlying our judgments is impartial, but factors exogenous to the moral system, such as the feelings, attitudes, and stereotypes we form toward certain groups or individuals, may (either consciously or unconsciously) bias or distort our moral judgments and actions.

Alternatively, at its very core, our intuitive sense of justice may be far from impartial. According to Haidt's moral foundations theory, ingroup/loyalty is one of the five psychological foundations of morality (See Haidt & Graham, 2007 for a description of the other four moral foundations). Those who support this theory use evolutionary arguments such as group selection and Inclusive Fitness Theory (Hamilton, 1964) to explain why people might be pre-wired to care more about the welfare of the ingroup than the outgroup when making moral evaluations; According to this view, because trust, cooperation, and loyalty to one's ingroup (and distrust of outgroup members) were evolutionarily advantageous traits, we developed an innate "evolutionary preparedness" to attend to social categories, and to weigh group identity and cohesion concerns when

making a moral judgment (i.e. to approve of actions that benefit the ingroup and disapprove of those who betray or fail to come to the aid of the ingroup) (Haidt & Joseph, 2007).

The latter theory, if correct, is disheartening. If our moral system is inherently biased to value the life of an ingroup member over that of an outgroup member (a notion which seems fundamentally at odds with our modern notions of "fairness" and "justice"), then we must fight against our innate sense of morality if we wish to live in a more just (i.e. impartial) world. Nevertheless, the question of whether we have a moral system that takes information about social categories as input (i.e. as part of our moral competence) is an empirical one. Indeed, both Rawls and Mikhail allow for the possibility that "when moral theorists attempt to solve the problem of descriptive adequacy, the set of moral judgments they originally take to be an accurate reflection of moral competence may change" (Mikhail, 2011, p. 55). It is possible, even within the framework of Universal Moral Grammar, that our blemished history of intergroup conflicts reflects an underlying moral system which is inherently biased to apply different rules to determine deontic status, depending on who is harmed or helped. This might occur via a parameter setting for who "counts" as a member of the ingroup (i.e. only those whom we consider to be "ingroup members" are within the scope of moral concern), a general prohibition against betraying one's group, and/or a principle specifying how to evaluate moral acts committed against ingroup versus outgroup members. On the other hand, social categorization and moral judgment processes may develop independently of each other, obeying different sets of rules. If this is the case, rather than reflecting core moral knowledge, the moral judgments that are influenced by group membership should be

viewed as performance errors due to extra-moral factors, such as a desire to comply with group norms, or a desire to rationalize the behavior of ourselves and our ingroup in order to maintain a positive self-image.

Social Categorization and Moral Judgment in Children

One of the ways we can begin to answer these questions empirically is by studying the intersection of moral judgment and intergroup reasoning in children. At what age do children become sensitive to their own group membership, and do they take group structure into account when making moral judgments? To what extent do young children apply the same moral principles to ingroup and outgroup members? Do they consider members of their ingroup to be of equal moral status to members of the outgroup? Surprisingly, until recently, relatively few developmental studies have examined these kinds of questions.

Intergroup preferences in children

Although the intersection between group cognition and moral judgment has been largely neglected in the developmental literature, children's knowledge of intergroup relationships (independent of moral judgment) has been more thoroughly investigated. Current evidence suggests that for some social categories, identification with one's ingroup may emerge quite early in development. For example, looking time studies have shown that children as young as five months old prefer to look at speakers of their native language over speakers of a foreign language (Kinzler, Dupoux, & Spelke, 2007), and three-month-old infants (but not newborns) prefer to look at racial ingroup members over racial outgroup members (Bar-Haim, Ziv, Lamy, & Hodes, 2006; Kelly, et al. 2005).

On their own, these findings provide only weak support for ingroup preferences, as looking time measures are limited in their ability to discriminate between increased looking time due to social preferences, and increased looking time due to perceptual familiarity. Particularly in the case of racial preference, increased looking time is most likely due to visual familiarity rather than an ingroup racial preference, as infants who reside in a community in which they have been exposed to both own-race and other-race members show no preference for their racial ingroup (Bar-Haim et al., 2006). Nevertheless, a series of studies by Katherine Kinzler and colleagues suggest that in contrast to race, language-based looking time differences may indeed reflect an early social preference for speakers of one's native language. They found that whereas 10-month-old infants interacted equally with own-race and other-race individuals (Kinzler & Spelke, 2011), infants preferentially choose to interact with native speakers of their native language over non-native speakers (e.g. English speaking children preferred to reach for a toy offered by an English speaker rather than a French speaker) (Kinzler et al., 2007). Furthermore, whereas white 2.5-year-old children gave toys equally to white and black individuals (Kinzler & Spelke, 2011), English-speaking children more frequently gave toys to native speakers rather than non-native speakers (Kinzler et al., 2007). These results suggest that language may be a particularly salient marker of social categories for young children, and a social preference for native speakers of their own language may emerge within the first year of life.

By the preschool years, children show clear ingroup preferences and biases across a wide variety of other social categories such as race (Aboud, 1988; Dunham, Baron, Banaji, 2008; Katz & Kofkin, 1997; Kircher & Furby, 1971; Kowalski & Lo, 2001),

gender (Alexander & Hines, 1994; Dunham, Baron, & Carey, 2011; Katz & Kofkin, 1997; Maccoby & Jacklin, 1987), age (French, 1984), and even novel categories such as shirt color, shared preferences, or drawing ability (Bigler, Brown, & Markell, 2001; Bigler, Jones, & Lobliner, 1997; Nesdale & Flesser, 2001; Nesdale, Durkin, Maass, & Griffiths, 2004; Patterson & Bigler, 2006). Although the research on novel categories suggests that ingroup biases are stronger when competition, and/or status differences between the groups are present (Mullen, Brown, & Smith, 1992), some evidence suggests that preschoolers demonstrate several forms of intergroup bias even in the case of minimal groups - a subset of novel groups for which group assignment is entirely arbitrary and uninformative, and no additional information about group status or competition is provided (Tajfel, 1971/2000).

For example, a recent study by Dunham and colleagues found that randomly assigning five-year-olds to minimal social groups (groups marked by t-shirt color) was sufficient to induce moderate to large ingroup bias on a variety of tasks (even when the groups were not verbally labeled), including explicit and implicit attitude measures, resource allocation tasks, behavioral attribution tasks (e.g. who would perform or be the recipient of positive events), and reciprocity prediction tasks (e.g. who would be more likely to share with the participant) (Dunham, Baron, & Carey, 2011). Thus, even in the absence of any relevant information about the ingroup and outgroup, and after only brief identification with a particular group, preschoolers' attitudes, patterns of interaction, and expectations about ingroup and outgroup members appear to be influenced by group membership. However, as I argue below, the question of whether children's *moral* judgment is influenced by group membership remains open.

Social groups and moral judgment in children

The developmental literature on the intersection of group cognition and moral judgment is relatively sparse, and somewhat contradictory. Some studies suggest that children expect agents to behave with favoritism towards their ingroup, and prefer agents who do so. For example, a recent study found that infants as young as 9 months old not only preferred agents who helped similar others (i.e. those who shared the infant's food preferences) but also preferred agents who harmed dissimilar others (Hamlin, Mahajan, Liberman, & Wynn, 2013). Similarly, a study by Dunham and colleagues (2011) found that preschoolers expected more generous and reciprocal behavior from members of their ingroup than from outgroup members. A study by Marjorie Rhodes (2012) also found that preschoolers who were introduced to novel social groups expected members of one group to harm members of the other group, but not to harm members of their own group (they did not use group information to make predictions about helpful actions).

However, some studies suggest that children's moral judgments of outgroup harm are more nuanced. For example, a study by Abrams and colleagues used a minimal group paradigm to examine how children between 5 and 11 years of age weigh both moral norms and group considerations when evaluating ingroup and outgroup members (Abrams, Rutland, Ferrell, & Pelletier, 2008). They found that when only group-based information was available (i.e. group identity information and loyalty information), children used both forms of group criteria in their evaluations: they evaluated ingroup members more favorably (on a 5-point "feeling" scale) than outgroup members, and evaluated both ingroup and outgroup members who were loyal to the child's ingroup more favorably than ingroup members who were disloyal (i.e. loyal to the outgroup) and

outgroup members who were loyal to their own group. In a second study, they found that when group identity information *and* moral (i.e. fairness) information was available, children again used both types of criteria to evaluate ingroup and outgroup members. However, children showed stronger differentiation between moral and immoral members than between ingroup and outgroup members. In other words, children primarily favored moral over immoral members, but to a lesser extent they also tended to favor moral ingroup members over moral outgroup members, and immoral ingroup members over immoral outgroup members. Furthermore, consistent with a domain-independence hypothesis, these two forms of evaluation (group-based and morality-based evaluation) were uncorrelated.

In addition to measuring children's evaluations of moral and immoral ingroup and outgroup members, Abrams and colleagues also measured children's multiple classification ability (the ability to categorize items along more than one dimension). They also asked participants how much they identified with their ingroup (e.g. "How do you feel about being a member of the diamond team?"), and asked them to predict how other members of their ingroup would evaluate moral and immoral ingroup and outgroup members. Interestingly, they found that better multiple classification ability was associated with lower intergroup bias and higher morality-based bias. They also found that the more strongly children identified with their ingroup, the more likely they were to show a positive correlation between their group-based behavioral predictions (e.g. predicting that other group members would evaluate ingroup members more favorably than outgroup members) and their own group-based evaluations. However, the correlation between *moral-based* behavioral predictions (i.e. predicting that other group

members would evaluate moral members over immoral members) and their own moral-based evaluations was unmediated by group identification. In other words, children who expected their group members to evaluate moral members more favorably than immoral members were more likely to do so themselves, regardless of how strongly they identified with the ingroup, whereas children who expected their group members to evaluate ingroup members more favorably than outgroup members were only more likely to show the same bias in their own evaluations if they strongly identified with their ingroup.

Abrams and colleagues suggest that these findings are consistent with social-cognitive domain theory (Killen, Lee-Kim, McGlothlin, & Stangor, 2002; Nucci & Turiel, 1978; Smetana, 1995; Turiel, 1983), which posits that group and morality-based judgment processes develop independently of each other, reflecting distinct social domains. Their findings hint that whereas group identity may be relevant for social-conventional reasoning in the group-based domain (e.g. making behavioral predictions, conforming to group norms and customs), it may not necessarily be relevant to the moral domain (Abrams et al., 2008). However, more research is needed to explore this hypothesis, particularly with younger children.

The Current Study

Although the study by Abrams et al. (2008) suggests that children use both group- and moral-based criteria (but rely more on moral-based criteria) to evaluate moral agents in the context of fairness and peer inclusion/exclusion, it is not known which type of criteria is given priority when evaluating morally complex acts involving double effects. In the current study, I addressed this question by investigating the extent to which

preschoolers apply the principle of double effect and the duty of rescue when the perceived in-group/out-group structure of the trolley problem is manipulated.

Children ages 3 to 5 years old were presented with modified versions of the dilemmas used in chapter 3, in which an agent must choose whether to prevent harm to five people, thereby causing harm to one person (either as a *side effect*, or as a *main effect*), or to do nothing (causing harm to five by *omission*). In the current study, the ingroup/outgroup structure of each dilemma was also manipulated, such that children either saw dilemmas in which the five people belonged to their ingroup (and the one person belonged to the outgroup), or the five people belonged to the outgroup (and the one person belonged to the child's ingroup).

Drawing upon previous work, I predicted that children would be sensitive to both group structure and causal/intentional structure (i.e. the PDE) in each dilemma, but that their moral judgments would be more sensitive to the latter. In other words, I predicted that children would primarily favor side effect dilemmas over main effect dilemmas, but would also (to a lesser extent) favor dilemmas in which their ingroup was saved and an outgroup member was harmed over dilemmas in which the outgroup was saved and an ingroup member was harmed (figure 4.1).

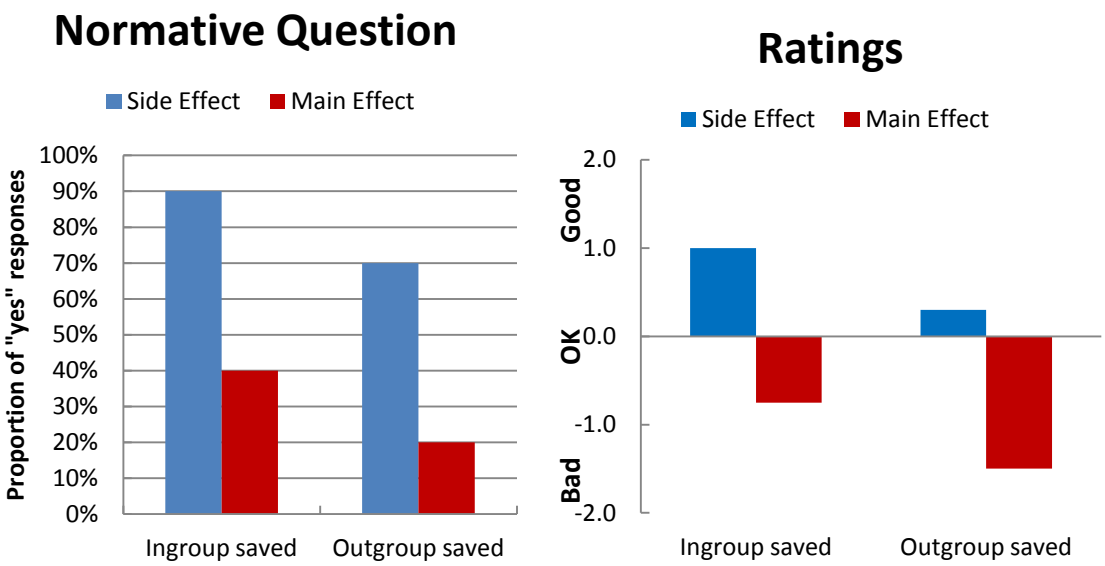

Figure 4.1. Predicted results. This figure shows the predicted pattern of responses to the normative question, "Should she have done that?" (left panel), and the ratings question (right panel).

Like Abrams et al. (2008), I chose to use minimal groups rather than socially recognized groups in order to remove the influence of factors other than group membership, such as knowledge of group norms, stereotypes, or histories that might swamp children's responses. I was also interested in whether the findings of the chapter 3 study extend beyond cases of property harm, so dilemmas in the current study also varied between subjects according to whether the threat involved a property violation (like the dilemmas in chapter 3) or assault (e.g. being frightened by an angry dog).

Method

Participants

Two hundred thirty-four preschoolers were seen, but 21 were eliminated from the study, 8 for failing to cooperate, and 13 for failing to pass the training phase. Of the remaining 213 participants, 103 (49 girls) were 3-year-olds between the ages of 36 and

48 months (M = 42.4, SD = 3.1), and 110 were 4- and 5-year-olds (43 girls) between the ages of 48 and 68 months (M = 55.0, SD = 5.1).

Design and Procedure

Children were tested individually in a quiet location at their preschool, or in the Rutgers Cognitive Development Lab. All testing sessions were videotaped for future scoring.

Training phase. Prior to testing, children were trained on the "Pink Scale" using a procedure similar to that in Chapter 3. Once participants were comfortable using all five points on the scale, they were given two training stories – one involving a harming action (i.e. hitting), and one involving a helping action (i.e. sharing). After each story, children were asked whether the moral agent in the story *should* have done what he/she did (*Normative question*), and then to rate his/her action using the Pink Scale (*Rating question*). Children who passed at least one of these measures (i.e. responded appropriately to both normative questions and/or both rating questions) continued to the hat choice phase.

Hat choice phase. In the hat choice phase, children were presented with two cone-shaped paper hats, one green, and one blue. They were told they would hear a story about "Blickets," who wore blue hats, and "Greebles," who wore green hats. Children were then told, *"You can wear a hat too! Which hat would you like? Do you want to be a Blicket or a Greeble?"* Once they made their selection, participants were given their hat to wear or hold while they heard the stories, and were reminded of their group affiliation: *"Here is your blue Blicket hat. Now you are a Blicket!*" Children were also told they could take the hat with them when the study was over.

Test phase. Based on their hat selection, participants were randomly assigned to either an *ingroup majority* condition (n = 104), or an *outgroup majority* condition (n = 109). Participants in the ingroup majority condition (figure 4.2, left panel) saw dilemmas in which an ingroup agent (a character wearing the same hat as the participant) chose whether to save five ingroup members at the cost of harming an outgroup member. Participants in the outgroup majority condition (figure 4.2, right panel) saw dilemmas in which an ingroup agent chose whether to save five outgroup members at the cost of harming an ingroup member. Thus, in each dilemma the moral agent always belonged to the child's ingroup, but the group majority she chose to save (or not save, in the case of omission) varied between subjects. At the beginning of each dilemma, participants were reminded of their group affiliation (e.g. "*This story is about Lisa, she is a Blicket just like you*."), and all characters were identified by their group ("Blickets" or "Greebles") throughout the story.

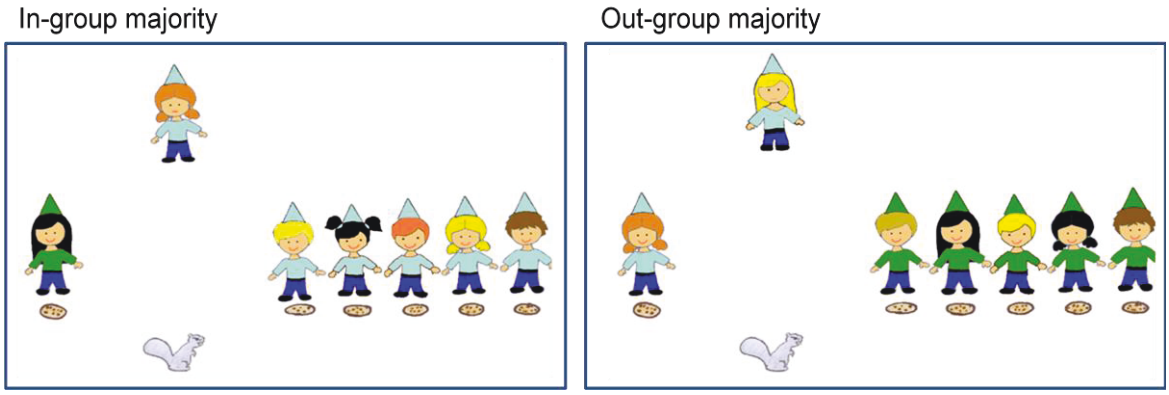

Figure 4.2. Two group conditions. The left panel represents the ingroup majority condition for a participant who picked a blue hat, and the right panel represents the outgroup majority condition for a participant who picked a blue hat.

Participants were also randomly assigned to either a *property harm* condition (n = 108), or an *assault* condition (n = 119). Participants in the *property harm* condition

(figure 4.3, left panel) saw computer-animated cartoons similar to those presented in the Chapter 3 study, in which the threat to the victim(s) involved a property violation (having their cookies eaten by a sneaky squirrel)[20]. Participants in the *assault* condition (figure 4.3, right panel) saw cartoons in which the victim(s) faced a credible threat of battery (i.e. being frightened by an angry dog who barks at people) (see Appendix D for a full description of each condition)[21].

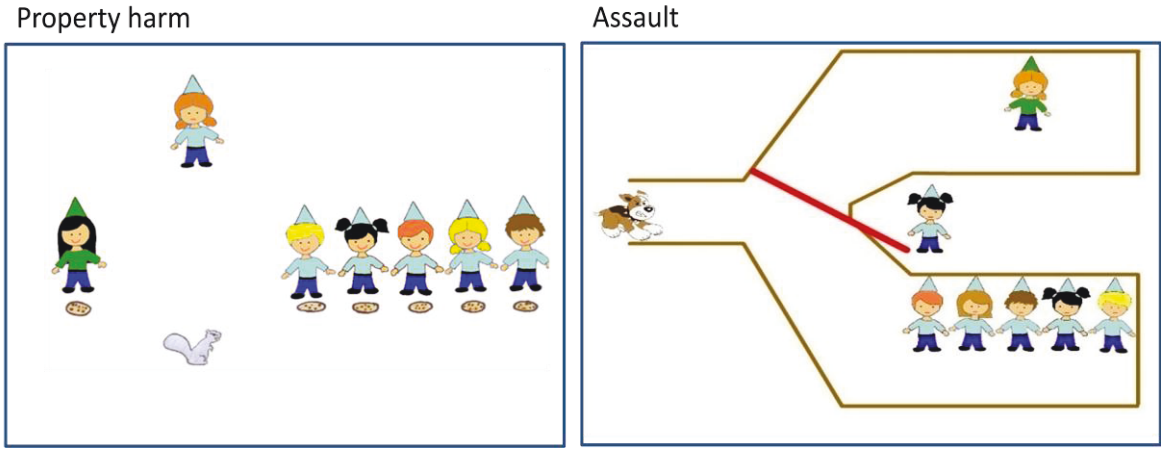

Figure 4.3. Property harm (left panel) and assault (right panel) conditions.

Similar to the procedure in Chapter 3, participants in each condition saw three computer-animated dilemmas: an *omission* dilemma, in which the agent chose not to intervene to prevent harm to five people; a *side effect* dilemma, in which one person was harmed as a foreseen side effect of saving five others; and a *main effect* dilemma, in which one person was harmed intentionally as a means of saving five others. In the current study, the omission dilemma was always presented as the first dilemma in the

[20] Events in the property harm cartoons were virtually identical to those in the Chapter 3 study, with the following exceptions: 1) each of the three cartoons involved a different set of characters and a different moral agent (Jane, Sally, or Lisa); 2) in each story, the agent and characters were identified by their group ("Blickets" or "Greebles"); 3) "munching" sound effects were added when the squirrel ate people's cookies; 4) the children who still had their cookies ate them at the end of the cartoon. See Appendix C for other changes to the script.

[21] Technically, the main effect dilemma in the assault condition also involved battery according to Mikhail's (2011) definition (see Appendix C, figure C9).

series (to illustrate what would occur if the agent failed to intervene, and to serve as an anchor for the other dilemmas), followed by either the side effect or main effect dilemma (counterbalanced between subjects). (See Appendix D for the full study design, as well as the number of participants assigned to each condition.)

For all three dilemmas, participants were asked control questions before the agent acted, to ensure that they understood what the squirrel/dog was about to do, and how the five children would feel if their cookies were eaten or the dog barked at them. Participants then watched each cartoon twice. During the second run-through, children were questioned to make sure they understood what the protagonist had done, what the squirrel/dog had done as a result of her action (or inaction), and how the victim(s) felt. If a child did not answer a control question correctly, he or she was corrected by the experimenter, and the relevant portion of the story was retold. The child's comprehension was then checked again, and the experimenter continued once the child responded appropriately. At the end of each dilemma participants were asked two test questions: 1) *Normative question*: Should [Jane/Sally/Lisa] have done that?; 2) *Rating question*: Can you show me on the Pink Scale? Was that good, bad, or just ok? A little good/bad or really good/bad? Finally, at the very end of testing, participants were asked a memory control question: *"Do you remember, are you a Blicket or a Greeble?"* Responses to all control and test questions were coded by at least two independent observers. If discrepancies occurred between the observations, a third coder was used.

Results

Because the number of people saved/harmed, as well as the identity of the person(s) saved/harmed differed between the omission dilemma and the other two dilemmas (and because the omission dilemma was designed with a different research question in mind), responses to the omission dilemmas were analyzed separately from the other two dilemmas. Results are therefore organized into two sections, beginning with the comparison between side effect and main effect dilemmas (in which the majority was always saved), and ending with the omission dilemma (in which the majority was always harmed). Within each section, results for each of the two measures (normative question, ratings) are presented in turn.

Data Analysis

All Ratings data were analyzed with ANOVA's. Responses to the normative question for side effect and main effect dilemmas were analyzed using various types of logistic regression, including the generalized estimating equations (GEE) procedure. GEE is an extension of the generalized linear model that allows for analysis of repeated measurements with binary (or discrete, or continuous) outcomes. Because GEE uses a quasi-likelihood estimation procedure for modeling correlated responses, GEE models require the specification of an appropriate working correlation matrix structure to account for the within-subject correlations. For the models presented in this chapter, there were no strong differences between Unstructured, Independent, and Exchangeable correlation structures. Therefore, all model estimates presented in this chapter (and in Appendix E) assume an unstructured correlation matrix (the most general structure).

Because GEE is not a likelihood-based method, the AIC (Akaike Information Criteria) statistic cannot be used to measure goodness of fit. Instead, the QIC (Quasi

likelihood under the Independence Model Criterion) and the QICC (Corrected Quasi-likelihood under Independence Model Criterion) are used to evaluate goodness of fit. Like the AIC, the lower these numbers are, the better the fit of the model.

Parameter estimates are presented in terms of log odds (β) and the odds ratio (Exp(β)). The reference categories for the dependent variable "normative question" was set to -1 ("No"). Thus, the odds ratio for each variable should be interpreted as the odds of saying "yes" for one level of the variable over the other, holding all other variables constant (i.e. at their reference level).

1. Did Preschoolers Show the Double-Effect Effect?

Normative question

A series of logistic regression analyses fitted with the generalized estimating equations method (GEE) were used to test the main and interactive effects of dilemma (2: side effect, main effect), majority group (2: ingroup, outgroup), harm (2: property, assault), age (2: 3-year-olds, 4- and 5-year-olds), and order (2: OSM, OMS) on responses to the normative question. Fit statistics for three GEE models are presented in table 4.1.

