THE PRINCIPLE OF DOUBLE EFFECT AND PRESCHOOLERS' MORAL INTUITIONS

by

KATHARINE SAUNDERS

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Psychology

written under the direction of

Alan Leslie

and approved by

_____

_____

_____

New Brunswick, New Jersey

October, 2010

ABSTRACT OF THE THESIS

The Principle of Double Effect and Preschoolers' Moral Intuitions

By KATHARINE SAUNDERS

Thesis Director:
Alan Leslie

A thought experiment known as the trolley problem has frequently been used to examine whether adult moral judgments follow utilitarian principles; for example, maximize happiness for the greatest number. Previous research has shown that adults find it morally permissible to harm one person to save five others *only* when the harm to one is a *foreseen side effect* of saving five, but not when the harm to one is deliberately intended as the *means* to saving five. However, when asked to justify this moral intuition, few adults can articulate the principles guiding their judgment, suggesting that our moral judgments may be rooted in unconscious processes that evaluate the causal and intentional properties of an event. The current study tested 54 children on age-appropriate trolley dilemmas to examine whether young children make this distinction between harm as side effect and harm as a main effect. Our research indicates that like adults, preschoolers favor a utilitarian outcome when harming one person is simply a foreseen side effect of saving five others, but not when harming one person is the main effect of saving five.

Acknowledgements

I would like to thank my advisor, Dr. Alan Leslie, and my committee members, Dr. Rochel Gelman and Dr. Gretchen Chapman for their help and support. I would also like to thank Melissa Kibbe, Deena Weisberg, Lu Wang, Anton Scherbakov, and all the members of the Cognitive Development Lab for their invaluable support and assistance in completing the research presented in this thesis.

Table of Contents

List of Figures

Illustrations

Figures

Tables

Introduction

How do we decide whether an action is right or wrong? Although this is a relatively new empirical question in developmental psychology, it has been a theoretical question of central importance to philosophers for centuries. Advocates of utilitarian ethics propose that the moral worth of an action should be determined by evaluating the costs and benefits of its outcome; morally acceptable actions are those that produce the greatest good for the greatest number of people. In this view, an action that results in harm to another may be considered morally acceptable as long as it is in the interest of the greater good. Deontologists, on the other hand, argue that some actions can be considered morally wrong, even when those actions maximize happiness for the greatest number of people.

A thought experiment known as the trolley problem has frequently been used to examine whether our judgments are actually consistent with either of these ethical platforms (e.g. Foot, 1967; Thomson, 1970). The "bystander" version of the trolley problem presents a bystander with a moral dilemma: five people are chained to a train track, and a runaway trolley is about to hit them. If the bystander does nothing, the five people will be killed (i.e. harm by omission). Alternatively, he can divert the trolley onto a side track where one person is standing. If he diverts the trolley onto the side track, one person will die, but five people will be saved. Recent research has found that people find it morally permissible (and even obligatory) to divert the trolley to the side track, sacrificing one person to save five (Cushman et al., 2006; Green, et al., 2001; Hauser, et al., 2006).

However, when presented with the "footbridge" variation of this moral dilemma, people show the reverse judgment. In the footbridge dilemma, the bystander is standing on a footbridge overlooking the train track, with a fat man standing next to him. In order to stop the train from killing the five people, the bystander must push the fat man off the footbridge and in front of the oncoming train. The weight of the fat man will stop the train, saving the five people but killing the fat man. Although this results in the same tradeoff as in the trolley problem – sacrificing one to save five – adults judge this scenario as impermissible (Cushman et al., 2006; Green, et al., 2001; Hauser, et al., 2006).

The fact that people make a moral distinction between these two cases suggests that our judgments do not conform to strict utilitarian or consequentialist principles. Instead, it has been proposed that our moral judgments align with the *principle of double effect*: a harmful act for the greater good is morally permissible only when the harm is a foreseen side effect of intervening for the greater good, but not when it is deliberately intended as the means to the greater good (e.g. Fischer and Ravizza, 1992; Hauser et al., 2007a; Kamm, 1998b; Mikail, 2000; Thomson, 1970). In other words, we make a moral distinction between harm as a *side effect* and harm as a *main effect*.

Recent studies have attempted to isolate potential differences between these two trolley dilemmas, as well as vary the content of each story (Cushman et al., 2006; Hauser et al., 2007a). Even when factors such as physical contact with the victim and the degree to which the harm was caused by action or omission are controlled, adults show surprising consistency in their responses to these types of scenarios, regardless of gender, age, education level, ethnicity, religion, and culture (to the extent that these principles

have been tested cross-culturally) (Cushman et al., 2006; Green et al., 2001, 2004; Hauser, 2006; Hauser et al., 2006; O'Neill & Petrinovich, 1998; Petrinovich et al., 1993). Furthermore, when asked to justify their moral distinction between these two scenarios, adults are unable to articulate the principle of double effect, suggesting that although this principle guides their judgments, it is not accessible to conscious reasoning[1] (Cushman et al., 2006; Hauser et al., 2007a).

Development of moral cognition

These findings have some interesting implications for the development of moral cognition. For decades, it was assumed that moral judgments were the result of careful deliberation over consciously held moral principles. Thus, classic studies of children's moral cognition, such as those conducted by Jean Piaget and Lawrence Kohlberg, focused primarily on the content of children's moral justifications rather than on their moral judgments. These researchers concluded that children's moral reasoning was considered inferior to that of adults because children's justifications of their moral intuitions were less sophisticated (Kohlberg, 1981, 1984; Piaget, 1932/1965). However, in light of the recent research on adult moral intuitions, insufficient moral justifications may not necessarily reflect moral incompetence. Furthermore, the finding that adults are frequently unable to articulate the principles guiding their moral judgments (Cushman et al., 2006; Haidt, 2001; Hauser et al., 2007a) suggests that adults' moral judgments are not necessarily the result of enculturation; instead, our moral judgments may be rooted in unconscious processes that evaluate the causal and intentional properties of an event. If this is true, the building blocks for moral cognition may be present even in infancy.

[1] Note, I use the term "reasoning" as cognitive scientists use it, to refer to any inference process, conscious or unconscious.

Although Piaget (1932/1965) initially claimed that only children over the age of seven can interpret others' behavior in terms of intentions, more recent evidence suggests that infants in the first year of life already consider an action as goal-directed (e.g. Csibra & Gergely, 1998; Csibra et al., 1999, 2003; Gergely & Csibra, 2003; Leslie, 1994; Premack, 1990; Premack & Premack, 1997; Woodward, 1998) and can distinguish intentional actions from accidental or unsuccessful actions (Behne et al., 2005; Gergely et al., 1995; Woodward, 1999). An emerging body of evidence also suggests that as young as 15-months, infants can attribute mental states such as knowing, believing, and pretending to an agent (Onishi & Baillargeon, 2005; Onishi et al., 2007; Southgate et al., 2007). Although few studies have directly investigated moral cognition in infancy, preliminary evidence suggests that infants can distinguish harmful actions from helpful actions (Kuhlmeier et al., 2003; Hamlin et al., 2007). A recent study by Hamlin, Wynn, and Bloom (2007) showed that when presented with a choice between two agents, 6- and 10-month-olds infants consistently preferred (i.e. reached for) agents who had been identified as "helpers" over agents who had been identified as "hinderers." They also preferred "helper" agents to "neutral" agents (those who neither helped nor hindered), and preferred "neutral" agents to "hinderer" agents.

