THE MORAL FACULTY IN ADOLESCENTS: INVESTIGATING AGE DIFFERENCES IN THE APPLICATION AND JUSTIFICATION OF THREE PRINCIPLES OF HARM

A Thesis

Submitted to the Graduate School of the University of Notre Dame in Partial Fulfillment of the Requirements for the Degree of Master of Arts

by

Paul C. Stey

_______________________________
Daniel K. Lapsley, Director

Graduate Program in Psychology
Notre Dame, Indiana
July 2011
THE MORAL FACULTY IN ADOLESCENTS: INVESTIGATING AGE DIFFERENCES IN THE APPLICATION AND JUSTIFICATION OF THREE PRINCIPLES OF HARM

Abstract

by

Paul C. Stey

The study of moral cognition has been dominated by Kohlberg’s stage theory, which emphasizes conscious deliberation as the source of moral judgments. More recently, researchers have proposed that moral judgments may be the product of certain implicit principles. The present study investigated whether 9th- and 12th-grade students would conform their judgments to three principles of harm, and the degree to which these principles would be available to conscious reflection. The present study also investigated the possibility of age-related differences with respect to the application and justification of these principles of harm. Results indicated that participants across age groups conformed their judgments to the principles of harm under investigation. However, participants had difficulty providing justifications for their judgments. This suggests these principles of harm operate outside of conscious control. Results also revealed age differences in the ability to provide justifications for one’s judgments—with older participants providing more sufficient justifications.
pour ma famille
# TABLE OF CONTENTS

- LIST OF TABLES ........................................................................................................ iv
- LIST OF FIGURES .................................................................................................... v
- ACKNOWLEDGMENTS .............................................................................................. vi
- INTRODUCTION ........................................................................................................ 1
- METHOD .................................................................................................................... 25
- RESULTS .................................................................................................................... 36
- DISCUSSION ............................................................................................................ 53
- REFERENCES ............................................................................................................. 60
- APPENDIX A ............................................................................................................. 67
- APPENDIX B ............................................................................................................. 78
LIST OF TABLES

Table 1. Attributes of Test Scenarios.................................................................26
Table 2. Scenario Pairs Corresponding to Principles of Harm.................................27
Table 3. Descriptive Statistics for 9th-Grade Group on Test Scenarios......................37
Table 4. Descriptive Statistics for 12th-Grade Group on Test Scenarios.....................38
Table 5. Summary of Paired Sample T-tests on Scenario Pairs.................................39
Table 6. Summary of Participants’ ISTEP English/Language Arts Scores....................78
LIST OF FIGURES

Figure 1. Schematic Illustration of Scenario 1; Ned (Hauser et al., 2007)………………14
Figure 2. Schematic Illustration of Scenario 2; Oscar (Hauser et al., 2007)…………….15
Figure 3. Diagram of A and B Scenarios Illustrating the 9 Test Pairs…………………..30
Figure 4. Overall Percentage of Sufficient Justifications by Principle for 9th and 12th Graders…………………………………………………………………………………….44
Figure 5. Number of Sufficient and Insufficient Justifications for the Frank and Bob Scenario Pair for 9th and 12th Graders…………………………………………………….46
Figure 6. Number of Sufficient and Insufficient Justifications for the Frank and Ned Scenario Pair for 9th and 12th Graders…………………………………………………….47
Figure 7. Number of Sufficient and Insufficient Justifications for the Ned and Oscar Scenario Pair for 9th and 12th Graders…………………………………………………….47
Figure 8. Number of Sufficient and Insufficient Justifications for the Chris and Dan Scenario Pair for 9th and 12th Graders…………………………………………………….48
ACWKNOWLEDGMENTS

Were it not for the help and support of many wonderful people, this project could not have been completed. I am grateful for the opportunity to express my heartfelt thanks and appreciation to those individuals. First and foremost, I am thankful beyond words for all the help provided at every stage of this project by my advisor and mentor Dan Lapsley. It would be difficult to overstate the degree to which Dan’s generosity and support has helped me with the completion of this project. Without Dan’s guidance, this project would have remained merely an idea—more than likely a crude one. I am also deeply indebted to Dan for his help and patience in revising many early drafts of this document. If it seems to read well at times, it is thanks to Dan.

