UNIVERSITA' DEGLI STUDI DI TRIESTE

Sede Amministrativa del Dottorato di Ricerca

XX CICLO DEL
DOTTORATO DI RICERCA IN PSICOLOGIA

CONTINUITY AND DISCONTINUITY IN MORAL REASONING:

THE 'SIDE EFFECT' EFFECT AND UTILITARIANISM IN YOUNG CHILDREN AND ADULTS

(Settore scientifico-disciplinare: M-PSI/04)

DOTTORANDA

SANDRA PELLIZZONI

COORDINATORE DEL COLLEGIO DEI DOCENTI

CHIAR. MO PROF. TIZIANO AGOSTINI

UNIVERSITA' DEGLI STUDI DI TRIESTE

RELATORE

CHIAR. MO PROF. LUCA SURIAN

UNIVERSITA' DEGLI STUDI DI TRENTO

CORRELATORE

CHIAR. MO PROF. MICHAEL SIEGAL

UNIVERSITA' DEGLI STUDI DI TRIESTE

Index

6. Appendix

Acknowledgment

1. The definition of morality

It is difficult to offer a simple definition of morality that finds a consensus among philosophers, psychologists, and the many commentators in the area. Philosophers were the first to wonder about the moral life of humans. One of the most eminent thinkers who addressed moral problems as the focus of his work was Hume. He argued that 'Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them' giving an important rule to the feelings in relation to the moral judgment (Hume, 1739-1740/1969, p. 462). Hume saw reason as a tool used by the mind to obtain and process information about events in the world or about relations among objects. He argued that a person in full possession of reason yet lacking moral sentiment, would have difficulty choosing any ends or goals to pursue and would look like what we now call a psychopath (Hume, 1777/1960). The social intuition model proposed by Haidt (2001) was an example of the Hume's influence on psychology. Haidt's work was oriented to emphasizing that quick, automatic evaluations (called intuitions) are the base of moral evaluation. Further he underlies the importance of cultural influences, deemphasizing the role of private individual reasoning (Haidt, Koller, & Diaz, 1993).

Another important philosopher who extensively wrote about morality was Kant (1785/1959). His ethical theory was proposed in order to give strong meaning to the primacy of rationality itself in morality. He stated that any rational agent could and should figure out the morally correct thing to do by applying the categorical imperative: 'I should never act in such a way that I could not also will that my maxim should be a universal law' (1785/1959, p. 18). The Kantian approach largely influenced the way to study moral psychology taken by Piaget (1932), Kohlberg (1969), Turiel (2006), and Hauser (2006). Classical research in developmental psychology proposed theories (Kohlberg, 1969; Piaget, 1932) based on the description of different levels of children's ability to justify them answers to moral scenarios. The moral development, according to this account, seems to reflect general cognitive development.

During the Sixties the cognitive approach open the road to different ways of looking at moral judgment. Some researchers start to doubt that justification could be the core aspect of the human judgment and they argue that decision, solution, or other evaluation appears suddenly and effortlessly in consciousness, without any awareness of the mental processes that led to the outcome (Bastick, 1982; Simon, 1992). In recent years, in the context of advances in brain imaging and cognitive neuroscience, many cognitive scientists and empirically oriented philosophers have turned their attention to old questions about morality. Among the issues that have been actively discussed are: the nature of the cognitive mechanisms subserving various aspects of moral cognition, whether those mechanisms are present early in life (Turiel, 2006), the extent which these are specified (Dwyer 1999, 2006; Greene & Haidt 2002; Haidt 2001; Hauser 2006; Nichols, 2004; Nichols & Mallon, 2006), the relation between the moral judgment and other cognitive function, for example intentionality (Knobe, 2003a, 2003b, 2005, 2006; Leslie, Knobe, & Cohen, 2006). Taken together, these questions have a great importance for

moral psychology in general and for the developmental psychology in particular. Does the moral sense develop through discontinuous or rather through continuous patterns? Is mortality a specific domain? Which relations exist between morality and other cognitive function such as theory of mind? These are only some of the questions that are emerging in the beginning research on morality connected out by developmental psychologists. The aim of this thesis is to shed some light on the early ability compute information about moral scenarios. In particular this work has three main goals: (1) To analyze the relation between intentionality judgments and moral evaluation in the SEE as shown by children and adults, (2) To evaluate differences between children and adults computation of information on measures of utilitarian moral judgments, (3). To shed light on the continuity vs discontinuity issue regarding moral development.

The introduction of the thesis will address crucial questions concerning moral judgment:

1.1 What is the nature of moral judgment?

1.2 How does moral reasoning develop?

1.3 Which are relationship between moral judgment and intentionality?

The following provides a historical view of issues and the rationale for the programme of research described in this thesis.

1.1 What is the nature of moral judgment?

Research on the development of moral reasoning was stimulated by Piaget's seminar monograph 'Le jugement moral chez l'enfant' (1932). In this book, Piaget shed light on stages through which a child may acquire a mature moral sense. His cognitive developmental approach provided a conceptual framework for the study of growth of the moral thought. According to Piaget, all development emerges from action: individuals construct and reconstruct their knowledge of the world as a result of interactions with the environment. Based on his observations of children's application of rules when playing, the Swiss researcher determined that morality, too, could be considered a developmental process based on mental structure.

Piaget's researches focused specifically on the moral lives of children, studying the way Swiss children aged 5-13 years play games in order to learn more about their beliefs about right and wrong. Interestingly, Piaget's research concerned not only moral thought itself but also on how children could understand the relation between intentionality and the moral

judgment, integrating information about the magnitude of damage and the information about the intention of an agent. Following 'his clinical' interviews, he asked children of different ages to compare short vignettes about two children. One of the story-character described as having produced a large negative effect (e.g., many broken cups) while performing an act born of positive intentions (e.g. helping his mother wash up); the other was described as having produced a minor negative event (one broken cup), despite his negative intentions. The child was asked to decide which of the two was the naughtier. This choice was followed by an inquiry into the reason of the choice. The main research findings showed that younger children made their choice on the basis of the outcome, whereas older children evaluated morally on the basis of intentions with a transition at around 7-8 years.

From his observations and interview, Piaget, inspired by Kantian thought, concluded that children develop their moral sense through different level of ability: from 'heteronomous' to a 'autonomous' stage. Being in the 'heteronomous' stage of moral reasoning means having a type of thought characterized by a strict adherence to rules and duties, and obedience to authority. This heteronomy results from two factors: (1) the child's cognitive structure and (2) the child's social relationship with adults. Concerning the first point, the author suggests that young children's way of reasoning is characterized by egocentrism. In other words young children are unable to simultaneously take into account their own view of things with the perspective of someone else. This egocentrism leads them to project their own thoughts and wishes onto others. It is also associated with the unidirectional view of rules and power associated with 'heteronomous' moral thought, and various forms of 'moral realism.' Moral realism is associated with 'objective responsibility'. Using this expression, Piaget meant that children fail to differentiate among types of outcomes (an undifferentiated view of consequences, material damage, and observable deviations) and judge by consequences rather than by intentions. Moral realism is also associated with the young child's belief in 'immanent justice.' This is the expectation that punishments automatically follow acts of wrong-doing.

The second contributor to heteronomous moral thinking in young children is their relative unequal social relationship with adults and the mutual influences of their peers. In the natural authority relationship between adults and children, power is handed down from above. The relative powerlessness of young children, coupled with their egocentrism, feeds into a heteronomous moral orientation. However, through interactions with other children in which the group seeks to play together in a way all find fair, children find this strict heteronomous adherence to rules sometimes problematic in that they must take into account peers' point of view. As children consider these situations, they develop towards an 'autonomous' stage of

moral reasoning, characterized by the ability to consider rules critically, and selectively apply these rules based on a goal of mutual respect and cooperation. The ability to act from a sense of reciprocity and mutual respect is associated with a shift in the child's cognitive structures from egocentrism to perspective taking. Coordinating one's own perspective with that of others means that what is right needs to be based on solutions that meet the requirements of fair reciprocity. Thus, Piaget viewed moral development as the result of interpersonal interactions through which individuals work out resolutions which all deem fair.

The second important author that embraced rationalist approach to moral judgment was Lawrence Kohlberg (1969). He modified and elaborated Piaget's work, and laid the groundwork for the current debate within psychology on moral development. Consistent with Piaget, he proposed that children form ways of thinking through their experiences including understandings of moral concepts such as justice, rights, equality and human welfare. Kohlberg proposed that the process of attaining moral maturity took longer and was more gradual than Piaget had proposed. His most famous task involved children as young as 10 years, adolescent, and adults the 'Heinz and drugs dilemma'

Heinz's wife was dying of cancer. Doctors said a new drug might save her. The drug had been discovered by a pharmacist in Heinz's town but he was charging ten times more than what it cost him to make. Heinz couldn't afford to buy the drug, so he asked friends and relatives to lend him the money. But he only raised half the money needed. He told the pharmacist his wife was dying and asked him to sell the drug cheaper, or asked if he could pay the rest of the money later. The pharmacist said no, as he had discovered the drug and wanted to make money out of it. Heinz then became so desperate he broke into the pharmacy and stole some of the drug.

Kohlberg then asked participants:
- Should Heinz have stolen the drug?
- Would it change anything if Heinz did not love his life?
- What if the person dying was a stranger, would it make any difference?

On the basis of persons' justifications of answers, Kohlberg identified six stages of moral reasoning grouped into three major levels. Each level represented a fundamental shift in the social-moral perspective of the individual.

In the first and the second stages of moral development, grouped together in a period called

pre-conventional level, moral judgment is characterized by an individual and concrete perspective. In Stage 1 (obedience and punishment), children focus on avoiding breaking rules that produce punishment. Children obey its own sake and try to avoid the physical consequences of punishment for harm to persons and property. As in Piaget's framework, Stage 1 reasoning is characterized by ego-centrism and an inability to consider the perspectives of others. At Stage 2 (self interest), there is the early emergence of moral reciprocity. The major rule becomes, 'If someone hits you, you hit them back.' This stage focuses on the instrumental, pragmatic value of an action. At Stage 2, children follow the rules only when it is to advance someone's immediate interests. What is right is what's fair in the sense of an equal exchange, a deal, an agreement. At Stage 2, there is an understanding that everybody has his (her) own interest to pursue and these conflicts, so that right is relative (in the concrete individualist sense).

The second period proposed by Kohlberg was the conventional level. Children in this level are characterized by a basic understanding that norms and conventions are necessary to uphold society. They tend to be self-identified with these rules, and uphold them consistently, viewing morality as acting in accordance with what society defines as right. In Stage 3 (interpersonal accord and conformity) individuals are aware of shared feelings, agreements, and expectations which take primacy over individual interests. Persons define what is right in terms of what is expected by people close to one's self, and in terms of the stereotypic roles that define being good - e.g., a good brother, mother, teacher. Being good means keeping mutual relationships, such as trust, loyalty, respect, and gratitude. The perspective is that of the local community or family. There is not as yet a consideration of the generalized social system. In Stage 4 (authority and social order obedience), persons are able to delineate what is right in terms of local norms and role expectations to defining right in terms of the laws and norms established by the larger social system. This is the 'member of society' perspective in which one is morally by fulfilling the actual duties defining one's social responsibilities. Those at stage 4 believe that must obey the law except in extreme cases in which the law comes into conflict with other prescribed social duties. Obeying the law is seen as necessary in order to maintain the system of laws which protect everyone.

Finally, the post conventional period is characterized by reasoning based on principles, using a 'prior to society' perspective. Individuals' reasoning is based on the principles which underlie rules and norms, but reject a strict application of a rule or norm. In Stage 5 (social contract), individuals are viewed as holding different opinions and values. Laws are regarded

as social contracts rather than rigid dictums. Those that do not promote the general welfare should be changed when necessary to meet utilitarian values (the greatest good for the greatest number of people). This is attained through majority decision, and inevitably compromise. Democratic government is ostensibly based on stage five reasoning. In Stage 6 (universal ethical principles), moral reasoning is based on abstract reasoning using universal ethical principles. Laws are valid only insofar as they are grounded in justice, and that a commitment to justice carries with it an obligation to disobey unjust laws. Rights are unnecessary as social contracts are not essential for deontic moral action. Decisions are not met hypothetically in a conditional way but rather categorically in an absolute way. This can be done by imagining what one would do being in anyone's shoes, who imagined what anyone would do thinking the same. The resulting consensus is the action taken. In this way action is never a means but always an end in itself; one acts because it is right, and not because it is instrumental, expected, legal or previously agreed upon. While Kohlberg insisted that stage six exists, he had difficulty in finding participants who consistently used it. It appears that people rarely if ever reach stage six of Kohlberg's model. Thus, there is an understanding that elements of morality such as regard for life and human welfare transcend particular cultures and societies and are to be upheld irrespective of other conventions or normative obligations. Despite this focus on moral reasoning from a rational perspective, recently there has been an increasing emphasis on the roles of emotion and intuition, whit the claim that justification is not the core of moral judgment.

One person who extensively underlines the rule played by emotion in determining moral reasoning is Haidt (2001). His theory is based on the premise that even mature adults could have strong intuitions about moral action, without been able to explain why and therefore without rational deliberation. He gave adults scenarios similar to the following:

> Julie and Mark are brother and sister. They are traveling together in
> France on summer vacation from college. One night they are staying alone in
> a cabin near the beach. They decide that it would be interesting and fun if
> they tried making love. At the very least it would be a new experience for
> each of them. Julie was already taking birth control pills, but Mark uses a
> condom too, just to be safe. They both enjoy making love, but they decide
> not to do it again. They keep that night as a special secret, which makes them
> feel even closer to each other.
>
> What do you think about that?
> Was it OK for them to make love?

Most people who read the above story immediately say that it was wrong for the siblings to make love, and they then begin searching for reasons (Haidt, Bjorklund, & Murphy, 2000). They justify their answers speaking about dangers of inbreeding, about the possibility that someone could be hurt, and many people say something like: 'I don't know, I can't explain it, I just know it's wrong.' Observing situation like this, where even fully mature adults are sometimes unable to provide any sufficient justification for strongly felt moral intuitions ('moral dumbfounding'), Haidt proposed that persons are not always use rational deliberation to solve moral dilemma but rather intuition. Based on these results, Haidt (2001) proposed 'The social intuitionist model' arguing that fast and automatic intuitions are the primary source of moral judgments, conscious deliberations play little causal role, and that persons use them mostly to construct post hoc justifications for judgments that have already occurred. The use of the term 'social' reflects the role of cultural norms and culturally shaped emotions in having a substantial impact on the domain of morality and the process of moral judgment (Haidt, Koller, & Dias, 1993).

While the debate over the role of intuition and justification is still active, some have proposed a hybrid theory that links intuitionism and rationalist approaches. According to such account, unconscious emotions and some form of principled and deliberate reasoning regulates human's moral life. This view has most recently embraced by Greene with fMRI studies (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001) by Damasio on neurologically impaired patients (Anderson, Bechara, Damasio, Tranel, & Damasio, 1999; Damasio, 1994; Koenigs, Young, Adolph, Tranel, Cushman, Hauser, & Damasio, 2007; Tranel, Bechara, & Damasio, 2000).

Greene el al., (2001) reported an interesting series of studies presenting fMRI data supporting a theory of moral judgment according to which both 'cognitive' and 'emotional' processes play crucial and sometimes mutually competitive roles. In these studies, participants were given scenarios similar to the following:

The Trolley problem (Thomson, 1986):
You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman. If you do nothing the trolley will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to

hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman.

Is it appropriate for you to hit the switch in order to avoid the deaths of the five workmen?

The Footbrige problem (Thomson, 1986):
A runaway trolley is heading down the tracks toward five workmen who will be killed if the trolley proceeds on its present course. You are on a footbridge over the tracks, in between the approaching trolley and the five workmen. Next to you on this footbridge is a stranger who happens to be very large. The only way to save the lives of the five workmen is to push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if you do this, but the five workmen will be saved.

Is it appropriate for you to push the stranger on to the tracks in order to save the five workmen?