The full model included all potential main effects, and all potential two- and three-way interactions among the variables (see Appendix E for the full model results). The main effects model included only the potential main effect of each variable (see Appendix E for the main effects model results). The final reduced model included all main effects, as well as two significant interactions that contributed to the fit of the model. Results for this model are presented in table 4.2 (see Appendix E for the parameter estimates table). This model was selected as the best-fitting model using an approach similar to the forward stepwise procedure used in other regression programs:

Effects from the full model were entered into the reduced model one at a time. At each step in the process, effects that did not improve the fit of the model were removed. For example, although the dilemma*age*order interaction was significant in the full model, this interaction did not improve the fit of the reduced model and was therefore not included in the final reduced model.

Table 4.1
Goodness-of-fit statistics for three models

Model	**QIC**	**QICC**
Full model	573.493	570.628
Main effects model	556.880	554.354
Final reduced model	551.707	550.413
Information criteria are in small-is-better form.		

Table 4.2
Model effects for the reduced model (N = 213)
Dependent variable = Normative question

Model Effect	**Wald's χ2**	***df***	***P***
(Intercept)	.038	1	.846
Dilemma	38.456	1	.000
MajorityGroup	3.254	1	.071
Harm	2.209	1	.137
AgeGroup	4.922	1	.027
Order	5.084	1	.024
Dilemma * Harm	6.947	1	.008
Dilemma * Order	4.934	1	.026

Results of the final reduced model revealed that dilemma, age, and order were significant predictors of children's responses to the normative question, but that majority group was not a significant predictor. A significant interaction between dilemma and harm, as well as an interaction between dilemma and order were also found.

Like in Chapter 3, the effect of dilemma was consistent with the PDE. While 62% of all participants approved of the agent's action in side effect dilemmas (Binomial test, $N = 13$, $p < .001$, two tailed), only 38% of participants approved of the agent's action

in main effect dilemmas (Binomial test, $N = 213$, $p < .001$, two-tailed). The effect of age indicated that in general, younger children tended to say "yes" (56%) more often than older children (44%) (see figure 4.4).

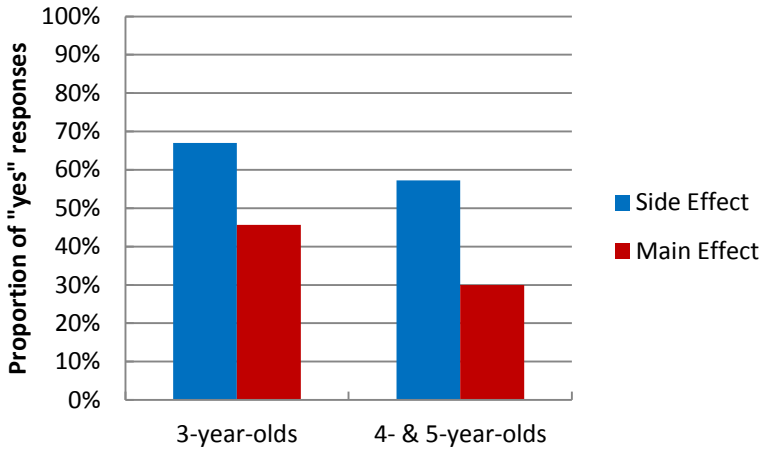

Figure 4.4. Children's normative judgments as a function of age and dilemma. This figure shows the percentages of children in each age group who answered "yes" to the normative question for all side effect and main effect dilemmas. As predicted, children in both age groups showed a pattern of responses that was consistent with the PDE.

The effect of order indicated that participants gave a higher proportion of "yes" responses overall when they saw the side effect dilemma before the main effect dilemma than when they saw the main effect dilemma before the side effect dilemma. However, a significant interaction between dilemma and order revealed that only side effect dilemmas (but not main effect dilemmas) were subject to this order effect (figure 4.5); Simple effects showed that participants were significantly more likely to approve of the side effect dilemma if they had seen it before the main effect dilemma (72%, Binomial test, $N = 107$, $p < .001$), but their proportion of "yes" responses in the side effect dilemma were at chance if they had seen it after the main effect dilemma (52%, Binomial test, $N = 106$, $p > .05$), $\chi^2[1, N = 213] = 9.210$, $p = .002$; however, participants were equally unlikely to approve of the main effect dilemma, regardless of the order in which the

dilemmas were presented (OSM: 39%, OMS: 36%, *p*'s < .05), $\chi^2[1, N = 213] = .219$, $p = .640$.

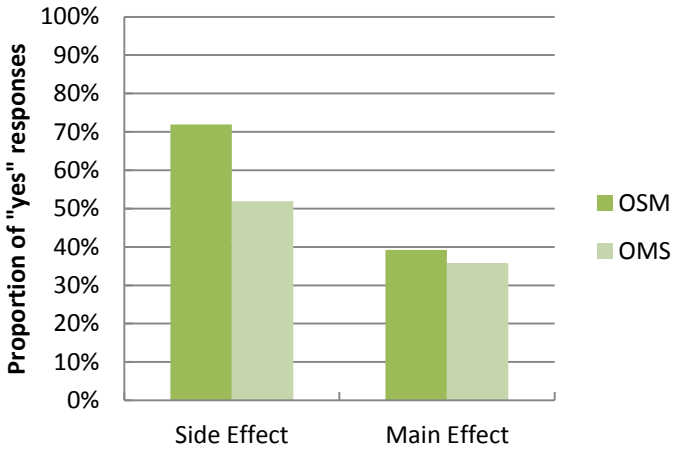

Figure 4.5. Children's normative judgments as a function of dilemma and order. This figure shows the percentages of children who answered "yes" to the normative question in each dilemma as a function of whether the side effect dilemma was presented as the second dilemma (OSM) or the last dilemma (OMS).

A significant interaction between dilemma and harm (figure 4.6) revealed that participants were significantly more likely to respond "yes" for side effect dilemmas involving a property violation (71%, Binomial test, $N = 108$, $p < .001$) than for side effect dilemmas involving assault (52%, Binomial test, $N = 105$, $p = .700$), Wald $\chi 2[1, N = 213] = 7.991$, $p = .005$; however, participants were equally unlikely to respond "yes" to both property and assault main effect dilemmas (37% and 38%, respectively, *p's* < .05), $\chi 2[1, N = 213] = .130$, $p = .719$.

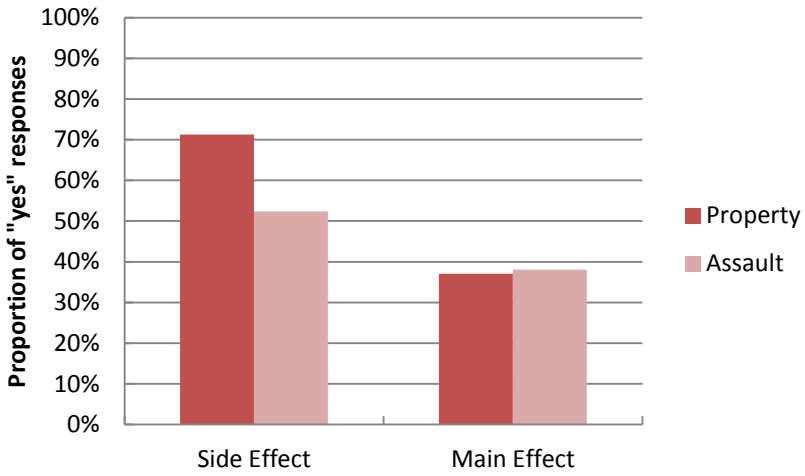

Figure 4.6. Children's normative judgments as a function of dilemma and harm. This figure shows the proportion of children who responded "yes" to the normative question as a function of dilemma (side effect, main effect) and harm (property, assault).

Figure 4.7 shows the percentages of children in each group majority condition who answered "yes" to the normative question in side effect dilemmas and main effect dilemmas. Although children's responses were in the predicted direction, with a higher percentage of children responding "yes" when the ingroup majority was saved (55%) than when the outgroup majority was saved (45%), this group majority effect was not significant. The interaction between dilemma and group majority was also not significant and was not included in the reduced model.

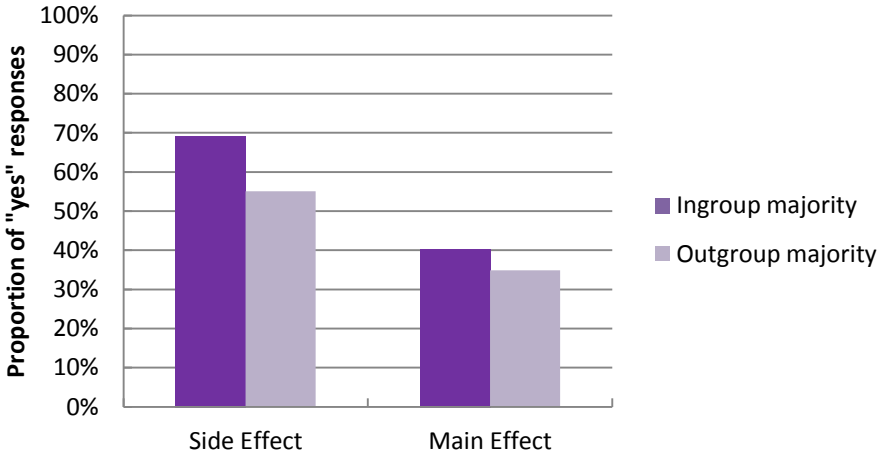

Figure 4.7. Children's normative judgments as a function of dilemma and majority group. This figure shows the percentages of children in each dilemma who responded "yes" to the question, "Should she have done that?" as a function of whether the child's ingroup or outgroup was saved.

Ratings

Children's ratings were scored as -2 for "really bad," -1 for "a little bad," 0 for "just ok," +1 for "a little good," and +2 for "a really good." Ratings for the side effect and main effect dilemmas were analyzed using a mixed ANOVA with dilemma (2: side-effect, main effect) as a within-subjects factor, and majority group (2: ingroup, outgroup), harm (2: ingroup, outgroup), age group (2: 3-year-olds, 4- and 5-year-olds), and order (2: side-effect first, main effect first) as between-subjects factors. Factorial results are presented in table 4.3.

Table 4.3
Factorial analysis of variance for children's ratings

Source	*df*	*F*	η_p^2	*p*
Between-Subjects				
Intercept	1	.244	.001	.622
Group	1	2.718	.014	.101
Harm	1	.060	.000	.807
AgeGroup	1	3.920	.020	.049*
Order	1	19.026	.088	.000***
Group * Harm	1	.925	.005	.337
Group * AgeGroup	1	1.284	.006	.259
Group * Order	1	1.985	.010	.160
Harm * AgeGroup	1	.082	.000	.775
Harm * Order	1	.001	.000	.978
AgeGroup * Order	1	.621	.003	.432
Group * Harm * AgeGroup	1	.260	.001	.611
Group * Harm * Order	1	.384	.002	.536
Group * AgeGroup * Order	1	.006	.000	.938
Harm * AgeGroup * Order	1	.136	.001	.713
Group * Harm * AgeGroup * Order	1	.167	.001	.683
Dilemma * Group * Harm * AgeGroup * Order	1	1.393	.007	.239
Error	197	(2.721)		
Within-Subjects				
Dilemma	1	49.186	.200	.000***
Dilemma * Group	1	.443	.002	.506
Dilemma * Harm	1	10.071	.049	.002**
Dilemma * AgeGroup	1	1.094	.006	.297

Dilemma * Order	1	6.383	.031	.012*
Dilemma * Group * Harm	1	5.722	.028	.018*
Dilemma * Group * AgeGroup	1	.000	.000	.991
Dilemma * Group * Order	1	.008	.000	.931
Dilemma * Harm * AgeGroup	1	.270	.001	.604
Dilemma * Harm * Order	1	.483	.002	.488
Dilemma * AgeGroup * Order	1	6.074	.030	.015*
Dilemma * Group * Harm * AgeGroup	1	.071	.000	.790
Dilemma * Group * Harm * Order	1	.776	.004	.380
Dilemma * Group * AgeGroup * Order	1	1.794	.009	.182
Dilemma * Harm * AgeGroup * Order	1	4.851	.024	.029*
Error	197	(1.506)		

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$

There were significant main effects of dilemma, order, and age, as well as significant interactions between dilemma and harm, dilemma and order, and dilemma, age, and order. Although the main effect of majority group was not significant, there was a three-way interaction between dilemma, majority group, and harm. A four-way interaction between dilemma, harm, age, and order was also found, but was uninterpretable (this effect accounted for only 2.4% of the variance).

As expected, a large effect of dilemma demonstrated that children gave significantly higher ratings for side effect dilemmas (M = 0.38, SE = 0.10) than for main effect dilemmas (M = -0.47, SE = 0.10). This effect was confirmed non-parametrically (Wilcoxon signed-ranks test, N =213, Z = 5.960, $p < .001$, r = .41). A small effect of age reflected a tendency for three-year-olds to give higher average ratings (M = 0.13, SE = 0.15) than older children (M = -0.21, SE = 0.15). The effect of order indicated that participants who saw side effect dilemmas prior to main effect dilemmas gave significantly higher ratings overall (M_{osm} = 0.31, SE = 0.14) than participants who saw main effect dilemmas prior to side effect dilemmas (M_{oms} = -0.40, SE = 0.15). However,

a significant interaction between dilemma and order indicated that this order effect was produced by participants' ratings in the side effect dilemma; participants rated side effect dilemmas significantly higher if they saw them second than if they saw them last, $F(1, 197) = 26.158$, $p < .001$, $\eta_p^2 = .117$, whereas participants' ratings for the main effect dilemma when it was presented second did not differ significantly from participants' ratings when it was presented last, $F(1, 197) = 2.769$, $p = .052$, $\eta_p^2 = .019$.

Inspection of the data (figure 4.8) suggested that the three-way interaction between dilemma, age, and order was significant because older children who saw the side-effect dilemma last (after the main effect dilemma) did not rate it significantly higher than the main effect dilemma. This was confirmed by running two separate two-way ANOVAs, one for each age group. While the interaction between dilemma and order was significant for four- and five-year-olds, $F(1, 108) = 12.556$, $p = .001$, $\eta_p^2 = .104$, it was not significant for three-year-olds, $F < .001$. Simple effects revealed that older children who saw the side effect dilemma second (after the omission dilemma), rated the side effect dilemma significantly higher than the main effect dilemma, $F(1, 197) = 44.010$, $p < .001$, $\eta_p^2 =$), whereas older children who saw the side effect dilemma last (after the main effect dilemma) showed no difference in their ratings for side effect and main effect dilemmas, $F(1, 197) = 2.443$, $p =$, $\eta_p^2 =$).

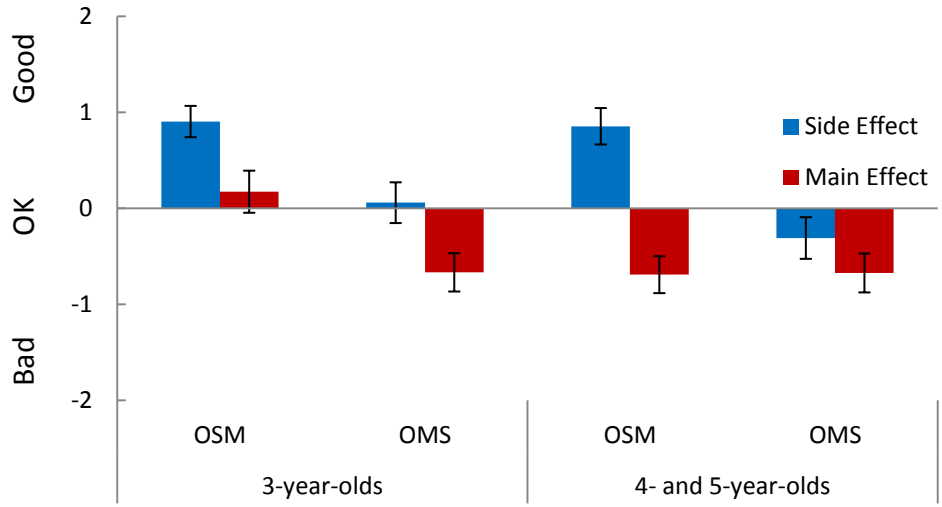

Figure 4.8. Children's ratings as a function of dilemma, age, and order. This figure shows children's average ratings in each dilemma as a function of age group and order of presentation (OSM = omission, followed by side effect, followed by main effect; OMS = omission, followed by main effect, followed by side effect). Error bars show standard error of the mean.

The interaction between dilemma and harm (figure 4.9) revealed that for side effect dilemmas, participants in the property harm condition gave significantly higher ratings (M = 0.58, SE = 0.15) than participants in the assault condition (M = 0.16, SE = 0.14), $F(1, 197) = 4.558$, $p = .034$, $\eta_p^2 = .023$, but for main effect dilemmas, participants' ratings in the property harm condition (M = -0.62, SE = 0.15) did not differ significantly from ratings in the assault condition (M = -0.31, SE = 0.14), $F(1, 197) = 2.769$, $p = .098$, $\eta_p^2 = .014$.

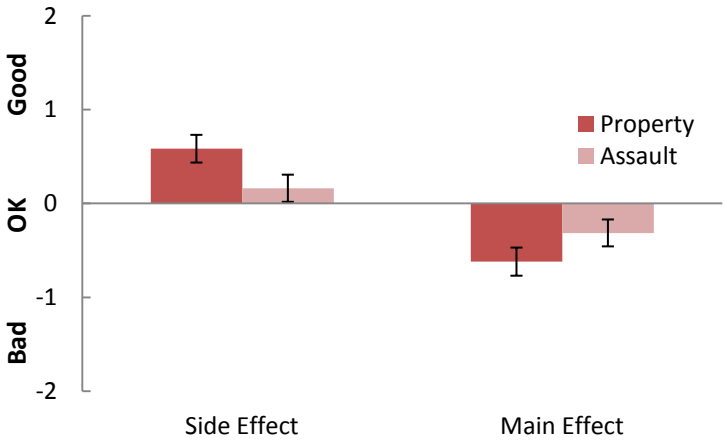

Figure 4.9. Children's ratings as a function of dilemma and harm. Error bars show standard error of the mean.

As illustrated in figure 4.10, the three-way interaction between dilemma, group, and harm revealed that there was a significant interaction between dilemma and majority group in the assault condition, $F(1, 103) = 4.773$, $p = .031$, $\eta_p^2 = .044$, but not in the property harm condition, $F(1, 106) = 1.381$, $p = .243$, $\eta_p^2 = .013$. In the assault condition, side effect dilemmas were rated significantly higher than main effect dilemmas in the majority outgroup condition (when the outgroup was saved), $F(1, 197) = 11.884$, $p = .001$, $\eta_p^2 = .057$, but not in the majority ingroup condition (when the ingroup was saved), $F(1, 197) = .147$, $p = .702$, $\eta_p^2 = .001$. There was also a significant interaction between harm and majority group for the side effect dilemma, $F(1, 209) = 4.510$, $p = .035$, $\eta_p^2 = .021$, but not for the main effect dilemma, $F(1, 209) = .454$, $p = .501$, $\eta_p^2 = .002$. For side effect dilemmas, property harm conditions in which the ingroup was saved were rated significantly higher than property harm conditions in which the outgroup was saved, $F(1, 197) = 5.197$, $p = .024$, $\eta_p^2 = .026$, but no effect of majority group was found in assault conditions (*p*'s > .05).

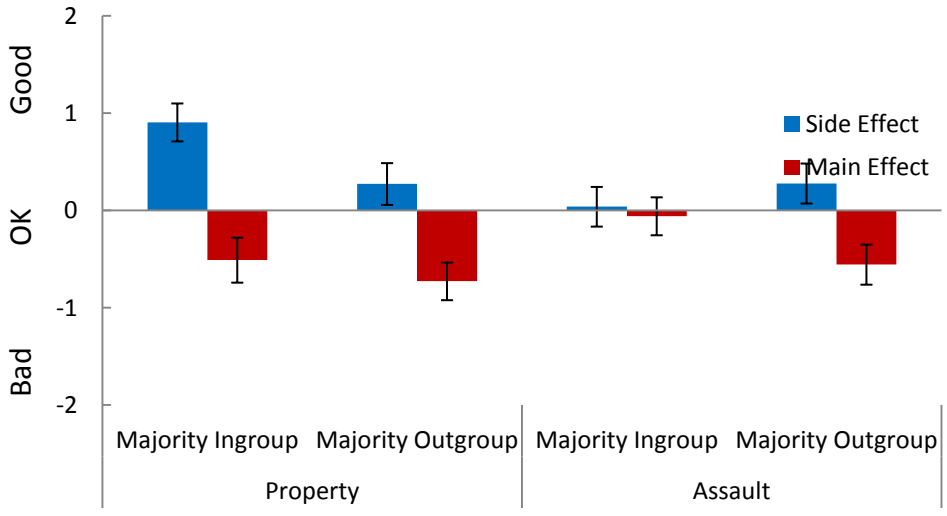

Figure 4.10. Children's ratings as a function of dilemma, group, and harm. This figure shows children's average ratings for each dilemma as a function of harm type (property, assault) and whether the child's ingroup or outgroup was saved. Error bars show standard error of the mean.

2. How did participants judge omission?

Normative question

Overall, a significant majority of preschoolers (68%) judged the decision to "just stand there" in the omission condition as impermissible (Binomial test, $N = 213$, $p < .001$, two tailed). A binary logistic regression analysis was used to test the main and interactive effects of majority group (2: ingroup, outgroup), harm (2: ingroup, outgroup), and age group (2: 3-year-olds, 4- and 5-year-olds) on normative responses in the omission dilemma. Using a forward stepwise procedure, the best-fitting model included a main effect of age, Wald $\chi^2(1, N = 213) = 7.807$, $p = .005$, $\beta = .840$, odds ratio (OR) = 2.315, but no main or interactive effects of majority group or harm were included in the model (ps > .05). Inspection of the data (figure 4.11) indicated that, like in chapter 3, the effect of age reflected a significant tendency to judge omission as impermissible only in the older age group: whereas 76% of 4- and 5-year-olds responded "no" to the normative question (Binomial test, N = 110, $p < .001$), only 58% of 3-year-olds did so (Binomial test, $N = 103$, $p = .115$). This model had an overall correct prediction rate of 67.6%. A test of this model against a constant-only model was statistically significant, $p = .005$.

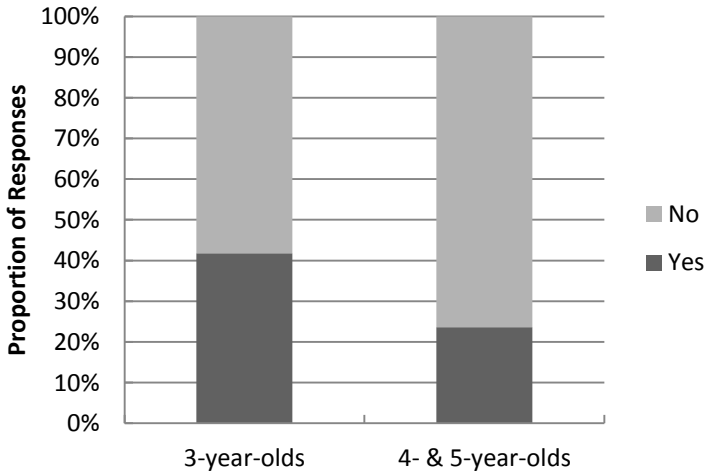

Figure 4.11. Omission: Children's normative judgments by age. This figure shows the distribution of children's responses to the normative question in the omission dilemma by age group.

Ratings question

A 2(age: 3-year-olds, 4- and 5-year-olds) x 2(majority group: ingroup, outgroup) x 2(harm: property, assault) ANOVA was computed to analyze ratings for the omission dilemma. A significant effect of majority group revealed that dilemmas in which the participant's ingroup was harmed (i.e. ingroup majority conditions) were rated significantly lower (M = -0.38, SE = 0.12) than dilemmas in which the outgroup was harmed (i.e. outgroup majority conditions) (M = 0.10, SE = 0.13), $F(1, 205) = 8.869$, p = .003, η_p^2 = .041. This effect was qualified by a significant three-way interaction between group, age, and harm (figure 4.12), $F(1, 205) = 6.316$, p = .013, η_p^2 = .030. Younger children gave significantly lower ratings when their in-group was harmed (vs. when the outgroup was harmed) only in the assault condition $F(1, 205) = 9.655$, p = .002 (but not in the property harm condition, F < .001), whereas older children gave significantly lower ratings when their in-group was harmed (vs. when the outgroup was harmed) only in the property harm condition, $F(1, 205) = 4.749$, p = .03 (but not the assault condition, p > .05). No main effects of age or harm were found (p's > .05).