I also owe a debt of thanks to the other three members of my thesis committee: Julie Turner, Nicole McNeil, and Josh Diehl. I am exceedingly grateful to each of them for their support and helpful comments on this project. Their suggestions and feedback at various stages helped shape what would become the final project as well as this document. I am truly grateful for their insights, and I consider myself exceptionally lucky to have worked with a committee made up of four scholars of the highest caliber. I owe Nicole a special thanks for agreeing to schedule my proposal meeting within 10 days of when she was due to give birth.

Mary McKeever deserves special thanks as well. Mary was an undergraduate research assistant in our lab, and expressed interest in helping on this project. I am very
grateful for Mary’s assistance with data collection, transcribing the interviews, and acting as the second coder for the qualitative data. I often feel that Mary got more than she bargained for. I would also like to thank Lori Peterson. Lori kindly agreed to lend her lovely voice to the audio narration of the ethical dilemmas used in this study. I also owe special thanks to Sara Fulmer for her patience in answering a barrage of methodological questions during the preparation of this study, and for her thorough and detailed feedback on the final draft of this document.

I am also deeply grateful to all the students, teachers, and staff at Clay High School in South Bend, Indiana for allowing me to be such a constant and disruptive presence in their school. I am especially indebted to Principal Mansour Eid who allowed us to conduct our study in his school, and further complicate his already busy days. I am also extremely grateful to Rick Roberts, Patrick Cain, and Jené Hartman, who were all kind enough to allow us to sample students from their classrooms. They have very difficult jobs, and I am tremendously thankful for their generosity in allowing me to increase their burden. I am also grateful to Joseph Good for all his help in compiling the participants’ standardized test scores, and for doing so on very short notice. Additionally, I am very appreciative for the hospitality and seemingly infinite patience of Kay Radecki, and Karen Fiechtner, who graciously provided me with a quiet space to conduct the actual data collection; it was really a pleasure to work alongside them. And, of course, I am truly grateful to all the individual students who were willing to take part in this study.

Finally, I am exceedingly grateful to my parents, George C. Stey and Marie-Madeleine Stey, who have supported me and instilled in me a love of learning and the belief that I could achieve what I set out to accomplish. It is certain that were it not for
their unwavering support and sacrifice, I would not be where I am today. I owe them everything.
INTRODUCTION

One of the central questions in the field of moral psychology relates to how individuals make decisions about the rightness or wrongness of actions. That is, how do individuals arrive at moral judgments? This question has received a great deal of attention throughout the history of the study of morality and moral development, yet a consensus remains elusive. The dominant view for several decades was Lawrence Kohlberg’s stage theory of moral development (Kohlberg, 1969). Kohlberg’s view emphasizes the role of conscious reasoning as the source of moral judgments (Arnold, 2000; Hauser, Cushman, Young, Jin, & Mikhail, 2007). In recent years, however, there has been growing interest in a number of different accounts of the origins of moral judgments. In particular, there is interest in theories that emphasize non-conscious or intuitive cognitive processes as the source of moral judgments (Green & Haidt, 2002, Haidt, 2001; Hauser, 2006; Kamm, 1998a; Petrinovich, O’Neill, & Jorgensen, 1993). The present study will investigate one such account of intuitive moral judgments. This particular account of intuitive moral judgments—as well as Kohlberg’s theory of moral reasoning—will be discussed in light of dual-process models of cognition (Chaiken & Trope, 1999; Sloman, 1996; Stanovich & West, 2000).

In order to situate this project within recent work in moral psychology, we will first examine Kohlberg’s moral stage theory as well as several lines of criticism of his theory. We will see that Kohlberg’s theory has been criticized for its singular emphasis
on conscious reasoning as the source of moral judgments. We will then turn our attention to dual-process models of cognition. Dual-process models typically claim that both conscious and non-conscious processes underlie cognition and behavior, and that both must be the targets of investigation; this view is now so common as to be normative in the study of cognition (Bargh, 1994; Chaiken, Liberman, & Eagly, 1989; Chaiken & Trope, 1999; Devine, 1989; Petty & Cacioppo, 1986; Trope, 1986). Next, we will briefly examine dual-process models of cognition as they relate to the moral domain. This will lead to a discussion of Hauser’s (2006) view on the origin of moral judgments, as well as recent evidence supporting this account. Finally, the current study, which attempts to further extend these findings, will be presented.

Kohlbergian Model of Moral Development

There has seldom been a theory that influences its respective field to the extent that Kohlberg’s stage theory has influenced the field of moral psychology. Kohlberg’s stage theory of moral development was heavily influenced by Jean Piaget’s views on cognitive development as well as Immanuel Kant’s normative ethical views (Lapsley, 1996; Narvaez, 2005). The former provided a framework for understanding development, and the latter provided the content that would come to represent the highest level of moral reasoning.