The dilemmas, first studied in philosophy, have opened up a lively debate on the nature utilitarian judgment. When asked to solve these dilemmas, many persons state that it is appropriate to switch the direction of the trolley in order to avoid the deaths of the five workmen, but few person say that it is appropriate to push the stranger on the tracks. Why this asymmetry happens is still controversial, but some interesting information comes from an fMRI study. Greene et al. (2001) suggest that the crucial differences between the two dilemma lies in the amount of emotional salience of the dilemmas, in that the footbridge dilemma is 'personal' and the trolley dilemma is 'impersonal'. Different parts of the brain are activated when the dilemmas are given to the participants. In particular, 'personal' moral dilemmas are accompanied by brain areas associated with emotional reasoning while 'impersonal' dilemmas are accompanied by the activation of areas associate with cognitive process such as working memory abstract reasoning, and problem solving. The findings were complemented by reaction time (RT) suggesting that some moral dilemmas elicit response conflicts between negative emotional responses and countervailing process (the participants who judged 'appropriate' the action to push the stranger were significantly slower than responses of

inappropriate), which are hypothesized to be cognitive in nature. In another study, Greene et al. (2004) used 'difficult personal' moral dilemma similar to the following:

> Enemy soldiers have taken your village. They have order to kill all remain civilians. You and some of your town people have sought to refuge in the cellar of a large house. Outside, you hear the voice of soldiers who have come to search the house for valuables. Your baby begins to cry loudly. You cover his mouth to block the sound. If you remove your hand from his mouth his crying will summon the attention of the soldiers who will kill you, your child, and the others hiding out in the cellar. To save yourself and the other you must smother your child to death.
>
> Is it appropriate for you to smother your child in order to save yourself and the other town people?

The results show that, while solving this type of conflict, brain regions involved in mediating response conflicts increased the activity. This finding supports the Kohlbergian claim that high level cognition process are used in the resolution of difficult dilemmas that provides a challenge Haidt's model (2001). Further, reaction time data raises doubt about the possibility that moral judgment is unreflective. Participants' RT indicates that their deliberation require considerable time. These took over 10 and in some cases 20 seconds to complete, even if participants were not asked to explain the reason of the choice. The increased cognitive activity suggests that cognitive control processes can override emotional responses, favouring personal moral violations when the benefits outweigh the costs. In summary, brain regions associated with abstract reasoning and cognitive control (including the dorsal prefrontal cortex and anterior cingulated cortex) are recruited to solve difficult personal moral dilemmas. These researches support the hypothesis that both cognitions and emotions play an important role in producing the moral evaluation. Research on brain damage show that prefrontal cortex play an important role in producing a proper moral behaviour (Damasio, 1994; Koenigs, Young, Adolph, Tranel, Cushman, Hauser, & Damasio, 2007). Anderson et al. (1999) compared patient with an early onset of brain damage (prefrontal cortex lesions occurring before 16 months) with adult-onset patients. The main differences between the two groups seem to be in defective on social and moral reasoning, suggesting that the acquisition of complex social conventions and moral rules have followed brain lesions.

Some recent models propose a new view to think of morality: the nativist model. Rawls (1971) proposed that, analogous to the Chomskyan linguistic faculty, humans could be endowed with unconscious moral principles. He wrote 'A conception of justice characterizes our moral sensibility when the everyday judgments we make are in accordance with its principles.' He went on to sketch the connection to language:

'A useful comparison here is with the problem of describing the sense of grammaticalness that we have for the sentences of our native language. In this case, the aim is to characterize the ability to recognize well-formed sentences by formulating clearly expressed principles which make the same discriminations as the native speaker. This is a difficult undertaking which, although still unfinished, is known to require theoretical constructions that far outrun the ad hoc precepts of our explicit grammatical knowledge. A similar situation presumably holds in moral philosophy. There is no reason to assume that our sense of justice can be adequately characterized by familiar common sense precepts, or derived from the more obvious learning principles. A correct account of moral capacities will certainly involve principles and theoretical constructions which go beyond the norms and standards cited in every day life [Rawls, 1971 pp. 46-47].'

Following Rawls' insights (1971), Hauser (2006) proposed that our moral judgments are the product of unconscious psychological processes, and are thus, intuitive (Chomsky, 1986; Dwyer, 1999; Dwyer, 2006; Hauser, 2006; Mikhail, 2000). In terms of a linguistic analogy, this view does not deny cultural variation. Rather, it predicts variation based on how each culture switches on or off particular parameters. In a recent study, Cushman, Young, and Hauser (2006) proposed principles that govern our moral faculty:
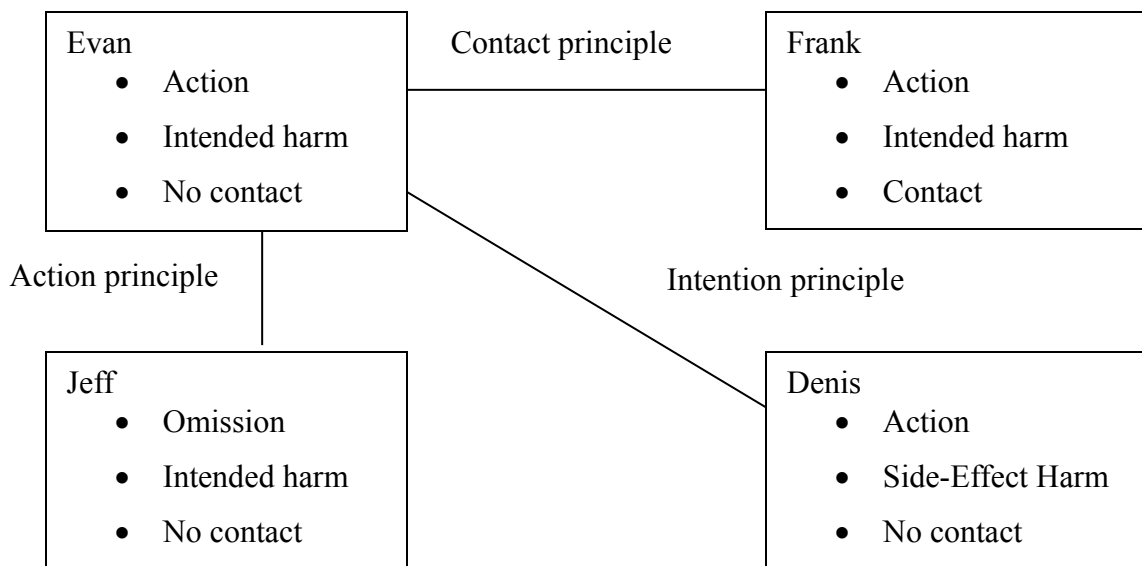
The Action Principle: Harm caused by action is morally worse than equivalent harm caused by omission.

The Intention Principle: 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: Using physical contact to cause harm to another is morally worse

than causing equivalent harm to another without physical contact.

The action principle or 'omission bias' is well-researched in the psychology literature (Baron & Ritov, 2004; Spranca, Minsk, & Baron, 2003). The relevance of the action principle is also recognized in the philosophical literature (Quinn, 1989; Rachels, 1975). The intention principle or the 'Doctrine of the Double Effect' has received intense study by philosophers (Foot, 1967; Nagel, 1986), but less by psychologists (Mikahil, 2002; Royzman & Baron, 2002). The contact principle implies that physical contact is neither a necessary nor a sufficient condition for a personal moral dilemma. Having established that subjects make use of a principle (Thomson, 1986), is possible to investigate whether this principle is available to conscious reflection during justification. Cushman et al. (2006) investigated these three principles by varying the crucial element of the stories (action vs omission, intended harm vs. side effect harm and contact vs. no contact)

Evan
- Action
- Intended harm
- No contact

Contact principle

Frank
- Action
- Intended harm
- Contact

Action principle

Intention principle

Jeff
- Omission
- Intended harm
- No contact

Denis
- Action
- Side-Effect Harm
- No contact

The authors proposed to participants scenarios similar to the following:

Scenario 1: Denise is a passenger on a trolley whose driver has just shouted that the trolley's brakes have failed, and who then fainted of the shock. On the track ahead are five people; the banks are so steep that they will not be able to get off the track in time. The track has a side track leading

off to the right, and Denise can turn the trolley onto it. Unfortunately there is one person on the right hand track. Denise can turn the trolley, killing the one; or she can refrain from turning the trolley, letting the five die.

Test question: Is it morally permissible for Denise to switch the trolley to the side track?

As predicted, subjects judged harmful actions as morally worse than harmful omissions (the Action principle), harm intended as the means to an end morally worse then harm foreseen as the side-effect of an end (the Intention principle), and harm involving physical contact as morally worse than harm without contact (the Contact principle). However, unlike judgments, subjects' justifications differed greatly by principle. The majority of participants were able to provide sufficient justifications for their judgments when asked to evaluate the case of the action principle. Although patterns of judgments were consistent with terms of rational model, this was not the case for intention principle. Less than a third of subjects were able to provide sufficient justifications. Consistent with an intuitionist model of morality subjects reliably generated a pattern of moral judgments in keeping with the intention principle, but were generally incapable of articulating it, to the point of expressing confusion when confronted with their own judgments or even denying their judgments altogether. The intention principle clearly plays a role in moral judgment, but most likely in the form of unconscious - or at least inexpressible - knowledge. However with respect to justifications of responses to the Contact principle, subjects were typically able to articulate the relevant principle used, but unwilling to endorse it as morally valid. Thus, although subjects are able to articulate the principle behind their reasoning, they often reject it as morally invalid. A possible explanation is that the contact principle guides moral judgments according to the intuitionist model during judgment, and that a process of post-hoc reasoning at the justification stage allows subjects to deduce the principle behind their judgments.

This analysis of the literature provides a basic outline of the different models that seek to explain the process of moral judgment. A crucial issue in the development of a moral sense concern whether there exists a stage discontinuity, as proposed by traditional rational models, or more continuous pattern as proposed by intuition and nativist models. In adjudicating between alternative models, developmental psychology has a fundamental role in observing when in development and what children are computing when asked to evaluate moral scenarios.

1.2 How does moral reasoning develop?

Piaget (1932) and Kohlberg (1969), whose theories were summarized in the previous section, posit stage-like discontinuities as a distinctive element of moral development. Further, both authors' conceptualize moral ability as domain general, in other words the development of moral sense is determined by the general level of cognitive development. Their view was challenged by the domain theory advanced by Elliot Turiel and colleagues (Nucci, 1982; Nucci, 2001; Nucci & Turiel 1993; Nucci, Turiel, & Encarnacion-Gawrych, 1983; Smetana, 1981; Smetana, 1993; Smetana, Toth, Cicchetti, Bruce, Kane, & Daddis, 1999; Turiel, 1983; 2006). Within Turiel's theory, a distinction is drawn between the child's developing concepts of morality, and other domains of social knowledge, such as social convention. According to domain theory, the child's concepts of morality and social convention emerge out of the child's attempts to account for qualitatively differing forms of social experience associated with these two classes of social events. In Turiel's studies, children were presented with dilemmas or stories which described the actions of an individual performing a moral transgression. Children are asked whether the actions of the protagonist are wrong, and whether they would still be wrong if there was no rule prohibiting it. In general, findings reveal that even children as young as 4 years believe that certain forms of behaviors (e.g. hitting, stealing or deceiving) are unacceptable at all times, whereas other social rules (e.g. addressing a teacher by her first name) are situation or context specific (Turiel, 1983, p.41). Due to this, the core features of moral cognition are centered around considerations on the well-being of persons. Morality is structured by concepts of harm, welfare, and fairness. In contrast, actions that are matters of social convention have no intrinsic interpersonal consequences. For example, there is nothing intrinsic hurting if a person eats using his hands instead of the cutlery. What makes it a bad thing is the existence of socially agreed upon rules. These conventions, while arbitrary in the sense that they have no intrinsic status, are nonetheless important to the smooth functioning of any social group. Conventions provide a way for members of the group to coordinate their social exchanges through a set of agreed upon and predictable modes of conduct. Concepts of convention, then, are structured by the child's understandings of social organization. These studies are interesting especially because they are proposing a continuous and domain specific hypothesis about the moral development already present at two years old and independent from culture.

Some sceptics (Shweder, Much, Mahapatra, & Park, 1997; Haidt et al., 1993) doubt that moral judgment has a universal component emphasizing, the role played by social

environment on moral evaluation. Haidt et al., (1993) showed that moral value are culturally driven and suggest that cultural norms and culturally shaped emotions have a substantial impact on the domain of morality and the process of moral judgment. Shweder et al., (1997) proposed a model based on three codes of moral thought and discourse, which cultures elaborate and rely on to different degrees. In the ethics of autonomy (similar to Turiel's moral domain), the self is conceptualized as an individual preference structure, and the point of moral regulation is to increase choice, autonomy, and control. In the ethics of community, the self is conceptualized as the holder of an office or role in a larger interdependent and collective enterprise. This code requires duty, respect, obedience to authority, and actions consistent with one's gender, caste, age, or other components of social role. In the third moral code, the ethics of divinity, the self is conceptualized as a spiritual entity motivated to avoid pollution and attain purity and sanctity. Acts that are disgusting or degrading to one's spiritual nature are condemned, even if they involve no harm to others.

Recently some of the rules proposed by Turiel to base his theory started to be challenged by some new findings. In particular a study proposed by Kelly, Stich, Haley, Eng and Fessler (2007) provide some data showing that persons not are always displaying the general role that harm norm are general actions judged wrong here and now as in other times and in other cultures. In their study, designed to determine whether participants judged rules prohibiting harmful behavior to be temporally universal participants were given scenarios similar to these:

A) Three hundred years ago, whipping was a common practice in most navies and on cargo ships. There were no laws against it, and almost everyone thought that whipping was an appropriate way to discipline sailors who disobeyed orders or were drunk on duty. Mr. Williams was an officer on a cargo ship 300 years ago. One night, while at sea, he found a sailor drunk at a time when the sailor should have been on watch. After the sailor sobered up, Williams punished the sailor by giving him 5 lashes with a whip.
Is it OK for Mr. Williams to whip the sailor?

B) Mr. Adams is an officer on a large modern American cargo ship in 2004. One night, while at sea, he finds a sailor drunk at a time when the sailor should have been monitoring the radar screen. After the sailor sobers up, Adams punishes the sailor by giving him 5 lashes with a whip.

Is it OK for Mr. Adams to whip the sailor?

In response to these couple of scenario 52% of participants said that it was OK to whip a drunken sailor 300 years ago (scenario A), but only 6% said it was OK to do it today (scenario B). Other scenarios were presented to participants to determine whether participants judged rules prohibiting harmful behavior to be temporally universal.

(A) It is against the law for teachers to spank students. Ms. Williams is a third grade teacher, and she knows about the law prohibiting spanking. She has also received clear instructions from her principal not to spank students. But when a boy in her class is very disruptive and repeatedly hits other children, she spanks him.
Is it OK for Ms. Williams to spank the boy?

(B) Now suppose that it was not against the law for teachers to spank students, and that Ms. Williams' principal had told her that she could spank students who misbehave if she wanted to.
Is it OK for Ms. Williams to spank the boy?

Again the result were he results were quite striking: 8% of participants said it was OK to spank the boy in response to question (scenario A) and 48% said it was OK to spank the boy in response to question (scenario B).

Turiel's theory cannot explain why moral rules can change across time but his conceptualization of moral ability as domain specific and continuous provides a novel means to conceptualize about the cognitive bases of morality. Hauser's proposal (2006) deals again the issue of domain specificity and continuity on moral reasoning. This theory offers a base to reexamine research on children's ability to understand moral principles from an early age and to discover cultural difference.

1.3 Which are relationship between moral judgment and intentionality?

May be due to the renew attention for the moral psychology, the interest on the debate on the reciprocal influence of intentionality and moral judgment emerged. Throughout this century, philosophers and psychologists have tried to explain features of our moral psychology by appealing to features of human capacity for understanding other minds or

'mindreading'. Piaget and his followers placed enormous weight on the ability for perspective taking, of imagining oneself to have the mental states of another (e.g., Damon 1977; Kohlberg, 1984; Selman, 1980; Rawls 1971). Over the last two decades, there has been considerable empirical and conceptual progress in work on moral psychology and in work on Theory of mind (or mindreading). The moral psychology tradition has looked at the nature and development, as we sow above, while and the ToM tradition has explored the capacity for attributing mental states to others and predicting others' behavior (Baron-Cohen, Leslie, & Frith, 1985; Bartsch & Wellman 1995; Goldman, 1989; Gopnik & Wellman, 1994; Gordon, 1986; Harris, 1992; Leslie, 1994). Although each tradition has flourished, work on moral psychology and work on mindreading has been pursued largely independently. Some exception can be found in works in developmental psychology. Nuñez and Harris (1998) found that three-year-olds assign more blame for intentional behaviors, while Siegal & Peterson (1998) found that three-year-olds can make distinguish between the intentional uttering of falsehoods, innocent mistakes and negligent mistakes. Taken together these works point to an early link between 'theory of mind' reasoning and moral judgment, in which intentional/unintentional judgments subsequently influence good/bad judgments.

A recent study propped by Knobe on adults (2003a), replicated with 4 and 5 years old children (Leslie, Knobe, & Cohen, 2006), has shown that the connection can also run in the opposite direction: from morality to 'theory of mind.' Adults were given vignettes about an agent who brings about a 'foreseen side-effect', the agent could foreseen the side effect but he does not care about. The scenarios were similar to these:

> The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, and it will also help/harm the environment.' The chairman of the board answered, 'I don't care at all about helping/harming the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was helped/harmed.