Research with preschoolers has found that when the consequence of a harmful action is held constant across conditions, children as young as 3 years old assign more blame to actors with bad intentions than to actors with good intentions (e.g. Nelson-Le Gall, 1985) and assign more blame to actors who have caused harm intentionally than to actors who have caused harm accidentally (eg. Nunez & Harris, 1998; Wellman et al., 1979). Siegal and Peterson (1998) found that 3-year-olds also show a distinction between

falsehoods that were uttered intentionally, because of an innocent mistake, or because of a negligent mistake.

However, when younger children are asked to integrate conflicting information regarding the intention of an actor and the consequence of his action (Zelazo et al., 1996), the actor's knowledge (Yuill & Perner, 1988), or the causal nature of the act (Zelazo et al., 1996), they tend to default to a consequence-oriented pattern of moral judgments. This ability to integrate information about the mental state of an actor (i.e. theory of mind) and the consequences of his action when forming a moral judgment appears to become more robust as children enter their late preschool and early Kindergarten years (Yuill & Perner, 1988, Zelazo et al., 1996).

Preschoolers have also been shown to distinguish between moral rules – those which are generalizable, unalterable, and independent of authority dictates (ex: hitting) – and conventional rules – those which are changeable and relative to the social context (ex: wearing pajamas to school) (Smetana, 1981). Furthermore, a study by Leslie, Mallon, and DiCorcia (2006) showed that preschoolers do not make this moral versus conventional distinction by merely appealing to whether the "victim" shows signs of distress; indeed, preschoolers can recognize when an agent's distress is unwarranted, and do not morally condemn actions that result in such "cry baby" displays of distress.

While these and many other studies suggest that causal and intentional aspects of an event *do* enter into children's moral judgments of that event, few studies have investigated this issue in the context of double effects or utilitarian outcomes. One exception is a study conducted by Leslie, Knobe, and Cohen (2006) that examined whether a morally-valenced side effect influenced preschoolers' judgments of whether

the side effect was caused intentionally. Previous studies had shown that adults judge a morally bad foreseen side effect (one which the agent foresees but does not care about) as intentional, and a morally good 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 intention (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. The authors found that children indeed exhibited the side-effect effect: bad side effects were judged as intentional, and good side effects were judged as unintentional. They also found that particularly for the younger children, this effect was only observed in children who understood the concept of "not caring" about something.

The only known study investigating children's moral judgments of utilitarian outcomes is a recently published study by Pellizzoni, Siegal, and Surian (2010). The authors compared children's responses to two scenarios: a footbridge dilemma and a bystander dilemma. In both scenarios, a ball was rolling towards five people on a track, and a bystander had to decide whether to act to prevent the collision. In the footbridge dilemma, the bystander had to push a big person in front the rolling ball in order to save the five people. In the bystander dilemma, the bystander had to pull a string, which would redirect the ball away from the five people and onto another track, on which only one person was standing. In each scenario children were asked to decide what the character should do – act, or not act. A second experiment using the same stimuli asked a

slightly different question: what was the right thing for the character to do – act or not act. Both experiments showed that preschoolers advocated action in the bystander scenario but did not advocate action in the footbridge scenario. The authors concluded from this finding that preschoolers are sensitive to 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. Green et al., 2001; Hauser, 2006; Hauser et al., 2007a; Pellizzoni et al., 2010). However, there is an alternative explanation for their results. Although physical contact with the victim systematically varied between stories, so did the degree to which the harm to victim was caused intentionally as a *means* to saving five people or as a *side effect* of saving five people. Thus, it is unclear whether children's moral judgments of these stories were guided by the contact principle, the principle of double-effect, or both principles.

The current study

The aim of current study was to investigate the extent to which preschoolers' moral intuitions align with the principle of double effect. We ask at what age children understand the concept of "the greater good," and at what age they begin to morally distinguish between causing harm as a side effect and causing harm as a main effect. In order to avoid conflating the contact principle and the principle of double-effect, we presented children with trolley scenarios in which characters made no physical contact with the victims. Stories varied according to whether one individual was harmed as a means of saving five (i.e. harm as a main effect), as a foreseen side effect of saving five (i.e. harm as a side effect), or whether five were harmed because a character failed to intervene to save them (i.e. harm by omission). The omission scenario was included to

directly illustrate what would occur if the character failed to intervene. The question of whether young children understand that a failure to act can be viewed as morally wrong was also of particular interest.

Given the findings of Pelizzoni, Siegal, and Surian (2010), as well as the other studies described above, we expected that preschoolers' moral judgments would indeed conform to the principle of double effect: children would view scenarios in which harm was caused as a main effect as morally impermissible, but would view scenarios in which harm was caused as a foreseen side effect as permissible. Furthermore, if children considered it obligatory to act in the side effect scenario but forbidden to act in the main effect scenario, then omission scenarios in which a character failed to intervene would be considered bad, but not as bad as main effect scenarios.

We based these hypotheses on the evidence suggesting that not only outcome information but also causal and intentional information enters into children's judgments of an event. However, conflicting evidence suggested several alternative hypotheses, particularly for the 3-year-olds. Several studies have found that younger children produce consequence-driven judgments when presented with complex moral dilemmas (Piaget, 1932/1965; Hebble, 1971; Shultz et al., 1986; Yuill & Perner, 1988; Zelazo et al., 1996; Karniol, 1978). Thus, it was possible that 3-year-olds might generate purely utilitarian judgments, and view any action that resulted in the greatest good for the greatest number as morally good. Alternatively, they might base their judgments on a more Hippocratic "do no harm" principle, where any action that resulted in any harm at all would be judged as bad.

Three possible response patterns to trolley scenarios

Hypothesis	Side Effect	Omission	Main Effect
Double Effect	Good	Bad	Bad
Utilitarian	Good	Bad	Good
Do no harm	Bad	Bad	Bad

Table 1: Responses to 3 trolley scenarios according to three different hypotheses.

Method

Participants

Participants were fifty-four children from various New Jersey preschools. One child was excluded because she was too shy to participate. Another six children who failed to understand both test questions during training were dropped. For the purpose of analyses, the remaining forty-seven children were assigned to two age groups: 23 three-year-olds (12 girls) between the ages of 36 and 47 months (M = 42.7 months, SD = 3.6 months), and 24 children older than three (14 girls) between the ages of 48 and 76 months (M = 55.8 months, SD = 7.4 months).