It is also important to point out a third major influence on Kohlberg’s work. Kohlberg was deeply affected by the atrocities of the Holocaust, and it may have been this experience that motivated Kohlberg to pursue his ultimate goal: defending ethical universalism and defeating ethical relativism (Lapsley, 1996; Power, 1991). Kohlberg’s program of research emerged at a time when the dominant paradigms in psychology (i.e.,
psychoanalysis and behaviorism) could not undermine the assertions of Nazi morality (Lapsley, 1996). Kohlberg would make it his life’s work to develop a scientific account on the basis of which it would be possible to condemn the atrocities of the Holocaust as objectively wrong; this was his goal, and he approached this task armed with the Piaget’s genetic epistemology, and Kant’s categorical imperative.

**Piagetian Structuralism**

Kohlberg’s theory is often described as the application of Piaget’s cognitive developmental theory to the moral domain. At the center of Piaget’s cognitive developmental approach is the concept of cognitive stages, which have the following features (Piaget, 1960):

1. Different stages involve *qualitatively* different modes of thinking and problem solving.

2. The differences in modes of thought that are observed in different stages make up an invariant sequence; while progress through these modes of thought may slow or stop based on certain environmental factors, the underlying sequence remains unchanged.

3. Each of these different modes of thought forms a structured whole; that is, an underlying pattern of thought organization (e.g., concrete operations).

4. Cognitive stages represent hierarchical integrations; meaning, not only are later stages qualitatively different than earlier stages, later stages in many cases also incorporate features of earlier stages. The fact that formal operational thought is achieved does not imply that concrete operational thought will no longer be utilized. Instead, concrete operational thinking will continue to be present when the specific demands of the situation require it.

These four general features of Piaget’s theory formed the foundation on which Kohlberg would build his theory of moral development. Indeed, it would be difficult to overstate the extent to which Piaget influenced Kohlberg. In many ways, Kohlberg’s theory is an extension of Piaget’s work (Rest, 1983).
It is important to consider one final feature of Piaget’s theory of cognitive development. The crucial element is the notion that development moves in the direction of more ideal forms of equilibrium where cognitive operations are increasingly perfected. Each succeeding stage in the sequence describes a more refined mode of operation that is “better” than the operations of the previous stage. On this account the process of development possesses its own internal standard of adequacy. It is both better and good that a change process shows development.

This idea that development has its own internal standard of adequacy was thought crucial by Kohlberg for providing the psychological grounds for defeating ethical relativism. If Piaget could show that one system of thought (“formal operations”) was better than another (“concrete operations”), on psychological grounds, then it might prove possible to show why one philosophical system was superior to another.

**Kantian Deontology**

In addition to Piaget, the Kohlbergian model of moral development was also heavily influenced by the work of Immanuel Kant. This influence is most apparent in the later stages of Kohlberg’s six-stage model, but can also be felt even in the early stages. Kant’s ethical theory is perhaps one of the most widely known and widely studied theories of morality. Kant’s theory is (1) objective in the sense that it presupposes that moral truths exist, (2) rational in the sense that it understands these truths to be knowable through reasoning, and (3) deontological in its conception of moral action as one which is carried out based on duty. These three elements are interrelated, and become clear upon examining Kant’s fundamental moral principle: the categorical imperative (Kant, 1785/1959).
Kant presented several formulations of his categorical imperative. The first formulation states: “Act only according to that maxim whereby you can at the same time will that it should become a universal law without contradiction” (Kant, 1785/1959, p. 421). Consider the following example of how the categorical imperative is used as a litmus test for the moral qualities of a behavior. In this example, suppose the behavior in question is stealing. If we consider the maxim “steal whenever it suits you” we find—not surprisingly—that the categorical imperative cannot condone this particular maxim. First, because it is certainly not clear that anyone would will this to be a universal law; after all, if this maxim were instituted universally, and stealing was sanctioned, then one could also become the victim of theft, which is certainly unfavorable. Additionally, and perhaps more importantly for Kant, the institution of this maxim would lead to a logical contradiction. Specifically, if it were the case that one could steal whenever it was desirable, then it seems to follow that the concepts of personal property and ownership become meaningless; thus, the concept of stealing itself becomes meaningless—one cannot steal something that belongs to no one. Therefore, the maxim “steal whenever it suits you” produces a contradiction that the categorical imperative—and logic—cannot allow.