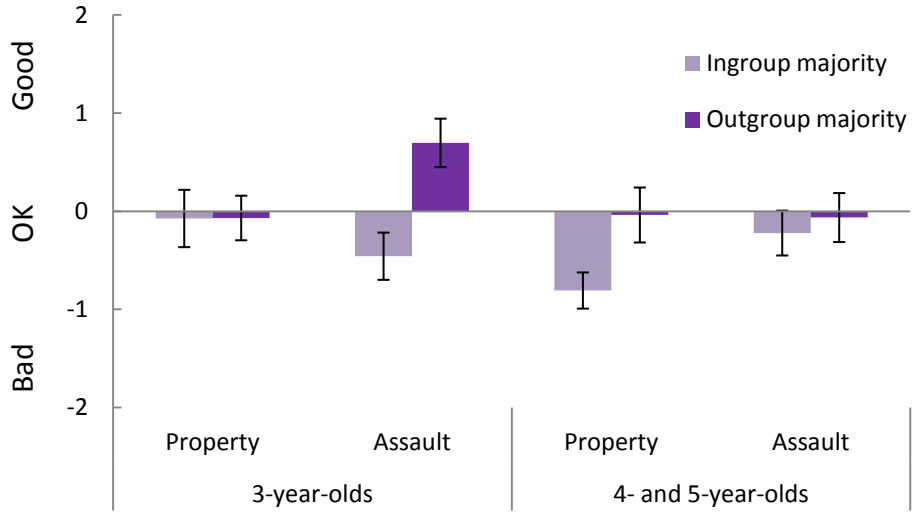

Figure 4.12. Omission: Children's ratings as a function of group, age, and harm. This figure shows children's average ratings in the omission condition as a function of age group, harm condition (property, assault), and whether the child's ingroup or outgroup was harmed. Error bars show standard error of the mean.

Discussion

Children's moral judgments in the current study were very similar to those in chapter 3, and were once again consistent with the PDE. Even in a minimal group context, preschoolers judged scenarios in which an individual was harmed as a foreseen side effect of saving five others as permissible or "good", but judged scenarios in which an individual was intentionally harmed as a means of saving five others as impermissible or "bad". Furthermore, this effect accounted for the largest proportion of the variance relative to all other effects. Thus, despite the possible temptation or bias to only disapprove of actions that cause harm to one's ingroup, preschoolers appeared to universally apply the PDE.

Furthermore, children tended to show this pattern of judgments regardless of whether the threat to the five people involved a property violation or assault. However, this pattern was weaker in the assault condition, as evidenced by lower ratings and a lower proportion of positive normative judgments in the side effect/assault dilemma (relative to the side effect/property harm dilemma). These results suggest that children may have been more ambivalent about the agent's action in the side effect/assault dilemma due to the threat of battery in this dilemma. It is also possible that children did not consider the good effects to clearly outweigh the bad effects in the side effect/assault dilemma due to a safety-in-numbers mentality: whereas the one person was all by herself

when she was frightened by the angry dog, the five people had each other (and therefore might have been less frightened by the dog).

In the omission dilemma, children's judgments were also consistent with the results from Chapter 3: on the normative measure, four- and five-year-olds responded that the agent should *not* have "just stood there," but three-year-olds responded at chance on this measure. Given three-year-olds' ambivalence about this dilemma in the previous chapter (and the possible reasons for their difficulty with omission, as discussed in Chapter 3), it is not surprising that they performed at chance in the current study as well. It is also important to note that in the Chapter 3 study, the omission dilemma was always presented second, whereas in the current study, the omission dilemma was always presented first. Thus, in the current study, children never saw a specific alternative action the agent could have taken in the omission dilemma[22]. Nevertheless, the fact that children as young as four years old continued to disapprove of the choice to "just stand there" in the current study suggests that children in this age group were indeed comparing the agents' inaction to some permissible means of intervention (presumably one in which some or all of the people were saved). Furthermore, their disapproval of inaction in this dilemma, even when only outgroup members were harmed, suggests that for children as young as four-years-old, the duty to rescue applies not just to their ingroup, but also to outgroup members. However, future research is needed to resolve whether rescue is indeed perceived as a *duty* (i.e. a moral obligation) when no unjustified costs are

[22] The researcher did state that the agent could do "something" (but decides "not to do anything"). However, it was never specified what "something" the agent could have done, and children never saw an alternative action/outcome sequence. Thus, this dilemma may have been even more cognitively demanding than in the previous study.

involved, or merely a preferred course of action (particularly in the case of omissions that are not life-threatening).

Like in Chapter 3, the current study showed evidence of order effects on children's moral judgments. On the normative measure, children were more likely to approve of the side effect dilemma if they had seen it second (in the OSM order) than if they had seen it last (in the OMS order). On the ratings measure, older children also showed this effect of order in the side effect dilemma, and three-year-olds tended to judge the both the side effect and main effect dilemmas more favorably in the OSM order. These findings are consistent with a pattern of order effects that has been observed in adults' moral judgments of trolley problems: ratings of permissible acts (based on first trial judgments) tend to be lower when preceded by impermissible acts that are similar in structure (i.e. produce the same effects) (Wiegmann, Okan, & Nagel, 2012).[23]

Despite the availability of social category information in the current study, I found only weak evidence of ingroup bias in children's moral judgments. In the two double-effect dilemmas (side effect and main effect dilemmas), no significant effects of group were found on the normative measure, and only a small effect of group was found on children's ratings in the side effect property harm dilemma. In the omission dilemma, some evidence of ingroup bias was found in children's ratings, but no effects of group

[23] For example, studies by Lombrozo (2009), Petrinovich and O'Neill (1996), Schwitzgebel and Cushman (2012), and Wiegmann et al. (2012) all found that adults who saw the bystander problem first gave it higher permissibility ratings than those who saw it after the footbridge problem. Both Lombrozo (2009) and Schwitzgebel and Cushman (2012) have suggested that this pattern of order effects may reflect a general desire to maintain consistency between judgments of acts perceived as similar. This is also a plausible explanation for the order effects found in the current study. Like the bystander and footbridge problems, the side effect and main effect dilemmas were perceptually similar, as they both involved acts (rather than omissions) that produced identical outcomes. Thus, anchoring effects, as well as the desire for consistency between these two dilemmas may have contributed to the order effects found in the current study.

were found on children's normative judgments. Why did I find so little evidence of group effects on children's moral judgments, particularly in double-effect scenarios?

One possibility is that the minimal groups used in the current study were not sufficient to induce ingroup favoritism in children. However, this explanation seems unlikely. Previous research has shown that the mere presence of a minimal ingroup/outgroup structure is sufficient for inducing moderate to large effects of ingroup favoritism in five-year-olds on a variety of non-moral measures, even in the absence of group labels (Dunham et al., 2011). In the current study, not only were the groups marked visually and with a verbal label (characters were identified as "Blickets and Greebles" throughout the study), participants were also reminded of their group affiliation at the beginning of each dilemma. Indeed, one could argue that the group information was more salient in these dilemmas relative to the intentional structure of the action, which had to be inferred. Furthermore, the fact that children's ratings were more susceptible to group influences in the omission dilemma - possibly because the agent did not actively cause harm in this dilemma - suggests that children were aware of their own group membership, but this information had little bearing on their moral judgments of double-effect scenarios.

It is also worth noting that the significant effects of group appeared only in children's ratings, but no effects of group were found in children's normative judgments. Furthermore, with the exception of three-year-olds' ratings in the assault omission condition (and given three-year-olds' ambivalence about the omission dilemma in general, this exception was not surprising), the *valence* of children's ratings did not appear to vary with ingroup/outgroup structure. This leads us to the tentative conclusion

that children selectively attended to morally-relevant criteria (i.e. the causal and intentional structure of each scenario) over non-moral group-based criteria, particularly when making categorical moral judgments. It also suggest that a dichotomous forced choice measure (e.g. yes/no, permissible/impermissible) may be a better measure for capturing unbiased deontic judgments (i.e. "considered judgments"), whereas an evaluative measure that produces a scale or rank-order response may be more susceptible to performance errors.

This interpretation of the current results is consistent with the moral grammar hypothesis, which assumes that group-based concerns are outside of the moral system, but can exert a distorting influence on moral competence. It is also consistent with Jonathan Haidt's moral foundations theory insofar as it assumes that harm/fairness concerns and group loyalty concerns reflect two distinct domains: one concerned with the universal application of principles of justice and fairness, and one concerned with intergroup relationships, group identity, and conformity to group norms and goals (i.e. group loyalty) (Haidt & Graham, 2007; See also social domain theory (Killen et al., 2002; Nucci & Turiel, 1978; Smetana, 1995; Turiel, 1983) and developmental subjective group dynamics theory (Abrams, Rutland, & Cameron, 2003; Abrams et al., 2008)). However, in contrast to Haidt's theory, the current results suggest that only the latter domain is part of the moral competence. However, this interpretation of the current results is speculative, and calls for further inquiry. It remains to be seen whether future research will continue to uphold this impartial theory of moral competence.

V. Investigating the the role of group structure in adults' moral judgments

Introduction

In the previous chapters, I showed that children as young as three years old exhibit a pattern of judgments consistent with the principle of double effect (PDE) and the Rescue Principle, even when evaluating double-effect dilemmas that involve no intentional battery or physical contact between agent and patient. These results tie in nicely with a growing number of studies showing that adults from a wide range of social and cultural demographics also show a pattern of judgments prescribed by the PDE, even though they are unaware of the principle guiding their judgments (e.g. Cushman et al., 2006; Hauser et al., 2007a; Mikhail, 2002). In Chapter 4, I showed that for preschoolers, this abstract pattern of reasoning holds even in a minimal group context, across different kinds of acts and moral transgressions. In the current study, I investigate whether adults' intuitions about the minimal group trolley problems used in Chapter 4 are similar to those of children. Although several modified versions of the trolley problem have been tested in adult populations, ours are the first to use moral violations other than battery, the first to use minimal groups, and the first to manipulate the identity of both the moral agent and the moral patients in the trolley problem.

Ingroup bias and moral judgment in adults

Perhaps not surprisingly, adults show experimental ingroup bias, even for minimal ingroups, across a variety of non-moral measures including resource allocation tasks, behavioral prediction tasks (Locksley Ortiz, & Hepburn, 1980; Tajfel & Turner, 2004), implicit attitude tasks (Ashburn-Nardo, Voils, & Monteith, 2001; Locksley et al.,

1980; Otten & Wentura, 1999), and trait attribution tasks (Locksley et al., 1980) (see Mullen et al., 1992 for a meta-analytic review of the minimal ingroup effect). However, the effect of perceived ingroup/outgroup structure on adult moral judgment is still an open question. In the following sections, I describe two lines of research that have begun to shed some light on this issue.

Agent-centered bias

One line of research has focused on in-group bias with respect to the moral *agent*; that is, whether moral judgments are sensitive to the identity of the moral agent. A series of studies by Valdesolo and DeSteno suggests that under certain conditions, individuals may discount the moral severity of their own or another ingroup agent's moral transgressions (Valdesolo & DeSteno, 2007), but that this effect is contingent on the degree of cognitive task load (Valdesolo & DeSteno, 2008). In Valdeso and DeSteno's (2007) study, participants were told that they would be assigned to one of two conditions – a shorter/easier condition, or a longer/harder condition – and that another (anonymous) participant would be assigned to the other condition. Participants were given the choice of using a random generator to determine who would be assigned to the better condition (fair choice) or to select the better condition for themselves (unfair choice). They were then asked to evaluate the fairness of their action. In a follow-up study, participants were asked to judge the fairness of a neutral party, a (minimal) ingroup member, or an outgroup member who, when faced with the same choice, selected the better condition for themselves. The researchers found that participants judged their own action, as well as an ingroup members' action to select the better condition for themselves as less offensive than the same action made by an outgroup member or a neutral party.

Critically, however, a subsequent study by Valdesolo and DeSteno (2008) showed that when participants were under conditions of cognitive load, this effect disappeared; participants judged their own fairness transgression to be equally as unfair as the same transgression committed by another (neutral) party.

According to Valdesolo and DeSteno, these findings suggest that the tendency to judge the actions of ingroup agents less harshly than the actions of outgroup agents is due to higher-order reasoning processes geared toward justification and rationalization:

> In this case, the intuitive system would favor a more ‘‘moral” judgment in accord with a basic fairness norm (i.e., showing self-interest is not appropriate), but conscious control systems might work to generate a more ‘‘immoral” judgment (i.e., showing self-interest is permissible) that nevertheless may serve to protect one’s self-image” (Valdesolo & DeSteno, 2008, p. 1335).

Consequently, when participants are under increased cognitive constraints, they give less biased (i.e. more fair/rational) judgments because their higher-order reasoning processes are no longer able to override the intuitive moral judgment that violating fairness norms is impermissible.

Although Valdesolo and DeSteno frame this interpretation of their results as consistent with a dual-process theory of moral judgment, it is also consistent with the moral grammar hypothesis proposed by John Mikhail (2011). As discussed in chapter 4, when accounting for ingroup bias in moral judgment, proponents of this theory distinguish between “considered judgments” (a la John Rawls, 1971) and prejudice. Under this view, the moral judgments that participants gave under conditions of cognitive constraint would be closer to Rawls’s definition of a considered judgment – “judgments in which our moral capacities are most likely to be displayed without distortion” (Rawls,

1971, p. 47) - whereas judgments that had been distorted by motivated reasoning processes in the control condition would be viewed as prejudices.

Patient-centered bias

Another line of research has investigated whether the identity of the moral *patient* affects moral judgment. In particular, several studies have shown that under certain conditions, adults may shift or even reverse their responses on trolley problems when the identities of the parties being saved and sacrificed are manipulated (i.e. when they are no longer anonymous moral patients). For example, one of the first studies to test lay people's intuitions of the classic trolley problem found that participants were more inclined to save humans over animals, kin over non-kin, friends over strangers, and politically neutral individuals over politically abhorrent individuals (i.e. Nazis) (Petrinovich, O'Neill, & Jorgensen, 1993). A more recent study also found that participants were unwilling to flip the switch in the Bystander problem if the individual on the side track was very young, a family member, or a significant other (Bleske-Rechek, Nelson, Baker, Remiker, & Brandt, 2010). These results have often been framed in the context of Hamilton's (1964) Inclusive Fitness Theory, which posits that there are evolutionary advantages to favoring those who are genetically related to us, have their reproductive lives ahead of them, or are likely to provide reproductive opportunity.

However, inclusive fitness is not always necessary to induce such bias. In some cases, the stereotypes associated with a certain social category are sufficient to alter adults' pattern of moral intuitions. For example, a recent study by Cikara, Farnsworth, Harris,& Fiske (2010) found that 84% of participants judged the act of shoving the man in the footbridge problem as acceptable when the man on the footbridge was identified as

a stereotypically low-warmth, low-competence person (e.g. a homeless man), and the five people on the track were identified as stereotypically high-warmth, high-competence individuals (e.g. Americans). This is in stark contrast to the traditional footbridge problem, where 88% of participants judged the act of shoving an innocent/anonymous bystander as *im*permissible (Hauser et al., 2007a). Cikara et al. interpret these findings as evidence for an ingroup bias in moral reasoning; people perceive high-warmth, high-competence individuals as part of their ingroup, whereas they perceive low-warmth, low-competence individuals as part of the outgroup.

Some research suggests that adults' sensitivity to the perceived ingroup/outgroup structure of the trolley problem may also be heightened when participants strongly identify with the ingroup, or when the ingroup ideology supports one outcome over another. For example, Swann, Gomez, Dovidio, Hart, and Jetten (2010) found that Spaniards whose personal identities were fused with their national identity were more willing to sacrifice themselves to save five fellow Spaniards or Europeans, but not five Americans. A study by Uhlmann, Pizarro, Tannenbaum, and Ditto (2009) also found that adults tended to give responses on the trolley problem that were consistent with their political affiliation; Americans who identified as politically conservative (but not those who identified as liberal) were more likely to accept the foreseen but unintended deaths of Iraqi civilians over the deaths of American civilians, and liberals (but not conservatives) were more likely to endorse harm as a main effect in the Footbridge problem when the victim had a stereotypically white name (Chip Ellsworth III) than when the victim had a stereotypically black name (Tyrone Payton), even though both

liberals and conservatives explicitly rejected race as a valid basis for judgment when asked directly.

Limitations of the current evidence

As discussed in chapter 4, some have interpreted this patient-centered bias as evidence of a group-sensitive morality, in which the welfare of the ingroup (either consciously or unconsciously) takes moral precedence over that of the outgroup (Haidt & Graham, 2007). According to this view, we are intrinsically less sensitive to transgressions committed against outgroup members, or transgressions committed by ingroup members. If this is the case, individuals' compliance with moral principles such as the PDE should vary depending on who is being saved or harmed, and who is doing the saving or harming. However, there are several limitations to the studies described above that make it difficult to determine whether this is in fact the case.

First, none of the studies described above directly compared responses in the bystander problem to responses in the footbridge problem, so the extent to which the double-effect effect is mediated by perceived ingroup/outgroup structure in adults remains unclear. Second, it is debatable whether the measures used in these studies capture *moral* judgment in particular, as opposed to other kinds of social judgment. For example, the studies by Petrinovich et al. (1993), Blesk-Rechek et al. (2010), and Swann, Gomez, Dovidio, Hart, & Jetten (2010) asked participants to take the perspective of the protagonist (e.g. *"Would you flip the switch in this situation?"*), rather than evaluate the actions of another party (e.g. *"Is it morally permissible for X to throw the switch?"*) or the act itself (e.g. *"Is it morally permissible to throw the switch?"*). This is problematic, particularly considering that participants often give very different responses when asked

to make a behavioral prediction (e.g. what *they* would do in a given situation) than when asked to judge whether an action is morally permissible (Borg et al., 2006; Royzman & Baron, 2002). Of the studies that did ask for some sort of judgment, their measures were often varied and ambiguously phrased (e.g. *"Is…X…acceptable or unacceptable?" "Is …X…justified or unjustified?" "How much do you agree or disagree with…X?"*). Thus, it is unclear how participants interpreted many of these questions, or whether any of these judgments reflect the processes that underlie moral judgment specifically.

The trolley studies described above are also limited in scope because they used already-established social groups that differed along dimensions other than mere membership in a particular group. Consequently, it is possible that these findings do not generalize to other groups or to judgments of intergroup harm in general, but merely reflect learned behavior towards specific social groups based on pre-existing social norms/stereotypes, a history of conflict or competition between groups, or other prior statistical patterns of association. If our moral sense is indeed designed to take group structure into account, it is likely that such a system requires little prior experience with the social categories in question in order to generate biased moral judgments. That is, perceived ingroup/outgroup structure alone may be sufficient to induce group-sensitive moral judgments. Indeed, as mentioned previously, minimal groups are sufficient to induce ingroup bias in adults on a variety of non-moral tasks (see Mullen et al., 1992 for a meta-analytic review of the minimal ingroup effect in adults). Therefore, further research is necessary to determine whether mere membership in a particular social group affects moral judgments when other correlating factors (such as knowledge of group norms, stereotypes, or histories) are no longer present. Furthermore, the use of minimal

groups would allow researchers to rule out the possibility that the group effects found in previous real-group studies merely reflect a swamping of strictly moral judgment by highly salient pre-potent extra-moral values.

Finally, while all of these trolley studies manipulated the identity of the moral *patient* (e.g. the person on the side track), none of them manipulated the identity of the moral *agent* (e.g. the person deciding whether to flip the switch). Further research is needed to determine whether and how these two potential sources of ingroup bias interact to affect moral judgment in the trolley problem.

The current study

I addressed these issues in the current study using the same moral dilemmas I presented to preschoolers in Chapters 3 and 4. To date, all trolley variations previously tested on adult populations have used "life-and-death" scenarios involving homicide and battery. It was therefore important to confirm that adults' moral intuitions (like those of preschoolers) accord with the PDE and the duty of rescue even when evaluating more mundane dilemmas involving moral hazards such as having one's cookies eaten and being frightened by an angry dog. Like in Chapter 4, the ingroup/outgroup structure of each dilemma was also manipulated, such that participants either saw dilemmas in which the five people belonged to their ingroup (and the one person belonged to the outgroup), or the five people belonged to the outgroup (and the one person belonged to the participant's ingroup). I also included an additional parameter: whether the moral *agent* belonged to the participant's ingroup or outgroup. Accordingly, participants saw conditions in which both the identity of the moral agent and the moral patient(s) were systematically manipulated.

Finally, I asked participants to answer an additional question at the end: whether the harm in each dilemma was brought about on purpose[24]. A critical assumption of the moral grammar hypothesis is that when people encounter trolley problems, they unconsciously compute the intentional structure of the agent's action plan; the bystander problem is judged permissible because participants infer that the bystander intends only the good effects of his action, but does not intend the bad effect (i.e. the side effect), whereas the footbridge problem is judged impermissible because participants infer that the bystander intends the bad effect as a means to achieving the good effects. Although the agent's intentions are never explicitly stated in the trolley problem, this assumption that we represent the intentional structure of the trolley problem in this manner (as opposed to an alternative interpretation, in which the agent's end/goal is to harm the one, and saving the five is merely a foreseen side effect) underlies almost all adult trolley studies to date. However, only a few studies have explicitly asked participants about the bystander's intention of the agent in each of these scenarios (Levine, Leslie, & Mikhail, 2013, January). In the current study, I predicted that if participants were indeed basing their deontic judgments on the intentional structure of these dilemmas, they would be more likely to infer that the agent intended the bad effect in the main effect dilemma (as a means), but did not intend the bad effect in side effect dilemma.

Method

Participants

[24] Only a portion of participants were asked this question, as it was only added to the protocol half-way through testing.

Two hundred forty-one students were recruited through the subject pool at Rutgers University and participated for course credit. The group consisted of 138 women, 99 men, and 4 participants whose gender was not recorded.

Design and Procedure

The procedure closely followed that in Chapter 4, with the following changes. Prior to testing, participants were told they would hear stories that were initially designed for children, and that their responses to the stories would be compared to those of preschoolers. No pre-screening was used to determine participants' fluency with the Pink Scale, and no control questions were used to check for comprehension during testing. Like in the Chapter 4 study, participants were asked to select a hat color, and were then randomly assigned to see dilemmas involving either a *property* violation or *assault*. Critically, participants were also randomly assigned to one of four group conditions: *ingroup agent/ingroup majority*, *ingroup agent/outgroup majority*, *outgroup agent/ingroup majority*, *outgroup agent/ingroup majority (*See Appendix F for a complete list of conditions and the number of participants in each condition). Participants in the two *ingroup agent* conditions (figure 5.1, top two panels) saw dilemmas identical to those in Chapter 4, in which a member of the participant's ingroup (a character wearing the same hat as the participant) chose whether to save five people who belonged either to the participant's ingroup (top left panel) or outgroup (top right panel). Participants in the *outgroup agent* conditions (figure 5.1, bottom two panels) saw dilemmas in which an outgroup member (a character wearing a different color hat from the participants') chose whether to save five people who belonged either to the participant's ingroup (bottom left panel) or outgroup (bottom right panel).

At the end of each dilemma participants were asked the same test questions as in the previous chapters (*Normative question, Ratings question)*. A subset of adults were also asked an additional *Purpose question*: Did Jane make the Blicket(s)/Greeble(s) sad on purpose?. This question was added to the protocol half-way through testing as a way to confirm that participants do in fact view the harm in the main effect dilemma as more intentional than the harm in the side effect dilemma[25].

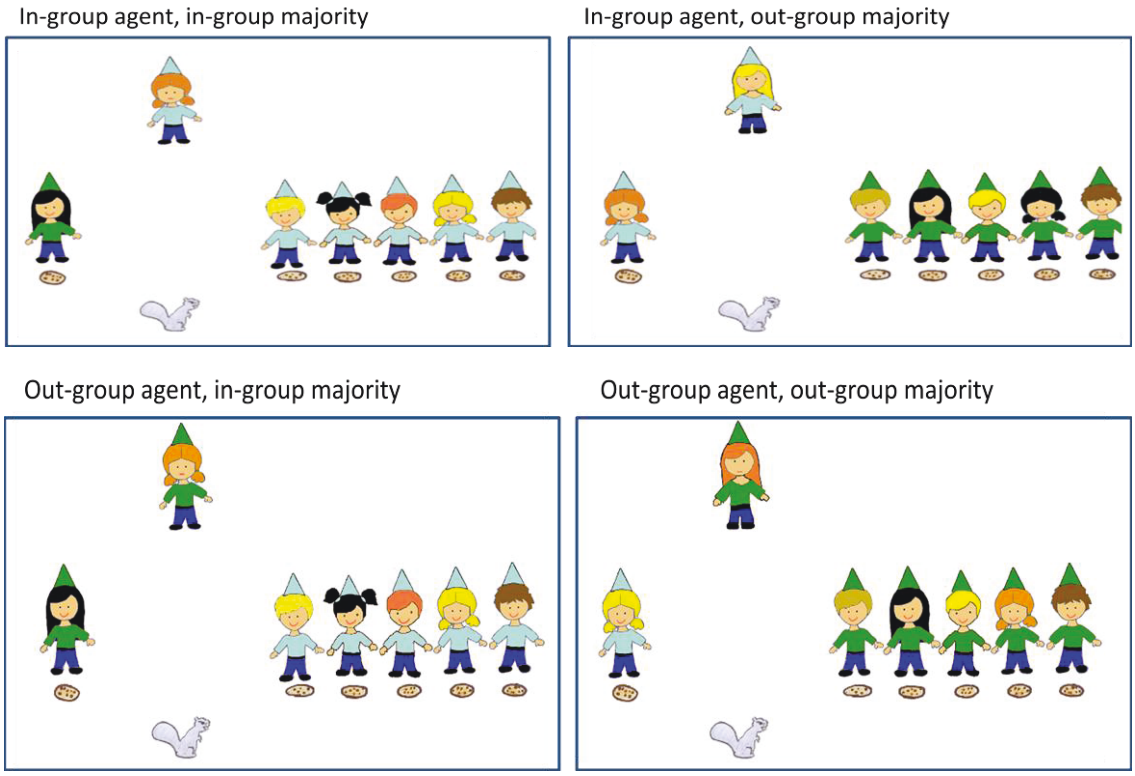

Figure 5.1. Four group conditions. This figure shows the four group conditions to which a participant who picked a blue hat could be assigned. The top two panels represent the ingroup agent conditions. The bottom two panels represent the outgroup agent conditions.