Materials and Procedure

Children were tested individually in a quiet location at their preschool. All testing sessions were videotaped for future scoring. Each child viewed three computer-animated cartoons on a laptop monitor. All three cartoons involved a main character faced with a moral dilemma over whether to save five people by harming another. The cartoons varied (within-subjects) according to whether the harm was caused intentionally as a *main effect* of saving five, as a foreseen *side effect* of saving five, or as a result of *omission*. Participants first responded to control questions probing their understanding of each story. They were then asked to judge whether the main character should have done

what she did, and to rate how good or bad the actions (or inactions) of the main character were on a five-point scale from "really bad" to "really good."

Training. Prior to testing, children were introduced to the "Pink scale."

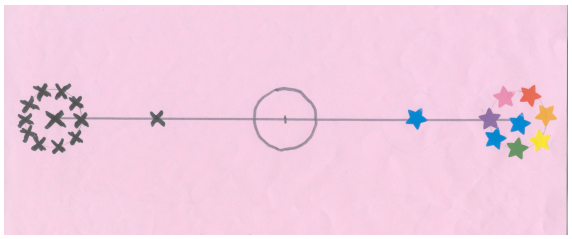


Figure 1. The Pink Scale. Children were told to point to the stars when something was good, the X's something was bad, and the circle when something was "just ok." They were told that one star/X meant "a little good/bad," and lots of stars/X's meant "really good/bad."

The pink scale showed a cluster of stars on one end of the scale, and a cluster of X's at the other end. Children were told that the stars meant that something was "really good," and the X's meant that something was "really bad." A circle in the middle of the scale indicated that something was "just ok." A single star and a single X located at midpoints in-between the circle and the two extremes indicated that something was "a little good" or "a little bad," respectively. After the scale was explained to them, children were asked to show how much they liked various items such as ice cream, eating bugs, and water, by pointing to the appropriate point on the scale.

Once participants were comfortable using the scale, they practiced using the scale to rate the morality of an action. Each child was told two training stories, each accompanied by three pictures presented in sequence. The first picture in each story introduced two characters, 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 did, and then to rate A's actions using the Pink Scale.

Testing. All participants who passed the training were presented with three test stories (within subjects) in semi-random order (between subjects). All three stories began with an image of the main character, Jane, at the top of the screen, with five children below her to the right (each with a cookie), one child below her to her left (also with a cookie), and a snarling squirrel at the bottom of the screen facing the cookies of the five children on the right. Participants were told that the sneaky squirrel was going to eat the cookies of the five unsuspecting children on the right, which would make them very sad. But Jane could see what was about to happen. What would Jane do next? Participants were then shown one of three cartoon animations: harm as a *side effect*, harm as a *main effect*, or harm by *omission*.

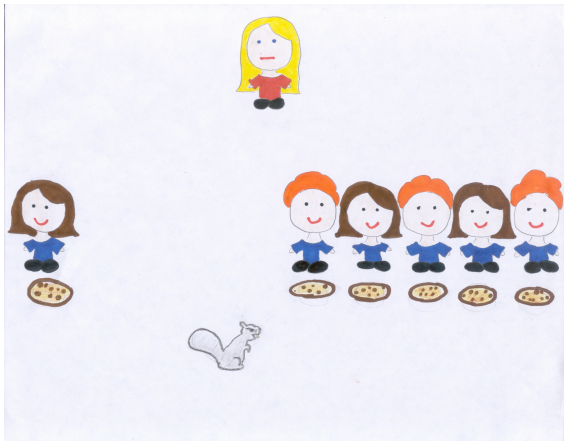


Figure 2. Testing introduction. "This is Jane. Jane is in the park today. And there are some other people in the park: all these people over here (right), and one person over here (left). And look! They all have cookies. But uh oh! There is a sneaky squirrel, and he is hungry! He wants to eat *those* cookies (right). If he eats those cookies, all those people will be very sad. Jane sees the sneaky squirrel, and she knows he wants to eat those cookies. Let's see what Jane does next."

In the *side effect* animation, participants watched Jane place a wall between the squirrel and the five cookies. As Jane anticipated, the squirrel then ate the cookie on the left instead, making the owner of that cookie very sad.

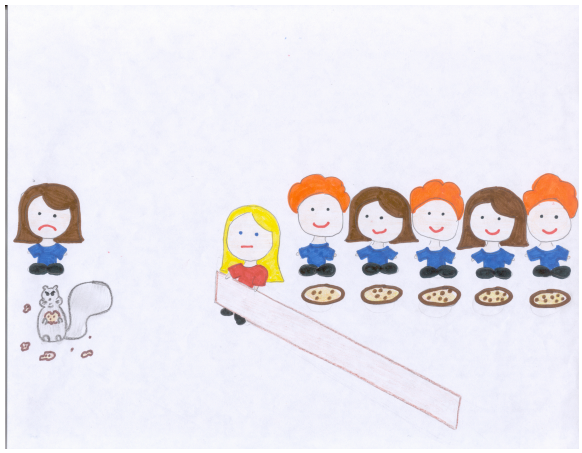


Figure 3. Harm as a side effect. "Well, Jane knows that if she puts up a wall here (right), the squirrel can't get to *those* cookies. So the squirrel will go here (left) and eat *her* cookie instead. But look. Now *she* is sad. Let's watch that again."

In the *omission* animation, the child was told that Jane chose to do nothing, so the squirrel ate all five cookies on the right (one by one), and the five children were very sad.

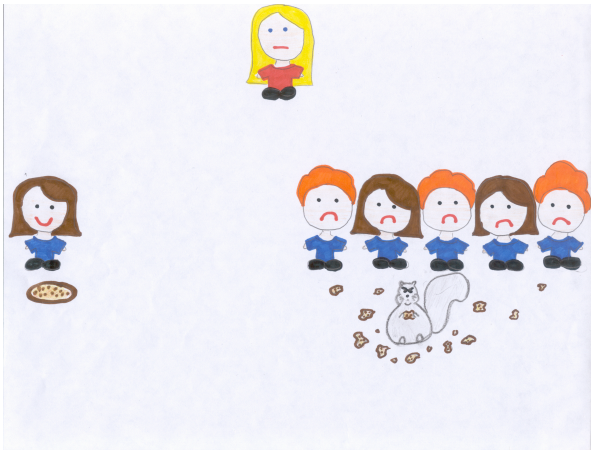


Figure 4. Harm by omission. "Well, Jane doesn't do anything. She just does nothing. So the squirrel goes over here (right) and eats all those cookies. And look. Now those people are sad. Let's watch that again."

In the *main effect* animation, the child was told that Jane decided to take the cookie on the left and feed it to the squirrel (making the cookie's owner very sad), to prevent the squirrel from eating the cookies on the right.