> Subjects were then asked, 'Did the chairman help/harm the environment intentionally?'

Both in the adults and in the developmental sample participants state that the effect was intentional when the side effect turns out to be negative but not when it was positive. The

authors propose that persons appear to be sensitive to the moral valence of the effect itself, with morally bad effects being considered 'intentional' and morally good effects being considered 'unintentional'.

But does the Knobe effect bear merely on people's performance with the concept of intentional action or does it cast some light on people's conceptual competence with this concept?

The opinion differs: Knobe (2006) and Nichlos and Ulatowski (2007) are answering yes while Nadelhoffer (2006a, 2006b), Alicke (2000; 2006), Machery (forthcoming).

According to Knobe (2006),

'We are now in a position to offer a new hypothesis about the role of moral considerations in people's concept of intentional action. The key claim will be that people's intentional action intuitions tend to track the psychological features that are most relevant to praise and blame judgments. But - and this is where moral considerations come in - different psychological features will be relevant depending on whether the behavior itself is good or bad. That is to say, we use different psychological features when we are (a) trying to determine whether or not an agent deserves blame for her bad behaviors from the ones we use when we are (b) trying to determine whether or not an agent deserves praise for her good behaviors (p. 225-226)'

Some other models addressed the issue of individual differences to better understand SEE. Nichols and Ulatowski (2007) authors argues that the interpretive diversity of 'intentional' show different style to use concept Anyway Nichols and Ulatowski, concur with Knobe on the fact that the concept of intentionality is shaped by its role in blaming and praising.

Other authors are skeptic about the possibility that the SEE casting any light on people's conceptual competence with the notion of intentional action. They believe that the effect results from factors beyond what is constitutive of possessing this concept, a misunderstanding of the meaning of the test question (Adams and Steadman, 2004a; 2004b, Malle, 2004; McCann, 2005), negative emotions (Nadelhoffer, 2004; 2006a; 2006b) or cost-benefit evaluation (Machery, forthcoming) The literature about the 'side-effect' effect , since now, haven't investigate the effect in relation to the role played by the epistemic status of the agent.

This doctoral is aimed to explore, trough four groups of experiments, the relation between the moral judgment and the intentionality judgment both in adults (cap. 2.1) and children (cap. 2.2), the utilitarian evaluation both in adults (cap. 3.1) and children and (cap. 3.2) and to test the possibility that continuous pattern are present in the development.

2. Investigation on the The 'Side-Effect' effect

2.1 The 'Side Effect' effect in adults

Consider the following story:

> The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm [help] the environment.' The chairman of the board answered, 'I don't care at all about harming [helping] the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was harmed [helped]. Did the chairman help [harm] the environment intentionally?

How would you answer?

If you share most adults' intuitions (Knobe, 2003a), in the version of the story in which the new program had negative consequences on the environment, then you are likely to judge that the chairman harmed the environment intentionally (Knobe, 2003). Now, suppose you read the version of the story in which the new program helps the environment (i.e., the version whose passages are indicated in brackets). In this case, if you share the intuitions of most individuals, you are likely to judge that the chairman did not help the environment intentionally (see Knobe, 2003a). This asymmetry is surprising: in both versions of the scenario, the chairman is described as having the same main goal (i.e., to make money), the same amount of information and the same lack of interest for the possible side-effect. The only element that differs in the two versions is the valence (negative vs. positive) of the side-effect. When the valence is negative, most individuals concluded that the side-effect had been produced intentionally (82% of subjects in Knobe's study); when it is positive, few individuals concluded that the side-effect had been produced intentionally (23% of subjects in Knobe's study). This phenomenon, already present in 4- and 5-years-olds (Leslie, Knobe & Cohen, 2006), has been defined 'side effect effect' (henceforth, SEE). The SEE suggests that, moral evaluation affects the everyday attribution of intentionality, contrary to the common view according to which the latter determines the former. Indeed, millennia of philosophical reflection and decades of psychological investigation have used the degree of intentionality attributed to a given action as a crucial criterion in order to decide whether such an action is praise- or blameworthy (e.g., Aristotle,1998; Heider, 1958; Schlenker, Britt, Pennington, Murphy, & Doherty, 1994). However, before drawing general conclusions about the nature of everyday assignment of intentionality, the nature and the possible sources of the SEE have to be closely scrutinized.

In a recent article, Knobe (2006, pp. 225-226) states, 'We are now in a position to offer a new hypothesis about the role of moral considerations in people's concept of intentional action. The key claim will be that people's intentional action intuitions tend to track the psychological features that are most relevant to praise and blame judgments. But - and this is where moral considerations come in - different psychological features will be relevant depending on whether the behavior itself is good or bad. That is to say, we use different psychological features when we are (a) trying to determine whether or not an agent deserves blame for her bad behaviors from

the ones we use when we are (b) trying to determine whether or not an agent deserves praise for her good behaviors.'

Some other authors have addressed the issue of individual differences to better understand SEE. Nichols and Ulatowski (2007) proposed that there are stable individual differences in the way participants interpret the term 'intentionally'. In their experiment it is shown that the 1/3 of the persons used a flexible strategy in answering the help and harm scenarios. The strategy is flexible in that participants interpret the word 'intentional' in different way depending on the context, appealing to the motives in the help condition and to the foreknowledge in the negative one. On the other hand some of the participants show a coherent strategy to answer, in terms of foreknowledge (1/3) or motives (1/3), when asked to evaluate the intentionality both in harm and help scenario. Nichols and Ulatowski argue that the 'interpretive diversity' of the word 'intentionally' opens up a new explanation for Knobe's results although they concur with Knobe on the fact that the concept of intentionality is shaped by its role in blaming and praising.

Now we come to a crucial question: does the effect cast any light on people's competence with the concept of intentional action? Some authors (see Adams & Steadman, 2004; Machery, forthcoming; Malle, 2004; McCann, 2005; Nadelhoffer, 2006a, 2006b) doubt that it is the case. They believe that the asymmetric use of this concept in the probes within a pair, for example in the harm and help cases, results from factors beyond what is constitutive of possessing this concept, such as a incorrect comprehension of the test question, presence of negative emotions, or cost-benefit evaluation.

Adams and Steadman (2004a; 2004b), for example, have suggested that individuals who read the negative side effect scenario are likely to disapprove of the chairman's indifference to the consequences of his actions. Participants would like to blame the chairman's indifference and, knowing that evaluations of blame are more effective at discouraging such acts if the chairman is said to have done the action intentionally, they judge the harming side effect as intentionally produced. In other words, participants could have meant 'The chairman is responsible for the damage' evaluating the negative side effect as produced intentionally. If Adams and Steadman were right it would be easy to say why participants evaluate as intentional the side effect only when it is negative but not when it is positiveIn the positive scenario participants do not feel the need to ascribe any guilt or express disapproval and therefore they do not take the questions as requiring a judgment of responsibility. A position, similar to the one proposed by Adams and Steadman, has been embraced by McCann (2005) and Malle (2004). They believed that the behavior of the chairman, in the harm case, must be sanctioned, and for this reason is necessary to consider this act as product intentionally. According to these theories the misinterpretation of the test question seem to be the real cause of the SEE. Is this the explanation of the asymmetry? In the Experiments 1 and 2 we will try to evaluate the hypothesis proposing the test question in a way that could not give raise to the 'rich interpretation' involving attribution of responsibility.

Some authors argue that an explanation of the SEE can be found in the crucial role to the emotion aroused in help and harm scenarios. According to these theories, intentionality considerations are influenced by affect driven evaluation of the situation. Haidt (2001) underscores the role played by emotions with respect to moral evaluation. Nadelhoffer (2004; 2006a; 2006b) suggests that the SEE is best explained in terms of the distorting effects of people's feelings of blame. Nadelhoffer, in particular, uses the term 'biasing effect' of emotions to explain this effect. He writes that 'affective or emotional responses ... inappropriately bias our otherwise rational judgments' (2006a, p. 214).

Machery (forthcoming) has proposed the 'trade-off hypothesis' to explain the SEE. He states, 'The hypothesis under consideration is that people make asymmetric judgments in the stories used by Knobe and others because they conceptualize the negative side-effect (e.g., harming the environment) as a cost that the agent incurs in order to reap a benefit (e.g., making profits). Because costs are intentionally incurred in order to reap a benefit, people judge that the foreseen, negative side-effect was intentionally brought about.' To provide evidence for his hypothesis, he carried out two experiments. In the first experiment, he gave subjects two scenarios concerning the intentionality of the agents: in the first scenario, the agent paid an extra dollar to get something to drink, in the other scenario the agent obtained a free cup while buying something to drink. Subjects were significantly more likely to judge that the agent intentionally paid an extra dollar to get something to drink instead of having won the cup intentionally. In a second experiment, he gave participants two other scenarios, the worker case and the dog case, that were based on the trolley dilemma. Subjects were more likely to judge that the agent intentionally brought about the side-effect in the worker case than the side-effect in the dog case, because they are more likely to conceptualize the side-effect in the worker case (causing the death of the worker on the side tracks) than the side-effect in the dog (saving a dog in addition to five workers) as a cost incurred in order to reap a greater benefit. Machery takes these data to be confirmation of the 'trade-off hypothesis.'

A different explanation of the effect comes from Alike (2000, 2007). He develops what he calls the Culpable Control Model (CCM) of blame attribution. He argues that the primary factor for the ascription of blame is *personal control*: 'the freedom to effect desired behaviors and outcomes or to avoid undesired ones (Alike, 2000)'. The model differs from traditional rationalist ones in the emphasis it places on the claim that personal control judgment and blame are influenced by the relatively unconscious, spontaneous evaluation of the mental, behavioral and consequence elements. According to Alike, these responses and relatively unconscious responses can be triggered by both the evidential structural linkage information (concerning personal control) and other factors such as the person's appearance, reputation and social status. According to the CCM, the SEE is influenced by evaluations of the callous disregard for the damage caused to the environment when persons are asked about the intentionality of the act.

The experimental research seems to have demonstrated that in people's folk concept of intentionality, performing an action intentionally, requires the presence of five components: a desire for an outcome; beliefs about an action that leads to that outcome; an intention to perform the action; skill to perform the action; and awareness of fulfilling the intention while performing the action (Malle & Knobe, 1997). Is the SEE radically change this model? Malle (2004) propose that, independently from the nature of the SEE, the naïve concept of intentionality still persist: in fact it is necessary to refer to that concept to say that persons are biased. In other word it is necessary to speak about the five components to say that people are not taking into account one of that.

The aim of the reported studies was twofold. First, we have tried to establish whether the SEE is real or a mere epiphenomenon, due to participants' tendency to interpret the question about the intentionality of the negative SE as a question about the responsibility of

the actor (Experiments 1, 2 and 3). Second, we have tried to establish the boundaries of the SEE, by investigating whether intentionality judgments are really affected by the evaluation of the outcome or whether they depend, more appropriately, on the evaluation of the actor's intentions and beliefs (Experiments 4 and 5).

2.1.1 Experiment 1: 'Responsibility Implicature'

According to Adams and Steadman (2004), McCann (2005) and Malle (2004) when the SE is negative individuals seem to attribute responsibility to the agent. Following this hypothesis, the SE would disappear if the test question were explicitly asking for an intentionality judgment, but not for a responsibility evaluation. To evaluate this hypothesis, we conducted an experiment using a test question that participants could not interpreted as a responsibility question, in fact both the options implicate this attribution. The experiment was designed to compare the answers produced by participants when the test question was standard (Knobe 2003 a) with the answers indicating that the responsibility for the SE was attributed to the to the chairman and the participants have to take position in relation to the agent's intentionality. If Adams and others were right the participants should judge the SEE as unintentional, both the positive and negative stories, when the test question is not ambiguous.

Method

Participants

The sample included 80 adults (14 male, $M = 23$ years, $SD = 4.1$ years ranging from 20 to 42 years). In all the experiments, participants were Italian native speakers from the Northern part of the country. They had various socioeconomic and educational backgrounds. Participants took part voluntarily in the experiments and received no pay for their participation.

Procedure

Participants were randomly assigned to one of the four conditions that result from the combination of two types of side effect valence (positive vs negative) and two types of test question (standard vs responsibility implicature). Participants had to read either the harm or the help version of the chairman scenario (see Knobe 2003).

In the standard question versions, they had to answer the following question:

Did the chairman help/harm the environment intentionally?

In the explicit question version, they had to answer the following question:

Two judges evaluate the situation:

Paul: 'The chairman is responsible for the damage caused to the environment because he produced it intentionally' [the chairman is responsible for the help given to the environment because he produced it intentionally]

Daniel: 'The chairman is responsible for the damage caused to the environment but he did not produce it intentionally' [the chairman is responsible for the help given to the environment but he did not produced it intentionally]

According to you, which one is the better judgment?

In this as well as in all the other experiments, participants were tested individually in a silent room, they were fully debriefed, and they indicated their answer by ticking a box at the end of the story. The presentation orders of the alternatives were counterbalanced across participants.

Results and discussion

In all experiments, we scored participants' responses as Knobe (2003) did.

Table 2.1 presents the percentage of intentionality attribution produced in the four conditions of Experiment 1. As found in Knobe (2003a), in the standard question conditions, the negative side effect was judged intentional more often than the positive side effect, χ^2 (1, $N = 40$) = 25.6, $p <$ 001. In the explicit question conditions the results were similar: participants accepted Paul's judgment (the attribution of intentionality to the side effect) more often in the negative than in the positive condition, χ^2 (1, $N = 40$) = 28.9, $p < 0.001$).

The two conditions in which the SE was negative did not differ from each other, Fisher Exact Probability Test, *p* = 1. Likewise, the two conditions in which the SE was positive did not differ from each other, Fisher Exact Probability Test, *p* = .99.

Table 2. 1: Percentage of Intentionality Attributions in Experiment 1 through 5

Condition	Side Effect	
	Positive	Negative
Experiment 1		
Standard	10	90
Explicit question	10	95
Experiment 2		
No implicature scenario	15	85
Experiment 3		
Caring attitude	-	69
Experiment 4		
External Ignorance	10	70
Internal Ignorance	20	54
Experiment 5		
False Belief	-	22

The results do not support the Adams and others' hypothesis. Participants continued to attribute intentionality to the chairman even when the factors of intentionality and responsibility were separated. We conduct a second experiment in order to verify if giving relevance to the responsibility of the chairman, making explicit mention about the fact that the agent has to pay a 5.000 euros sanction independently from the intentionality attribution, could make any changes in the participants evaluation. If Adams and other were right the participants should judge the SEE as unintentional, both the positive and negative stories,

when the test question is not ambiguous.

2.1.2 Experiment 2: 'Responsibility Implicature'

Method

Participant

The sample included 59 adults participated in the experiment (20 males, $M = 26$ years, $SD = 3.4$ years, ranging from 19 to 35).

Procedure

Participants were randomly assigned to one of the two conditions: responsibility implicature with a positive side effect (n = 20), responsibility implicature with a negative side effect (n = 39). The procedure was the same as in the Experiment 1 except for the alternative proposed as answers to the test question.

Two judges evaluate the situation:

Paul: The chairman must pay a fine of 5.000 euro [The chairman will receive a reward of 5.000 euros]: he is responsible for the damage caused to the environment because he produced it intentionally [he is responsible for the help given to the environment because he produced it intentionally]

Daniel: The chairman must pay a sanction of 5.000 euro [The chairman will receive 5.000 euro]: he is responsible of the damage caused to the environment but he did not produce it intentionally [he is responsible for the help given to the environment but he did not produced it intentionally]

According to you, which one is the better judgment?

Result and discussion

On the responsibility implicature test question, few persons judged the side effect as intentionally produced when the valence was positive (Daniel's judgment) while the majority judged the side effect as intentionally produced when the valence was negative (Paul's

judgment). The difference was statistically significant, χ^2 (1, $N = 59$) = 26.9, $p < 0.001$.

Taken together the results of Experiment 1 and 2 show that the participants do not seem to mistakenly interpret question about the attribution of intentionality with question on attribution of responsibility (Adams & Steadman, 2004a; 2004b),

Alike (2007) sustains that participants, influenced by the callous disregard for the damage, assign intentionality to the negative SE (2007). In other words, participants would assign intentionality to the negative side effect not because of the negative evaluation itself but rather because of the attitude of the agent toward the side effect. If Alike's hypothesis is correct in Experiment 3 the negative SE should not be judged as intentional.

2.1.3 Experiment 3: 'Caring attitude'

Method

Participants

The adult sample included 30 participants, of which (5 males, $M = 23.5$, $SD = 3.1$ years, ranging from 20 to 31 years).