Results

Like in Chapter 4, results are organized into two sections, beginning with the comparison between side effect and main effect dilemmas (in which the majority was

[25] At the very end of the study, the experimenter reminded participants of their responses to the first two questions in each condition, and asked them explain why they gave the judgments they did. These explanations were transcribed by a coder, but were not analyzed statistically in this chapter.

always saved), and ending with the omission dilemma (in which the majority was always harmed). Within each section, results for each of the three measures (normative question, ratings question, purpose question) are presented in turn.

1. Did Participants Show the Double-Effect Effect?

Normative question

A series of logistic regression analyses fitted with the generalized estimating equations method (GEE) were used to test the main and interactive effects of dilemma (2: side effect, main effect), agent group (2: ingroup, outgroup), majority group (2: ingroup, outgroup), harm (2: property, assault), order (2: OSM, OMS), and gender on responses to the normative question. Fit statistics for three GEE models (the full model, the main effects model, and the final reduced model) are presented in tables 5.1. The full model included all potential main effects, all potential two-way interactions, and all potential three-way interactions among the variables listed above, with the exception of the three-way interaction between dilemma*majority*order (the GEE did not converge when this interaction was included, since there were no observations of "yes" responses to the main effect dilemma when it was presented last and the ingroup was harmed) (see Appendix F for the full model results). The main effects model included only the potential main effect of each variable (see Appendix F for the main effects model results). The final reduced model was selected as the best-fitting model using the same approach described in Chapter 4. Results for this model are presented in table 5.2 (see Appendix F for the parameter estimates table).

Table 5.1
Goodness-of-fit statistics for three models

Model	QIC	QICC
Full model	498.295	496.065

Main effects model	472.53	470.645
Final reduced model	455.770	454.934
Information criteria are in small-is-better form		

Table 5.2
Model effects for the reduced model (N = 237)
Dependent variable = Normative question

Model Effect	**Wald's χ2**	***df***	***p***
(Intercept)	31.484	1	.000
Dilemma	98.053	1	.000
Harm	1.165	1	.280
AgentGroup	1.903	1	.168
MajorityGroup	2.787	1	.095
Order	6.804	1	.009
Dilemma * Harm	11.461	1	.001
Dilemma * Order	13.595	1	.000

As predicted by the PDE, results indicated that dilemma was a significant predictor of responses to the normative question. Of the 241 participants, 63% approved of the protagonist's action in side effect dilemmas (Binomial test, $N = 241$, $p < .001$, two tailed), while only 11% of participants approved of the protagonist's action in main effect dilemmas (Binomial test, $N = 241$, $p < .001$, two-tailed). However, as illustrated in figure 5.2, no main or interactive effects of agent group or majority group significantly contributed to the fit of the reduced model.

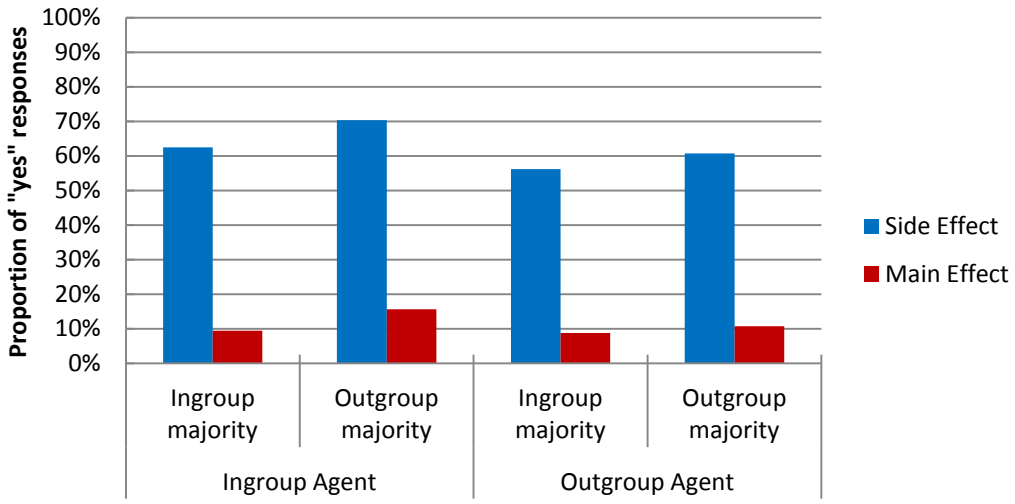

Figure 5.2. Normative judgments as a function of agent group, majority group, and dilemma. This figure shows the percentages of participants in each agent group/majority group condition who answered "yes" to the question, "Should she have done that?" as a function of dilemma (side effect, main effect).

Harm was also a significant predictor of normative responses in the main effects model, indicating that participants in the property harm condition gave "yes" responses more frequently than participants in the assault condition. However, a significant interaction between dilemma and harm in the final reduced model revealed that this effect was driven by the side effect dilemmas (figure 5.3); Simple effects indicated that like the preschoolers in Chapter 4, adult participants were significantly more likely to respond "yes" for side effect dilemmas involving a property violation (75%, Binomial test, N = 122, $p < .001$), than for side effect dilemmas involving assault (50%, Binomial test, N = 119, $p > .05$),Wald $\chi 2[1, N = 237] = 15.319$, $p < .001$. However, participants showed no significant difference in their proportion of "yes" responses to main effect dilemmas involving property and assault (9% and 13%, respectively, $p's < .001$), $\chi 2[1, N = 237] = 1.127$, $p = .288$.

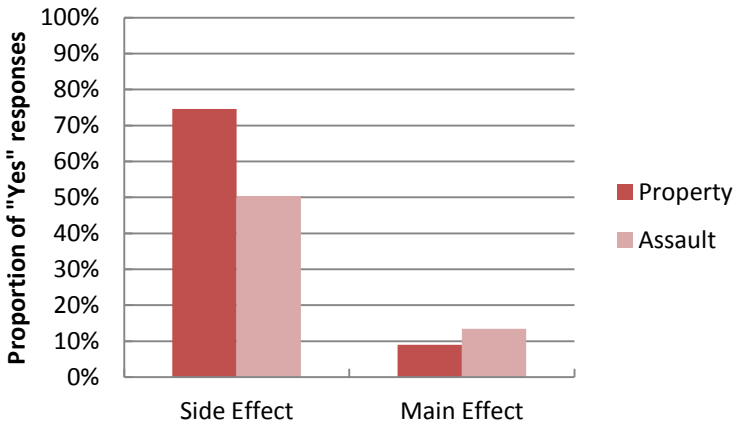

Figure 5.3. Normative judgments as a function of dilemma and harm. This figure shows the proportion of participants who responded "yes" to the normative question by dilemma harm condition (property, assault).

A significant interaction between dilemma and order also revealed that participants were significantly more likely to approve of the main effect dilemma if they had seen it before the side effect dilemma (after the omission dilemma) than if they had seen it after the side effect dilemma (18% and 4%, respectively), $\chi^2[1, N = 237] = 16.256$, $p < .001$, but their approval of the side effect dilemma did not differ between the two order conditions, $\chi^2[1, N = 237] = .480$, $p = .488$ (see figure 5.4).

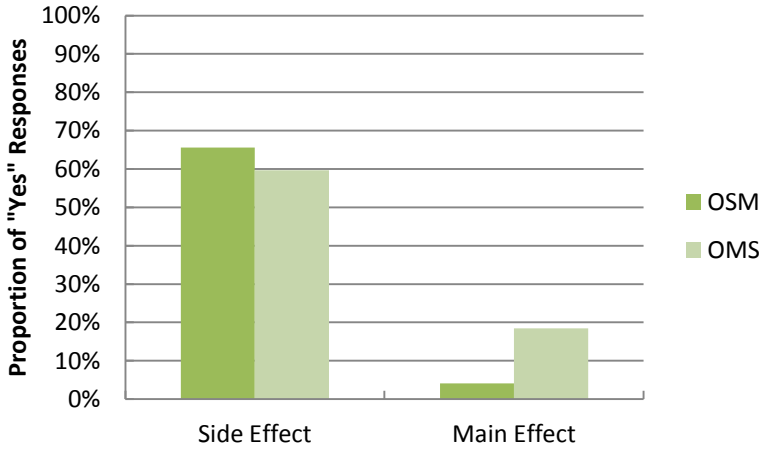

Figure 5.4. Normative judgments as a function of dilemma and order. This figure shows the proportion of participants who responded "yes" to the normative question as a function of dilemma and order of presentation (OSM = omission, followed by side effect, followed by main effect; OMS = omission, followed by main effect, followed by side effect).

Ratings question

Pink Scale ratings were scored as -2 for "really bad," -1 for "a little bad," 0 for "just ok," +1 for "a little good," and +2 for "a really good." Preliminary analyses indicated that there were no significant gender differences, so gender was dropped from further analyses. Ratings for the side effect and main effect dilemmas were analyzed using a mixed ANOVA with dilemma (2: side-effect, main effect) as a within-subjects factor, and agent (2: ingroup, outgroup), majority (2: ingroup, outgroup), harm (2: ingroup, outgroup), and order (2: side-effect first, main effect first) as between-subjects factors.

As expected, a large effect of dilemma indicated that participants gave significantly higher ratings for side effect dilemmas ((M = -0.02, SE = .07) than for main effect dilemmas (M = -1.32, SE = .06), $F(1, 225) = 357.861$, $p < .001$, $\eta_p^2 = .614$. This effect was confirmed non-parametrically (Wilcoxon signed-ranks test, $N = 241$, $Z = -11.047$, $p < .001$, $r = .71$). A significant effect of majority group was also found, but in the opposite direction from the predicted outcome. Participants who saw dilemmas in which their own majority group was saved (and a member of the outgroup was harmed) gave significantly lower ratings (M = -0.85, SE = .08) than participants who saw dilemmas in which the outgroup majority was saved (and a member of their own group was harmed) (M = -0.50, SE = .08), $F(1, 225) = 10.493$, $p = .001$, $\eta_p^2 = .045$.

This effect was qualified by a small but significant three-way interaction between dilemma x majority group x order (figure 5.5), $F(1, 225) = 4.646$, $p = .032$, $\eta_p^2 = .020$. For participants who saw the side effect dilemma second (immediately following the omission dilemma), the preference for dilemmas in which the outgroup was saved was

significant for the side effect dilemma, $F(1, 225) = 4.612$, $p = .033$, $\eta_p^2 = .020$, but not for the main effect dilemma, $F(1, 225) = 1.548$, $p = .215$, $\eta_p^2 = .007$. Conversely, for participants who saw the main effect dilemma second, the preference for dilemmas in which the outgroup was saved was significant for the main effect dilemma, $F(1, 225) = 12.140$, $p = .001$, $\eta_p^2 = .051$, but not for the side effect dilemma, $F(1, 225) = 1.086$, $p = .298$, $\eta_p^2 = .005$. No main or interactive effects of agent group were found (p's > .05).

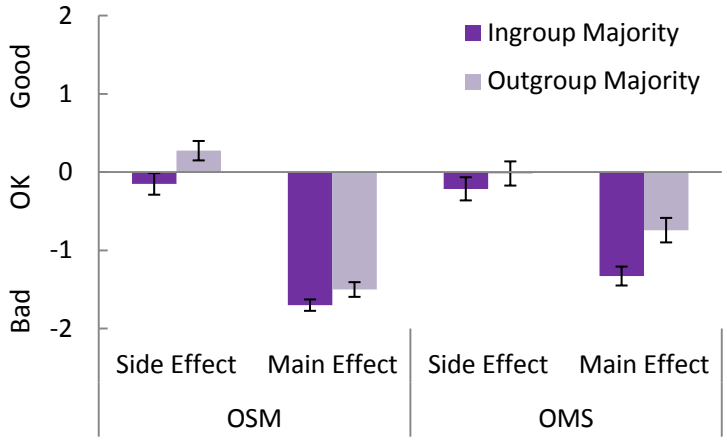

Figure 5.5. Ratings as a function of majority group, dilemma, and order. This figure shows participants' ratings for each dilemma (side effect, main effect) as a function of majority group condition (ingroup saved, outgroup saved) and order (OSM = omission, followed by side effect, followed by main effect; OMS = omission, followed by main effect, followed by side effect). Error bars show standard error of the mean.

An effect of harm indicated that participants tended to rate property harm dilemmas significantly higher than assault dilemmas, $F(1, 225) = 11.839$, $p = .001$, $\eta_p^2 = .050$. However, this effect was qualified by a significant interaction between dilemma x harm, $F(1, 205) = 14.132$, $p = .001$, $\eta_p^2 = .059$ (figure 5.6): for side effect dilemmas, participants in the property harm condition gave significantly higher ratings ($M = 0.29$, $SE = .09$) than participants in the assault condition ($M = -0.34$, $SE = .10$), $F(1, 225) = 20.581$, $p < .001$, $\eta_p^2 = .084$, but for main effect dilemmas, participants in the property harm condition ($M = -1.26$, $SE = .09$) did not give significantly different ratings on

average from participants in the assault condition (M = -1.39, SE = .09), $F(1, 225)$ = 1.016, p = .315, η_p^2 = .004.

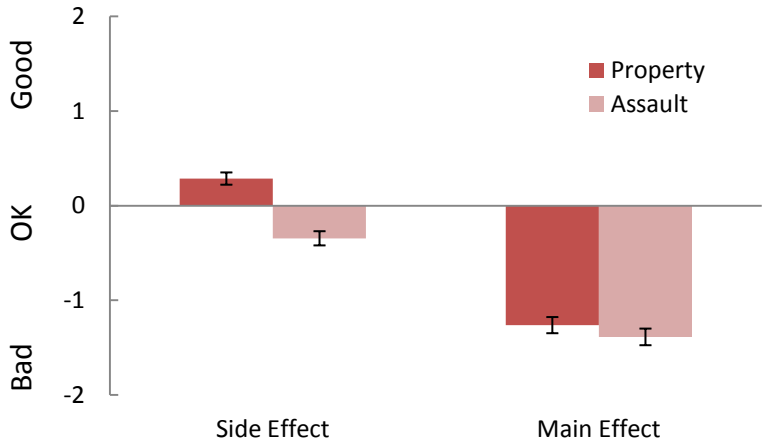

Figure 5.6. Ratings as a function of dilemma and harm. Error bars show standard error of the mean.

A significant interaction between dilemma and order was also found, $F(1, 225)$ = 29.833, $p < .001$, η_p^2 = .117. As illustrated in figure 5.7, participants rated main effect dilemmas significantly higher if they saw them second (M = -1.04, SE = .10) than if they saw them last (M = -1.60, SE = .06), $F(1, 225)$ = 23.468, p < .001, η_p^2 = .094. However, participants' ratings for side effect dilemmas when they were presented second (M = 0.06, SE = .09) did not differ significantly from participants' ratings for side effect dilemmas when they were presented last (M = -0.12, SE = .11), $F(1, 2250$ = 1.670, p = .198, η_p^2 = .007.

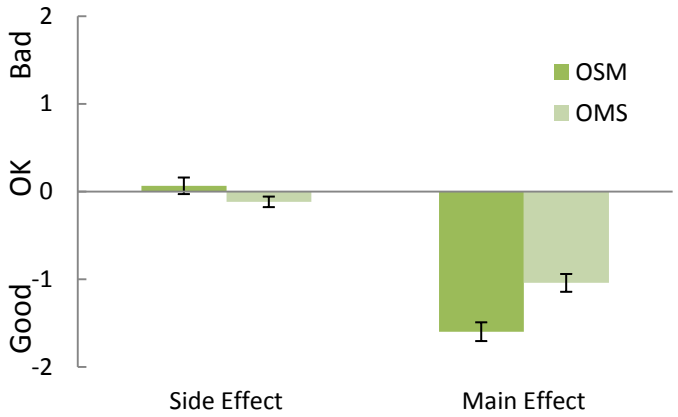

Figure 5.7. Ratings for as a function of dilemma and order. This figure shows participants' average ratings for side effect and main effect dilemmas as a function of order of presentation (OSM = omission, followed by side effect, followed by main effect; OMS = omission, followed by main effect, followed by side effect).

Purpose question

A series of logistic regression analyses fitted with the generalized estimating equations method (GEE) were used to test the main and interactive effects of dilemma (2: side effect, main effect), agent group (2: ingroup, outgroup), majority group (2: ingroup, outgroup), harm (2: property, assault), order (2: OSM, OMS), and gender on responses to the purpose question. Fit statistics for the full model, the main effects model, and the reduced model are presented in tables 5.3. Results for the final reduced model are presented in table 5.4 (see Appendix F for the reduced model parameter estimates table, the full model results, and the main effects model results).

Table 5.3
Goodness-of-fit statistics for three models

Model	**QIC**	**QICC**
Full model	429.808	427.994
Main effects model	419.342	417.415
Final reduced model	401.008	399.812
Information criteria are in small-is-better form		

Table 5.4
Model effects for the reduced model (N =190)
Dependent variable = Purpose question

Model Effect	**Wald's χ2**	***df***	***p***

(Intercept)	1.795	1	.180
Dilemma	77.110	1	.000
MajorityGroup	9.409	1	.002
Harm	29.013	1	.000
Order	3.862	1	.049
Dilemma * Order	21.314	1	.000

As expected, results indicated that dilemma was a significant predictor of responses to the purpose question (figure 5.8). Of the 190 participants who responded to the purpose question, 61% said that the negative outcome in the main effect dilemma was brought about on purpose (Binomial test, $N = 190$, $p < .01$, two-tailed), whereas only 23% said that the negative outcome in the side effect dilemma was brought about on purpose (Binomial test, $N = 190$, $p < .001$, two-tailed).

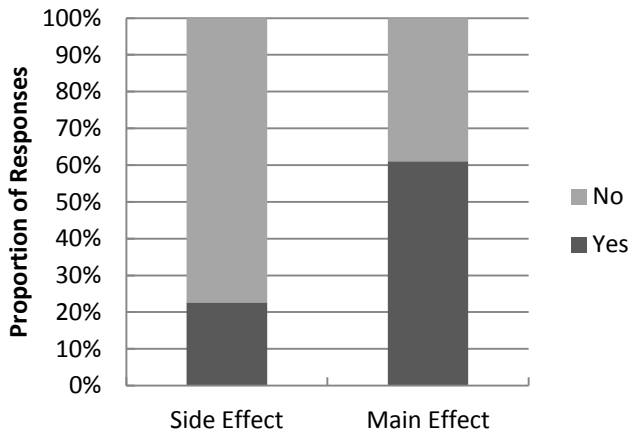

Figure 5.8. Purpose judgments by dilemma. This figure shows the distribution of participants' responses to the question, "Did she make the Blicket(s)/Greeble(s) sad on purpose?" for each dilemma.

Harm, majority group, and order were also significant predictors of responses to the purpose question. The effect of harm indicated that participants were more likely to say "yes" in dilemmas involving assault (60%) than in dilemmas involving a property violation (32%). The effect of majority group reflected a tendency for participants to say "yes" to the purpose question more frequently when their own majority group was saved

(and a member of the outgroup was harmed) (49%) than when the outgroup majority was saved (and a member of their own group was harmed) (35%). The effect of order indicated that in general, participants were more likely to say "yes" when they had seen the main effect dilemma after the side effect dilemma (47%) than when they had seen the main effect dilemma before the side effect dilemma (36%). However, a significant interaction between dilemma and order revealed that this order effect was primarily driven by responses to main effect dilemmas. Participants were significantly more likely to say "yes" to the purpose question in the main effect dilemma if they had seen it after the side effect dilemma (76%, Binomial test, $N = 95$, $p < .001$, two-tailed) than if they had seen it before the side effect dilemma (46%, Binomial test, $N = 95$, $p = .538$, two-tailed), $\chi^2[1, N = 237] = 16.694$, $p < .001$, but their responses in the side effect dilemma did not differ between the two order conditions (OSM: 19%, OMS: 26%), $\chi^2[1, N = 237] = .365$, $p = .546$. No significant effects of agent group were found.

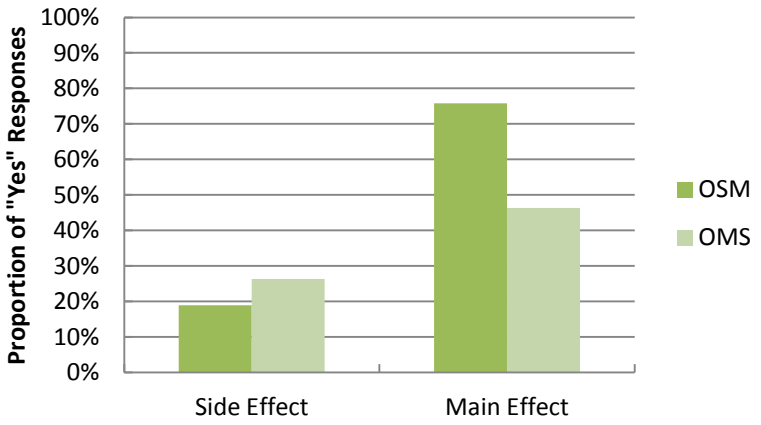

Figure 5.9. Purpose judgments as a function of dilemma and order. This figure shows the proportion of participants who responded "yes" to the normative question as a function of dilemma and order of presentation (OSM = omission, followed by side effect, followed by main effect dilemma; OMS = omission, followed by main effect, followed by side effect dilemma).

2. How did participants judge omission?

Normative question

Overall, 212 of the 241 participants (88%) responded "no" to the omission dilemma (Binomial test, $N = 241$, $p < .001$, two tailed). Preliminary analyses indicated that there were no significant gender differences in omission normative responses, so gender was dropped from further analyses. Omission normative responses were fitted with a binary logistic regression, with agent group (ingroup, outgroup), majority group (ingroup, outgroup), and harm (property, assault) as between-subjects factors. Using a forward stepwise procedure, the best-fitting model included a main effect of agent group, Wald $\chi^2(1, N = 213) = 4.415$, $p = .036$, $\beta = .926$, odds ratio (OR) = 2.524. Inspection of the data (figure 5.8) revealed that participants were more likely to respond "yes" to the omission dilemma if they were in the ingroup agent condition (16%) than if they were in the outgroup agent condition (7%). No main or interactive effects of majority group or harm were included in the best-fitting model (ps > .05). This model had an overall correct prediction rate of 88.2%. A test of this model against a constant-only model was statistically significant, $p = .026$.

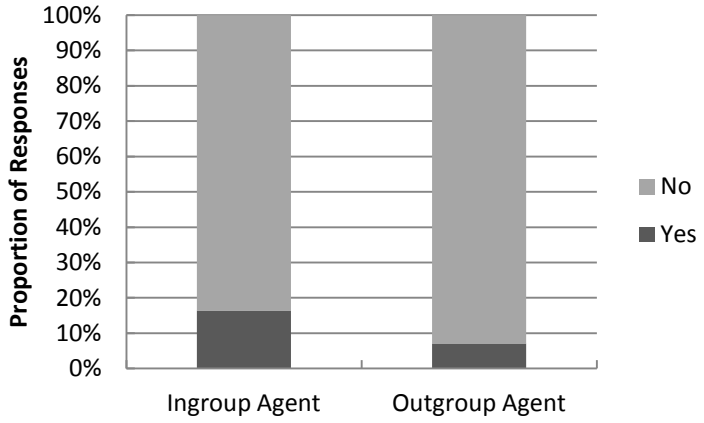

Figure 5.10. Omission: Normative judgments by agent group. This figure shows the distribution of participants' responses to the normative question in the omission dilemma as a function of whether the moral agent belonged to the participant's ingroup or outgroup.

Ratings question

Overall, participants rated the omission dilemma negatively, with an average rating of -1.07 (*SE* = .77), which was significantly different from chance $t(240) = 21.452$, $p < .001$. Preliminary analyses indicated that there were no significant gender differences in participants' omission dilemma ratings, so gender was dropped from further analyses. A 2(agent: ingroup, outgroup) x 2(majority: ingroup, outgroup) x 2(harm: property, assault) ANOVA was computed to analyze ratings for the omission dilemma. A small effect of agent group revealed that ratings for ingroup agents were rated slightly higher (*M* = -0.98, *SE* = .07) than ratings for outgroup agents (*M* = -1.16, *SE* = .07) (figure 25), $F(1, 233) = 4.938$, $p = .027$, $\eta_p^2 = .021$. No main or interactive effects of majority group or harm condition were found (*p*s > .05).