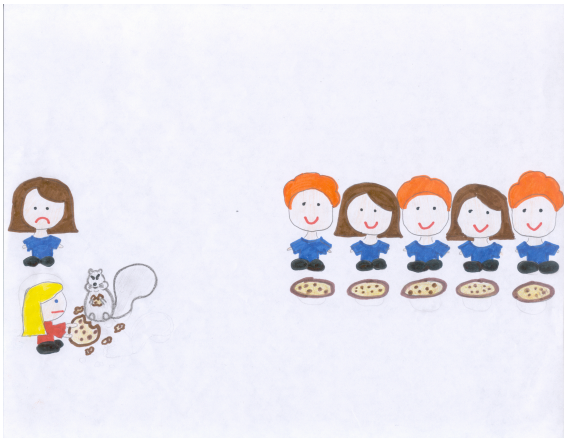


Figure 5. Harm as a main effect. "Well, Jane knows that if she takes *her* cookie (left) and gives it to the squirrel, the squirrel will eat *that* cookie instead of *those* cookies (right). But look. Now *she* is sad. Let's watch that again."

Children were shown all three cartoons in semi-random order.[2] Before each new cartoon, 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 stories, participants were asked control questions before each animation to ensure that they understood what the squirrel was about to do, and how the five children would feel if their cookies were eaten[3]. Participants then watched each animation twice. During the second run-through, children were questioned to make sure they understood what Jane had done, what the squirrel had

[2] Stories were counterbalanced such that the "harm by omission" story was always presented as the second story in the series, to ensure that participants understood that there was at least one alternative to doing nothing. The other two stories alternated as the fist and last stories presented.

[3] See Appendix. Note: no "knowledge" control question was asked because a prior pilot study revealed that children understand that Jane foresees the consequences of her actions in all three stories. 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.

done as a result of her action (or inaction), and how the victim(s) felt when their cookie(s) were eaten. Participants were then asked two test questions:

Test question 1: Should Jane have done that?

Test question 2: 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?)

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."

Results

Preliminary analyses indicated that children at all age levels correctly understood all the stories. There were also no significant gender differences, so gender was dropped from future analyses.

Test Question 1: Should Jane have done that?

Responses to Test Question 1 were analyzed using a mixed model logistic regression with story (3: omission, side effect, main effect) as a within-subject factor, and age (2: 3-year-olds, 4- and 5-year-olds) and order (2: side effect first, main effect story first) as between subject factors. Figure 6 shows the percentages of children who answered "yes" and "no" to question 1 in each story. While most children (83%) approved of Jane's decision in the side effect story, only a few did so in the omission story (38%) or the main effect story (30%), $\chi^2(2, N = 47) = 19.28$, $p < .001$. Pairwise comparisons using NcNemar's test showed that significantly more children approved of

Jane's decision in the side effect story than in the omission story ($\chi^2(1, N = 47) = 13.79$, $p < .001$) or the main effect story ($\chi^2(1, N = 47) = 17.46$, $p < .001$).

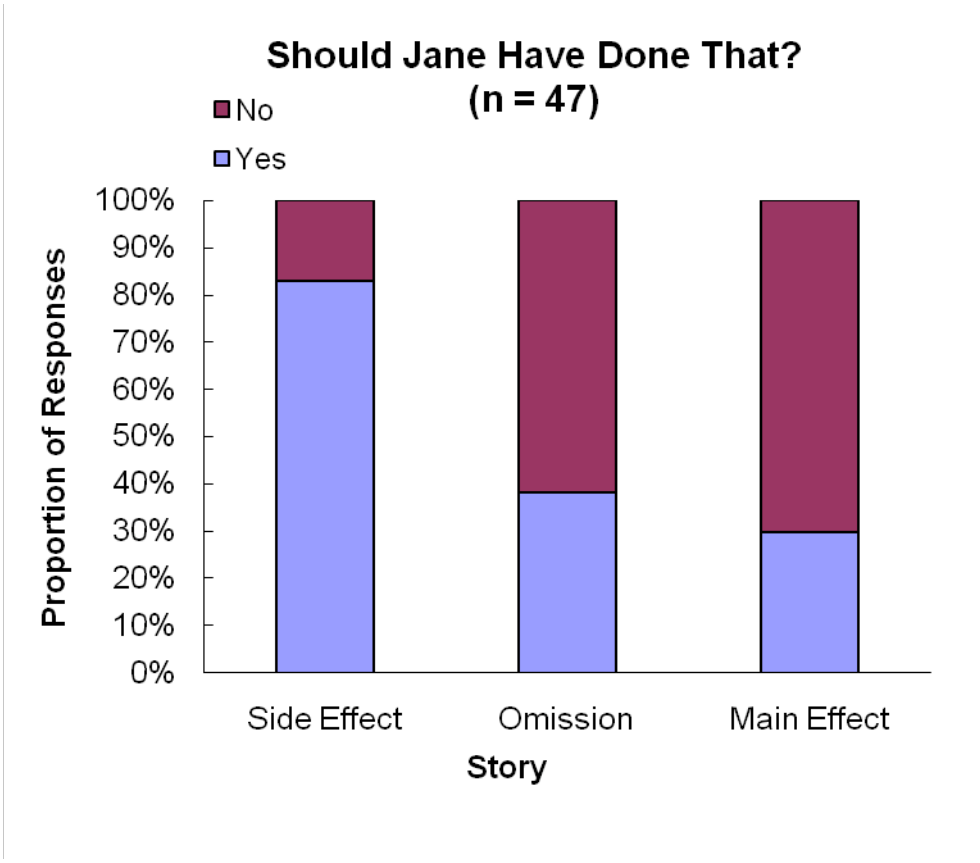


Figure 6. Children's responses to test question 1: Should Jane have done that? Significantly more children responded "yes" in the side effect story than in the omission or main effect stories.

Response patterns for individual children are broken down in Table 1. Of the 47 participants, 4 children responded that Jane should not have acted in either the side effect story or the main effect story, and 10 advocated action in both stories. A majority of 29 children said that Jane should have done what she did in the side effect story but should not have done what she did in the main effect story. Only 4 participants showed the reverse judgment. Similarly, 25 children approved of Jane's decision in the side effect story but not in the omission story, whereas only 4 participants showed the reverse

judgment. 14 children condoned Jane's decision in both the side effect and omission stories, and 4 children did disapproved of Jane's decision in both stories.

Of the 47 participants, the highest number of children (20) approved of Jane's decision in the side effect story but disapproved of her decision in the omission and main effect stories. In contrast, only 4 participants showed the opposite pattern of judgment. The second most frequent response pattern was to say "yes" in both the side effect and omission stories and "no" in the main effect story (9 children). No children showed the opposite pattern of judgment. No children said "yes" to the side effect and main effect stories and "no" to the omission story, and only 4 children said "no" for all three stories.