Procedure

Participants were given only the harm version of Knobe's stories. The sentence 'I don't care at all about helping/harming the environment. I just want to make as much profit as I can' was substituted with 'I am sorry if the environment will be harmed, but we must increase profits: Let's start the new program.' In this experiment we use the standard test question proposed by Knobe (2003, see also Experiment 1).

Results and discussion

The majority of the participants stated that the damage was produced intentionally. The careless attitude of the chairman did not change the intentionality attribution of the participants. According to Malle and Knobe (1997) one of the five elements that is necessary to determine intentionality evaluation is the belief about an action that leads to that outcome. We evaluated the possibility that the agent's epistemic status could have a crucial influence in determining the SEE. The models that underline the role of emotions in determining the SEE (Haidt, 2001; Nadelhoffer, 2004; 2006a; 2006b) should not be influenced by the manipulation on agent's epistemic status (belief) because the negative effect still persists. If the emotion based models were right the changes on the epistemic status of the agent should not modify

the intentionality evaluation, on the other hand if Malle and Knobe (1997) were right, when the agent could not foreknow the consequence of SE the intentionality evaluation should decrease.

2.1.4 Experiment 4: 'Internal and External Ignorance'

In the following experiment, we manipulated the belief of the agent (level of ignorance) about the side effect using scenarios characterized by different level of chairman's foreknowledge of the SE. In the internal ignorance scenario, the agent could not foreknown the result of the action, because he was not informed about that. By contrast, in the External ignorance scenario nobody could foreknow the side effects of the agent's action (Kahneman, & Tversky, 1982) because it was generally unknown.

Method

Participants

The sample included 99 adults (25 male, $M = 21,9$ year, SD 2.3 ranging from 18 to 36)

Procedure

In Experiment 4, participants were randomly assigned to one of the four conditions: two valence types (positive vs negative) and two types of ignorance (internal vs external). The scenarios given to the participants were the following:

Internal Ignorance Scenario

The vice-president of a company went to the chairman of the board and said 'We are thinking of starting a new program. It will help us increase profits'. The vice president knows that the program would have harm the environment but he didn't told to the chairman.

The chairman of the board answered, 'I want to make as much profit as I can. Let's start the new program.' They started the new program. And the environment was harmed/helped.

Did the chairman help/harm the environment intentionally?

External Ignorance Scenario

The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase

profits but we do not know what it could be its impact on the environment.'

The chairman of the board answered: 'I want to make as much profit as I can. Let's start the new program'. They started the new program. And the environment was harmed/helped.

Did the chairman help/harm the environment intentionally?

Results

In the Internal ignorance condition, few persons evaluated the help produced the effect as intentionally when the valence was positive, while nearly half of the participants evaluated the damage as produced intentionally when the valence was negative. The differences were not significant, χ^2 (1, $N = 48$) = 3.38, $p = 0.06$. In the External ignorance condition, few participants evaluated the help as intentionally produced when the valence was positive, while the majority of the participants evaluated the effect as intentionally determined when the valence was negative. The difference was statistically significant, χ^2 (1, $N = 55$) = 21.86, $p < 0.001$. The two conditions in which the SE was negative did not differ from each other, χ^2 (1, $N = 48$) = 3.9, $p = 0.08$.

We explain the fact that half of the subjects continued to judge the harm produced as intentional in terms of 'semifactual' reasoning. Participants could have thought 'the chairman stated that his first interest is to do as much profit as he can. Hence, even if he could have known the harmful effect of his action, he could not have changed it.' Observing that the pattern of answer, it is not possible to exclude the influence of emotions in determining the effect. For this reason we gave participants (Experiment 5) scenarios in which chairman had a false belief about the consequences of the side effect. In this new experiment the absence of belief is more explicit, and the semifactual reasoning is blocked. If Malle and Knobe (1997) were right, when the agent has a false belief on the consequence of SE the intentionality evaluation should decrease.

2.1.5 Experiment 5: 'False beliefs'

Method

Participants

The sample included 62 adults (35 male, *M* = 21, *SD* = 4.6, ranging from 18 to 45).

Procedure

Participants were randomly assigned to one of the two false belief conditions: in one condition the participants were asked to answer a question on the chairman's belief first and then a question about the intentionality; in the second condition, participants were asked to answer a question about intentionality first and then one about chairman's belief. The scenarios were given to the participants were the following:

The vice-president of a company went to the board chairman and said: 'We are thinking of starting a new program. It will help us increase profits and it will also help the environment.' The vice president knew that the program would have damage the environment but he told a lay to avoid the rejection of the program. In the past the chairman was fined for damage produced to the environment and he does not want the accident to be repeat again. The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was helped. The program was enact and it provoke a big damage for the environment think that astonish the chairman because he belief that the program would have helped the environment.

Subjects were then asked
- The chairman knows that the program would have...
Helped the environment
Damaged the environment

- 'Did the chairman harm the environment intentionally?'

Results

The results show that, in the belief-first condition, 29 participants judged the chairman's belief as helping while 9 participants judged the chairman's belief as harmful. Evaluating only

the participants who attributed the helping belief to the chairman (n = 20), we found that 4 only persons evaluated the side effect as product intentionally.

In the intentionality-first condition, 9 participants attributed intentionality to the chairman and 2 attributed the belief of harming to the chairman. Evaluating only the participants that attributed a help belief to the chairman (n = 22), only 7 participants judged the side-effect as intentional. Responses in the conditions did not differ significantly, Fisher Exact Probability Test, $p = 0.17$, so the two group were collapsed.

General discussion

The aim of this investigation was to examine the relation between pragmatic evaluation, caring attitude, epistemic status and intentionality judgments in relation to the SEE. Experiments 1 and 2 evaluated the explanation of the SEE proposed by Adams and Steadman (2004a; 2004b), McCann (2005) and Malle (2004). In both experiments, we tried to disambiguate the test question by blocking ambiguous interpretation of the test question. The participants continued to attribute intentionality to the chairman when the side effect was negative but not when it was positive. Our results show that pragmatic factors play a minor role, if any, in the SEE.

In the Experiment 3, we examined the chairman's attitude as a factor that could determine the intentionality attribution (Alike, 2006), we realized that the majority of the participants continue to attribute intentionality to the chairman. This result shows that, in terms of the elements that could produce the SEE, the attitude is not really meaningful. In the last two experiments (Experiments 4 and 5) we evaluated the role played by epistemic mental state in producing the intentionality attribution. The data showed that people decreased intentionality judgments when the chairman was described as not having foreknowledge or having the false belief about the consequences of his action. Data from Experiment 5 highlight that the source of knowledge that inform people about the intentionality of an action is the chairman's mental states.

The crucial question opened in the introduction refers to the possibility that the SEE is casting any light on people's competence with the concept of intentional action. Knobe (2006) and Nichols and Ulatowski (2007) claim that responses to measures of the SEE reflect participants' grasp of the concept of intentionality. By contrast, other authors propose sustain that the effect results from factors beyond what is constitutive of possessing this concept (Adams & Steadman; 2004a; 2004b; Machery, forthcoming; Nadelhoffer, 2006). Our results suggest that intentionality evaluation of the SE, in the harm case, are influenced by epistemic status (Experiment 5) and less negative effect of the action (Experiment 4) but not the agent's declarations of uncaring (Experiment 3). In our opinion, the fact that the SEE disappear when it is explicitly underlined the chairman does not have the belief to produce a negative side effect, seem to corroborate Malle's proposal. In other words, the SEE is not changing the model on the naïve concept of intentionality proposed by Malle and Knobe (1997)

In the next chapter, we evaluate the possibility that the intentionality judgments of preschoolers and adults are influenced in the same way.

2.2 Foreknowledge, Caring, and the 'Side-Effect' effect in young children

In the context of the recent burgeoning interest in moral psychology (Dwyer, 2006; Hauser, 2006), a major focus for research has concerned the side-effect effect (SEE). This effect occurs in adults asked to consider scenarios in which agents dismiss information about the harming or helping side effects of their actions. Harmful side effects of these actions are judged as having been produced intentionally whereas the helpful side effects are not. The asymmetry is surprising because, in both the negative and positive scenarios, agents are explicitly described as dismissive ('not caring') about the side effects of their actions. In an intriguing series of studies, Knobe (2003a, 2003b, 2004) gave adults stories in which the issue of not caring was stated in relation to positive or negative side effects. For example, in one situation, a company vice-president was described as having told the board chairman, 'We are thinking of starting a new program. It will help us increase profits, and it will also help/harm the environment.' The chairman of the board answered, 'I don't care at all about helping/harming the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was helped/harmed. The test question was 'Did the chairman help/harm the environment intentionally?' Judgments of whether or not the chairman brought about the side effect intentionally depended crucially upon whether or not this was helping or harmful. Participants often – but not always – judged the helpful side effect as unintentional. If the effect was harmful, it was often judged as intentional. In this case, there had been a clear reference to declarations of uncaring that the environment could be harmed by the company's action. Leslie, Knobe and Cohen (2006) reported a similar asymmetry in 4- and 5-years-olds. In one condition in which the outcome was negative, children were told, 'Here is a boy named Andy, and he's over at his house. And here is a girl named Janine, and she's over at her house. And look what Andy has with him, he has a ... [frog]. Now Andy loves frogs, but Janine hates frogs. Now can you remember, does Andy love frogs? Does Janine love frogs? Andy wants to bring the frog over to Janine's house, but she will get upset. Why will she get upset? Now listen very carefully. Andy does not care that Janine will get upset. He is going to bring the frog over anyway. Does Andy care that Janine will get upset? So Andy brings the frog over to Janine's house and she gets upset. Now I have a question for you. Does Andy make Janine upset on purpose?' In another condition, the story structure was the same except that the second character likes the frog and there was a happy outcome. The children answered that the side effect was brought about 'on purpose' when it was negative but not when positive,

suggesting a link that runs from moral outcomes to theory of mind (judgments of purpose). This finding complements research that has pointed to a link that runs from theory of mind to negative judgments of morally reprehensible outcomes based on an agent's state of knowledge (Nuñez & Harris, 1998; Piaget, 1932) that includes judgments of negligence in both preschoolers (Siegal & Peterson, 1998) and adults (Walster, 1966). Leslie et al.'s (2006) studies represent an initial examination indicating continuity between young children and adults in showing a SEE. The aim of our investigation was to clarify the circumstances under which children demonstrate this effect. In Leslie et al.'s work, statements about not caring on the part of the agent could have been interpreted to imply that the agent did have knowledge of the outcome but chose to dismiss this possibility. In fact, as Nichols and Ulatowski (2007, p. 360) have observed, the key issue for intentionality judgments may concern foreknowledge. If so, participants could interpret declarations of not caring as evidence of a deliberate intention based on possessing the foreknowledge that harm could occur. Adults may regard statements such as 'The chairman harmed the environment intentionally' and 'The chairman knowingly incurred the risk of harming the environment' to merit equivalent intentionality judgments (Adams & Steadman, 2004a; 2004b). In this respect it is unclear whether Leslie et al.'s SEE findings concern a side effect that is foreknown but disavowed – one that is similar to a negligent mistake with a negative outcome that was identifiable beforehand – and a side effect similar to a pure or honest mistake in which the consequences are genuinely brought about without foreknowledge. To clarify the issue of foreknowledge and declarations of not caring in creating the SEE and asymmetry in intentionality judgments, we carried out three experiments involving preschoolers from urban, middle-class areas of northeastern Italy. They participated with informed parental consent and no child was included in more than one experiment. In Experiment 6, we manipulated the state of knowledge (possessing or lacking foreknowledge) of an uncaring agent whose behavior resulted in either a positive or negative outcome. We predicted that 'on purpose' intentionality judgments would be restricted to situations in which the action of an uncaring agent created a negative side effect rather than a positive one.

Experiment 6: 'Lack of caring'

Method

Participants

These were 89 children divided in two age groups: 4-year-old (n = 34, 14 female; *M* = 4 years, 6 months, *SD* = 3.16 months; range = 48 to 59 months) and 5-year-old (n = 55, 25

female; M = 5 years, 5 months, SD = 3.52 months; range = 60 to 71 months). Each child was tested in a quiet schoolroom during two 15-minute sessions separated by a month interval. Five other 4-year-olds were excluded: one because of specific language disability and four who failed control questions.

Procedure

Training phase. To ensure that the children understood thought bubbles that were to be used in the test stories, each child was told a story about twin brothers who were about to celebrate their fifth birthday illustrated using a sequence of three pictures: a boy with a dog inside a thought bubble to represent 'thinking about a dog,' a boy with a dog on a leash represent actually having a dog, and a girl drawn with an empty thought bubble to represent not knowing the information about the agent's state of mind. The story was retold to children who answered incorrectly on comprehension control questions. No child in the experiment failed the questions twice, consistent with previous studies on preschoolers' understanding of thought bubbles (Kerr & Durkin, 2004; Wellman, Hollander, & Schult, 1996).

Test Phase. Each child was told eight stories, two of which were each used to represent four possible combinations: knowing agent-positive outcome, unknowing agent-positive outcome, knowing agent-negative outcome, and unknowing agent-negative outcome (see Appendix). Half of the children were told four stories with the positive outcome first (two with agent foreknowledge and two without agent foreknowledge). Then in a second testing session about a month later, they received four negative outcome stories (again two with foreknowledge and two without). The remaining children received the four negative outcome stories in the first testing session and the positive ones in the second one. The order of presentation in the stories of agents possessing or lacking foreknowledge was counterbalanced across children. In all eight cases, the agent was described as not caring about the outcome as in Leslie et al.'s (2006) investigation. Illustrations were used to aid in storytelling with children asked control questions to monitor attention and understanding of the basic plot. The stories had two female and two male characters and two different types of animal (a frog and a gerbil) that one character, the agent, was described as having presented to a second character. Answers to each of the test questions were given a score of 1 if the child judged the side effect as intentional and a score of 0 if judged as unintentional. Thus for each of the two stories that represented the four combinations of foreknowledge and outcome, a child received a score on a 0-2 scale.

Results and discussion

Preliminary analyses indicated that there were no main or interaction effects attributable to the story theme, order of the story presentation, order of story valence, or order of test questions. Therefore responses were analyzed in a 2 (age group: 4- vs. 5-years-old) X 2 (valence outcome: positive vs. negative) X 2 (foreknowledge: present vs. absent) ANOVA, with age group as a between-subjects factor and valence outcome and foreknowledge as within-subjects factors. As shown in Figure 2.1, a SEE pattern emerged. There were significant main effects for valence outcome, $F(1, 87)= 53.51$, $p < 0.001$, $\eta^2 = .38$, and foreknowledge, $F(1, 87) = 22,31$, $p < 0.001$, $\eta^2 = 20$. More 'on purpose' judgments were given for negative than for positive side effects and for agents possessing rather lacking foreknowledge. The valence outcome X foreknowledge interaction effect was also significant, $F(1, 87) = 10.76$, $p < 0.01$, $\eta^2 = .38$, and foreknowledge, $F(1, 87) = 22,31$, $p < 0.001$, $\eta^2 = .20$. More 'on purpose' judgments were given for negative than for positive side effects and for agents possessing rather lacking foreknowledge. The valence outcome X foreknowledge interaction effect was also significant, $F(1, 87) = 10.76$, $p < 0.01$, $\eta^2 = .11$. When the outcome was negative, more on purpose judgments were assigned to the agent with foreknowledge ($M = 1.39$, $SD = .77$) than without foreknowledge ($M = .94$, $SD = .78$), $F(1, 87) = 23.08$, $p < 0.001$, $\eta^2 = .21$. By contrast, when the outcome was positive, the difference was insignificant ($M = 0.49$, $SD = .72$ and $M = 0.39$, $SD = .68$ respectively). There was considerable consistency in responses across conditions. The correlations between intentionality judgments of the agents with and without foreknowledge were significant (for positive outcome stories, $r = .73$, $p < .01$; for negative outcome stories, $r = .46$, $p < .01$). Altogether, 34 of 89 children unambiguously displayed an asymmetrical pattern in their intentionality judgments of the stories involving foreknowledge by judging the agents as having acted intentionally when causing harm on both 'knowing agent-negative outcome' stories but not having acted intentionally on both 'knowing agent-positive outcome' stories. Though about half (47.7%) of the judgments of agents without foreknowledge whose actions resulted in a negative outcome were intentional, there were indications that the children were not responding by chance alone. In the negative scenarios, 19 of the 25 children who judged in both stories that the agent without foreknowledge acted on purpose also judged that the agent with foreknowledge in both stories also acted on purpose.