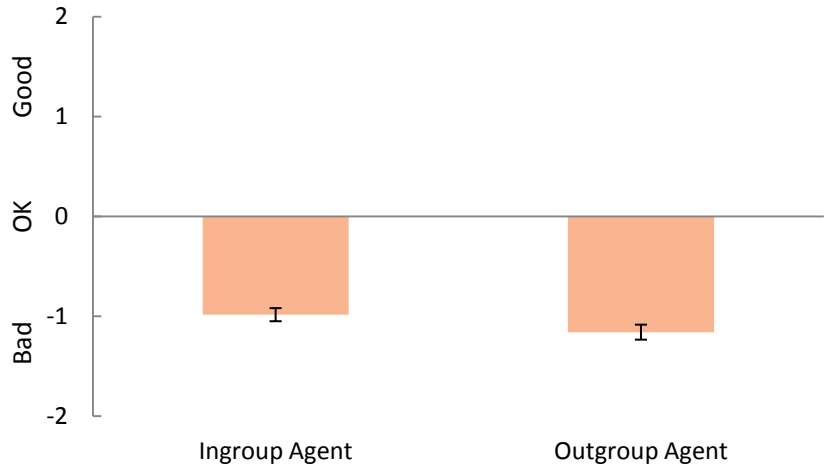

Figure 5.11. Omission: Participant's ratings by agent group. Error bars show standard error of the mean.

Purpose question

Responses to the purpose question for the omission dilemma were fitted with a series of generalized linear models, with agent group (ingroup, outgroup), majority group (ingroup, outgroup), harm (property, assault) and gender as between-subjects factors. Tables 5.5-5.6 present results for the main effects model, which was also the best-fitting model (see Appendix F for the main effects parameter estimates table, and for results of

the full model). Overall, only 36 out of the 190 participants (19%) who were asked the purpose question responded that the agent in the omission dilemma made the majority sad on purpose (Binomial test, $N = 190$, $p < .001$), two-tailed). A main effect of majority group revealed that participants were more likely to say "yes" to the purpose question when the outgroup was harmed (26%) than when the ingroup was harmed (12%), and a main effect of harm indicated that participants were also more likely to say "yes" in dilemmas involving assault (28%) than in dilemmas involving property harm (14%). Men were also significantly more likely to say "yes" (26%) than women (14%). No main or interactive effects of agent group were found.

Table 5.5
Model effects for the main effects model (N = 190)
Dependent variable = Purpose question

Model Effect	Wald's χ2	*df*	*p*
(Intercept)	45.610	1	.000
AgentGroup	.574	1	.449
MajorityGroup	6.493	1	.011
Harm	5.308	1	.021
Gender	5.164	1	.023

Table 5.6
Goodness-of-fit statistics for the main effects model (N = 190)
Dependent variable = Purpose question

Test	Value	*df*	*p*
Overall model evaluation (against intercept-only model)			
Omnibus Likelihood ratio Chi-Square	17.894	4	.001
Goodness-of-fit test			
Deviance	6.295	10	
Pearson Chi-Square	4.705	10	
Model fitting criteria			
Log likelihood	-19.773		
Akaike's Information Criteria (AIC)	49.547		
Finite Sample Corrected AIC (AICC)	49.873		
Bayesian Information Criterion (BIC)	65.782		
Consistent AIC (CAIC)	70.782		

a. Information criteria are in small-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Discussion

The dilemmas presented in the current study represent the first tests of adult intuitions on "trolley" problems that do not involve cases of battery or homicide. Nevertheless, the findings of the current study revealed a familiar pattern of intuitions: dilemmas in which an agent intended harm to one person as a means of saving five people were judged impermissible, whereas dilemmas in which an agent caused harm to one person as a foreseen side effect of saving five people were judged as permissible. This effect accounted for most of the variance in adults' responses, and was consistent not only with previous findings in the adult literature, but also with preschoolers' responses in Chapter 4. Thus, even when evaluating acts as mundane as taking another person's cookie, adults appear to respect Aquinas's principle.

Indeed, adults' intuitions in the current study were surprisingly consistent with preschoolers' intuitions in general. Not only did they show the same pattern of judgments across all three dilemmas, they also showed a similar degree of ambivalence about the side effect dilemma in the assault condition, most likely for the same reasons discussed in Chapter 4[26]. Like 4- and 5-year-olds, adults also disapproved of the decision to "just stand there" in the omission dilemma, suggesting that, like four- and five-year-olds, adults also view inaction in this dilemma as morally problematic. Like preschoolers, adults also showed anchoring effects in their moral judgments of side effect

[26] Participants' justifications provided some support for the safety-in-numbers explanation discussed in Chapter 4. Several participants referred to the fact that the victim was alone, and would therefore be more frightened by the dog. For example: "there are more on the bottom so they would be less scared than one all alone"; "the Blicket was alone so she would be more lonely and sad than the Greebles."

and main effect dilemmas, such that their judgments on the final trial were often biased in the direction of their previous judgments. However, whereas anchoring effects were primarily found in preschoolers' judgments of the side effect dilemma in Chapter 4 (e.g. the side effect dilemma was judged less positively when it followed the main effect dilemma), in the current study, adults' moral judgments were primarily affected by order in the main effect dilemma (e.g. the main effect dilemma was judged less negatively when it followed the side effect dilemma)[27].

Unlike previous studies which used real (i.e. already-established) social groups to manipulate the ingroup/outgroup structure of trolley problems, the current study is the first to use minimal groups, and the first to manipulate the identities of both the moral agent and the moral patients in a double-effect scenario. I found that, contrary to the results of some real-group studies, adults' categorical moral judgments did not vary with either the identity of the moral agent or the identities of the moral patients in a minimal group context. This suggests that, like their preschool counterparts, adults selectively attended to moral criteria (i.e. the causal and intentional properties of the action) over group-based criteria, particularly when making categorical moral judgments[28]. Although adults did show some evidence of agent effects in their ratings of the omission dilemma,

---

[27] This is somewhat surprising, as previous studies with adults have typically shown order effects on judgments of permissible acts (rather than impermissible acts). Nevertheless, we suspect that the order effects found in the current study are similarly due to a desire to maintain consistency between judgments of perceptually similar dilemmas (and perhaps also due to selective accessibility, whereby the anchor-consistent information between the two dilemmas becomes more salient).

[28] Interestingly, although group structure had a minimal effect on their judgments, participants frequently referenced group in their justifications. For example, they often referred to the characters in the story as "her people," "her own," "her tribe," "her friends," "someone not like her," and even referenced group as an explanation for her actions: "she was trying to protect her Blickets"; "She shouldn't have hurt the Greeble, no matter how much she hated it." This reinforces our claim that group structure was indeed perceptually salient to adults, but was not relevant to the PDE. It further suggests that whereas the PDE is not accessible to conscious introspection, group structure is consciously accessible, and perhaps even serves as a post-hoc justification for one's judgments.

as well as a majority group (i.e. patient) effect in their ratings of double-effect dilemmas, these effects of group were very small. Furthermore, the effect of majority group was in the opposite direction from the group effects found in previous studies: dilemmas in which an outgroup member was harmed and the ingroup majority was saved were rated as *worse* than dilemmas in which an ingroup member was harmed and the outgroup majority was saved.

One possible explanation for this puzzling result is that adults in the current study were motivated by a desire to not appear prejudiced, which caused them to overcompensate for potential bias in their ratings. The results of at least one previous study support this explanation. Recall that in the study by Uhlmann et al., (2009), individuals who identified as politically liberal were more likely to endorse harm as a main effect in the footbridge problem if the victim had a stereotypically white name (Chip) than if the victim had a stereotypically black name (Tyrone), even though participants explicitly rejected race as a valid basis for judgment. Uhlmann et al. interpreted this result as reflecting liberal antipathy toward anti-Black prejudice: "Our Chip-Tyrone manipulation presented liberals with choices likely to alert their sensitivities to issues of racial inequality, and they responded more negatively when asked to sacrifice a Black life than a White life." (p. 484). Similarly, the minimal group manipulations in the current study may have alerted participants to social taboos such as prejudice and discrimination when rating the dilemmas.

Overall, the current results revealed either small effects or no effects of group structure on adults' moral judgments in a minimal group context. These results differ from both the results of real-group trolley studies, and the well-established minimal

group effects that have been observed in a number of adult non-moral tasks (Mullen et al., 1992). I therefore suggest that the current findings provide further support for the theory that group concerns fall outside the domain of moral judgment, and that the effects of group structure found in real-group studies reflect a swamping of strictly moral judgment by highly salient pre-potent extra-moral values (e.g. emotional/motivational factors, conformity to group norms, etc.).

Finally, in the current study I asked participants to judge whether the agent in each dilemma made the victim(s) sad on purpose. I hypothesized that if the PDE was indeed guiding their judgments, participants should infer that the harmful effect of the agent's action (making the one person sad) was part of her intended action plan in the main effect dilemma, but not in the side effect dilemma. This hypothesis was supported: participants showed a tendency to say that the agent made the victim sad on purpose in the main effect dilemma, but did not make the victim sad on purpose in the side effect dilemma. Thus, despite the fact that the agent's intention in each dilemma was never explicitly stated in the current study (nor is it stated in the standard trolley problems), participants' intentionality inferences were consistent with the mental representations assumed to underlie the PDE. However, there is some evidence that moral judgments can sometimes influence participants' intentionality inferences (Knobe, 2003; Leslie et al., 2006a). Thus, it is not clear in the current study whether participants' attributions of intentionality played a causal role in their moral judgments, or whether participants' moral judgments subsequently influenced their intentional inferences in these dilemmas[29].

---

[29] However, there is no reason to suppose that the side-effect effect is at work in our double-effect dilemmas. First, in our side effect dilemmas the good main effects clearly outweigh the bad side effect (at

In the omission dilemma, despite their disapproval of the agent's inaction, participants tended to say that the agent did not make the five victims sad on purpose. This finding suggests that at least in the omission dilemma, participants were not simply basing their intentional inferences on the valence of their moral judgments. Furthermore, it is consistent with Mikhail's theory of how the mind computes the intentional structure of acts and omissions (i.e. how it computes ends, means, and side effects). According to Mikhail, unless we are given sufficient evidence to the contrary, we default to a "presumption of good intentions." In other words, we assume the agent "is a person of good will, who pursues good and avoids evil" (Mikhail, 2011, p. 173). It is for this reason that we assume that the ultimate end/goal of the agent in double-effect scenarios (both in the current study, and in the traditional trolley problems) is to save the five people, rather than to harm the one person, even though this is never explicitly stated.

In the omission dilemma, the agent's mental state and end/goal is even more ambiguous. She *knows* that the five people will be harmed if she does not act, so why does she choose not to act? Because she wants the five people to be harmed? Because she simply does not care whether the five people are harmed are not? Because she cannot think of anything else to do? Because she is lazy or afraid? Because she believes the five people can fend for themselves? Given this ambiguity, according to Mikhail's model, participants should default to an assumption of benevolent intentions, as they did in the current study. However, while the current results support this theory, further research is

---

least in the property harm condition), whereas in the side-effect effect scenarios, the main effects are purely self-serving (the CEO gets to make a profit, in the adult version, and the little boy gets to be with his frog, in the children's version). Furthermore, preschool data shows that the side-effect effect is conditioned on the agent *not caring* about the side effect (Leslie et al., 2006a). However, in our dilemmas there is reason to assume that the agent *does* care about the bad side effect, but acts anyway in the interest of the greater good.

needed to determine whether participants are indeed representing the agent's action plan using the "good intentions" default. Do children also use the good intentions default? If so, what is the nature of the evidence that is sufficient to overcome this default? (see Levine & Leslie (2013, October) for ongoing work on this topic).

VI. General discussion

In Chapter 1, I introduced the trolley problem, a classic thought experiment in moral philosophy that has recently become a popular experimental tool for investigating the psychological foundations of moral judgment. I described several prominent theories of moral cognition that have used a class of trolley problems as a means of exploring the cognitive mechanisms underlying our moral judgments. In particular, I outlined a computational theory proposed by John Mikhail (2011) that makes an intriguing but controversial claim – namely, that we are evolutionarily endowed with a neuro-cognitive mechanism (UMG) that is comprised of an innate set of abstract principles, rules, and concepts that guide our acquisition of moral knowledge.

The studies presented in this dissertation explored this claim by focusing on two principles that have a long history within the fields of philosophy, psychology, religion, and law, and have been suggested to be part of our universal moral grammar (Mikhail, 2011): the principle of double effect and the duty of rescue. Although several studies have shown that the PDE is universally operative in adults' moral judgments (but is non-instrospectable), the studies in this dissertation are among the first to test preschoolers' knowledge of these principles, and the first to use "double effect" dilemmas (dilemmas involving an action that produces more than one effect) that do not involve bodily harm, either in the developmental literature or the adult literature.

Across three experiments, I uncovered four major findings: 1) Both preschoolers and adults showed a strong and stable pattern of intuitions consistent with the PDE, even when evaluating dilemmas involving abstract moral violations such as trespass to

personal property and assault. That is, dilemmas in which an individual was intentionally harmed as a means to saving five other people were judged as impermissible, whereas dilemmas in which an individual was harmed as a foreseen side effect of saving five people were judged as permissible. 2) 4- and 5-year-olds and adults, but not three-year-olds, disapproved of an agent's choice *not* to act when she could have reasonably intervened to prevent harm to others. 3) In the case of minimal groups, ingroup loyalty had little to no effect on either preschoolers' or adults' moral judgments in these dilemmas. 4) At least for adults, the pattern of intentional inferences predicted by Mikhail's UMG model holds: participants showed a tendency to say that the agent made the victim sad on purpose in the main effect dilemma, but did not make the victim sad on purpose in the side effect dilemma or the omission dilemma.

In the following sections I frame these findings in the context of three theoretical questions: 1) Are our moral intuitions primarily driven by emotions or principles? 2) Is our moral sense innate? 3) Is our moral sense impartial?

1. Are our moral intuitions primarily driven by emotions or principles?

In Chapter 1, I described two theoretical approaches to studying moral intuition: the cognitive/computational approach, which operates on the assumption that a number of complex, universal, and possibly innate moral principles are systematically guiding our moral intuitions below the level of conscious awareness (e.g. Dwyer, 1999; Hauser, 2006; Mikhail, 2011; Rawls, 1971), and the emotions-based approach, which emphasizes the causal role of emotions in moral judgment and development (e.g. Damasio, 1994; Greene & Haidt, 2002; Haidt, 2001, 2003; Moll & de Oliveira-Souza, 2007; Nichols, 2004; Prinz, 2004, 2007). Whereas the emotions-based approach relies on purely perceptual

models to explain our moral intuitions (i.e. a perceptual feature of the stimulus triggers the accompanying emotion, which triggers the intuition), the computational approach postulates an "intervening step…a pattern of organization that is imposed on the stimuli by the mind itself" (Mikhail, 2007, p. 145). Although this computational approach does not deny the role of emotions in moral judgment, it assumes that such emotions are triggered by the application of internally represented moral principles and/or the abstract mental representations of the stimulus over which those principles are defined.

The studies presented in this dissertation provide evidence in support of the computational approach. The preschoolers and adults in these studies showed a pattern of intuitions that is difficult to explain by merely appealing to perceptual features in the stimulus. Greene et al.'s (2001/2004) personal/impersonal distinction does not account for the current findings, as all of the dilemmas in the current studies would qualify as "impersonal" dilemmas according to Greene et al.'s (2001, 2004) definition. Even Greene et al.'s (2009) definition of personal force would not apply to the current property harm dilemmas, as there was no direct physical contact between agent and patient, or even any "force" that directly impacted the victims in these dilemmas[30]. Furthermore, as discussed in chapter 3, simpler rules such as "don't take things that aren't yours" are unlikely to explain preschoolers' intuitions in these dilemmas, as even young children recognize that taking possession of another person's property is wrong only when that person does not give his or her consent (Rossano et al., 2011, Neary et al., 2009).

I suggest instead that children in the current studies demonstrated tacit knowledge of three principles: the prohibition of intentional trespass to personal property (similar in

[30]However, our main effect assault dilemma would fall under Greene's definition of personal force, as well as under Mikhail's definition of intentional battery.

principle to Mikhail's (2002) prohibition of intentional battery), the PDE, and the duty of rescue[31]. However, further research is needed to explore the extent to which children possess full knowledge of these principles. A common formulation of the PDE states that an action is permissible if it meets the following criteria (Mikhail, 2011):

1) the prohibited act itself is not directly intended
2) the good but not the bad effects are directly intended
3) the good effects outweigh the bad effects
4) no morally preferable alternative is available

However, in the current studies, only the first criteria was manipulated. Nevertheless, we can reasonably assume that participants inferred that the good but not the bad effects were directly intended in the side effect dilemma (adults' responses to the purpose question support this assumption. Also see Levine & Leslie (2013, October) for ongoing work on this topic). We can also reasonably assume that participants perceived the good effects to have outweighed the bad effects at least in the property harm dilemmas, or they would not have judged the side effect dilemma as permissible. Indeed, I suspect that the reason both adults and preschoolers responded less positively to the side effect dilemma in the assault condition (relative to the property harm condition) is because the good effects in the assault condition did not clearly outweigh the bad effects. As evidenced by some of the adults' verbal justifications (which were not reported in this dissertation), many participants reported that they thought the victim who was by herself would be

[31] And perhaps also the prohibition of intentional infliction of emotional distress (in the case of the main effect assault dilemma) and the prohibition of negligent infliction of emotional distress (in the case of the omission assault dilemma)

more frightened by the angry dog (because she was alone) than the five people would be (because they had each other).

Returning to the first criteria, I suggest that the most likely explanation for the pattern of intuitions observed in the double effect dilemmas in the current studies is that both preschoolers and adults possess tacit knowledge of the distinction between harm as an intended means and harm as a foreseen but unintended side effect. This is supported by the fact that both age groups (preschoolers and adults) showed the same pattern of intuitions across all three studies – a pattern which has consistently been observed in adults across a wide range of demographics (Cushman et al., 2006; Greene et al., 2001, 2004; Hauser et al., 2007; Mikhail, 2002; O'Neill & Petrinovich, 1998; Petrinovich et al., 1993), even when other potentially relevant factors are carefully controlled, such as the degree of physical contact between the agent and the victim, the temporal order of the good and bad effects, whether the act is personal or impersonal, and whether a new threat has been introduced or an existing threat has been redirected (Mikhail, 2002; Cushman et al., 2006; Hauser et al., 2007). Furthermore, adults' responses to questions about the agent's intention in each dilemma were consistent with the intentional structure predicted by the PDE. Although it is possible that the principles underlying children's and adults' intuitions are different (a possibility that requires further inquiry), postulating two separate models to explain the same pattern of judgments in children and adults is unwarranted given that we have theoretical reasons to suppose that the same set of principles is operative in both children and adults.

Interestingly, although the side effect dilemmas did not clearly meet the fourth "better alternative" criteria, participants were nevertheless willing to judge the side effect

dilemma as permissible. That is, even though it is plausible that participants could have come up with a better alternative to putting up the gate (presumably one in which all six people were saved instead of only five), even adult participants judged the act of putting up the gate as morally permissible. With respect to preschool participants, this is not particularly surprising. Although preschoolers are capable of generating a single/constrained counterfactual (e.g. "what if she had done X instead?"), they have difficulty generating open/relatively unconstrained counterfactuals (e.g. "what else could she have done?") (e.g. Beck, Robinson, Carroll, & Apperly, 2006). However, in the case of adult participants, although they are clearly capable of generating better alternatives, they may have assumed (for the purpose of the story) that there was no better alternative in the side effect dilemma. Nevertheless, this raises interesting questions for future research: under what conditions do people spontaneously generate better alternatives, and how are the possible alternatives constrained (in particular, how do we identify the least harmful alternative)? After all, both adults and 4- and 5-year-olds judged the omission dilemma negatively, presumably because they recognized there was a better alternative to "doing nothing."

Admittedly, the principle(s) underlying participants' responses in the omission dilemma are less clear. Although my results hint that children as young as four-years-old consider it a moral obligation not only to avoid harm, but also to help those in need, this conclusion is speculative. According to Mikhail's formulation of the rescue principle, preventing harm is obligatory unless doing so requires unjustified costs, such as risk to one's own safety, or violating other higher-ranking moral principles. This involves inferring that the agent *could* have intervened in a permissible/low-risk manner, but chose

not to. In other words, in order to judge "doing nothing" as impermissible, one must keep in mind two acts (and their effects) simultaneously: the current act (knowingly harmful omission) and its least harmful alternative. However, as discussed in previous chapters, it is not clear *which* least harmful alternative to "doing nothing" participants were contemplating, or whether this alternative involved "unjustified" costs (nor is it clear what would constitute "unjustified costs" in this scenario). Nevertheless, I tentatively suggest that three-year-olds' difficulty in this dilemma was due to the cognitive demands of having to spontaneously generate an appropriate least harmful alternative that met these criteria and then compare it to its respective omission.

In Chapter 3, I suggested that more research is needed to explore preschoolers' understanding of the duty of rescue – in particular, the conditions under which children consider it morally obligatory to intervene on another's behalf, and the development of children's understanding of choice in the context of moral judgment. I also suggested that future work should investigate children's knowledge of the moral distinction between harming by omission and harming by commission. Our finding that children and adults tended to rated the omission dilemma slightly higher than the main effect dilemma, despite the fact that more people were harmed in the omission dilemma, suggests this distinction is operative in their judgments. However, future work should explore this more directly.

2. Is our moral sense innate? [32]

In Chapter 2, I presented some preliminary evidence for the argument of the poverty of the stimulus in the moral domain (Mikhail, 2002). The results of the studies

[32] I use the term "innate" here in the way it is typically used in cognitive development, to mean "emerges without exposure to relevant information in the environment."

presented in Chapters 3-5 provide additional support for the hypothesis that the moral input children receive is insufficient to account for their moral knowledge. If children were indeed guided by the PDE in the current studies, as I suggest, it is unlikely that anyone has explicitly taught three-year-olds this principle. Indeed, even adults are incapable of articulating the PDE when asked to justify their judgments of trolley problems (Mikhail, 2002; Cushman et al., 2006; Hauser et al., 2007)[33]. This suggests that preschoolers likely grasp this principle by way of something other than cultural transmission. Furthermore, given that such a principle relies on abstract structural descriptions that are not directly observable in the stimulus, it is unlikely that domain-general learning can account for children's acquisition of this principle.

Similar reasoning can also be applied to the prohibition of intentional trespass to personal property. Indeed, a growing body of work on children's knowledge of ownership (as distinct from physical possession) and property rights supports Mikhail's (2007) hypothesis that "the intuitive jurisprudence of young children is complex and exhibits many characteristics of a well-developed legal code." (p. 143). For example, children in the preschool years can reason about who owns what, infer control of permission/exclusion from property status, and distinguish between transfers of ownership versus borrowing/lending or theft (e.g. Beggan, 1992; Berti et al., 1982; Friedman & Neary, 2008; Neary et al., 2009; Rossano et al., 2011). We are therefore lead to the tentative conclusion that much of children's moral knowledge may be the result of innate principles.

---

[33] Although we did not ask preschoolers for their moral justifications in these studies, we can reasonably assume that three-year-olds would also have difficulty providing logically adequate justifications for their intuitions.

3. Is our moral sense impartial?

In Chapter 4, I discussed the question of moral impartiality. While some theories contend that the moral domain concerns not only issues of harm, care, and fairness, but also issues of ingroup loyalty (Haidt & Graham, 2007), others draw a principled distinction between "considered judgments" – judgments in which our moral capacities are most likely to be displayed without distortion" (Rawls, 1971, p. 47) – and prejudices (Mikhail, 2011). Under this view, although we may behave in ways that are prejudiced or discriminatory, such behavior is not an accurate reflection of our underlying moral competence. In other words, the moral grammar underlying our judgments is impartial, but factors exogenous to the moral system, such as the feelings, attitudes, and stereotypes we form toward certain groups or individuals, may (either consciously or unconsciously) bias or distort our moral judgments and actions.

The results presented in Chapters 4 and 5 support the latter hypothesis. Despite the availability of social category information in these studies, ingroup loyalty had little to no effect on either preschoolers' or adults' moral judgments. That is, both preschoolers' and adults' intuitions (particularly their categorical judgments) overwhelmingly relied on morally-relevant factors (e.g. property violation, assault, intended harm vs. foreseen side effects etc.) over group-based information. To be clear, I do not suggest that preschoolers and adults did not discriminate between the minimal groups used in the current studies or that they did not identify with their minimal ingroup. On the contrary, the small effects of group found on children's ratings in the omission dilemma, as well as the small group effects found on adults' moral judgments indicate that participants were indeed aware of the group structure in these dilemmas and their

own group affiliation. Furthermore, previous research has shown that the mere presence of a minimal ingroup/outgroup structure is sufficient for inducing moderate to large effects of ingroup favoritism in adults (see Mullen et al., 1992 for a review), and even in preschoolers (e.g. Dunham et al., 2011) on a variety of non-moral tasks, even in the absence of group labels[34]. I do not dispute these claims, or similar claims that children from a very early age are sensitive to social category information, particularly when making sense of others' behavior, and when generating behavioral predictions (e.g. Dunham et al., 2011; Rhodes, 2012). Instead, I suggest that these group-based concerns are outside of the moral system, but can exert a distorting influence on moral competence, particularly when group information is highly salient and massively valued compared with the rest of the elements in the task, such as in the case of "real" groups that have a history of conflict, norms, stereotypes, etc. However, future work should explore this hypothesis further. In particular, future work should investigate the role of group structure on moral judgment in the case of social categories such as native language, which has been shown to induce group preferences in infants as young as five months old (Kinzler et al., 2007).