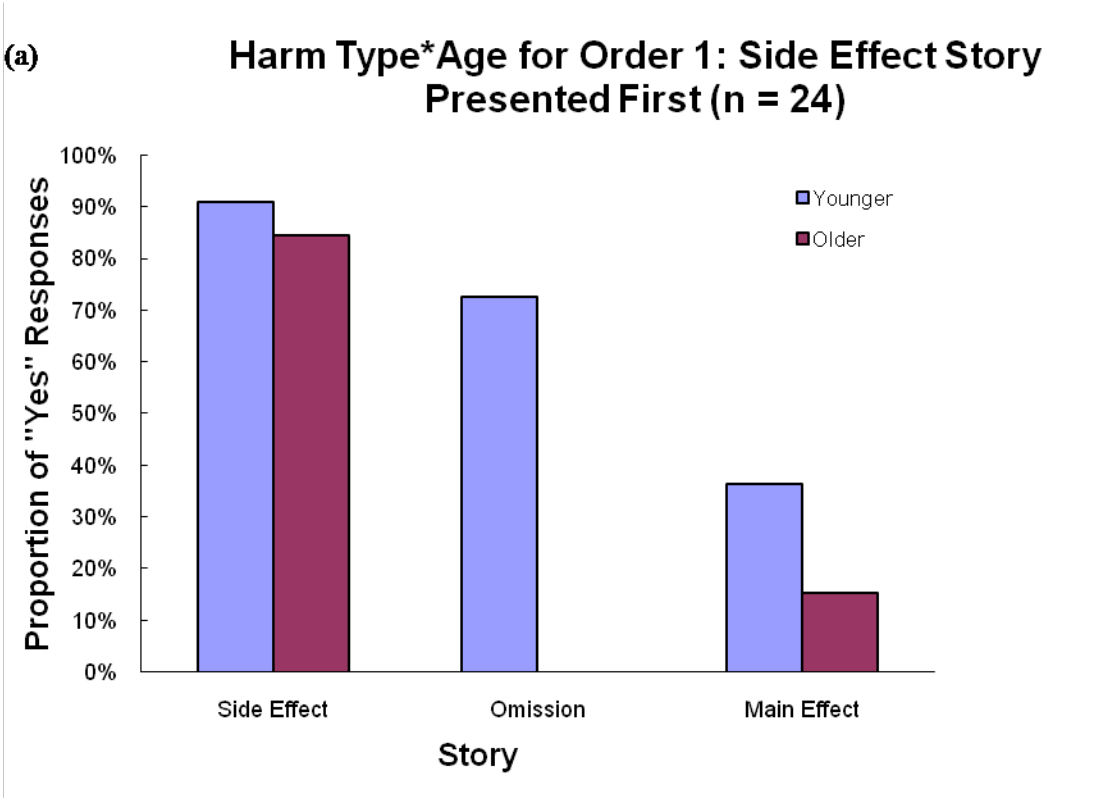
Table of responses to "should Jane have done that?"

# of Children	Side Effect	Omission	Main Effect
4	No	No	No
0	No	No	Yes
0	No	Yes	No
4	No	Yes	Yes
5	Yes	Yes	Yes
9	Yes	Yes	No
5	Yes	No	Yes
20	Yes	No	No

Table 2. Table of responses to Test Question 1: Should Jane have done that?" Number of children whose responses corresponded to each of 8 possible patterns.

A main effect of age was also found; across all stories, children younger than 4 (61%) tended to respond "yes" more frequently than children 4 and older (40%), $\chi^2(1, N = 47) = 7.28$, $p = .007$. However, this effect was driven by the 3-way interaction between story, age, and order presented below, $\chi^2(2, N = 47) = 6.75$, $p = .03$. Figure 8 shows younger and older children's responses for order 1, in which the side effect story was presented first (panel A), and order 2, in which the main effect story was presented first

(panel B). When the side effect story was presented first, 73% of children under 4 approved of Jane's decision in the omission condition, whereas 0% of older children did so (Fisher's Exact test, $p = .02$). There were no significant differences between younger and older children when the main effect story was presented first. Thus, the main effect of age was predominantly due the difference between younger and older children's responses in the omission condition for order 1.



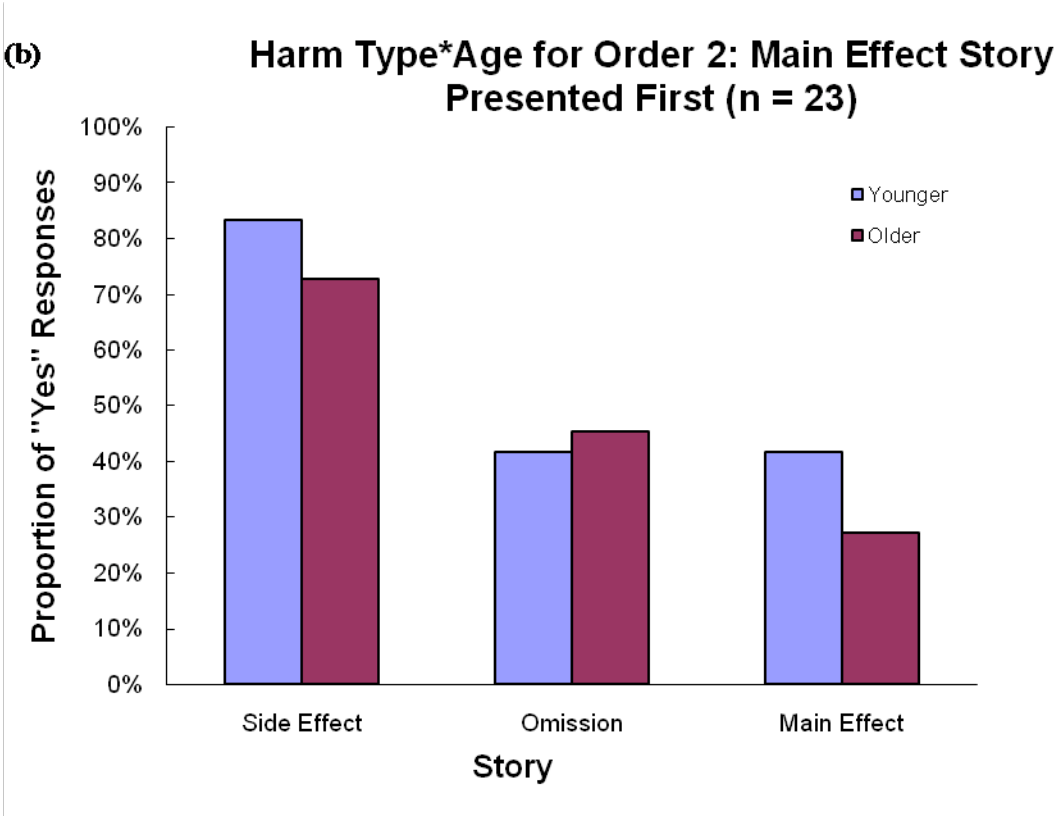


Figure 7. Proportion of "yes" responses to Test Question 1 as a function of age for Order 1: Side effect story presented first (panel A) and Order 2: Main effect story presented first (panel B).

Test Question 2: Pink Scale ratings.