Figure 2.1. Children's intentionality judgments in positive and negative scenarios of Experiment 6 for uncaring agents whose foreknowledge of the outcome was present or absent
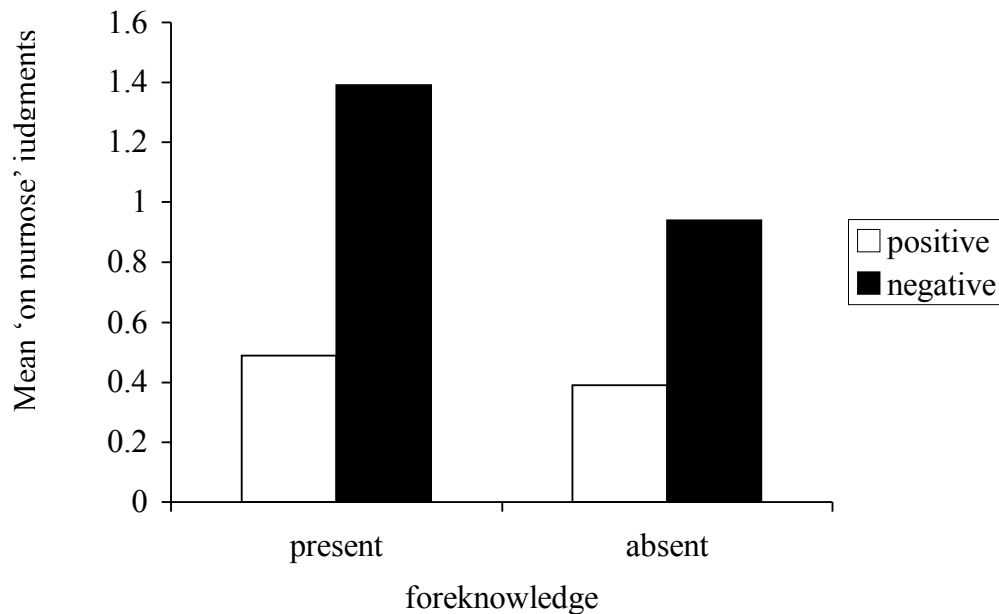
As Leslie et al. (2006) found for the SEE, preschoolers often judge agents' actions as intentional when there is a disavowed (uncaring) negative outcome but not when there is a positive outcome. However, the presence of substantial individual differences supports the notion of 'interpretative diversity' of intentionality proposed by Nichols and Ulatowski (2007) that involves both foreknowledge and motive. On the one hand, children's judgments often focused on the presence or absence of foreknowledge in an uncaring agent ('Andy does not care that Janine will get upset'). Children were more likely to judge as intentional a negative effect produced by an agent with foreknowledge than one produced by an agent without foreknowledge. On the other hand, although intentionality judgments decreased significantly when the agent was described as not knowing about the effect, a substantial number of children displayed the SEE even in the absence of foreknowledge. They displayed a judgmental pattern that appeared to be mainly based on agent motive in relation to declarations of uncaring. In Experiment 7, we sought to test the strength of the asymmetrical pattern shown in Experiment 6 by determining whether omitting information about the agent's state of caring would influence children's intentionality judgments. We predicted that judgments would be influenced by knowledge only when the outcome was foreknown and negative.

Experiment 7: 'Lack of knowledge'

Method

Participants

These were 46 children divided in two age groups: 4-years-old (n = 24, 11 female; *M* = 4 years and 5 months, *SD* = 4.20; range = 48 to 59 months) and 5-years-old (n = 22, 18 female; *M* = 5 years and 4 months, *SD* = 3.5 months; range = 60 to 71 months). Each child was tested in two 15-minute sessions separated by a month interval. One other 4-year-old who failed control questions was excluded.

Procedure

This was the same as in Experiment 6 except that no mention was made of the disinterest of the actor ('not caring') in the outcome.

Results

Preliminary analyses indicated that there were no main or interaction effects attributable to the story theme, order of the story presentation, order of story valence, or order of test questions. Therefore the results were analyzed in a 2 (age group: 4- vs. 5-years-old) X 2 (valence outcome: positive vs. negative) X 2 (foreknowledge: present vs. absent) ANOVA with age group as a between-subjects factor and valence outcome and foreknowledge as within-subject factors. There were significant main effects for valence outcome, $F(1, 44)$ = 27.46, $p < 0.001$, η^2 =.38, and foreknowledge, $F(1, 44)$ = 96.11, $p < 0.001$, η^2 = 69, together with a significant valence outcome X foreknowledge interaction effect, $F(1, 44)$ = 64.47, $p < 0.01$, η^2 = .59. As predicted, intentionality judgments were influenced by foreknowledge only when the outcome was negative. As shown in Figure 2.2 with regard to negative outcomes, children were significantly more likely to assign on purpose judgments when foreknowledge was present (*M* = 1.54, *SD* = .75) than absent (*M* = .17, *SD* = .48), $F(1, 44)$ = .81, $p < .81$, =.001. When the reaction was positive, the difference was insignificant (*M* = 0.30, *SD* = 0.69 and *M* = .28, *SD* = .58 respectively), $F(1, 44)$ = 111.83, $p < 0.001$, η^2 = .71. In the case of negative outcomes, we compared children's responses in Experiments 6 and 7. A 2 (age group) X 2 (caring state: uncaring vs. unspecified) X 2 (foreknowledge: present vs. absent) ANOVA with age group and caring state as between-subjects factors and knowledge state as a within-subjects factor revealed a significant main effects for caring state, $F(1, 131)$ = 8.55, $p < 0.01$, η^2 =.06, and foreknowledge, $F(1, 131)$ = 136.36, $p < 0.001$, as well as a significant caring state X foreknowledge interaction effect, $F(1, 131)$ = 36.07, $p < 0.001$, η^2 =.22. When there was foreknowledge of the potential harm was absent, intentionality was attributed

significantly more often to the uncaring agent than to the agent for whom caring was unspecified, $F(1, 134) = 36.59$, $p < 0.001$. When foreknowledge was present, the difference between the agents was insignificant, $F(1, 134) = 1.15$, $p > .28$. There were no significant main or interaction effects involving age group.

Figure 2.2. Children's intentionality judgments in positive and negative scenarios of Experiment 7 for agents with unspecified caring whose foreknowledge of the outcome was present or absent.
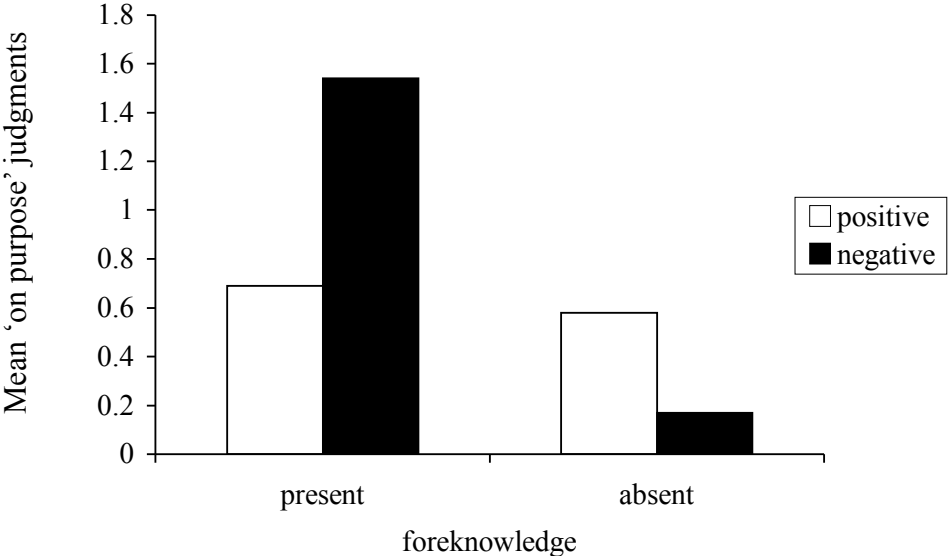
Experiment 8: 'False belief'

In both Experiments 6 and 7, there was substantial asymmetry in intentionality judgments with children often responding that negative effects were brought about on purpose. Only in the case of an agent described as without foreknowledge and whose state of caring was unspecified did children generally respond that the effect was not brought about on purpose. Therefore, the aim of Experiment 8 was to examine the effects on intentionality attributions of an agent explicitly described as having a false belief about the outcome of his or her actions. In contrast to Experiment 7 in which the negative outcome of a knowledgeable agent's actions was judged as purposeful, we predicted that, in the case of a agent with a false belief, children would be unlikely to attribute intentionality regardless of outcome and that, even if the outcome was negative, children would be likely to exonerate the agent.

Participants

These were 52 children divided into two age groups: 4-year-old (n = 26, 15 female; *M* = 4 years, 6 months, *SD* = 2.7 months; range = 51 to 59 months) and 5-year-old (n = 26, 11 female; *M* = 5 years, 5 months, *SD* = 4.4 months; range = 60 to 71 months). Each child was tested in an experimental session lasting about 15 minutes. No children were excluded because of failure on control questions.

Procedure

This was similar to Experiments 6 and 7 except that children received four stories instead of eight (two with a positive outcome and two with a negative outcome). In these stories, we described the agent as having a false belief about the other character's reaction that was in reality negative in the case of an agent who had believed that it would be positive and was in reality positive in the case of an agent who had believed that it would be negative. As in Experiment 7, there was no mention of the uncaring state of the agent. To simplify the story presentation, no indication was given about the actual outcome in terms of the reaction of the second story character.

Results

There were few 'on purpose' responses to the agents with false beliefs: 19.2% in the positive outcome condition and 11.5% in the negative outcome condition. Preliminary analyses indicated that there were no main or interaction effects attributable to the story theme, order of the story presentation, order of story valence or order of test questions. The results were analyzed in a 2 (age group: 4- vs. 5-years-old) X 2 (valence outcome: positive vs. negative) repeated measures ANOVA. There was a main effect of age, $F(1, 50) = 5.65$, $p < 0.05$, $\eta^2 = .10$, but no main or interaction effect involving outcome, F's < 1.7. In keeping with findings that preschoolers' success on standard false belief tasks increases with age (Callaghan et al., 2005; Siegal, 2008), 5-year-olds were significantly more likely than 4-year-olds to absolve the story character with a false belief of intentionality regardless of whether the outcome was positive or negative (5 years: *M* = .28, *SD* = .58; 4 years: *M* = .66, *SD* = .83, respectively). There was a significant correlation between the children answer to the false belief stories with a positive and negative outcome ($r = .36$, p < 0.05). To compare the results of Experiment 8 with those of the foreknowledge present (true belief) condition of Experiment 7, we carried out a 2 (age group) X 2 (belief: true vs. false) X 2 (valence

outcome: positive vs. negative) ANOVA with age group and belief as between subjects factors and valence outcome as a within-subjects factor. There was a significant main effect for outcome, $F(1, 97) = 64.32$, $p < 0.001$, $\eta^2 = .39$ and a significant outcome X belief interaction effect, $F(1, 97) = 16.45$, $p < 0.001$, $\eta^2 = .14$. As shown in Figure 2.3, when the outcome was negative, children were significantly more likely to judge as intentional the action of an agent with a true belief ($M = 1.55$, $SD = .74$) than the action of an agent with a false belief ($M = .38$, $SD = .69$), $F(1, 98) = 65.49$, $p > .001$, $\eta^2 = .40$. When the outcome was positive, the difference between the agent with a true belief ($M = .30$, $SD = .68$) and the agent with a false belief ($M = .54$, $SD = .80$) was insignificant, $F(1, 98) = 2.97$, $p > .08$, $\eta^2 = .03$.

Figure 2.3. Children's intentionality judgments in positive and negative scenarios of Experiment 8 for agents with unspecified caring whose foreknowledge of the outcome was true or false
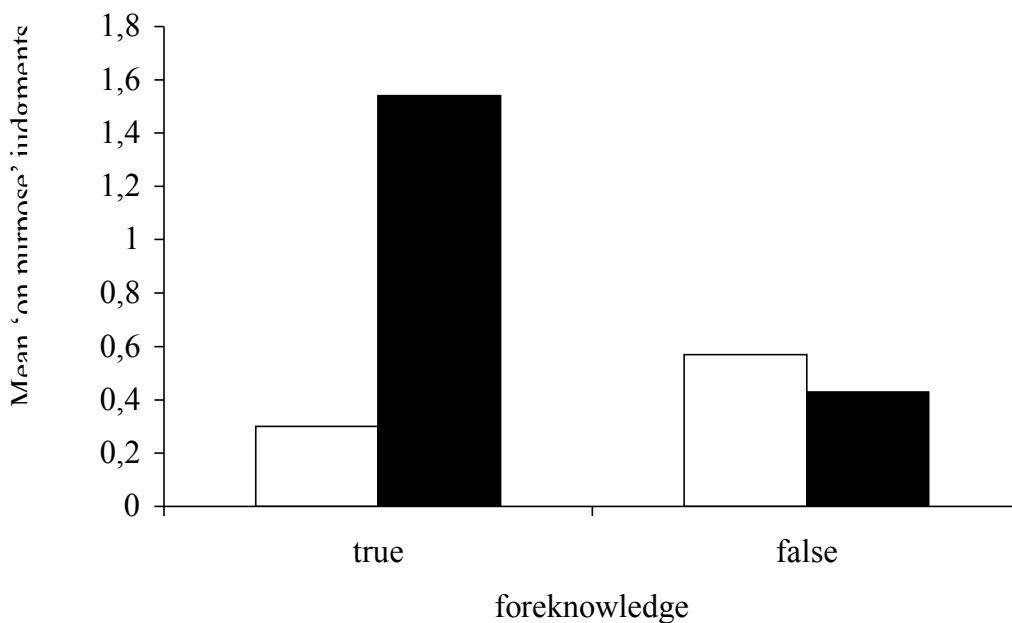
General Discussion

The aim of this investigation was to examine the influence of agent foreknowledge and uncaring declarations in children's asymmetrical intentionality judgments of actions with positive and negative outcomes. Experiment 6 compared 'on purpose' judgments of agents who either possessed foreknowledge of the effects of their actions or did not, following

declarations of not caring. When a careless agent knowingly produced a negative side effect, young children – as do adults (Knobe, 2004) – often judge this effect as intentional, consistent with the proposal of continuity in the SEE. A lack of foreknowledge produced a significant decline, but many children still persisted in responding that the negative effects were brought about intentionality. The pattern was asymmetrical in that, for the positive stories, the agent's action was mostly judged as unintentional whether foreknowledge was absent or present. In Experiment 7, unspecified caring on the part of the story perpetrator was contrasted with the SEE shown in Experiment 6. Only when a negative outcome was foreknown were there strong judgments of intentionality that contrasted with a corresponding lack of intentionality attributions for the effects caused by an agent who lacked foreknowledge. This finding was replicated in Experiment 8 in which children infrequently judged an agent with a false belief who produced a negative outcome as having acted purposely.

Our investigation with 4- and 5-year-olds was limited in that no detailed rationale was provided for the uncaring declarations of the agent and the judged effects were restricted to a single instance of a child's happiness or upset reaction. Bearing these issues in mind, our results suggest that many preschoolers as well as adults are capable of a rational analysis of situations involving helping or harming by using information about a protagonist's mental state, both in terms of caring and foreknowledge. An area in need of further study concerns the nature of continuity in intentionality judgment asymmetry and the extent to which considerations of caring and foreknowledge remains stable with age. In terms of the debate over the primacy of cognitive and emotional factors in moral judgment (Danovitch & Keil, 2008; Greene & Haidt, 2002; Haidt, 2001; Nichols & Mallon, 2006; Pizarro & Bloom, 2003), another important issue concerns how asymmetry relates to variations in moral orientations across cultures (Shweder, Much, Mahapatra, & Park, 1997), and the extent to which asymmetrical patterns can mediated by the strength of reactions involving emotions such as fear and disgust.

3. Utilitarian moral reasoning

For decades, moral cognition has been described as a developmental process that takes the form of an invariant sequence of stages (Kohlberg, 1969; Piaget, 1932). Yet while classic models posit stage-like discontinuities, more recent accounts have proposed that certain significant aspects of moral understanding are present in young children, at least from the age of 4 years (Nuñez & Harris, 1998; Siegal & Peterson, 1998; Turiel, 2006). For example, studies of the 'side-effect effect' in which a foreseen action produced by an uncaring protagonist results in a negative or positive outcome indicate that both children and adults are significantly more likely to judge the negative effect than the positive one as having been brought about intentionally (Knobe, 2005; Leslie, Knobe, & Cohen, 2006).