4. Conclusions

The experiments presented in this dissertation show that both adults and preschoolers exhibit a strong and stable pattern of intuitions that is consistent with the principle of double effect, even when evaluating dilemmas involving abstract moral violations such as trespass to personal property or assault, and even in the context of

[34] In the current study, not only were the groups marked visually and with a verbal label (characters were identified as "Blickets and Greebles" throughout the study), participants were also reminded of their group affiliation at the beginning of each dilemma. Indeed, one could argue that the group information was more salient in these dilemmas relative to the intentional structure of the action, which had to be inferred.

minimal group structure. I suggest that these findings provide support for the hypothesis that our moral knowledge is rational/principled, impartial, and most likely relies on an innate neuro-cognitive structure specific to our moral sense. If true, this hypothesis is likely to have broader implications in the areas of justice, public policy, and education. First, it suggests that children are already pre-wired to care about and attend to the morally relevant aspects in their environment, placing the roles of parents and educators in a new light. Rather than teaching children the concepts of "right" and "wrong" from scratch, or teaching children a list of rules they must learn such as "don't steal," it may be that parents and teachers serve a different but equally important function: to expose their children to the right kinds of moral input, and to help children *act appropriately* in light of their moral knowledge. While the research presented in this dissertation focused exclusively on children's moral knowledge (i.e. moral competence), future research is needed to investigate how this knowledge plays out in terms of the development of moral behavior (i.e. performance). Why do people act in moral and immoral ways, and what role does our environment play in the development of such behavior?

Second, the moral grammar hypothesis generates a somewhat surprising and disturbing prediction. If we can indeed model our moral knowledge in terms of computational principles, a computer program could hypothetically be written to execute the same moral operations that the mind/brain computes. Although this sounds like science fiction, it may perhaps become a reasonable concern in the near future. The Department of Defense has already begun to invest money in research aimed at modeling theory of mind and human morality, and drone fighters are being developed by the military. This raises significant ethical concerns for our line of work, and future research

should proceed with caution. Finally, while we may be able to eventually uncover the set of principles, concepts, and parameters that make up our moral knowledge, this knowledge tells us nothing about whether our psychologically endowed principles are the "right" ones. Such work will need to be taken up by moral ethicists and philosophers, particularly if we are headed towards a future in which drones will be modeled on human morality.

Appendix A
Twelve trolley problems

In a series of papers (Mikhail, 2002; Mikhail, 2007; Hauser et al., 2007a), and in his recent book (Mikhail, 2011), John Mikhail describes twelve different trolley problems that he adapted from the original trolley problems devised by Philippa Foot (1967), Gilbert Harman (1977), and Judith Thomson (1985). The following descriptions are taken from Mikhail's (2011) book, pages 106-109:

> **Bystander**. Hank is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Hank sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Hank is standing next to a switch, which he can throw, that will turn the train onto a side track, thereby preventing it from killing the men. There is a man standing on the side track with his back turned. Hank can throw the switch, killing him; or he can refrain from doing this, letting the five die. Is it morally permissible for Hank to throw the switch?
>
> **Footbridge**. Ian is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Ian sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast they will not be able to get off the track in time. Ian is standing next to a heavy object, which he can throw onto the track in the path of the train, thereby preventing it from killing the men. The heavy object is a man, standing next to Ian with his back turned. Ian can throw the man, killing him; or he can refrain from doing this, letting the five die. Is it morally permissible for Ian to throw the man?
>
> **Expensive Equipment**. Karl is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Karl sees what has happened: the driver of the train saw five million dollars of new railroad equipment lying across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the equipment. It is moving so fast that the equipment will be destroyed. Karl is standing next to a switch, which he can throw, that will turn the train onto a side track, thereby preventing it from destroying the equipment. There is a man standing on the side track with his back turned. Karl can throw the switch, killing him; or he can

refrain from doing this, letting the equipment be destroyed. Is it morally permissible for Karl to throw the switch?

Implied Consent. Luke is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Luke sees what has happened: the driver of the train saw a man walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the man. It is moving so fast that he will not be able to get off the track in time. Luke is standing next to the man, whom he can throw off the track out of the path of the train, thereby preventing it from killing the man. The man is frail and standing with his back turned. Luke can throw the man, injuring him, or he can refrain from doing this, letting the man die. Is it morally permissible for Luke to throw the man?

Intentional Homicide. Mark is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Mark sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Mark is standing next to a switch, which he can throw, that will turn the train onto a side track, thereby preventing it from killing the men. There is a man on the side track. Mark can throw the switch, killing him; or he can refrain from doing this, letting the men die. Mark then recognizes that the man on the side track is someone who he hates with a passion. "I don't give a damn about saving those five men," Mark thinks to himself, "but this is my chance to kill that bastard." Is it morally permissible for Mark to throw the switch?

Loop Track. Ned is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Ned sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Ned is standing next to a switch, which he can throw, that will temporarily turn the train onto a side track. There is a heavy object on the side track. If the train hits the object, the object will slow the train down, giving the men time to escape. The heavy object is a man, standing on the side track with his back turned. Ned can throw the switch, preventing the train from killing the men, but killing the man. Or he can refrain from doing this, letting the five die. Is it morally permissible for Ned to throw the switch?

Man-in-front. Oscar is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Oscar sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the

track in time. Oscar is standing next to a switch, which he can throw, that will temporarily turn the train onto a side track. There is a heavy object on the side track. If the train hits the object, the object will slow the train down, giving the men time to escape. There is a man standing on the side track in front of the heavy object with his back turned. Oscar can throw the switch, preventing the train from killing the men, but killing the man; or he can refrain from doing this, letting the five die. Is it morally permissible for Oscar to throw the switch?

Costless Rescue. Paul is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Paul sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Paul is standing next to a switch, which he can throw, that will turn the train onto a side track, thereby preventing it from killing the men. Paul can throw the switch, saving the five men; or he can refrain from doing this, letting the five die. Is it morally obligatory for Paul to throw the switch?

Better Alternative. Richard is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Richard sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Richard is standing next to a switch, which he can throw, that will turn the train onto a side track, thereby preventing it from killing the men. There is a man standing on the side track with his back turned. Richard can throw the switch, killing him; or he can refrain from doing this, letting the men die. By pulling an emergency cord, Richard can also redirect the train to a third track, where no one is at risk. If Richard pulls the cord, no one will be killed. If Richard throws the switch, one person will be killed. If Richard does nothing, five people will be killed. Is it morally permissible for Richard to throw the switch?

Disproportional Death. Steve is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Steve sees what has happened: the driver of the train saw a man walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the man. It is moving so fast that he will not be able to get off the track in time. Steve is standing next to a switch, which he can throw, that will turn the train onto a side track, thereby preventing it from killing the man. There are five men standing on the side track with their backs turned. Steve can throw the switch, killing the five men; or he can refrain from doing this, letting the one man die. Is it morally permissible for Steve to throw the switch?

Drop Man. Victor is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Victor sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the

brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Victor is standing next to a switch, which he can throw, that will drop a heavy object into the path of the train, thereby preventing it from killing the men. The heavy object is a man, who is standing on a footbridge over-looking the tracks. Victor can throw the switch, killing him; or he can refrain from doing this, letting the five die. Is it morally permissible for Victor to throw the switch?

Collapse Bridge. Walter is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Walter sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Walter is standing next to a switch, which he can throw, that will collapse a footbridge overlooking the tracks into the path of the train, thereby preventing it from killing the men. There is a man standing on the footbridge. Walter can throw the switch, killing him; or he can refrain from doing this, letting the five die. Is it morally permissible for Walter to throw the switch?

Appendix B
Chapter 3 stimuli

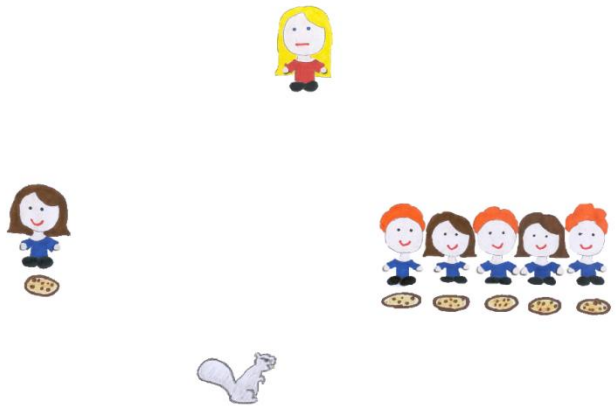

Figure B1. Introduction. This is Jane. Jane is in the park today. And there are some other people in the park too. There are lots of people over here, and there is one person over here. And look! They all have a snack. These people over here have a snack, and this person over here also has a snack. What does it look like they are eating? That's right! They all have cookies. But Uh oh! What is that? That's right. That is a sneaky squirrel. And do you know what he likes to do? He likes to eat other people's food! And he sees all those yummy cookies over there, so he is going over there to eat all of those cookies! That will make these people very sad.

Controls:

- So do you remember, where is the squirrel going to go? [Prompt: Whose cookies does he want to eat?]
- How will that make those people feel?

Well, Jane sees that sneaky squirrel, and Jane knows that the sneaky squirrel is going to eat those cookies and make those people sad. Let's see what Jane does.

Play animation

Figure B2. Side effect dilemma. Well, Jane puts up a wall. Jane knows that if she puts up a wall next to these people, the squirrel will go over here and eat this person's cookie instead. But now this person sad. Let's watch that again (repeat from the beginning).

Controls:

- So do you remember, what did Jane do in this story?
- And when Jane put up the wall, what did the squirrel do?
- How did that make this person feel?

Test:

- Do you think Jane should have put up the wall?
- Can you show me on the Pink Scale? Was putting up the wall a good thing to do, a bad thing to do, or an ok thing to do? Was it a little good/bad or very good/bad?

Omission dilemma

Ok! So I'm going to tell you the same story again, but this time Jane is going to do something a little different (repeat introduction).

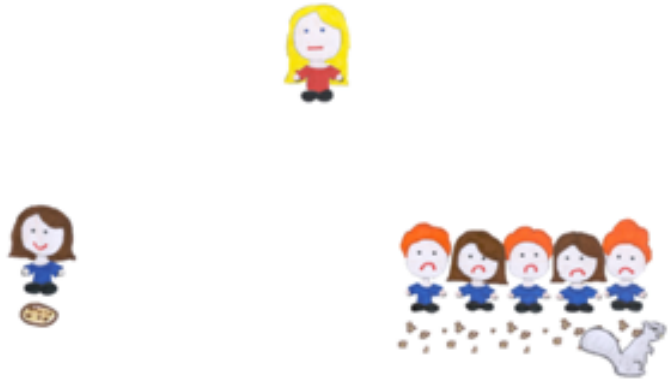

Figure B3. Omission dilemma. Well, Jane doesn't do anything. She just stands there and does nothing. Jane knows that if she just does nothing, the squirrel will go over here and eat all these people's cookies. And now they are all sad. Let's watch that again (repeat from the beginning).

Controls:

- So do you remember, what did Jane do in this story? (Prompt: did she do anything?)
- And what did the squirrel do?
- How did that make these people feel?

Test:

- Do you think Jane should have just done nothing?

- Can you show me on the Pink Scale? Was just doing nothing a good thing to do, a bad thing to do, or an ok thing to do? Was it a little good/bad or very good/bad?

Main effect dilemma

Ok! So I'm going to tell you the same story again, but this time Jane is going to do something a little different.

Figure B4. Main effect dilemma. Well, Jane takes this person's cookie and gives it to the squirrel. Jane knows that if she takes this person's cookie and gives it to the squirrel, the squirrel will eat this person's cookie instead. But now this person is sad. Let's watch that again (repeat from the beginning).

Controls:

- So do you remember, what did Jane do?
- And when Jane took this person's cookie, what did the squirrel do?
- How did that make this person feel?

Test :

- Do you think Jane should have taken that person's cookie?
- Can you show me on the Pink Scale? Was giving that person's cookie to the squirrel, a good thing to do, a bad thing to do, or an ok thing to do? Was it a little good/bad or very good/bad?

Appendix C

Chapter 3 supplementary results

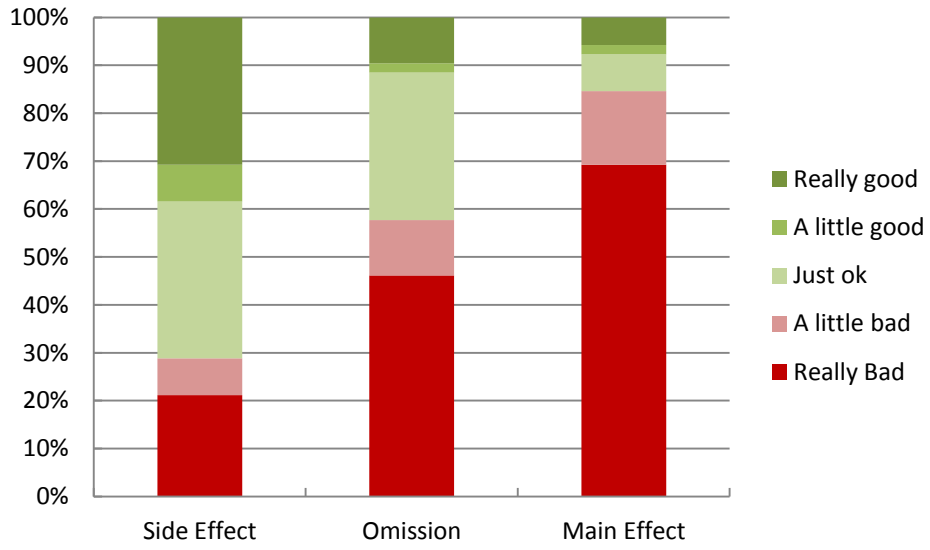

Figure C1. Distribution of children's ratings by dilemma.

Appendix D

Chapter 4 design and stimuli

Table D1

Chapter 4 study design: Independent Variables Combined to Form 16 conditions, and Number of Participants (N) in Each Condition

IV1: Majority Group	IV2: Harm	IV3: Age Group	IV4: Order	N
Ingroup Majority	Property	3 years	Omission, Side Effect, Main Effect (OSM)	14
			Omission, Main Effect, Side Effect (OMS)	13
		4 & 5 years	OSM	13
			OMS	13
	Assault	3 years	OSM	12
			OMS	12
		4 & 5 years	OSM	13
			OMS	14
Outgroup Majority	Property	3 years	OSM	14
			OMS	15
		4 & 5 years	OSM	13
			OMS	13
	Assault	3 years	OSM	12
			OMS	11
		4 & 5 years	OSM	16
			OMS	15

Group Assignment

Now I'm going to tell you a story about Blickets and Greebles. In this story, Blickets wear blue hats, and Greebles wear green hats. You can wear a hat too!

- **Which hat would you like? Do you want to be a Blicket or a Greeble?**

Here's your _____ hat. Now you are a _____! You can wear it during the story and take it home to you when we're done!

Property harm (Blicket ingroup majority example)

Summary of the procedure: The events in the three property harm animations were virtually identical to those in the Chapter 3 study, with the following exceptions: 1) each of the three stories involved a different set of characters and a different moral agent (Jane, Sally, or Lisa); 2) in each story, the agent and characters were identified by their group ("Blickets" or "Greebles"); 3) "munching" sound effects were added when the squirrel ate people's cookies; 4) the children who still had their cookies ate them at the end of the cartoon.

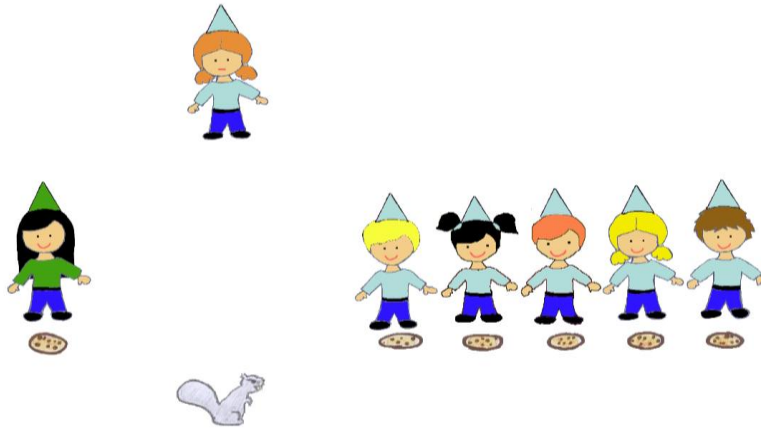

Figure D1. Introduction. This is Jane. Jane is a Blicket, just like you. And look! There are lots of Blickets over here, and one Greeble over here. And they each have a cookie. They are just about to eat their cookies when along comes a sneaky squirrel! The sneaky squirrel likes to eat other people's food. And look! The squirrel is looking at the Blickets' cookies. He wants to eat the Blickets' cookies! If the squirrel eats the Blickets' cookies, they will be very sad.

Controls:

- Where does the squirrel want to go? [prompt: whose cookies does he want to eat?]
- And if the squirrel eats all the Blickets' cookies, how will they feel?

Well, Jane sees the squirrel looking at the Blickets' cookies. Jane knows that sneaky squirrel is going to eat the Blickets' cookies and make them sad. Let's see what she does!

Play animation

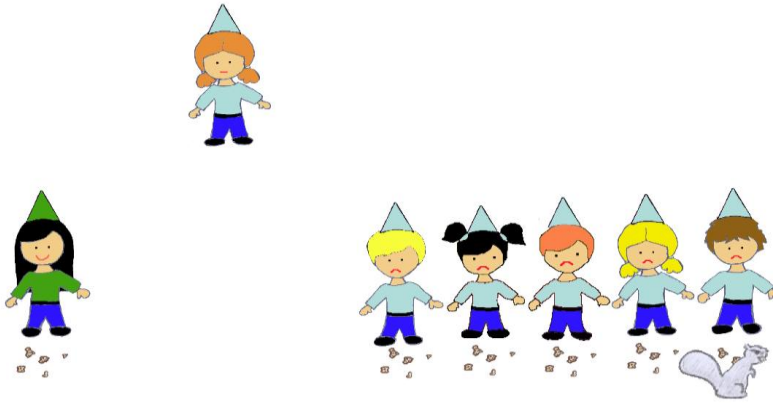

Figure D2. Omission dilemma. Well, she *could* do something, but Jane decides not to do anything. She just stands there. So the squirrel goes over there and eats all the Blickets' cookies and they are all sad. But the Greeble isn't sad because she gets to eat her own cookie. Let's watch that again (repeat from the beginning).

Controls:

- At the beginning of the story, where did the squirrel want to go? [Prompt: which cookies was he looking at?]
- What did Jane do? [Prompt: did she do anything?]

- Where did the squirrel go?
- How do the Blickets feel?
- Is the Greeble sad?

Test:

- In this story, Jane didn't do anything. She just stood there. Should she have done that?
- Can you show me on the Pink Scale? Was that good/bad/just ok? A little good/bad or very good/bad?

Side effect dilemma

Now I'm going to tell you a story about Sally. (Repeat introduction).

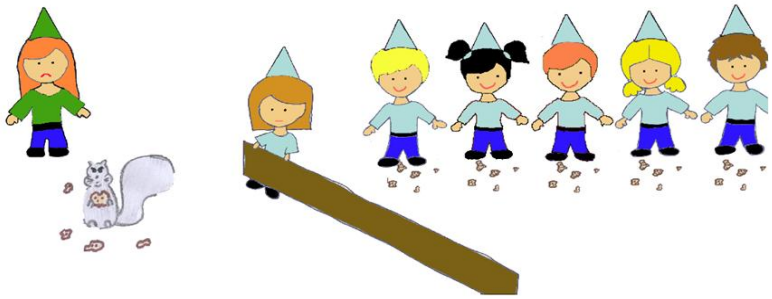

Figure D3. Side effect dilemma. Well, Sally has a gate with her, and she decides to put the gate right there. Sally knows that if she does that, the squirrel can't get to the Blickets' cookies, so he will eat the Greeble's cookie instead. Now the Greeble is sad. But the Blicekts aren't sad because they get to eat their own cookies. Let's watch that again. (Repeat from the beginning).

Controls:

- At the beginning of the story, where did the squirrel want to go? [Prompt: which cookies was he looking at?]
- What did Sally do?
- Where did the squirrel go?
- How does the Greeble feel?
- Are the Blickets sad?

Test:

- In this story, Sally used her gate. Should she have done that?
- Can you show me on the Pink Scale? Was that good/bad/just ok? A little good/bad or very good/bad?

Main effect dilemma

Now I'm going to tell you a story about Lisa. (Repeat introduction).

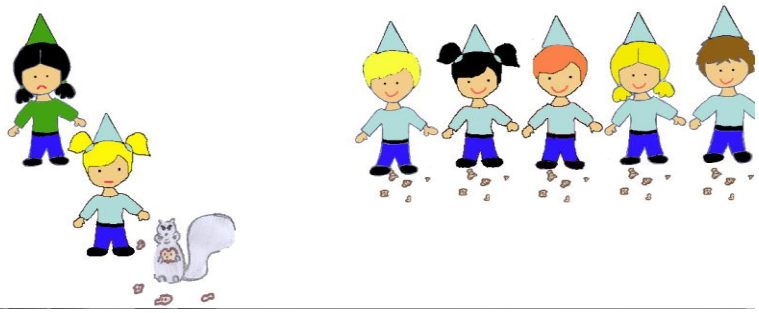

Figure D4. Main effect dilemma. Well, Lisa decides to distract the squirrel, so she takes the Greeble's cookie and feeds it to the squirrel. So the squirrel eats the Greeble's cookie instead. Now the Greeble is sad. But the Blickets aren't sad because they get to eat their own cookies. Let's watch that again. (Repeat from the beginning).

Controls:

- At the beginning of the story, where did the squirrel want to go? [Prompt: which cookies was he looking at?]
- What did Lisa do?
- Where did the squirrel go?
- How does the Greeble feel?
- Are the Blickets sad?

Test:

- In this story, Lisa took the Greeble's cookie and gave it to the squirrel. Should she have done that?
- Can you show me on the Pink Scale? Was that good/bad/just ok? A little good/bad or very good/bad?

Assault condition (Blicket ingroup majority example)

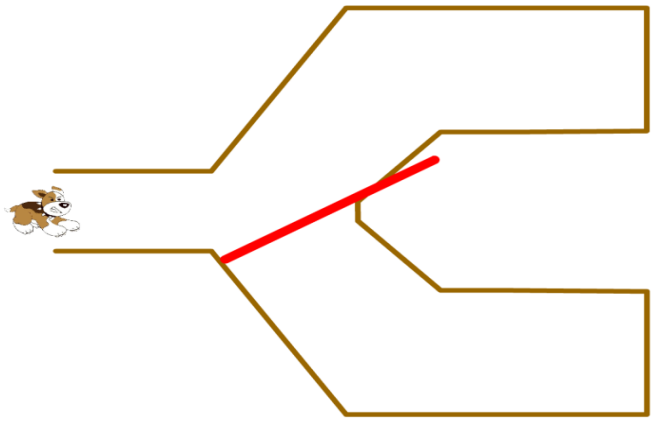

Figure D5. Introducing the gate. This is a mean angry dog. Right now, he has to decide whether to go up into that hallway, or down into that hallway. But look: there is a red gate there, so the dog can't go down that hallway, see? (*play animation of dog stopping at the gate*). So where do you think he is going to go instead? That's right. (*play animation*). The dog goes up there and barks a really mean scary bark!

(*Gate reappears blocking off the other hallway*). Now look, the gate is up there. If the gate is up there, where do you think the dog will go? That's right. (*play animation*). The dog goes down there and barks a really mean scary bark!

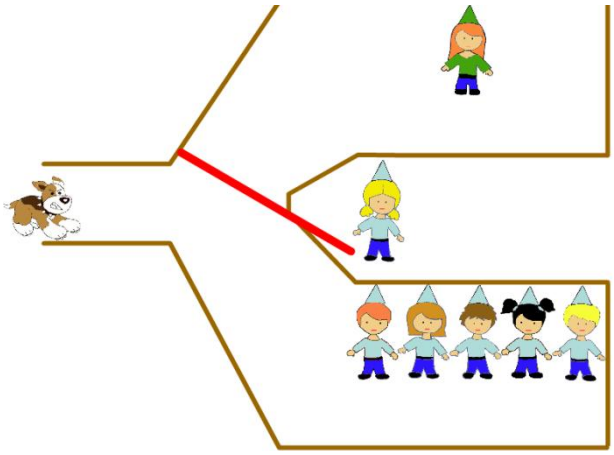

Figure D6. Introduction. This is Jane. Jane is a Blicket, just like you. And look! There are some other people in the story: there are lots of Blickets down here, and one Greeble up here. But uh oh, here comes that mean angry dog! And look: the red gate is up there, blocking that hallway. The dog can't go through that red gate, so the dog is going to go down and bark at the Blickets. If the dog barks at the Blickets, they will be very scared and sad.

Controls:

- Where will the dog go? [prompt: up or down?]
- And if the dog barks at the Blickets, how will they feel?

Well, Jane sees the angry dog. Jane knows that angry dog is going to go down and bark at the Blickets and make them scared and sad. Let's see what she does!