Pink Scale ratings were entered into a repeated measures ANOVA with story (3: side-effect, omission, main effect) x age (2: 3-year-olds, 4- and 5-year-olds) x order (2: side-effect first, main effect first). There was a significant main effect of story, $F(1, 43) = 44.115$, $p < .001$, but no main effect of age, $F(1, 43) = .119$, $p = .732$, or order, $F(1, 43) = .119$, $p = .732$. The main effect of story reflected a tendency to judge side-effect scenarios ($M = .21$, $SE = .21$) as more permissible than main effect scenarios ($M = -1.43$, $SE = .15$), $t(46) = 6.286$, $p < 001$. Omission scenarios ($M = -.87$, $SE = .16$) were judged as more permissible than main effect scenarios, $t(46) = 2.527$, $p = .015$, but less permissible than side-effect scenarios, $t(46) = 3.897$), $p < .001$.

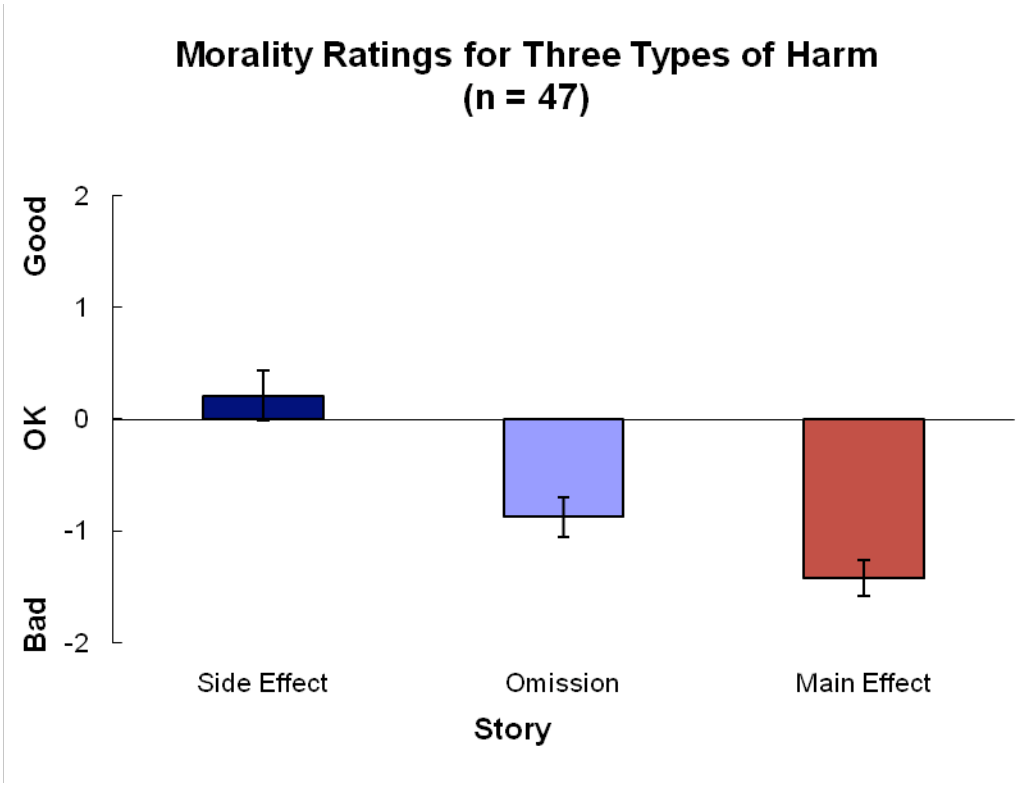
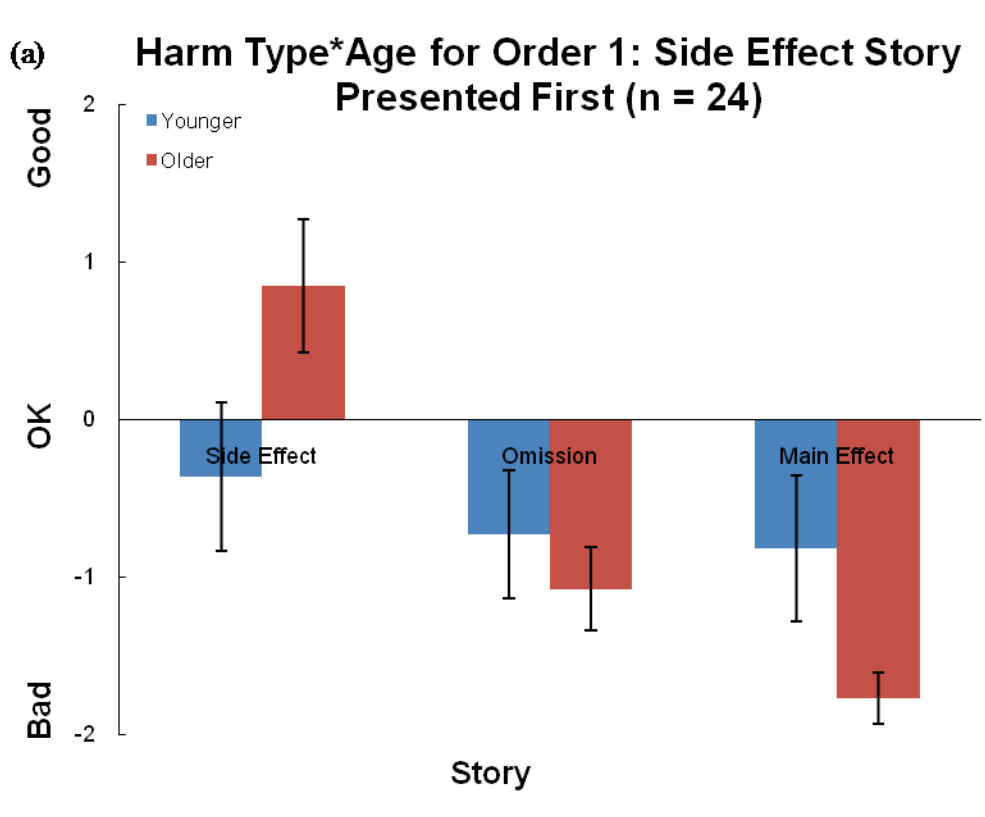


Figure 8. Pink Scale ratings for three types of harm. Preschoolers gave higher morality ratings for Jane's action in the side effect story than in the harm by omission story, and judged Jane's action in the main effect story more harshly than in the omission story.

A significant story x age x order interaction was also found, $F(1, 43) = 10.308$, $p = .003$. Inspection of the data in Figure 10 suggests that the pattern of judgments described above was more fragile in the younger age group. While children 4 and older showed a tripartite distinction between harm as a side effect, harm by omission, and harm as a main effect regardless of the order in which stories were presented (p's $< .01$), children younger than 4 showed no discrimination between the three types of harm when the side effect story was presented first (p's $> .05$). However, younger children rated harm as a main effect ($M = -1.75$, $SE = .179$) significantly lower than harm as a side effect (M = .33, SE = .432) when the main effect story was presented first, $t(11) = 4.614$,

p = .001. Unlike older children, younger children did not distinguish between harm by omission and harm as a main effect in either order (p's > .05).