However, to date, young children's moral understanding has not been investigated in connection with the burgeoning research on adults that concerns the factors, or principles, that affect the production of utilitarian moral judgments. These studies have typically asked adults to solve dilemmas such as the 'trolley' or 'footbridge' dilemmas that involve sacrificing one person in order to save five others (Foot, 1978; Thomson, 1986). By systematically varying some aspects of these dilemmas, Cushman, Young, & Hauser (2006) have found that adults rate harm caused by an action as less permissible than harm caused by an omission (the action principle), harm caused as a mean to an end as worse than harm caused as a foreseen side effect (the intention principle) and harm produced by physical contact worse than harm caused without involving physical contact (the contact principle; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Hauser, 2006; Hauser, Cushman, Young, Jin, & Mikhail, 2007; Hauser, Young, & Cushman, 2007; Nichols & Mallon, 2006). Contrary to what one would expect if persons strictly follow an utilitarian calculus, adults' adherence to these principles is found also when the cost-benefit ratio of the relevant decisions and actions is kept constant.

Departing from traditional models that emphasize the conscious nature of moral reasoning (Kohlberg, 1969; Piaget, 1932), Cushman et al. (2006) propose that only some of the processes that guide moral judgments are conscious. In this respect, Greene et al. (2001, 2004) have underscored the importance of alternative 'emotional' factors, such as the degree of the agent's personal involvement. They have reported that utilitarian evaluations are less frequent in personal than in impersonal scenarios. These proposals are in keeping with the tradition that emphasizes the role of emotional processes in moral evaluation (Haidt, 2001; Hume, 1777/1960).

So far these aspects of moral cognition have been investigated mainly in adults coming from English speaking countries. The Experiments 9 and 10 were aimed to observe the way Italian adults answers to the Trolley and the Footbridge dilemma, both in a versions for adults and in one suitable for children. The Experiments 11, 12 and 13 were designed to determine whether young children, like adults, are sensitive to utilitarian cost/benefit ratios and whether

they rely on the contact or intention principles when judging the permissibility of interventions.

3.1 Utilitarian moral reasoning in adults

3.1.1 Experiment 9: 'Utilitarian judgment in adults'

The Experiment 9 was designed to verify if the asymmetric pattern of answers, shown by English speaking participants, was present also in an Italian sample. In Experiment 9 we differentiate between the dilemma proposed by Greene et al. (2004) and by Cushman et al. (2006) because they proposed different way to address the test question: if in Greene et al.'s study the question was more neutral, they asked about the appropriateness of the action (Is it appropriate for you to push the stranger on to the tracks in order to save the five workmen?), the Cushman et al.'s (2006) proposal was oriented on the moral permissibility of the action (Is it morally permissible for Frank to shove the man?).

Method

Participants.

The adult sample included 82 participants (51 female, $M = 29.34$, $SD = 6.8$ years, ranging from 18 to 58 years). Participants were all Italian native speakers from the Northern part of the country, they have various socioeconomic and educational backgrounds. All subjects were tested individually in a silent room.

Procedure

The participants were randomly assigned to one of the four conditions that result from the combination of two types of test question (Greene version vs Cushman version) and two types of dilemma (Trolley vs Footbridge). Fifteen participants were asked to read an Italian version of the 'Trolley dilemma' proposed by Greene et al. (2001), while fifteen participants were asked to read an Italian translation of the 'Footbridge dilemma' Greene et al. (2001). Twenty-six participants were asked to read an Italian version of the 'Trolley dilemma' proposed by Cushman et al. (2006) while twenty-six to read an Italian version of the 'Footbridge dilemma' Cushman et al. (2006). The participants responded to the scenarios in a questionnaire format. They indicated the moral permissibility of one of two courses action to be taken by the agent by ticking a box at the end of the story. The presentation orders of the alternatives were counterbalanced across participants.

Results and discussion

In Greene's version of the dilemmas 1 participant (0.7%) judged as appropriate to do the action in the Footbridge scenario (pushing the stranger on the tracks), while 14 participants (93.3%) judged as appropriate to do the action (switching the direction of the trolley) in the Trolley dilemma. The differences were statistically significant, χ^2 (1, N = 30) = 22.5, $p <$ 0.001. In Cushman's version of the dilemmas 4 participants (15.4%) judged as morally permissible to do the action (showing the man), while 26 participants (100%) judged as permissible do the action (switching the direction of the trolley) in the Trolley dilemma. The difference was statistically significant, χ^2 (1, N = 52) = 38.1, $p <$ 0.001. The way participants answered to the dilemmas, proposed in Greene and Cushman versions, did not differ statistically, Fisher Exact Probability Test, all p's $>$.36.

As shown in Figure 3.1, the data from Experiment 9 give support to the documented asymmetric pattern of answers found in previous studies on adults (Greene et al., 2001; Cushman et al., 2006), showing that the Trolley and the Footbridge dilemma seem to elicit different types of cost-benefit evaluation.

Picture 3.1. Percentages of adults advocating intervention in the footbridge and trolley dilemmas in Experiment 9
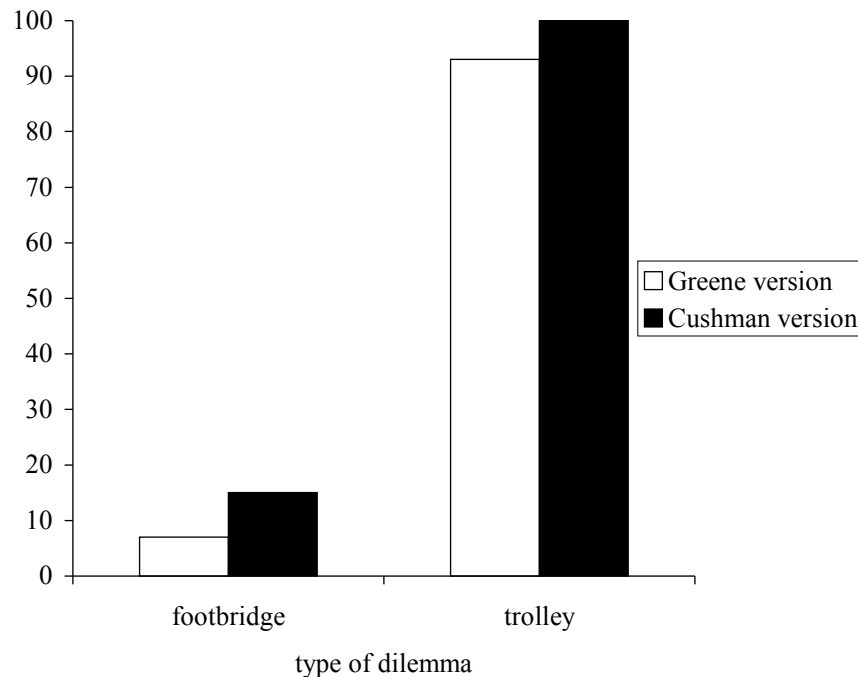
In Experiment 10 the participants were given a version of the Trolley and the Footbridge dilemma suitable for children. This was done particularly in order to have the possibility to compare children and adults on the same type of tasks.

3.1.2 Experiment 10: 'Utilitarian judgment in stories suitable for children'

Similarly to the Experiment 9, in the following Experiment adults were given a version of the dilemmas with an implicit test question (based on Greene proposal) and an explicit version of the test question (based on Cushman proposal).

Method

Participants

The adult sample included 56 participants (23 female, $M = 41.53$, $SD = 10.6$ years, ranging from 22 to 71 years).

Procedure

Participants were divided into four experimental groups that result from the combination of two types of test question (morally implicit vs morally explicit) and two types of dilemmas (Trolley vs Footbridge). Twelve subjects were asked to read an Italian version suitable for children of the 'Trolley dilemma' while sixteen subjects were asked to read an Italian version, suitable for children, of the 'Footbridge dilemma'. The test question, in both scenarios, was morally implicit (see Experiment 11).

Twelve subjects were asked to read an Italian version suitable for children of the 'Trolley dilemma' while sixteen were asked to read a version suitable for children of the 'Footbridge dilemma'. The test question, in both scenarios, was morally explicit (see Experiment 12). The participants responded to the scenarios in a questionnaire format. They indicated the appropriateness of one of two courses action to be taken by the agent by ticking a box at the end of the story. The presentation orders of the alternatives were counterbalanced across participants.

Results and discussion

When the test question was morally implicit, 5 participants (31 %) choose the option 'do the action' in the Footbridge scenario (pushing the fat man on the street), while 11 participants (91%) choose the option 'do the action' (switching the direction of the ball) in the Trolley dilemma. The difference was statistically significant, $\chi^2 (1, N = 28) = 10.2$, $p < 0.001$.

When the test question was morally explicit, 5 participants (31%) judged right to do the

action (pushing the fat man on the street), while twelve participants (100%) judged as right to do the action (switching the direction of the ball) in the Trolley dilemma. The difference was statistically significant, Fisher Exact Probability Test, $p < 0.001$.

The adult's answers did not statistically differ in relation to the type of test question (morally implicit or morally explicit) nor for the version of the story (suitable for adult or for children), Fischer Exact Probability Test, all $p's > 0.17$.

Experiment 10 show that the versions of the trolley and footbridge scenarios suitable for children seem to provoke an asymmetrical evaluation on adults' judgment similar to the one observed in Experiment 9. Kipping this in mind, the second part of the study has the aim to evaluate the way in which children answer to the footbridge and trolley dilemma.

3.2 Utilitarian moral judgments in young children

3.2.1 Experiment 11: 'What John should do?

Method

Participants

These were 93 children divided into age groups: 15 3-years-olds (6 girls) aged between 36 to 44 months ($M = 40.8$ months, $SD = 2.4$), 29 4-years-olds (9 girls) aged between 48 and 59 months ($M = 54.9$ months, $SD = 3.3$), and 49 5-years-olds (23 girls) aged between 60 to 71 months ($M = 65.6$ months, $SD = 3.5$). One other 4-year-old was excluded because of severe problems in language comprehension. The children in our studies attended kindergartens located in middle class urban areas. All received informed parental consent for participating. No child participated in more than one experiment.

Materials and Procedure.

Two wooden models were used to illustrate the dilemmas (Figure 3.2).

Footbridge dilemma: The model for the footbridge dilemma was 25.5 X 30 X 37 cm in size and consisted of a 45° inclined plane with a straight track and a footbridge above. At the end of the track were five Lego play-people. Standing on the footbridge were two other play-people: a small one (the main story character, John) and a big one (the potential victim). The dilemmas were given to the children as follows:

A big ball is running quickly along this track [the experimenter demonstrated the event using the model]. On the track, there are five persons that do not see the ball rolling down because they are looking in front of them.

Control Question 1: Do these persons see the ball rolling down the street?

The ball is rolling down on this street and it will soon badly hurt five persons. John is on a footbridge between the big ball running and the five persons. On the footbridge, close to John there is a big person. The person does not see the ball rolling down because he is looking in front of him.

Control Question 2: Does this person see the ball rolling down?

John knows that the only way to stop the big ball is to drop a very heavy weight into the street. John knows that if he pushes down the big person close to him, the big person will be badly hurt but the other 5 persons will be safe.

Test Question: What John should do? Push the person or not to push the person? Point to the picture showing what John should do.

Figure 3.2. Materials used to illustrate (a) the footbridge dilemma and (b) the trolley dilemma.

(a)   (b)
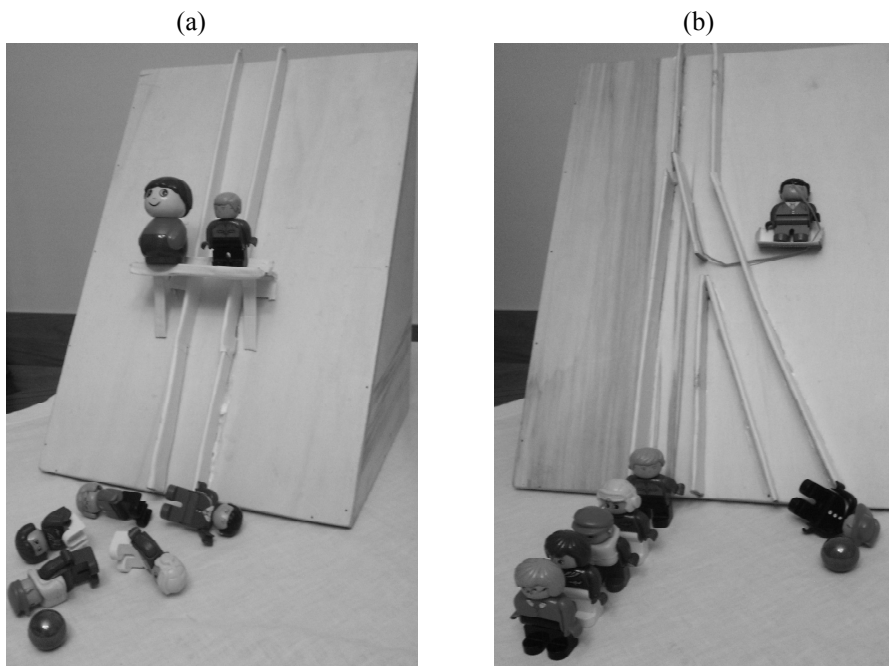
Trolley dilemma. The model for the trolley dilemma was identical in size but on the inclined plane a main track ran straight through the middle and a secondary track ran off the main one. At the junction of the tracks (23 cm from the top of the wooden structure) was a gate (10.5 cm in length). By pulling the string attached to the gate, access to the second part of the main track was blocked and the ball was diverted to roll down the

secondary one. The main story character, Albert, was shown holding the string. Five play-people were at the end of the main track and one was at the end of the secondary track. The dilemma was presented as follows:

> A big ball is running quickly along this track [the experimenter demonstrates the event using the model]. On the track, there are five persons that do not see the ball rolling down because they are looking in front of them.

> Control Question 1: Do these persons see the ball rolling down the street?

> The rolling ball will soon hurt five persons badly. Albert is near the street and he sees the big ball rolling down. He also sees the five persons that are going to be hurt by it. Albert knows that by pulling this string he can make the ball go into another track where there is one person. The person does not see the ball running because he is looking in front of him.

> Control Question 2: Does this person see the big ball rolling down?

> If the ball rolls onto this track, the person will be badly hurt, but the other five will be safe.

> Test Question: What Albert should do? Should he pull the cord or not pull the cord? Point to the picture showing what Albert should do.

The testing for the children took place in two sessions lasting approximately five minutes. Half the children were given the footbridge dilemma first and, about a month later, the trolley dilemma. The others received the dilemmas in the reverse order. The experimenter asked the children the test questions using a neutral tone of voice and illustrated the consequences of each choice using two drawings.

Results and Discussion

Preliminary analyses indicated that there were no significant differences attributable to gender of the participants, story theme, order of story presentation, or order of presentation of the test alternatives. As shown in Figure 3.3, most children advocated action in the trolley dilemma (87%), but only a few did so in the footbridge dilemma (27%). Of the 93 children, 11 responded that the agent should not act in both dilemmas and 24 said gave positive responses to both dilemmas. A majority of 57 children advocated that the protagonist should

act in the trolley dilemma but said that he should refrain from acting in the footbridge dilemma. In contrast, only a single child showed the reverse pattern, McNemar test, χ^2 (1, N = 93) = 52.15, $p < .0001$.

There was no significant difference among age groups (Fisher's exact test, all p's > .16). Similarly, on the footbridge dilemma, only 5 of 16 adults (31%) judged as appropriate do the action of pushing a man, while on the trolley dilemma, 11 of 12 (91%) judged that the action of switching the direction of the trolley was appropriate (Experiment 9). Adults' choices between the alternatives in the scenarios closely resembled those of the children. On either dilemma, there were no significant differences between children and adults, Fisher Exact Probability Test, p's > .14).