Play animation

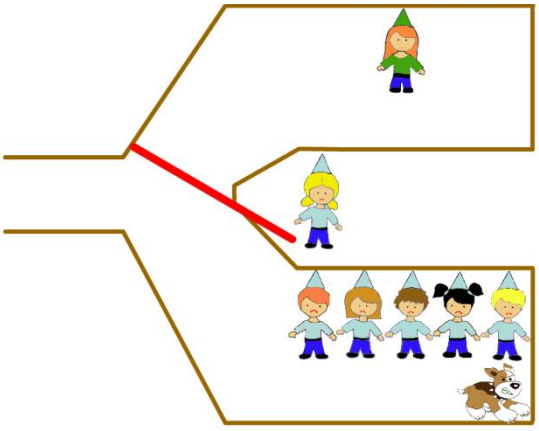

Figure D7. Omission dilemma. Well, she *could* do something, but Jane decides not to do anything. She just stands there. So the do goes down and barks at the Blickets and they are all scared and sad. But the Greeble isn't sad because the dog didn't bark at her. Let's watch that again (repeat from the beginning of the introduction).

Controls:

- At the beginning of the story, where was the dog going to go? [Prompt: up or down?]

- What did Jane do? [Prompt: did she do anything?]
- Where did the dog go?
- How do the Blickets feel?
- Is the Greeble sad?

Test:

- In this story, Jane didn't do anything. She just stood there. Should she have done that?
- Can you show me on the Pink Scale? Was that good/bad/just ok? A little good/bad or very good/bad?

Side effect dilemma

Now I'm going to tell you a story about Sally. (Repeat introduction).

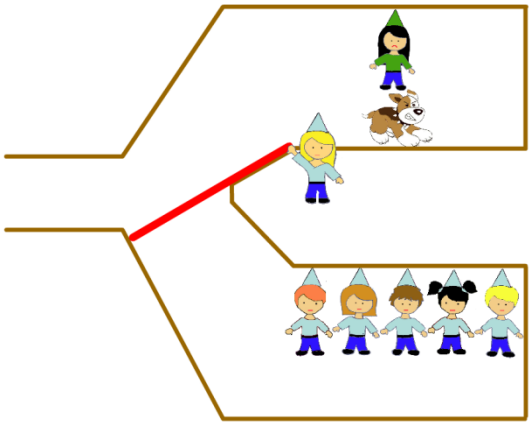

Figure D8. Side effect dilemma. Well, Sally decides to move the gate. Sally knows that if she moves the gate, the dog can't get to the Blickets, so he will go up and bark at the Greeble instead. Now the Greeble is scared and sad. But the Blickets aren't sad because the dog didn't bark at them. Let's watch that again. (Repeat from the beginning).

Controls:

- At the beginning of the story, where was the dog going to go? [Prompt: up or down?]
- What did Sally do?
- Where did the dog go?
- How does the Greeble feel?
- Are the Blickets sad?

Test:

- In this story, Sally moved the gate. Should she have done that?
- Can you show me on the Pink Scale? Was that good/bad/just ok? A little good/bad or very good/bad?

Main effect dilemma

Now I'm going to tell you a story about Lisa. (Repeat introduction).

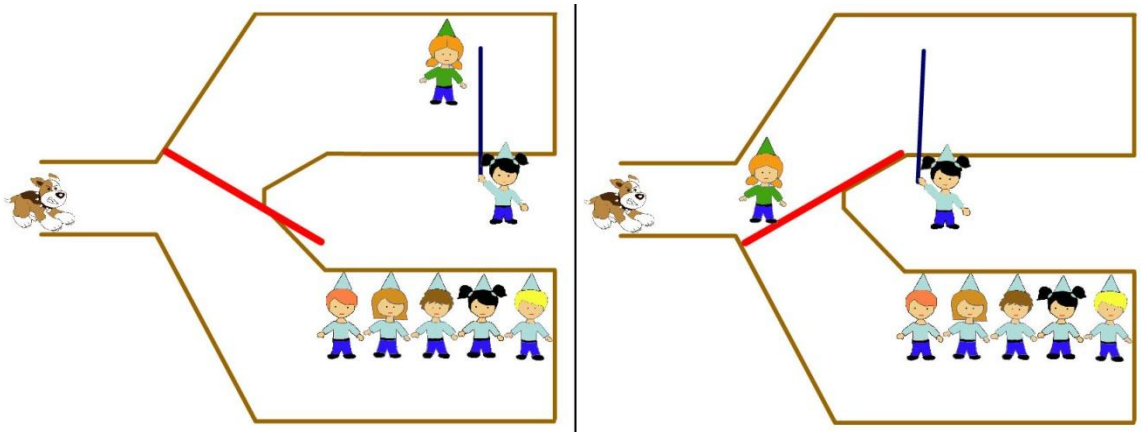

Figure D9. Main effect dilemma. Well, Lisa decides to use a blue stick to push the Greeble in front of the dog. Lisa knows that if she pushes the Greeble, the gate will close, and the dog will bark at the Greeble instead. Now the Greeble is scared and sad. But the Blickets aren't sad because the dog didn't bark at them. Let's watch that again. (Repeat from the beginning).

Controls:

- At the beginning of the story, where was the dog going to go? [Prompt: up or down?]
- What did Lisa do?
- Where did the dog go?
- How does the Greeble feel?
- Are the Blickets sad?

Test:

- In this story, Lisa used the blue stick. Should she have done that?
- Can you show me on the Pink Scale? Was that good/bad/just ok? A little good/bad or very good/bad?

Appendix E

Chapter 4 supplementary results

Table E1

Model effects for the full model (N = 213)

Dependent variable = Normative question

Model Effect	**Wald's χ2**	***df***	***p***
(Intercept)	.001	1	.974
Dilemma	38.128	1	.000
Group	4.041	1	.044
Harm	2.175	1	.140
AgeGroup	4.600	1	.032
Order	4.087	1	.043
Dilemma * Group	1.348	1	.246
Dilemma * Harm	7.860	1	.005
Dilemma * AgeGroup	1.101	1	.294
Dilemma * Order	4.342	1	.037
Group * Harm	.012	1	.914
Group * AgeGroup	.851	1	.356
Group * Order	.091	1	.763
Harm * AgeGroup	.107	1	.743
Harm * Order	1.354	1	.245
AgeGroup * Order	.147	1	.702
Dilemma * Group * Harm	.362	1	.548
Dilemma * Group * AgeGroup	2.037	1	.153
Dilemma * Harm * AgeGroup	.117	1	.733
Dilemma * Group * Order	.359	1	.549
Dilemma * Harm * Order	1.861	1	.173
Dilemma * AgeGroup * Order	4.185	1	.041
Group * Harm * AgeGroup	1.754	1	.185
Group * Harm * Order	.008	1	.929
Group * AgeGroup * Order	.201	1	.654
Harm * AgeGroup * Order	1.030	1	.310

Table E2

Parameter estimates for the main effects model (N = 213)

Dependent variable = Normative question

Predictor	**β**	***SE (β)***	**Wald's χ2**	***df***	***p***	**Exp(β)**
(Intercept)	-1.443	.2848	25.659	1	.000	.236
Dilemma (Side Effect)	1.057	.1709	38.285	1	.000	2.879
Dilemma (Main Effect)	----	----	----	----	----	1
MajorityGroup (Ingroup)	.433	.2353	3.389	1	.066	1.542

MajorityGroup (Outgroup)	----	----	----	----	----	1
Harm (Property)	.340	.2356	2.086	1	.149	1.405
Harm (Assault)	----	----	----	----	----	1
AgeGroup (3 years)	.522	.2368	4.857	1	.028	1.685
AgeGroup (4 & 5 years)	----	----	----	----	----	1
Order (OSM)	.522	.2351	4.921	1	.027	1.685
Order (OMS)	----	----	----	----	----	1

Main effects model: (Intercept), Dilemma, Group, Harm, AgeGroup, Order
Note. The last category of each predictor variable served as the reference category. For the dependent variable "Normative question", the reference category is “No.”

Table E3
Parameter estimates for the reduced model (N = 213)
Dependent variable = Normative question (yes)

Predictor	**β**	***SE* (β)**	**Wald's χ2**	***df***	***p***	**Exp(β)**
(Intercept)	-1.022	.2975	11.797	1	.001	.360
Dilemma (Side Effect)	.230	.2861	.647	1	.421	1.259
Dilemma (Main Effect)	----	----	----	----	----	1
MajorityGroup (Ingroup)	.431	.2391	3.254	1	.071	1.539
MajorityGroup (Outgroup)	----	----	----	----	----	1
Harm (Property)	-.097	.2892	.112	1	.738	.908
Harm (Assault)	----	----	----	----	----	1
AgeGroup (3 years)	.534	.2407	4.922	1	.027	1.706
AgeGroup (4 & 5 years)	----	----	----	----	----	1
Order (OSM)	.158	.2886	.300	1	.584	1.171
Order (OMS)	----	----	----	----	----	1
Dilemma (SE) * Harm (Assault)	.904	.3429	6.947	1	.008	2.469
Dilemma (SE) * Harm (Property)	----	----	----	----	----	1
Dilemma (ME) * Harm (Property)	----	----	----	----	----	1
Dilemma (ME) * Harm (Assault)	----	----	----	----	----	1
Dilemma (SE) * Order (OSM)	.763	.3434	4.934	1	.026	2.144
Dilemma (SE) * Order (OMS)	----	----	----	----	----	1
Dilemma (ME) * Order (OSM)	----	----	----	----	----	1
Dilemma (ME) * Order (OMS)	----	----	----	----	----	1

Final reduced model: (Intercept), Dilemma, MajorityGroup, Harm, AgeGroup, Order, Dilemma * Harm, Dilemma * Order
Note. The last level of each predictor variable served as the reference category. For the dependent variable “Normative question,” the reference category is “No.”

Appendix F

Chapter 5 Tables

Table F1

Chapter 5 study design: Independent Variables Combined to Form 16 conditions, and Number of Participants (N) in Each Condition

IV1: Agent	IV2: Majority	IV3: Harm	IV4: Dilemma Order	N
Ingroup Agent	Ingroup Majority	Property	Omission, Side Effect, Main Effect (OSM)	16
			Omission, Main Effect, Side Effect (OMS)	16
		Assault	OSM	16
			OMS	16
	Outgroup Majority	Property	OSM	16
			OMS	16
		Assault	OSM	16
			OMS	16
Outgroup Agent	Ingroup Majority	Property	OSM	14
			OMS	15
		Assault	OSM	14
			OMS	14
	Outgroup Majority	Property	OSM	16
			OMS	13
		Assault	OSM	14
			OMS	13

Double-effect tables

Table F2

Model effects for the full model (N = 237)
Dependent variable = Normative question

Model Effect	Wald's χ2	*df*	*p*	Model Effect	Wald's χ2	*df*	*p*
(Intercept)	24.901	1	.000	Dilemma * Agent * Majority	0.609	1	.435
Dilemma	63.935	1	.000	Dilemma * Agent * Harm	1.525	1	.217
Agent	0.23	1	.632	Dilemma * Agent * Order	0.095	1	.758
Majority	3.293	1	.070	Dilemma * Agent * Gender	0.28	1	.596
Harm	2.68	1	.102	Dilemma * Majority * Harm	0.54	1	.462
Order	8.169	1	.004	Dilemma * Majority * Gend	5.073	1	.024
Gender	4.739	1	.029	Dilemma * Harm * Order	1.208	1	.272
Dilemma * Agent	0.368	1	.544	Dilemma * Harm * Gender	1.631	1	.202
Dilemma * Maj	0.003	1	.957	Dilemma * Order * Gender	0.333	1	.564
Dilemma * Harm	3.409	1	.065	Agent * Majority * Harm	0.002	1	.965
Dilemma * Order	12.187	1	.000	Agent * Majority * Order	0.898	1	.343
Dilemma * Gend	0.752	1	.386	Agent * Majority * Gender	1.183	1	.277
Agent * Majority	0.76	1	.383	Agent * Harm * Order	0.676	1	.411
Agent * Harm	0.208	1	.648	Agent * Harm* Gender	0.024	1	.878

Agent * Order	0.776	1	.378	Agent * Order * Gender	0.221	1	.638
Agent * Gender	1.143	1	.285	Majority * Harm * Order	0.628	1	.428
Majority * Harm	0.04	1	.841	Majority * Harm * Gender	0.323	1	.570
Majority * Order	0.169	1	.681	Majority * Order * Gender	1.971	1	.160
Majority * Gender	0.21	1	.647	Harm * Order * Gender	3.988	1	.046
Harm * Order	0.168	1	.682				
Harm * Gender	1.255	1	.263				
Order * Gender	2.379	1	.123				

Table F3

Parameter estimates for the main effects model (N = 237)

Dependent variable = Normative question (yes)

Predictor	**β**	***SE (*β)**	**Wald's χ2**	***df***	***p***	**Exp(β)**
(Intercept)	-2.696	.438	37.966	1	.000	.067
Dilemma (Side Effect)	2.763	.239	134.100	1	.000	15.853
Dilemma (Main Effect)	----	----	----	----	----	1
AgentGroup (Ingroup)	.376	.254	2.188	1	.139	1.456
AgentGroup (Outgroup)	----	----	----	----	----	1
MajorityGroup (Ingroup)	-.363	.252	2.077	1	.150	.696
MajorityGroup (Outgroup)	----	----	----	----	----	1
Harm (Property)	.723	.263	7.519	1	.006	2.060
Harm (Assault)	----	----	----	----	----	1
Order (OSM)	-.279	.251	1.231	1	.267	.757
Order (OMS)	----	----	----	----	----	1
Gender (Male)	.578	.256	5.101	1	.024	1.783
Gender (Female)	----	----	----	----	----	1

Main effects model: (Intercept), Dilemma, Agent, Majority, Harm, Order, Gender

Note. The last category of each predictor variable served as the reference category. For the dependent variable "Normative question", the reference category is “No.”

Table F4

Parameter estimates for the reduced model (N = 237)

Dependent variable = Normative question (yes)

Parameter	**β**	***SE (*β)**	**Wald's χ2**	***df***	***p***	**Exp(β)**
(Intercept)	-1.484	0.4153	12.772	1	.000	0.227
Dilemma (Side Effect)	1.198	0.3043	15.513	1	.000	3.315
Dilemma (Main Effect)	----	----	----	----	----	1
AgentGroup (Ingroup)	0.361	0.2654	1.851	1	.174	1.435
AgentGroup (Outgroup)	----	----	----	----	----	1
MajorityGroup (Ingroup)	-0.423	0.2628	2.59	1	.108	0.655
MajorityGroup (Outgroup)	----	----	----	----	----	1
Harm (Property)	-0.425	0.4293	0.979	1	.322	0.654
Harm (Assault)	----	----	----	----	----	1

Order (OSM)	-2.033	0.5801	12.276	1	.000	0.131
Order (OMS)	----	----	----	----	----	1
Gender (Male)	0.578	0.2667	4.692	1	.030	1.782
Gender (Female)	----	----	----	----	----	1
Dilemma (SE) * Harm (Property)	1.507	0.4528	11.071	1	.001	4.512
Dilemma (SE) * Harm (Assault)	----	----	----	----	----	1
Dilemma (ME) * Harm (Property)	----	----	----	----	----	1
Dilemma (ME) * Harm (Assault)	----	----	----	----	----	1
Dilemma (SE) * Order (OSM)	2.203	0.5968	13.632	1	.000	9.056
Dilemma (SE) * Order (OMS)	----	----	----	----	----	1
Dilemma (ME) * Order (OSM)	----	----	----	----	----	1
Dilemma (ME) * Order (OMS)	----	----	----	----	----	1

Final reduced model: (Intercept), Dilemma, Agent, Majority, Gender, Dilemma*Harm, Dilemma*Order

Note. The last level of each predictor variable served as the reference category. For the dependent variable "Normative question," the reference category is "No."

Table F5

Model effects for the full model (N = 190)

Dependent variable = Purpose question

Model Effect	**Wald's χ2**	***df***	***p***	**Model Effect**	**Wald's χ2**	***df***	***p***
(Intercept)	.047	1	.829	Dilemma * Agent * Maj	.105	1	.746
Dilemma	36.709	1	.000	Dilemma * Agent * Harm	.128	1	.720
Agent	.194	1	.660	Dilemma * Agent * Order	.444	1	.505
Majority	2.055	1	.152	Dilemma * Agent * Gend	.725	1	.395
Harm	20.429	1	.000	Dilemma * Maj * Harm	2.133	1	.144
Order	1.932	1	.165	Dilemma * Maj * Order	5.110	1	.024
Gender	.305	1	.581	Dilemma * Maj * Gend	.513	1	.474
Dilemma * Agent	.435	1	.510	Dilemma * Harm * Order	1.191	1	.275
Dilemma * Maj	.469	1	.493	Dilemma * Harm * Gend	.004	1	.952
Dilemma * Harm	1.179	1	.278	Dilemma * Order * Gend	.049	1	.824
Dilemma * Order	10.529	1	.001	Agent * Majority * Harm	.991	1	.319
Dilemma * Gend	.914	1	.339	Agent * Majority * Order	.022	1	.883
Agent * Majority	1.890	1	.169	Agent * Majority * Gender	1.163	1	.281
Agent * Harm	.364	1	.546	Agent * Harm * Order	.993	1	.319
Agent * Order	2.511	1	.113	Agent * Harm* Gender	4.279	1	.039
Agent * Gender	.826	1	.364	Agent * Order * Gender	.001	1	.976
Majority * Harm	1.555	1	.212	Majority * Harm * Order	1.057	1	.304
Majority * Order	.270	1	.603	Majority * Harm * Gender	.776	1	.378
Majority * Gender	.152	1	.697	Majority * Order * Gender	.006	1	.940
Harm * Order	.885	1	.347	Harm * Order * Gender	1.787	1	.181
Harm * Gender	.953	1	.329				
Order * Gender	.037	1	.848				

Table F6
Parameter estimates for the main effects model (N = 190)
Dependent variable = Purpose question (yes)

Predictor	β	*SE (*β)	Wald's χ2	*df*	*p*	Exp(β)
(Intercept)	.746	.3510	4.521	1	.033	2.109
Dilemma (Side Effect)	-1.988	.2380	69.757	1	.000	.137
Dilemma (Main Effect)	----	----	----	----	----	1
AgentGroup (Ingroup)	-.032	.2672	.014	1	.906	.969
AgentGroup (Outgroup)	----	----	----	----	----	1
MajorityGroup (Ingroup)	.843	.2713	9.663	1	.002	2.324
MajorityGroup (Outgroup)	----	----	----	----	----	1
Harm (Property)	-1.568	.2941	28.429	1	.000	.208
Harm (Assault)	----	----	----	----	----	1
Order (OSM)	.713	.2709	6.930	1	.008	2.040
Order (OMS)	----	----	----	----	----	1
Gender (Male)	.085	.2718	.097	1	.756	1.088
Gender (Female)	----	----	----	----	----	1

Main effects model: (Intercept), Dilemma, AgentGroup, MajorityGroup, Harm, Order, Gender
Note. The last category of each predictor variable served as the reference category. For the dependent variable "Purpose question", the reference category is “No.”

Table F7
Parameter estimates for the reduced model (N =190)
Dependent variable = Purpose question

Parameter	**β**	***SE (*β)**	**Wald's χ2**	***df***	***p***	**Exp(β)**
(Intercept)	.415	.3241	1.640	1	.200	1.514
Dilemma(Side Effect)	-1.034	.2872	12.977	1	.000	.355
Dilemma(Main Effect)	----	----	----	----	----	1
MajorityGroup (Ingroup)	.890	.2901	9.409	1	.002	2.435
MajorityGroup (Outgroup)	----	----	----	----	----	1
Harm (Property)	-1.634	.3034	29.013	1	.000	.195
Harm (Assault)	----	----	----	----	----	1
Order (OSM)	1.572	.3336	22.196	1	.000	4.814
Order (OMS)	----	----	----	----	----	1
Dilemma(SE) * Order (OSM)	-2.029	.4395	21.314	1	.000	.131
Dilemma(SE) * Order (OMS)	----	----	----	----	----	1
Dilemma(ME) * Order (OSM)	----	----	----	----	----	1
Dilemma(ME) * Order (OMS)	----	----	----	----	----	1

Purpose reduced model: (Intercept), Dilemma, Harm, Majority, Order, Dilemma*Order
Note. The last category of each predictor variable served as the reference category. For the dependent variable “Purpose question,” the reference category is “No.”

Omission dilemma

Table F8
Model Effects for the full omission model (N = 190)
Dependent variable = Purpose question

Predictor	Wald's χ2	*df*	*p*
(Intercept)	40.488	1	.000
AgentGroup	.162	1	.687
MajorityGroup	5.626	1	.018
Harm	6.189	1	.013
Gender	6.200	1	.013
AgentGroup * MajorityGroup	.003	1	.959
AgentGroup * Harm	.058	1	.810
AgentGroup * Gender	.024	1	.877
MajorityGroup * Harm	1.301	1	.254
MajorityGroup * Gender	1.137	1	.286
Harm * Gender	.186	1	.666
AgentGroup * MajorityGroup * Harm	.146	1	.702

Dependent Variable: OnPurpose
Full omission model: (Intercept), AgentGroup, MajorityGroup, Harm, Gender, AgentGroup * MajorityGroup, AgentGroup * Harm, AgentGroup * Gender, MajorityGroup * Harm, MajorityGroup * Gender, Harm * Gender, AgentGroup * MajorityGroup * Harm
Note. The last category of each predictor variable served as the reference category. For the dependent variable "Purpose question," the reference category is "No."

Table F9
Goodness-of-fit statistics for the full omission model (N = 190)
Dependent variable = Purpose question

Test	Value	*df*	*p*
Overall model evaluation (against intercept-only model)			
Omnibus Likelihood ratio Chi-Square	20.329	11	.041
Goodness-of-fit test			
Deviance	3.861	3	
Pearson Chi-Square	3.143	3	
Model fitting criteria			
Log likelihood	-18.556		
Akaike's Information Criteria (AIC)	61.112		
Finite Sample Corrected AIC (AICC)	62.875		
Bayesian Information Criterion (BIC)	100.077		
Consistent AIC (CAIC)	112.077		

a. Information criteria are in small-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Table E10
Parameter estimates for the omission main effects model (N = 190)
Dependent variable = Purpose question (yes)

Predictor	β	*SE* (β)	Wald's χ2	*df*	*p*	Exp(β)

(Intercept)	-.740	.3790	3.811	1	.051	.477
AgentGroup (Ingroup)	-.310	.4095	.574	1	.449	.733
AgentGroup (Outgroup)	----	----	----	----	----	1
MajorityGroup (Ingroup)	-1.047	.4108	6.493	1	.011	.351
MajorityGroup (Outgroup)	----	----	----	----	----	1
Harm (Property)	-.923	.4006	5.308	1	.021	.397
Harm (Assault)	----	----	----	----	----	1
Gender (Male)	.899	.3954	5.164	1	.023	2.456
Gender (Female)	----	----	----	----	----	1

Reduced model: (Intercept), Dilemma, Agent, Majority, Harm, Order, Gender
Note. The last category of each predictor variable served as the reference category. For the dependent variable "Normative question", the reference category is “No.”

References

Aboud, F.E. (1988). *Children and prejudice.* Oxford, England: Blackwell.

Abrams, D., Rutland, A., Ferrell, J.M., & Pelletier, J. (2008). Children's judgments of disloyal and immoral peer behavior: Subjective group dynamics in minimal intergroup contexts. *Child Development, 79*(2), 444-461.

Abrams, D., Rutland, A., & Cameron, L. (2003). The Development of subjective group dynamics: Children's judgments of normative and deviant in-group and out-group individuals. *Child Development, 74*(6), 1840-1856.

Alexander, G.M., & Hines, M. (1994). Gender labels and play styles: Their relative contribution to children's selection of playmates. *Child Development, 65*(3), 869-879.

Anderson, S.W., Barrash, J., Bechara, A., & Tranel, D. (2006). Impairments of emotion and real-world complex behavior following childhood- or adult-onset damage to ventromedial prefrontal cortex. *Journal of the International Neuropsychological Society, 12*(2), 224–235.

Anderson, S.W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A.R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience, 2*(11), 1032-1037.

Aquinas, T. (1274/1988). Summa Theologica II-II, Q. 64, art. 7, 'Of killing'. In W.P. Baumgarth & R.J. Regan (Eds.), *On Law, Morality, and Politics* (pp. 226–227). Indianapolis/Cambridge: Hackett Publishing.

Ashburn-Nardo, L., Voils, C.I., & Monteith, M.J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81*(5), 789-799.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R.M. (2006). Nature and nurture in own-race face processing. *Psychological Science, 17*(2), 159-163.

Baron J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*(2), 74-85.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37-46.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A.R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex*, *6*(2), 215-225.

Beck, S.R., Robinson, E.J., Carroll, D.J., & Apperly, I.A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, *77*(2), 413-426.

Beggan, J.K. (1992). On the social nature of nonsocial perception: The mere ownership effect. *Journal of Personality and Social Psychology, 62*(2), 229-237.

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology, 41*(2), 328-337.

Berti, A.E., Bombi, A.S., & Lis, A. (1982). The child's conceptions about means of production and their owners. *European Journal of Social Psychology, 12*(3), 221-239.

Bigler, R.S., Jones, L.C., & Lobliner, D.B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development, 68*(3), 530-543.

Bigler, R.S., Spears Brown, C., & Markell, M. (2001). When groups are not created equal: Effects of group status on the formation of intergroup attitudes in children. *Child Development, 72*(4), 1151-1162.

Blair, R.J. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition, 57*(1), 1–29.