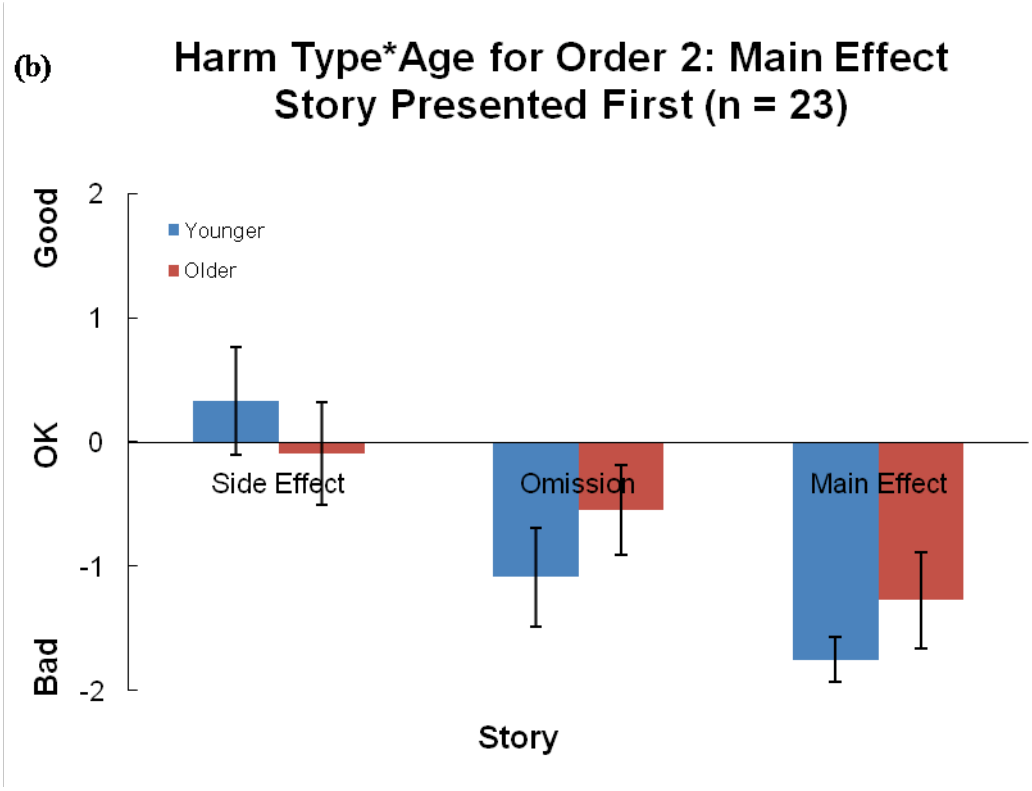


Figure 9. Pink Scale ratings for three types of harm as a function of of age for Order 1: Side effect story presented first (panel A) and Order 2: Main effect story presented first (panel B).

Discussion

In the current study, we tested preschoolers' moral intuitions of three dilemmas using two different measures: a normative measure (Should Jane have done that?), and a permissibility measure (Was that a good/bad/ok thing to do?). On both measures, we found that children showed a pattern of judgments that conformed to the principle of double effect: children considered it permissible to harm someone when that harm was a foreseen side effect of acting for the greater good, but not when that harm was the means to the greater good. This pattern of responses is in direct contrast with outcome-based accounts of children's moral judgments.

On the normative measure, children tended to approve of Jane's decision to intervene in the side effect story, but disapprove of her decision to intervene in the main effect story or to "just do nothing" in the omission story. However, 3-year-olds tended to approve of Jane's decision to "just do nothing" in the omission story when they had seen the main effect story presented first, but not when they had seen the side-effect story presented first.

On the permissibility measure, children tended to rate Jane's decision as "ok" in the side effect story, "a little bad" in the omission story, and "really bad" in the main effect story. However, 3-year-olds' responses on this measure were also sensitive to the order in which the stories were presented; while older children showed a 3-way distinction between the three types of harm, 3-year-olds only discriminated between harm as a side effect and harm as a main effect when the main effect story was presented first. When the side effect story was presented first, 3-year-olds showed no moral distinction between any of the 3 stories.

These findings suggest that while the principle of double effect is already well established in 4- and 5-year-olds, it is more fragile in younger children, and thus more sensitive to order effects and the measure being used. These findings build on the research of Pellizzoni, Siegal ,& Surian (2010), who found that children consider it acceptable to harm someone only when that harm results in the greatest good for the greatest number, and only when there is no physical contact with the victim.

Although structurally similar, our stimuli differ from those of Pellizzoni et al. in several notable ways. First, our stories involved no physical contact with the victim, or even any physical harm to the victim. Instead, our harm involved a violation of property

rights, and its negative effects were purely psychological (provoking sadness in the victim). This suggests that children are sensitive to the distinction between harm as a side effect and harm as a main effect, even when no physical harm is involved and the bystander has no physical contact with the victim. Second, our measures differ 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 character's action on a permissibility scale, providing us with a more sensitive measure of children's moral judgments. Finally, our study is the first to ask children to morally evaluate scenarios in which an actor failed to prevent harm from occurring. Research with adults has shown that although adults view actors who fail to prevent a harm as morally accountable, they tend to judge actors who actively cause a harm to occur more severely than actors who fail to prevent the same harm from occurring (Baron & Ritov, 2004; Cushman et al., 2006; Kamm, 1998a; Spranca et al., 1991). However, to our knowledge, no research has been done to explore children's understanding of the concept of harming by omission.

We should first note that we did not directly test for evidence of the omission bias described above, since we did not provide a comparison condition in which five people were harmed as a result of the character's action. Our initial motivation for including an omission story in the current study was to directly illustrate the counterfactual case implicit in the other two dilemmas. In both the side effect and main effect stories, the character was faced with the choice to act or not to act. Presenting children with the

omission scenario thus allowed them to directly observe and evaluate what would happen if the character failed to act in either of the other two stories.

Our results indicated that 4- and 5-year-olds consistently condemned harm by omission, although not as harshly as they condemned harm as a main effect. Three-year-olds, on the other hand, tended to default to the responses they had given in the previous story; they judged the omission scenario negatively when they had seen a "bad" scenario presented first (i.e. harm as a main effect), and positively when they had seen a "good" scenario presented first (i.e. harm as a side effect).

There are two important findings here. First, even though five people were harmed in the omission condition, compared to only one person harmed in the main effect story, 4- and 5-year-olds still rated the main effect story more negatively. This lends additional support to the theory that children's judgments are not purely consequence-driven, and instead are sensitive to the causal and intentional properties of action. Second, 3-year-olds appeared to be more sensitive to anchoring effects. There are a few reasons why harm by omission may have been a tricky case for younger children. In order to judge a character's decision to "just do nothing," the child must appreciate that not only a person's action, but also her inaction can be morally evaluated. To grasp this concept, the child must understand that the actor *could* have intervened, but chose not to – a level of hypothetical/counterfactual reasoning which is not necessarily required in the other trolley scenarios. Furthermore, in omission scenarios, the harm was neither caused by the actor nor prevented by the actor; thus, children could not assign moral responsibility based on simple causal rules. It is also difficult to infer the intentions of the actor in omission scenarios because there are few behavior cues as to the

actor's goal. Further research is needed to explore young children's understanding of harm by omission and at what age the action/omission distinction emerges.