Figure 3.3. Percentages of children in each age group and adults advocating intervention in the trolley and footbridge dilemmas in Experiments 11 and 9.
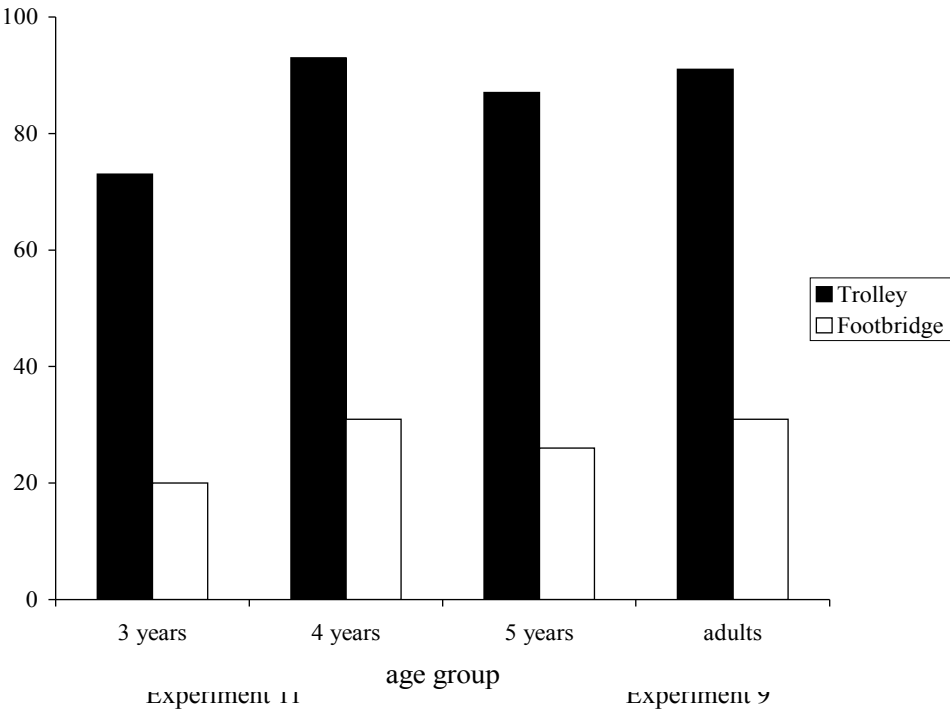
These results closely resemble those found in adult studies (e.g., Greene et al., 2001; Cushman et al., 2006). However, it may be suspected that the moral aspect of the test question is implicit in that the children in Experiment 11 were not overtly asked to indicate choices as right or wrong but rather than a generic question concerning what the protagonist should do. Possibly, the similarity with adult performance may reduce or vanish if children were asked more explicitly to say whether the hypothesized action

was right or wrong. Therefore, Experiment 12 assessed the robustness of the effects found in Experiment 11 under different testing conditions.

3.2.2 Experiment 12: 'What is the right thing for Jonh to do?'

Method

Participants

These were 62 children: 21 3-year-olds (11 girls) aged between 41 and 47 months (*M* = 44.8 months, *SD* = 1.8), 19 4-year-olds (13 girls) aged between 48 and 59 months (*M* = 53.5 months, *SD* = 3.9), and 22 5-years-old (18 girls) aged from ranged from months 60 to 71 months (*M* = 65 months, *SD* = 3.4). The children in our studies attended kindergartens located in middle class urban areas. All received informed parental consent for participating.

Procedure

The procedure was identical to Experiment 11 except, as in Cushman et al. (2006), the test question for the footbridge dilemma was more explicitly focused on the moral value of the action, 'What is the right thing for John to do? To push the person or not to push him? Point to the picture showing what it is right for John to do.' For the trolley scenario, the corresponding question was, 'What is the right thing for Albert to do? To pull the cord or not to pull it? Point to the picture showing what it is right for Albert to do.' The order in which the two possible response alternatives were mentioned was counterbalanced across children.

Results and Discussion

Preliminary analyses indicated that there were no significant differences attributable to gender of the participants, story theme, order of story presentation, or order of presentation of the test alternatives. The results were very similar to those in Experiment 11 and, again, were consistent with the principle of contact (Cushman et al., 2006) and the personal/impersonal distinction (Greene et al., 2001). As shown in Figure 3.4, most children advocated intervention in the trolley dilemma, but only a few of them did so in the footbridge dilemma (88% and 25%, respectively). Of the 62 children, 7 gave negative answers in both dilemmas, 16 gave positive answers in both dilemmas, 39 responded that it was right to act in the trolley dilemma and not to act in the footbridge task, and no child showed the reverse pattern, McNemar test, χ^2 (1, N = 62) = 37.03, p < .0001.

The age groups did not differ significantly in their responses to the two dilemmas (Fisher's exact test, all $p's$ > .31). Similarly, on the footbridge dilemma, only 5 of 16 adults (31%) judged as appropriate do the action of pushing a man, while on the trolley dilemma, 12

of 12 (100%) judged that the action of switching the direction of the trolley was appropriate (Experiment 10). Adults' choices between the alternatives in the scenarios closely resembled those of the children. On either dilemma, there were no significant differences between children and adults (Fisher Exact Probability Test, p's > .06).

Figure 3.4. Percentages of children in each age group and adults advocating intervention in the trolley and footbridge dilemmas in Experiment 12 and 10.
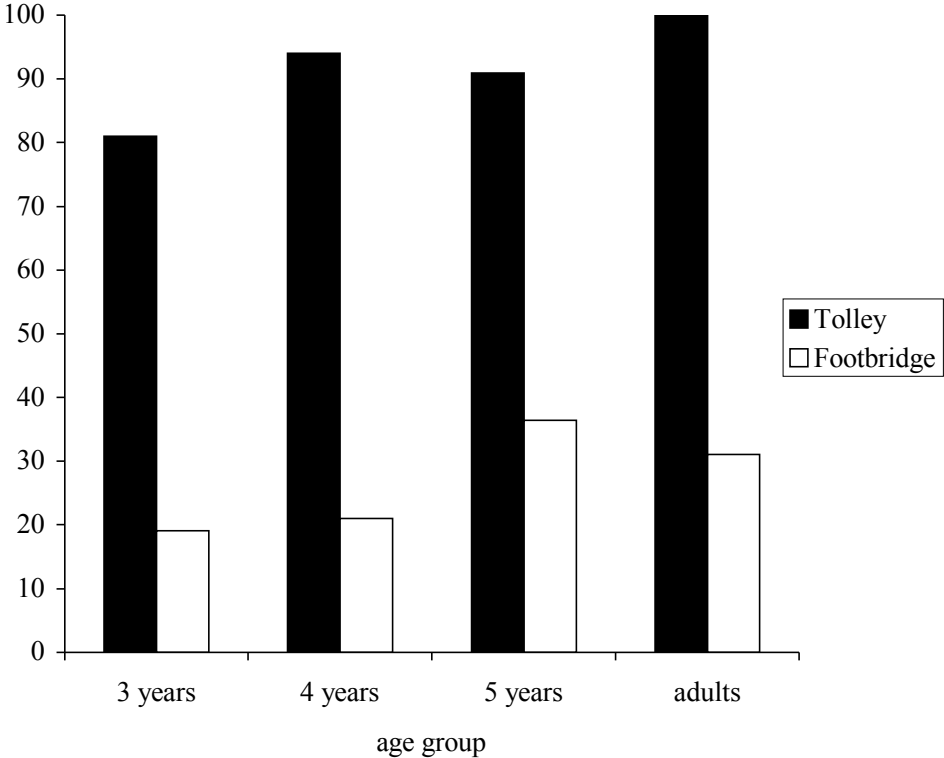
Advocating that it is right to refrain from acting in a personal scenario may indeed be due to a greater involvement of emotional processes, as suggested by Greene et al. (2001). However, the interpretation that children's advocacy of intervention in impersonal dilemmas reflects a utilitarian heuristic requires clarification. Rather than employing a cost/benefit analysis, children may have followed a heuristic such as 'something should be done when there is a situation that places people in danger.' This heuristic would then be blocked in the personal scenario because of emotional activation (Haidt, 2001), but remain available in the other scenario. To test this hypothesis, we carried out a third experiment in which adherence to a cost/benefit analysis in a modified trolley dilemma should lead to the rejection of the action rather

than its endorsement.

3.2.3 Experiment 13: 'Investigating ad-hoc something must be done euristic'

Method

Participants

These were 52 children divided into two age groups: 26 4-year-olds (15 girls) aged between 51 and 59 months ($M = 55.9$ months, $SD = 2.7$) and 26 5-years-old (11 girls) aged from ranged from months 60 to 72 months ($M = 66.1$ months, $SD = 4.4$). The children in our studies attended kindergartens located in middle class urban areas. All received informed parental consent for participating.

Procedure

In this experiment, only one story was told to the children. The story and the test question were identical to those used in the trolley dilemma of Experiment 12 except for one crucial difference: the action in this modified version as illustrated by the Lego people and pictures would save one person but would sacrifice five.

Results and discussion

Significantly more children overall (67%) choose 'not to act' than 'to act' in this 'inverted' Trolley dilemma, χ^2 (1, N = 52) = 6.23, $p < .05$ (see Figure 3.5). In contrast to Experiment 12 where the large majority of children (88%) judged that it was right to act to save five persons by sacrificing one, a significantly smaller number (32%) of children in Experiment 13 judged that it was right to act to save one person by sacrificing five, χ^2 (1, N = 114) = 34.14, $p < .01$.

Figure 3.5. Percentages of children in each age group and adults advocating intervention in the trolley and footbridge dilemmas in Experiment 12 and 13.
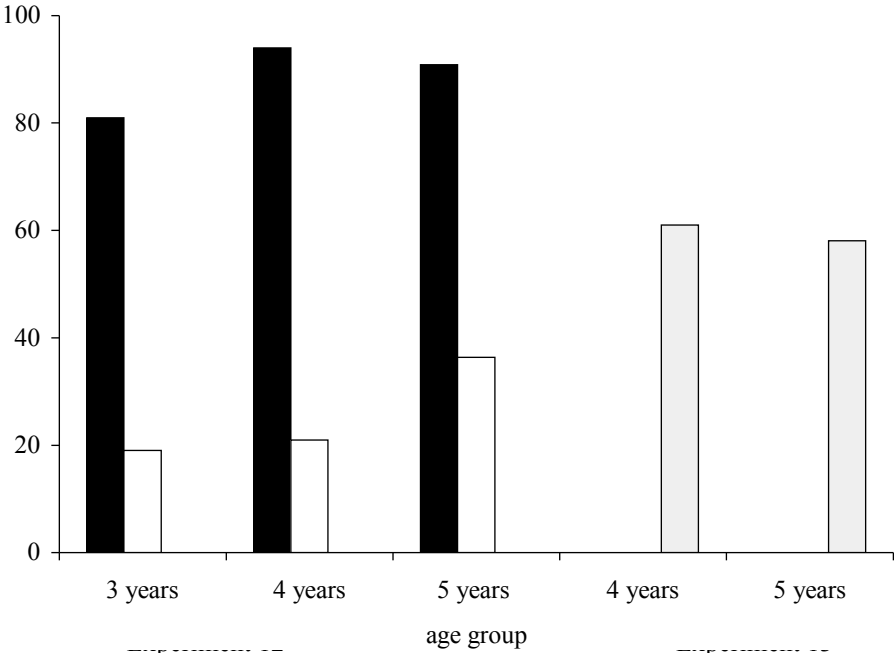
The age groups did not differ significantly in their responses, Fisher's Exact Probability Test, $p = .34$. In sum, consistent with utilitarian cost/benefit analyses of the consequences, children displayed opposite patterns of responses in Experiments 12 and 13.

General discussion

In the present investigation, we tested typically developing preschoolers on simple versions of the trolley and the footbridge scenarios, two moral dilemmas that have been extensively used in the recent literature on adult moral psychology (Cushman et al., 2006; Greene et al., 2001). We found that, like adults (Experiment 9 and 10), most children by the age of 3 or 4 years generate utilitarian judgments in the trolley dilemma, but not in the footbridge dilemma both when asked what an agent should do (Experiment 11) and when asked whether an action was the right thing to do (Experiment 12). Moreover, children did not simply follow a blind 'something must be done' heuristic in a trolley scenario (Experiment 13). They endorsed intervention in that dilemma only when it was justified by a cost/benefit analysis.

Our investigation was limited in that it did not address the issue of whether the strong difference in the frequency of utilitarian responses given by children to the dilemmas was due to their reliance on the contact principle or, alternatively, on the intention principle (Cushman et al., 2006) or whether it resulted from reactions to differences in the story characters' personal involvement (Greene et al., 2001). Another possible explanation for the pattern of results is that children relied on the distinction between an action that redirects a threat and an action that introduces a threat. Further research is required to clarify whether the similarity in the responses of young children and adults on utilitarian moral dilemmas reflects similarity in the cognitive processes and representations that underlie such judgements.

Continuity in moral cognition, as well as in other domains of cognitive and language development, suggests that the development of morality is strongly canalized and may be based on domain-specific developmental mechanisms (Dwyer, 1999; 2007; Hauser, 2006; Rawls, 1971), rather than domain-general mechanisms that require elaborate theory revision processes (Kohlberg, 1969). However, the children tested here were aged 3 to 5 years and constructivist models have been proposed for sophisticated competencies that appear very early in development (e.g., Gopnik & Meltzoff, 1997). Ultimately, the choice between competing explanations of acquisition in the moral domain will depend on a complex set of evidence not only from precocious skills, but also from learnability studies and neuropsychological investigations that will involve both typically developing children and atypical groups such as children with autism (Hauser, Young, & Cushman, 2007; Leslie, Mallon, & Di Corcia, 2006) or psychopathic tendencies (Blair, Perschardt, Budhani, Mitchell, & Pine, 2006).

4. Conclusion and new perspective

In the introduction to this thesis, three main questions were raised concerning the nature of moral psychology:

1.1 What is the nature of moral judgment?
1.2 How does moral reasoning develop?
1.3 Which are relationship between moral judgment and intentionality?

The aim of the research reported here was to provide a bases for addressing these questions by (1) analyzing the relation between intentionality judgments and moral evaluation in the SEE as shown by children and adults, and (2) evaluating differences between children and adults in participants' computation of information on measures of utilitarian moral judgments (3) addressing the continuity vs discontinuity issue regarding moral development.

4.1 Intentionality judgment and moral evaluations that result in the SEE

Concerning the first theme, the SEE is clearly present in young children. We based our research on previous studies carried out by Knobe (2003a) and Leslie et al. (2006) and replicated their results with an Italian sample of adults and preschoolers. The experimental work sought to explore conditions that determine the effect and, although our results are not conclusive, these can help to offer a new perspective on the theme in demonstrating asymmetry in moral judgments. The crucial aspect on which both adults and children base their intentionality evaluations is the agent's foreknowledge of the effect of the action. In situations where the agent had a false belief (Experiments 5 and 8) or did not have foreknowledge of the valence of the outcome (Experiments 4 and 7), participants often formulated negative intentionality judgments. For this reason, we stress the importance of computing information on the epistemic status of the agent to produce an intentionality judgment. As in previous research on the development of moral evaluation in relation to foreknowledge (Nuñez & Harris, 1998; Siegal & Peterson, 1998), our results suggest that early in life it is possible to understand and give meaning to an agent's behavior on the base of his/her epistemic status the in same way as adults (Aristotle 350/1998; Heider, 1958;

Schlenker, Britt, Pennington, Murphy, & Doherty, 1994).

According to Leslie et al. (2006, p. 426), 'As soon as children are old enough to correctly attribute 'not caring,' they begin showing the pattern characteristic of the adult 'side effect' effect.' In our research, we clarified that the nature of the SEE in terms of uncaring attitude of the agent. In research on adults (Experiment 3), the SEE persists when sentences such as 'I do not care if the environment will be harmed, but we must increase profits. Let's start the new program' is substituted with the sentence 'I am sorry if the environment will be harmed, but we must increase profits'. Similarly, children continued to produce the SEE when the sentence 'Andy does not care if Janine will get upset' is omitted (Experiment 7).

The surprising, counterintuitive aspect of the SEE lies in participants' asymmetrical attribution of intentionality when a disavowed side effect is negative but not when it is positive (Knobe, 2003a). From our results, it emerges that, if the agent foreknew the negative side effects of his actions, participants were likely to make an intentionality attribution. In other words, the effect emerges if the participants are informed about the agent's epistemic mental state, more than the careless attitude, of the agent. However, when participants are presented with ambiguous situations, in which the agent could not have foreknown the outcome (Experiment 4) or the agent makes declarations of uncaring (Experiment 6), about half of the participants attributed intentionality to the agent. As for adults, when children are unaware about the agent's foreknowledge of the outcome the sentence 'I do not care' has a strong impact on the attribution of intentionality.

Now we come to the crucial question: does the SEE bear merely on people's performance with the concept of intentional action or does it cast some light on competence with this concept? Knobe (2006) and Nichols and Ulatowski (2007) claim that responses to measures of the SEE reflect participants' grasp of the concept of intentionality. By contrast, others propose that the effect results from factors beyond what is constitutive of possessing this concept (Adams & Steadman; 2004a; 2004b; Machery, forthcoming; Nadelhoffer, 2006). Adams and Steadman (2004a; 2004b) McCann (2005) and Malle (2004) argue that the SEE result from participant's misinterpretation of the test question. In other word, participants could have believed that the behavior of the chairman, in the harm case, must be sanctioned, and for this reason is necessary to consider this act as product intentionally.