Blair, R.J. (1996). Brief report: Morality in the autistic child. *Journal of Autism and Developmental Disorders, 26*(5), 571-579.

Blair, R.J. (1997). Moral reasoning and the child with psychopathic tendencies. *Personality and Individual Differences, 22*(5), 731–739.

Blair, R.J., Jones, L., Clark, F., & Smith, M. (1995). Is the psychopath 'morally insane'? *Personality and Individual Differences*, *19*(5), 741-752.

Bleske-Rechek, A., Nelson, L., Baker, J., Remiker, M., & Brandt, S. (2010). Evolution and the trolley problem: People save five over one unless the one is young, genetically related, or a romantic partner. *Journal of Social, Evolutionary, and Cultural Psychology, 4*(3), 115-127.

Borg, J.S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience, 18*(5), 803-817.

Boseovski, J.J. (2010). Evidence for "Rose-colored glasses": An examination of the positivity bias in young children's personality judgments. *Child Development Perspectives, 4*(3), 212-218.

Boseovski, J.J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology, 42*(3), 500-513.

Boseovski, J.J., & Lee, K. (2008). Seeing the world through rose-colored glasses? Neglect of consensus information in young children's personality judgments. *Social Development, 17*(2), 399-416.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., & Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624-652.

Bybee, J.L., & Fleischman, S. (Eds.). (1995). *Modality in grammar and discourse*. Amsterdam: John Benjamins.

Carey, S. (1988). Conceptual differences between children and adults. *Mind & Language, 3*(3), 167-181.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*(2), 315-330.

Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve-and 18-month-olds copy actions in terms of goals. *Developmental Science*, *8*(1), F13-F20.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1986). *Knowledge of language: Its nature, origins, and use.* New York, NY: Praeger.

Cikara, M., Farnsworth, R.A., Harris, L.T., & Fiske, S.T. (2010). On the wrong side of the trolley track: Neural correlates of relative social valuation. *Social Cognitive and Affective Neuroscience, 5*(4), 404-413.

Cima, M., Tonnaer, F., & Hauser, M.D. (2010). Psychopaths know right from wrong but don't care. *Social Cognitive and Affective Neuroscience, 5*(1), 59-67.

Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment, Vol. 2: Standard issue scoring manual*. New York, NY: Cambridge University Press.

Csibra, G., & Gergely G. (1998). The teleological origins of mentalistic action explanation: A developmental hypothesis. *Developmental Science, 1*(2), 255-259.

Csibra, G., B ró, S., Koos, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27*(1), 111-133.

Csibra, G., Gergely, G., B r , S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, *72*(3), 237-267.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuitions in moral judgment: Testing three principles of harm. *Psychological Science, 17*(12)*,* 1082-1089.

Damasio, A. (1994). *Descartes' error*. Boston, MA: Norton.

Damasio, A.R., Tranel, D., & Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience, 13*(1), 89–109.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology, 38*(3), 352-362.

Dunham, Y., Baron, A.S., & Banaji, M.R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, *12*(7), 248-253.

Dunham, Y., Baron, A.S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development, 82*(3), 793-811.

Dupoux, E., & Jacob, P. (2007). Universal moral grammar: A critical appraisal. *Trends in Cognitive Sciences, 11*(9), 373-378.

Dwyer, S. (1999). Moral competence. In K. Murasugi & R. Stainton (Eds.), *Philosophy and Linguistics* (pp. 169-190). Boulder, CO: Westview Press.

Dwyer, S. (2006). How good is the linguistic analogy? In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind, volume 2: Culture and cognition* (pp. 237-256). New York, NY: Oxford University Press.

Eibl-Eibesfeldt, I. (1970). *Ethology: The biology of behavior*. Oxford, England: Holt, Rinehart, & Winston.

Fallon, A.E., Rozin, P., & Pliner, P. (1984). The child's conception of food: The development of food rejections with special reference to disgust and contamination sensitivity. *Child development*, 55(2), 566-575.

Fischer, J.M., & Ravizza, M. (1992). *Ethics: Problems and principles*. New York, NY: Holt, Rinehart & Winston.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5*, 5-15.

French, D.C. (1984). Childreńs knowledge of the social functions of younger, older, and same-age peers. *Child Development, 55*(4)*,* 1429-1433.

Neary, K.R., Friedman, O., & Burnstein, C.L. (2009). Preschoolers infer ownership from "control of permission". *Developmental Psychology*, *45*(3), 873-876.

Gergely, G., Bekkering, H., & Király, I. (2002). Developmental psychology: rational imitation in preverbal infants. *Nature*, *415*(6873), 755-755.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287-292.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165-193.

Grant, M.G., & Mills, C.M. (2011). Children's explanations of the intentions underlying others' behaviour. *British Journal of Developmental Psychology, 29*(3), 504-523.

Greene J.D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*(12), 517-523.

Greene, J.D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology, 45*(3), 581-584.

Greene, J.D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*(3), 364-371.

Greene, J.D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322-323.

Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E., & Cohen, J.D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144-1154.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., & Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389-400.

Greene, J.D., Sommerville, B.R., Nystrom, L.E., Darley, J.M., & Cohen, J.D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*(5537), 2105–2108.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4)*,* 814-834.

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852-870). Oxford, England: Oxford University Press.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20*(1)*,* 98-116.

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stitch (Eds.), *The innate mind*, *vol. 3* (pp. 367-392). New York, NY: Oxford University Press.

Haidt, J., Rozin, P., McCauley, C., & Imada, S. (1997). Body, psyche, and culture: The relationship of disgust to morality.*Psychology and Developing Societies, 9*, 107-131.

Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences, 16*(5), 701-713.

Haidt, J., Rozin, P., McCauley, C., & Imada, S. (1997). Body, psyche, and culture: The relationship between disgust and morality. *Psychology & Developing Societies, 9*(1)*,* 107-131.

Hamilton, W.D. (1964). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology, 7*(1), 17-52.

Hamlin, K.J., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not like me = bad: Infants prefer those who harm dissimilar others. *Psychological Science, 24*(4), 589-594.

Hamlin, J.K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development, 26*(1), 30-39.

Hamlin, J.K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences, 108*(50), 19931-19936.

Hamlin, J.K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*(7169), 557-560.

Hamlin, J.K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science, 13*(6)*,* 923-929.

Hare, R.D., & Quinn, M.J. (1971). Psychopathy and autonomic conditioning. *Journal of Abnormal Psychology, 77*(3), 223-235.

Hauser, M. (2006). *Moral minds: The unconscious voice of right and wrong*. New York, NY: Harper Collins.

Hauser, M., Cushman, F., Young, L., Jin, R.K., & Mikhail, J. (2007a). A dissociation between moral judgment and justification. *Mind & Language, 22*(1), 1-21.

Hauser, M.D., Young, L., & Cushman, F.A. (2007b). Reviving Rawls' linguistic analogy. In W. Sinnott-Armstrong (Ed.), *Moral psychology and biology.* New York, NY: Oxford University Press.

Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: What is it, who has, it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Hebble, P.W. (1971). The development of elementary school children's judgment of intent. *Child Development, 42*(4), 1203-1215.

Heyman, G.D., & Giles, J.W. (2004). Valence effects in reasoning about evaluative traits. *Merrill-Palmer Quarterly, 50*(1), 86-109.

Hobson, R.P. (1993). The emotional origins of social understanding. *Philosophical Psychology*, *6*(3), 227-249.

Hoffman, M. (2000). *Empathy and moral development: Implications for caring and justice*. New York, NY: Cambridge University Press.

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences, 13*(1), 1-6.

Huebner, B., Lee, J.J., & Hauser, M.D. (2010). The moral-conventional distinction in mature moral competence. *Journal of Cognition and Culture, 10*(1-2), 1-26.

Hume, D. (1739-1740/1978). *A treatise of human nature*. (P.H. Nidditch, Ed.). Oxford: Clarendon Press.

Johnson, S.C., Booth, A., & O'Hearn, K. (2001). Inferring the goals of a nonhuman agent. *Cognitive Development, 16*(1), 637-656.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman

(Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). Cambridge, England: Cambridge University Press.

Kamm, F.M. (1998a). Moral intuitions, cognitive psychology, and the harming-versus-not-aiding distinction. *Ethics, 108*(3), 463-488.

Kant (1785/1964). Groundwork of the metaphysics of morals. (H.J. Paton, Trans.). New York, NY: Harper Perennial.

Karniol, R. (1978). Children's use of intention cues in evaluating behavior. *Psychological Bulletin, 85*(1), 76-85.

Katz, P.A., & Kofkin, J.A. (1997). Race, gender, and young children. In S.S Luthar, J.A. Burack, D. Cicchetti, & J.R. Weisz (Eds.), *Developmental psychopathology: Perspectives on adjustment, risk, and disorder* (pp. 51-74). New York, NY: Cambridge University press.

Kelly, D.J., Quinn, P.C., Slater, A.M., Lee, K., Gibson, A., Smith, M., & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science, 8*(6), F31-F36.

Kiehl, K.A., Smith, A.M., Hare, R.D., Mendrek, A., Forster, B.B., Brink, J., & Liddle, P.F. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological Psychiatry, 50*(9), 677-684.

Killen, M., Lee-Kim, J., McGlothlin, H., & Stangor, C. (2002). How children and adolescents evaluate gender and racial exclusion. *Monographs of the Society for Research in Child Development*, *67*(4), i-vii, 1-119.

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

Killen, M., Pisacane, K., Lee-Kim, J., & Ardila-Rey, A. (2001). Fairness or stereotypes? Young children's priorities when evaluating group exclusion and inclusion. *Developmental Psychology, 37*(5), 587-596.

Kinzler, K.D., Dupoux, E., & Spelke, E.S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences, 104*(30), 12577-12580.

Kinzler, K.D., & Spelke, E.S. (2011). Do infants show social preferences for people differing in race? *Cognition, 119*(1), 1-9.

Kircher, M., & Furby, L. (1971). Racial preferences in young children. *Child Development, 42*(6), 2076-2078.

Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology, 16*(2), 309-324.

Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the Ultimatum Game. *The Journal of Neuroscience, 27*(4), 951–956.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, H., Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature, 446*(7138), 908–911.

Kohlberg, L. (1981). *Essays on moral development, volume 1: The philosophy of moral development.* New York, NY: Harper Row.

Kohlberg, L. (1984). *The psychology of moral development: The nature and validity of moral stages*. New York, NY: Harper & Row.

Kowalski, K., & Lo, Y.F. (2001). The influence of perceptual features, ethnic labels, and sociocultural information on the development of ethnic/racial bias in young children. *Journal of Cross-Cultural Psychology, 32*(4), 444-455.

Kuhlmeier, V., Wynn K., & Bloom P. (2003). Attribution of dispositional states in 12-month-olds. *Psychological Science, 14*(5), 402-408.

Lagattuta, K.H., & Sayfan, L. (2013). Not all past events are equal: Biased attention and emerging heuristics in children's past-to-future forecasting. *Child Development*, *84*(6), 2094-2111.

Lerner, J. S., Small, D.A., & Loewenstein, G. (2004). Heart strings and purse strings carryover effects of emotions on economic decisions. *Psychological Science, 15*(5), 337-341.

Leslie, A.M. (1987). Pretense and representation: The origins of" theory of mind." *Psychological review, 94*(4), 412.

Leslie, A.M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. Hirschfeld and S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119-148). New York, NY: Cambridge University Press.

Leslie, A.M., Knobe J., & Cohen A. (2006a). Acting intentionally and the side-effect effect: 'Theory of mind' and moral judgment. *Psychological Science, 17*(5), 422-428.

Leslie, A.M., Mallon, R., & DiCorcia, J.A. (2006b). Transgressors, victims, and cry babies: Is basic moral judgment spared in autism? *Social Neuroscience, 1*(3-4), 270-283.

Levine, S. and Leslie, A. (2013, October). Inferring intention in double-effect scenarios. Poster presented at the meeting of the Cognitive Development Society, Memphis, TN.

Levine, S., Leslie, A., and Mikhail, J. (2013, January). Meta-representations of moral action: Support for the universal moral grammar thesis. Poster presented at the Budapest Central European University Conference on Cognitive Development, Budapest, Hungary.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12-and 18-month-olds point to provide information for others. *Journal of Cognition and Development, 7*(2), 173-187.

Lockhart, K.L., Chang, B., & Story, T. (2002). Young children's beliefs about the stability of traits: Protective optimism? *Child Development, 73*(5), 1408-1430.

Locksley, A., Ortiz, V., & Hepburn, C. (1980). Social categorization and discriminatory behavior: Extinguishing the minimal intergroup discrimination effect. *Journal of Personality and Social Psychology, 39*(5), 773-783.

Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science, 33*(2), 273-286.

Lorenz, K. (1966). The role of gestalt perception in animal and human behavior. In L. White (Ed.), *Aspects of Form* (pp. 157-178). Bloomington, IN: Indiana University Press.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science, 16*(8), 601-608.

Maccoby, E.E., & Jacklin, C.N. (1987). Gender segregation in childhood. In H.W. Reese (Ed.), *Advances in child development and behavior, Vol. 20* (pp. 239-287). San Diego, CA: Academic Press.

McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology, 45*(3), 577-580.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental psychology, 31*(5), 838-850.

Mendez, M.F., Anderson, E., & Shapira, J.S. (2005). An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology, 18*(4), 193–197.

Mikhail, J. (2000). Rawls' linguistic analogy: A study of the 'generative grammar' model of moral theory described by John Rawls in *A theory of justice.* Ph.D. dissertation, Cornell University.

Mikhail, J. (2002). Aspects of the theory of moral cognition: Investigating intuitive knowledge of the prohibition of intentional battery and the principle of double effect. *Georgetown University Law Center Public Law & Legal Theory Working Paper No. 762385.* Available at: http://ssrn.com/abstract=762385.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Science, 11*(4), 143-152.

Mikhail, J. (2008). Moral cognition and computational theory. In W. Sinnott-Armstrong (Ed.), *Moral psychology, vol. 3: The neuroscience of morality: Emotion, brain disorders, and development* (pp. 81-91), Cambridge, MA: MIT Press.

Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment.* New York, NY: Cambridge University Press.

Mikhail, J., Sorrento, C., & Spelke, E. (1998). Toward a universal moral grammar. In M.A. Gernsbacher & S.J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (p. 1250). Mahwah, NJ: Lawrence Erlbaum Associates.

Miller, E.K., & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24(1),* 167-202.

Miller, W.I. (1997). *The anatomy of disgust*. Cambridge, MA: Harvard University Press.

Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions, and the utilitarian brain. *Trends in Cognitive Sciences, 11*(8), 319–321.

Moll, J., de Oliveira-Souza, R., Eslinger, P.J., Bramati, I.E., Mourão-Miranda, J., Andreiuolo, P.A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience*, *22*(7), 2730-2736.

Moll, J., de Oliveira-Souza, R., Moll, F.T., Ignacio, F.A., Bramati, I.E. Caparelli-Daquer, E.M., & Eslinger, P.J. (2005). The moral affiliations of disgust: A functional MRI study. *Cognitive and Behavioral Neurology, 18*(1), 68–78.

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology, 22*(2), 103-122.

Neary, K.R., Friedman, O., & Burnstein, C.L. (2009). Preschoolers infer ownership from "control of permission". *Developmental Psychology, 45*(3), 873-876.

Nelson, S.A. (1980). Factors influencing young children's use of motives and outcomes as moral criteria. *Child Development, 51*(3), 823-829.

Nelson-Le Gall, S.A. (1985). Motive-outcome matching and outcome foreseeability: Effects on attribution of intentionality and moral judgments. *Developmental Psychology, 21*(2), 332-337.

Nesdale, D., Durkin, K., Maass, A., & Griffiths, J. (2004). Group status, outgroup ethnicity and children's ethnic attitudes. *Journal of Applied Developmental Psychology, 25*(2), 237-251.

Nesdale, D., & Flesser, D. (2001). Social identity and the development of children's group attitudes. *Child Development, 72*(2), 506-517.

Nichols, S. (2002). How psychopaths threaten moral rationalism: Is it irrational to be amoral? *The Monist, 85*(2), 285–303.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment.* New York, NY: Oxford University Press.

Nucci, L.P. (2001). *Education in the moral domain.* Cambridge, England: Cambridge University Press.

Nucci, L.P., & Nucci, M.S. (1982). Children's social interactions in the context of moral and conventional transgressions. *Child Development, 53*(2), 403-412.

Nucci, L.P., Turiel, E., & Encarnacion-Gawrych, G. (1983). Children's social interactions and social concepts analyses of morality and convention in the Virgin Islands. *Journal of Cross-Cultural Psychology, 14*(4), 469-487.

Nucci, L.P., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development, 49*(2), 400-407.

Nucci, L., & Turiel, E. (1993). God's word, religious rules, and their relation to Christian and Jewish children's concepts of morality. *Child Development, 64*(5), 1475-1491.

Nucci, L., & Weber, E.K. (1995). Social interactions in the home and the development of young children's conceptions of the personal. *Child Development, 66*(5), 1438-1452.

Nunez, M., & Harris, P.L. (1998). Psychological and deontic concepts: Separate domains or intimate connection? *Mind & Language, 13*(2), 153-170.

O'Neill, P., & Petrinovich L. (1998). A preliminary cross cultural study of moral intuitions. *Evolution and Human Behavior, 19*(6), 349-367.

Onishi, K.H., & Baillargeon R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255-257.

Onishi, K.H., Baillargeon R., & Leslie, A.M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica, 124*(1), 106-128.

Otten, S., & Wentura, D. (1999). About the impact of automaticity in the Minimal Group Paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology, 29*(8), 1049-1071.

Patterson, M.M., & Bigler, R.S. (2006). Preschool children's attention to environmental messages about groups: Social categorization and the origins of intergroup bias. *Child Development, 77*(4), 847-860.

Pellizzoni, S., Siegal, M., & Surian, L. (2009). Foreknowledge, caring, and the side-effect effect in young children. *Developmental Psychology, 45*(1), 289-295.

Pellizzoni, S., Siegal, M., & Surian, L. (2010). The contact principle and utilitarian moral judgments in young children. *Developmental Science, 13*(2), 265-270.

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology, 17*(3), 145-171.

Petrinovich, L., O'Neill, P., & Jorgensen, M.J. (1993). An empirical study of moral intuitions: Towards an evolutionary ethics. *Ethology and Sociobiology, 64*(3), 467-478.

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

Pizarro, D.A., Uhlmann E., & Bloom P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology, 39*(6), 653-660.

Premack, D., & Premack, A.J. (1997). Infants attribute value+/- to the goal-directed actions of self-propelled objects. *Journal of Cognitive Neuroscience, 9*, 848-856.

Prinz, J. (2007). *The emotional construction of morals*. Oxford, England: Oxford University Press.

Prinz, J. J. (2004). *Gut reactions: A perceptual theory of emotion*. New York, NJ: Oxford University Press.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Quinn, W.S. (1989). Actions, intentions, and consequences: The doctrine of double effect. *Philosophy & Public Affairs, 18*(4), 334-351.

Rachels, J. (1975). Active and passive euthanasia. *New England Journal of Medicine*, *292*(2), 78-80.

Rhodes, M. (2012). Naïve theories of social groups. *Child Development, 83*(6), 1900-1916.

Robinson, E.J., & Beck, S. (2000). What is difficult about counterfactual reasoning? In P. Mitchell, & K.J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 101-119). Hove, England: Psychology Press/Taylor & Francis (UK).

Rossano, F., Rakoczy, H., & Tomasello, M. (2011). Young children's understanding of violations of property rights. *Cognition, 121*(2), 219-227.

Roth, D., & Leslie, A.M. (1998). Solving belief problems: Toward a task analysis. *Cognition, 66*(1), 1-31.

Royzman, E.B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research, 15*(2), 165-184.

Rozin, P. (1997). Moralization and becoming a vegetarian. *Psychological Science, 8*(2), 67–73

Rozin, P., Fallon, A., & Augustoni-Ziskind, M. (1985). The child's conception of food: The development of contamination sensitivity to" disgusting" substances. *Developmental Psychology*, *21*(6), 1075-1079.

Rozin, P., & Fallon, A. (1987). A perspective on disgust. *Psychological Review, 94*(1), 23-41.

Rozin, P., Haidt, J., & McCauley, C.R. (2008). Disgust. In M. Lewis & J. Haviland (Eds.), *Handbook of emotions, 3rd Edition* (pp. 757-776). New York, NY: Guilford Press.

Rozin, P., Hammer, L., Oster, H., Horowitz, T., & Marmora, V. (1986). The child's conception of food: differentiation of categories of rejected substances in the 16 months to 5 year age range. *Appetite*, *7*(2), 141-151.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*(4), 574-586.

Saver, J.L., & Damasio, A.R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia, 29*(12), 1241-1249.

Schnall, S., Haidt, J., Clore, G.L., & Jordan, A.H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin, 34*(8), 1096–1109.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language, 27*(2), 135-153.

Shultz, T., Wright, K., & Schleifer, M. (1986). Assignment of moral responsibility and punishment. *Child Development, 57*(1), 177-184.

Siegal, M., & Peterson, C.C. (1998). Preschoolers' understanding of lies and innocent and negligent mistakes. *Developmental Psychology, 34*(2), 332-341.

Sinnott-Armstrong, W. (2008). Framing moral intuitions. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 2: The cognitive science of morality: intuition and diversity* (pp. 47-76). Cambridge, MA: MIT Press.

Smetana, J. G., & Braeges, J.L. (1990). The development of toddlers' moral and conventional judgments. *Merrill-Palmer Quarterly*, 36(3), 329-346.

Smetana, J.G. (1981). Preschool children's conceptions of moral and social rules. *Child Development, 52*(4), 1333-1336.

Smetana, J.G. (1983). Social cognitive development: Domain distinctions and coordinations. *Developmental Review, 3*(2), 131-147.

Smetana, J.G. (1985). Preschool children's conceptions of transgressions: Effects of varying moral and conventional domain-related attributes. *Developmental Psychology, 21*(1), 18-29.

Smetana, J.G. (1995). Morality in context: Abstractions, ambiguities and applications. In R. Vasta (Ed), *Annals of child development: A research annual, Vol. 10* (pp. 83-130). London, England: Jessica Kingsley Publishers.

Southgate, V., Senju A., & Csibra G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*(7), 587-592.

Spranca, M., Minsk, E., Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27*(1), 76-105.

Sunstein, C.R. (2005). Moral heuristics. *Behavioral and Brain Sciences, 28*(4)*,* 531-541.

Swann, W.B., Gómez, Á., Dovidio, J.F., Hart, S., & Jetten, J. (2010). Dying and killing for one's group identity fusion moderates responses to intergroup versions of the trolley problem. *Psychological Science, 21*(8), 1176-1183.

Tajfel, H., & (1971/2001). Experiments in intergroup discrimination. In M. A. Hogg & D. Abrams (Eds.), *Intergroup relations: Essential readings* (pp. 178–187). London: Psychology Press.

Tajfel, H., & Turner, J.C. (2004). The social identity theory of intergroup behavior. In J. T. Jost & J. Sidanius (Eds.), *Political psychology: Key readings* (pp. 276–293). London: Psychology Press.

Thomson, J.J. (1971). Individuating actions. *Journal of Philosophy, 68*(21), 774-781.

Thomson, J.J. (1985). Double effect, triple effect and the trolley problem: Squaring the circle in looping cases. *Yale Law Journal, 94*(6)*,* 1395-1415.

Turiel, E. (1978). Social regulations and domains of social concepts. In W. Damon (Ed.), *New directions in child development, volume 1, social cognition* (pp. 45-74). San Francisco, CA: Jossey-Bass.

Turiel, E. (1983). *The development of social knowledge: Morality and convention.* Cambridge, England: Cambridge University Press.

Uhlmann, E.L., Pizarro, D.A., Tannenbaum, D., & Ditto, P.H. (2009). The motivated use of moral principles. *Judgment and Decision Making, 4*(6), 476-491.

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental Psychology, 45*(2), 534-543.

Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young children selectively avoid helping people with harmful intentions. *Child Development, 81*(6), 1661-1669.

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*(6), 476–477

Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy social groups and the flexibility of virtue. *Psychological Science, 18*(8), 689-690.

Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology, 44*(5), 1334-1338.

Wellman, H.M., Larkey, C., & Somerville, S.C. (1979). The early development of moral criteria. *Child Development, 50*(3), 879-873.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*(3), 271-294.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science, 16*(10), 780–784.

Wiegmann, A., Okan, Y., & Nagel, J. (2012). Order effects in moral judgment. *Philosophical Psychology, 25*(6), 813-836.

Wimmer, H., Hogrefe, G.J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. *Child Development*, 59(2), 386-396.

Woodward, A.L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*(1), 1-34.

Woodward, A.L. (1999). Infants' ability to distinguish between purposeful and nonpurposeful behaviors. *Infant Behavior and Development, 22*(2), 145-160.

Woodward, A.L., & Sommerville, J.A. (2000). Twelve-month-old infants interpret action in context. *Psychological Science*, *11*(1), 73-77.

Yuill, N. (1984). Young children's coordination of motive and outcome in judgements of satisfaction and morality. *British Journal of Developmental Psychology, 2*(1), 73-81.

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental Psychology, 24*(3), 358-365.

Zelazo, P.D., Helwig C.C., & Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development, 67*(5), 2478-2492.