Two accounts of moral cognition

Given that adults are generally unable to articulate the principle of double-effect, preschoolers may acquire this moral rule by way of a mechanism other than direct cultural transmission. In contrast to Kohlberg's view that moral judgments are the result of conscious reasoning, our findings are consistent with the idea that moral judgments are formed automatically and intuitively, and the principles that guide such judgments are present quite early in development (e.g. Haidt, 2001; Hauser et al., 2007b; Greene, 2001).

Several researchers have proposed an innate "moral faculty" akin to Chomsky's language faculty to explain this early acquisition of moral rules (Dwyer, 1999, 2006; Hauser, 2006; Hauser et al., 2007b; Mikhail, 2007; Rawls, 1971). Proponents of this "universal moral grammer" (UMG) theory argue that while other cognitive systems are certainly recruited for moral reasoning (such as theory of mind), there is nevertheless a modular system "specialized for recognizing certain problems as morally relevant and others as irrelevant, and then generating intuitions about possible moral outcomes" (Hauser, 2006, p. 219). Under this view, moral judgments of right or wrong (or whether something is obligatory, permissible, or forbidden) are similar to judgments of grammaticality or acceptability. Just as language competence enables speakers to intuitively produce and understand an infinite number of original utterances, this "moral instinct" allows us to form fluent, intuitive moral judgments about an infinite number of moral events, many of which we have never seen before. Furthermore, just as native speakers are unable to articulate the rules of grammar that inform their language

competence, many of our moral principles are also automatic, intuitive and inaccessible to conscious thought. Though emotions and conscious processes do enter into our moral judgments, these processes occur after (are triggered by) an unconscious appraisal system has provided an analysis of the causes, intentions, and consequences of action. Thus, emotions may interface with the moral faculty, causing performance differences in morality judgments and taking moral judgments as input to metacognitive processes of justification (Hauser, 2006).

Researchers such as Greene and Haidt have proposed an alternative "dual-process" theory of moral cognition. The dual-process model views moral judgments as the products of two competing systems: a domain-specific social-emotional system, and a species-specific general abstract reasoning system (Green & Haidt, 2002; Green et al. 2001, 2004). According to this account, these reason and emotion systems both generate moral intuitions, which are often in conflict with each other. The extent to which a particular moral event engages each of these systems determines the moral judgment. In this model, emotion and conscious processes are causal factors in our moral judgments – a strict contrast to the UMG theory presented above. Greene's neuroimaging 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 whether a subject is presented with a "personal" or "impersonal" moral dilemma (Green et al., 2001, 2004). However, the extent to which emotions and conscious reasoning play *causal* roles in moral judgment is as yet inconclusive. Further research is required to determine which, if either of these theories, best explains the development of moral cognition.

Future research

The current study focused on the principle of double effect, to the exclusion of other relevant principles such as the action/omission principle (active harm is worse than harm by omission) and the contact principle (harm involving direct contact with the victim is worse than harm involving no contact). We are especially interested in further examining young children's understanding of harm by omission. We hope to run several follow-up studies that systematically test these principles using trolley stories with more varied content and multiple measures (ex: expectations of reciprocation, obligatory vs. forbidden acts, punishment vs. reward, etc.).

While utilitarianism is certainly about minimizing harm, it is also about maximizing good for the greatest number. Few studies have investigated people's intuitions of the trolley scenario or other kinds of moral dilemmas in cases where a *benefit* (ex: money or food) is redirected towards few or many. Are we still obligated to intervene when no threat is involved? By systematically varying the number of people who are saved, helped, or harmed, as well as the amount of property being provided or damaged, we can gain a better understanding of how children and adults quantify harms and benefits.

Children's and adults' intuitions regarding utilitarian principles and property rights become particularly interesting in the context of group membership. We hope to further explore the role of in- and out-group factors in children's and adults' moral judgments of trolley scenarios. Are our intuitions the same when the fat man is a member of the out-group?

There is a burgeoning interest in the relationship between theory of mind and moral reasoning. We hope to explore this domain by examining the moral judgments of autistic individuals – a population known to be impaired in their ability to infer the mental states of others – on similar trolley scenarios. We also believe that investigating the building blocks of moral reasoning in infancy by using nonverbal measures such as anticipatory looking, preferential looking, and looking time based on violations of expectation will bring us closer to understanding the role of causal and intentional factors in generating valence judgments.

Investigating these and other questions will inform cognitive models of moral reasoning. While the current work exposes the principle of double effect as a moral intuition that children and adults seem to share, it goes far from providing an explanatory theory regarding the development of moral cognition. More research must be done to uncover the mechanism by which we generate such intuitions and acquire moral rules. This is a daunting but exciting enterprise, since so little is known about the moral judgments of children.

References

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

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

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

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

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

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

Csibra, G., Gergely, G., Biro, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of "pure reason" in infancy. *Cognition, 72*, 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.

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

Dwyer, S. (2006). How good is the linguistic analogy? In *The Innate Mind, Volume 2: Culture and Cognition* (Carruthers, P, Laurence, S. & Stich S., eds*),* pp. 237-256, New York: Oxford University Press.

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.

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

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of Age. *Cognition, 56*, 165-193.

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

Green, J., 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*, 389-400.

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

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

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

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.

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

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

Kamm, F.M. (1998b). *Morality, Mortality: Death and whom to save from it.* New York, NY: Oxford University Press.

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

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

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.

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

Leslie, A.M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In *Mapping the Mind: Domain Specificity in Cognition and Culture* (Hirschfeld, L. and Gelman, S., eds.), pp. 119-148, 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., & Di Corcia, J. A. (2006b). Transgressors, victims, and cry babies: Is basic moral judgement spared in autism? *Social Neuroscience, 1*(3-4), 270-283.

Mikhail, J.M. (2000). *Rawls' linguistic analogy: A study of the 'generative grammar' model of moral theory described by John Rawls in 'A theory of justice'.* in *Philosophy*. 2000, Cornell University: Ithaca. p. 375.

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

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.

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*, 349-367.

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

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

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., & Jorgensen, M.J. (1993). An empirical study of moral intuitions: towards an evolutionary ethics. *Ethology and Sociobiology, 64*, 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*, 653-660.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36*, 1-16.

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.

Rawls, J. (1971). *A Theory of Justice*. Cambride, MA: Harvard University Press.

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

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

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

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*, 76-105.

Thomson, J.J. (1970). Individuating actions. *Journal of Philosophy, 68*, 774-781.

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

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

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

Young, L., Cushman, F., Hauser, M, Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America, 104*(20), 8235-8240.

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.