Nadelhoffer (2004; 2006a; 2006b, p. 214) proposes that emotions are biasing the intentionality judgment in participants' processing of Knobe's scenarios. He states: 'affective or emotional responses ... inappropriately bias our otherwise rational judgments.' Machery (forthcoming) suggests the 'trade-off hypothesis' to explain the asymmetrical judgments related to the SEE. He thinks that the asymmetry is merely a product of how people

conceptualize the side-effect in the harm case. According to Machery (forthcoming), in the harm case, people take costs to be intentionally incurred in order to reap benefits and, in this case, they answer that the side effect has been intentionally brought about. To test his hypothesis Machery developed two new scenario, the worker case and the dog case, that are based on the trolley dilemma. Consistent with this hypothesis, subjects were more likely to judge that the agent intentionally brought about the side-effect in the worker case than the side-effect in the dog case, because they are more likely to conceptualize the side-effect in the worker case (causing the death of the worker on the side tracks) than the side-effect in the dog (saving a dog in addition to five workers) as a cost incurred in order to reap a greater benefit.

Our results suggest that the asymmetrical moral judgments of preschoolers and adults are based more on agent's epistemic status and somewhat less on the agent's declarations of uncaring (see Experiment 5 and 8 on false belief). These data show that intentionality evaluation rests on epistemic information together with the careless state. When the belief to produce damage is not present the intentionality evaluation does not seem to be influence by the valence of the effect. In our opinion, the fact that the SEE disappear when it is explicitly underlined the chairman does not have the belief to produce a negative side effect, seem to corroborate Malle's proposal that the SEE is not changing the model on the naïve concept of intentionality proposed by Malle and Knobe (1997).

We propose a model, illustrate in Figure 4.1, that better explains the SEE. The model shows the hierarchical order of the information processing in relation to intentionality evaluations that participants produce.

Figure 4.1 the scheme of the relation between the agent's epistemic in terms of careless attitude and intentionality judgment

Epistemic status	Present				Absent			
Careless	Yes	No			Yes	No		
Effect	positive	negative	positive	negative	positive	negative	positive negative	
Intentionality	no	yes	no	yes	no	yes	no no	
Judgment								

4.2 Utilitarian judgments

The second theme we addressed concerns how children compute utilitarian moral judgments on the trolley and the footbridge dilemmas. The rationalist approach (Piaget, 1932; Kohlberg, 1969) cannot explain why persons often evaluate that it is permissible to sacrifice one person to save five in the trolley case but not in the footbridge case. One candidate theory for explaining this asymmetry comes from Haidt (2001) who emphasized the role of emotion in determining moral reasoning. His theory maintains that intuitions and emotions are core aspects of moral judgment, while justifications provide only a post hoc evaluation. Often justifications cannot be provided at all in situations that Haidt terms 'moral dumbfounding.'

While the debate over the role of intuition or justification in relation to moral judgment is ongoing, it has been proposed that a dual route model of unconscious emotions and deliberate reasoning regulates humans' moral life. This is the type of position has been embraced by Greene using fMRI studies of normal adults (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001) and by Damasio using brain-damaged patients (Anderson, Bechara, Damasio, Tranel, & Damasio, 1999; Damasio, 1994; Tranel, Bechara, & Damasio, 2000). Another model that seeks to explain humans' moral ability comes from Hauser (2006). He argues that our moral judgments are made on the bases of principles (e.g., the Action principle, the Intention principle, the Contact principle) that are active but unconscious. Hauser proposed that using physical contact to cause harm to another is morally worse than causing equivalent harm to another without physical contact. A possible explanation for the trolley and footbridge dilemma is that the contact principle guides moral judgments according to the intuitionist model during judgment, and that a process of post-hoc reasoning at the time of justification allows subjects to deduce the basis for why they have made judgments (Cushman et al. 2006).

We based our research in this instance on Cushman et al.'s (2006) and Greene et al. 's (2001) studies and we extended the pattern of findings in this area with an Italian sample of adults (Experiments 9 and 10). As shown in previous investigations, the majority of participants in our research stated that it is permissible to change the direction of the trolley

but few stated that it is permissible to show a man from a footbridge. Interestingly, even preschoolers show the same type of asymmetrical judgment when asked to evaluate the footbridge and trolley dilemmas (Experiments 11, 12 and 13). Although from our data it is not possible to choose, within the models described above, the one that best accounts for differences in judgments on the footbridge and trolley dilemmas, it appears that children and adults show the same pattern of answers on the footbridge and the trolley scenarios.

4.3 Continuity hypothesis and moral reasoning

In line with some recent studies on the continuity issue concerning the moral judgment (Hamlin, Wynn, & Bloom, 2007; Nuñez & Harris, 1998; Siegal & Peterson, 1998; Turiel, 2006), our data show that children seem to compute moral stimuli early in life producing asymmetrical moral evaluation on the trolley and footbridge scenarios. Moreover, children seem to analyze the agents' actions in term of mental state and frame these in terms of knowledge, caring attitude and outcome. These data seem to confirm the findings in terms of an early capacity to compute moral information (Nuñez & Harris,1998; Siegal & Peterson, 1998) and, to some extent, supports the proposal that human minds are endow with an innate moral faculty (Dwyer, 2007; Hauser, 2006; Rawls, 1981).

However, the area needs further studies concerning, for example, the nature of continuity in intentionality judgment and the extent to which considerations of caring and foreknowledge remains stable with age. Further research is necessary to compare the development of a moral sense in typically developing children with children who have developmental problems such as autism, conduct disorder or Down Syndrome. Studies of such groups may help to determine the extent to which theory of mind and emotional understanding plays a decisive role in moral evaluation. Another interesting theme would be to study the various ways in which the cultural context (Shweder, Much, Mahapatra, & Park, 1997) determine responses on the SEE measure and footbridge and trolley problems. A comparison of societies that stress individualist values with those that stress collectivist ones is likely to lead to differences in responses on measures of utilitarian moral judgment based on diverse connotations of the term 'intentionally.'

Moral psychology is a discipline of both intrinsic and practical interest; uncovering the determinants of moral judgment and behavior is fascinating in its own right. A better understanding of these determinants may help us to better understand what educational and policy interventions may facilitate good conduct and ameliorate bad conduct.

References

Adams, F., & Steadman, A. (2004a). Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis, 64*, 173-181.

Adams, F., & Steadman, A. (2004b). Intentional action and moral considerations: Still pragmatic. *Analysis, 64,* 264-267.

Alicke, M.D., (2000). Culpable control and psychology of blame. *Psychological Bulletin, 126,* 556-574.

Alicke, M.D. (2006). Blaming badly. *Journal of Cognition and Culture*, 6, Special Issue - On Folk Conceptions of Mind, Agency, and Morality.

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,* 1032-1037.

Aristotle (1998). *The Nicomachean Ethics* (D. Ross, Trans). New York: Oxford University Press. (Original work published 350 B.C.)

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

Baron-Cohen, S., Leslie, A.M., & Frith, U. 1985. Does the Autistic Child Have a Theory Of Mind? *Cognition, 21*, 37 46.

Bartsch, K., & Wellman, H. 1995. *Children Talk about the Mind*. New York*:* Oxford University Press.

Bastick, T. (1982). *Intuition: How we think and act*. Chichester, England: Wiley.

Blair, R. J. R., Perschardt, K. S., Budhani, S., Mitchell, D. G. V., & Pine, D. S. (2006). The development of psychopathy. *Journal of Child Psychology and Psychiatry, 47*, 262-275.

Callaghan, T.C., Rochat, P., Lillard, A., Claux, M.L., Odden, H., Itakura, S., Tapanya, S., & Singh, S. (2005). Synchrony in the onset of mental-state reasoning. *Psychological Science, 16,* 378–384.

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

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,* 1082-1089.

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

Damon, W. (1977). *The social world of the child.* San Francisco: Jossey-Bass.

Danovitch, J. H., & Keil F. C. (2008). Young Humeans: The role of emotions in children's evaluation of moral reasoning abilities. *Developmental Science, 11*, 33-39.

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: Oxford University Press.

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

Foot, P. (1978). *Virtues and vices*. Oxford: Blackwell.

Goldman, A. (1989). Interpretation Psychologized. *Mind and Language*, *4,* 161-185.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, Mass.: Bradford, MIT Press.

Gopnik, A., & Wellman, H. (1994). The theory-theory. In L. Hirschfeld and S. Gelman (eds.), *Mapping the Mind: Domain Specificity in Cognition and Cultur*e (pp. 257-293). New York: Cambridge University Press.

Gordon, R. (1986). Folk psychology as simulation. *Mind and Language, 1,* 158-171.

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

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 judgement. *Neuron, 44,* 389-400.

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

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

Haidt, J., Bjorklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. *Unpublished manuscript*, University of Virginia.

Haidt, J., Koller S.H., & Dias, M. (1993). Affect, culture and morality or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*, 613-628.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, 450, 557-559.

Harris, P. (1992). From Simulation to Folk Psychology: The Case for Development. *Mind and Language, 7*, 120-144.

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

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

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

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.

Hume, D. (1960). An enquiry concerning the principles of morals. La Salle, IL: Open Court. (Original work published 1777).

Hume, D. (1969). A treatise of human nature. London: Penguin. (Original work published 1739-1740)

Kahneman, D. & Tversky, A. (1982). Variants of uncertainty. Cognition, 11, 143-157.

Kant, I. (1959). Foundation of the metaphysics of morals (L. W. Beck, Trans.). Indianapolis,

IN: Bobbs-Merrill. (Original work published 1785).

Kelly, D., Stich, S., Haley, K., Eng, S., & Fessler, D. (2007). Harm, affect and the

moral/conventional distinction. Mind and Language, 22, 117-131

Kerr, S., & Durkin, K. (2004). Understanding of thought bubbles as mental representations in children with

autism: Implications for theory of mind. *Journal of Autism and Developmental Disorders, 34*, 637-648.

Knobe, J. (2003a). Intentional Action and Side Effects in Ordinary Language. *Analysis, 63*,

190-193.

Knobe, J. (2003b). Intentional Action in Folk Psychology: An Experimental Investigation.

Philosophical Psychology, 16, 309-324.

Knobe, J. (2004). Intention, Intentional Action and Moral Considerations. *Analysis, 64,* 181-187.

Knobe, J. (2005). Theory of mind and moral cognition: Exploring the connections. *Trends in Cognitive Sciences,*
9, 357-359.

Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical*
Studies, 130, 2, 203-231.

Koenigs, M., Young, L., Adolph, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the

prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*, 908-911.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to

socialization. In D. A. Goslin (Eds.), *Handbook of socialization theory and research*

(pp. 347-480). Chicago: Rand McNally.

Kohlberg, L. (1984). *The Psychology of Moral Development: The Nature and Validity of*

Moral Stages. New York: Harper & Row.

Leslie, A. (1994). ToMM, ToBY and Agency: Core Architecture and Domain Specificity. In

L. Hirschfeld & S. Gelman (eds.) *Mapping the mind* (pp. 119-148). Cambridge,

England: Cambridge University Press

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: 'Theory of mind' and moral judgement. *Psychological Science, 5*, 421-427.

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

Machery, E., (forthcoming). Understanding the Folk Concept of Intentional Action: Philosophical and Experimental Issues. *Mind and Language*.

Malle, B. F., (2004). The moral dimension of intentionality judgments. *Technical Reports of the Institute of Cognitive and Decision Sciences*, No. 04-2, Eugene, Oregon.

Malle, B. F. (2006). Intentionality, morality, and their relationship in human judgment. *Journal of Cognition and Culture, 6*, 87-113.

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33,* 101-121.

Malle, B. F., & Nelson, S. E. (2003). Judging Mens Rea: The Tension Between Folk Concepts and Legal Concepts of Intentionality. *Behavioral Sciences and the Law, 21,* 563-580.

McCann, H. (2005). Intentional Action and Intending: Recent Empirical Studies. *Philosophical Psychology, 18*, 737-748.

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'. *Unpublished PhD thesis*, Cornell University, Ithaca.

Mikahil, J. (2002). Aspects of a theory of moral cognition: Georgetown University Law Center.

Mill, J. S. (1974). *On Liberty*. London: Penguin Classics. (Original work published 1858).

Nadelhoffer, T. (2004). Praise, Side Effects, and Intentional Action. The *Journal of Theoretical and Philosophical Psychology 24*, 196-213.

Nadelhoffer, T. (2006a). Bad acts, blameworthy agents, and intentional actions: Some problems for juror impartiality. *Philosophical Explorations, 9,* 203-219.

Nadelhoffer, T. (2006b). On trying to save the simple view. *Mind & Language, 21,* 565-586.

Nagel, T. (1986). Ethics. In T. Nagel (Ed.), *The view from nowhere*. Oxford: Oxford University Press.

Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment.* Oxford England: Oxford University Press.

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition, 100,* 530-542.

Nichols, S., & Ulatowski, J. (2007). Intuitions and Individual Differences: The Knobe Effect Revisited. *Mind and Language, 22,* 346-365.

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

Nucci, L., & Herman, S. (1982). Behavioral disordered children's conceptions of moral, conventional, and personal Issues. Journal of Abnormal Child Psychology, 10, 411-425.

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

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, 1475-1491.

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

Piaget, J. (1932). *Le jugement moral chez l'enfant. Paris:* Libraiirie Fëlix Alcan.

Pizarro, D. A. & Bloom, P. (2003). The intelligence of the moral intuitions: comment on Haidt 2001. *Psychological Review, 110,* 193-196.

Quinn, W. S. (1989). Actions, intentions, and consequences: The doctrine of doing and

allowing. *The Philosophical Review, 145*.

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

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

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

Schlenker, B. R., Britt, T. W., Pennington, J., Murphy, R., & Doherty, K. (1994). The triangle model of responsibility. *Psychological Review, 101*, 632-652.

Selman, R. (1980). The Growth of Interpersonal Understanding. New York: Academic Press.

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The 'big three' of morality (autonomy, community, and divinity), and the 'big three' explanations, of suffering. In A. Brandt & P. Rozin (Eds.), Morality and Health (pp. 119-169). New York: Routledge.

Siegal, M. (2008). *Marvelous minds: The discovery of what children know.* New York: Oxford University Press.

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

Simon, H. A. (1992). What is an 'explanation' of behavior? *Psychological Science, 3,* 150-161.

Singer, P. (1981). *The Expanding Circle: Ethics and Sociobiology.* New York: Farrar, Straus & Giroux.

Singer, P. (1993). Practical Ethics. Cambridge: Cambridge University Press.

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

Smetana, J. (1993). Understanding of social rules. In M. Bennett (ed.) The development of social cognition: The child as psychologist (pp. 111-141). New York: Guilford Press.

Smetana, J., S. Toth, D. Cicchetti, J. Bruce, P. Kane, & C. Daddis. (1999). Maltreated and nonmaltreated preschoolers' conceptions of hypothetical and actual moral transgressions. Developmental Psychology, 35, 269-281.

Spranca, M., Minsk, E., & Baron, J. (2003). Omission and commission in judgment and

      choice (pp. 31): University of Pennsylvania.

Thomson, J. J. (1986). *Rights, restitution, and risk: Essays in moral theory*. Cambridge, MA:

      Harvard University Press.

Tranel, D., Bechara, A., & Damasio, A. (2000). Decision making and the somatic marker

      hypothesis. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (pp. 1047-1061).

      Cambridge: MIT Press.

Turiel, E. (1983). The development of social knowledge. Cambridge: Cambridge University

      Press.

Turiel, E. (2006). The development of morality. In W. Damon (Series Ed.) & N. Eisenberg

      (Vol. Ed.), Handbook of child psychology: Vol. 3: Social, emotional, and personality

      development (5th ed., pp. 863–932). New York: Wiley.

Walster, E. (1966). Assignment of responsibility for an accident. Journal of Personality and

      Social Psychology, 3, 73-79.

Wellman, H. M., Hollander, M., & Schult, C. A. (1996). Young children's understanding of thought bubbles and

      of thoughts. *Child Development, 67,* 768-788.

Appendix

Knowing Agent-Negative Outcome story

This is the story of a child named Andy and his frog/gerbil. Andy loves his frog and he wants to keep it always with him.

Here is a boy named Andy, and he's over at his house [pointing to the drawing]. And here is a girl named Janine, and she's over at her house [pointing to the drawing]. And look what Andy has with him, he has a [frog/gerbil- let child answer]. Now Andy loves frogs/gerbil, but Janine hates frogs/gerbils. Now can you remember, does Andy love frogs? Does Janine love frogs?

Andy wants to bring the frog over to Janine's house if Andy brings the frog over, Janine will get upset. Why will Janine get upset? Now listen very carefully.

Andy knows that Janine hates frogs/gerbils and that she will be really upset when she will see the frog/gerbil. Andy does not care if Janine will get upset. He is going to bring the frog over anyway.

Knowledge control question: Does Andy know that Janine will be very upset to see the frog/gerbil?

Caring control question: Does Andy care that Janine will get upset?

So Andy brings the frog over to Janine's house and she gets upset. Now I have a question for you.

On Purpose test question: Does Andy make Janine upset on purpose or without wanting to?

(Italian: *Andrea ha fatto arrabbiare Gaia apposta o senza volere?*)