The example above illustrates the ways in which Kant’s theory is objective, rational, and deontological; reasoning about a certain behavior produces an objective judgment about the qualities of the behavior, and, therefore, a mandate about one’s duty with respect to this behavior (e.g., “do not steal”). These three elements of Kantian ethics (i.e., objectivity, rationality, and duty) are critical to Kohlberg’s theory and serve as the pillars on which Kohlberg’s later stages of reasoning are built. For Kohlberg, moral
reasoning is the means by which individuals arrive at moral judgments; moral reasoning is how we decide which actions are right and which actions are wrong. Kohlberg and colleagues defined moral reasoning as a “conscious process of using ordinary moral language” (Kohlberg, Levine, & Hewer, 1983, p. 69). At its highest levels, mature moral reasoning begins to resemble the thought process of a philosopher—in particular, a Kantian. That is, it is very similar in form to the example discussed above; one evaluates the moral qualities of an action based upon certain rational criteria (i.e., the categorical imperative). As we will see below, the emphasis that Kohlbergians place on a deliberate and conscious process of reasoning has recently drawn considerable criticism.

**Kohlberg’s Stages**

Kohlberg’s stage theory of moral development contains six stages. The stages are grouped into three levels: preconventional, conventional, and postconventional (or principled). These levels can be thought of in terms of their sociomoral perspective; that is, how the self is related to the moral expectations of society (Arnold, 2000; Colby & Kohlberg, 1987). Each level contains two stages, the second of which represents a more adequate articulation of the sociomoral perspective.

In the early stages, reasoning is instrumental and relativistic; it is oriented towards punishment and praise. As individuals progress through the stages, the forms of reasoning they use begin to take into consideration the norms of the groups and societies of which they are a part. In the later stages, individuals’ views of morality become closer approximations of Kant’s normative ethical theory. Thus, an individual using Stage 6 reasoning will approach moral dilemmas in much the same way as a trained ethicist—in particular, a Kantian.
Criticisms of Kohlberg’s Model. Without question, Kohlberg’s theory influenced a generation of researchers. As with any major theoretical view, however, his theory has faced a great deal of criticism over the years. Kohlberg’s model has been criticized both for its theoretical grounding, as well as its empirical claims. Most relevant for the present discussion is the criticism of the Kohlbergian claim that reasoning is the source of moral judgments.

One of the essential characteristics of the cognitive-developmental paradigm as it relates to morality (Kohlberg, 1981, 1984; Piaget, 1932/1965; Rest, 1983; Turiel, 1998) is the human ability to reason (Arnold, 2000). This has become a point of contention in the field of moral psychology. Indeed, the Kohlbergian claim that moral judgments are solely the product of reasoning has come under criticism from a great number of sources (Arnold, 2000; Greene & Haidt, 2002; Haidt, 2001; Hauser, 2006; Macnamara, 1990; Mikhail, 2007; Petrinovich et al., 1993). At least one reason for the dissatisfaction with Kohlberg’s model of moral rationality is that it seems to shrink the moral domain to encompass only that which can be explicitly expressed. This is the paradox that results from Kohlberg’s exclusive focus on reasoning. Judgments and actions that are not prefaced by consciously invoking the moral law cannot be considered moral. Thus, the moral landscape becomes limited to only those “hard case” dilemmas that lend themselves to conscious and effortful deliberation.

Another reason to question Kohlberg’s notion that moral judgments are the product of a conscious process of reasoning comes from recent evidence in the cognitive sciences. Lapsley and Hill (2008) observed that Kohlberg’s singular focus on establishing a theory of moral development that would oppose ethical relativism may have had the
inadvertent consequence of disconnecting this line of research from advances in other areas of psychology. It is interesting to note that when Kohlberg’s theory rose to prominence it drew upon one of the foremost theories in cognitive development. Yet, in the last few decades the Piagetian paradigm has been eclipsed by new ways of understanding cognitive processing and intellectual development. These new models have not yet penetrated our understanding of moral cognition and moral development. The following section discusses these new models of cognition.

**Dual-Process Models of Cognition**

One of the critical advances in the field of cognitive psychology has been the understanding that a substantial portion of our daily lives is influenced by non-conscious processes (Bargh & Chartrand, 1999). Moreover, current research indicates that there exist two systems of processing (Chaiken & Trope, 1999; Sloman, 1996; Stanovich & West, 2000), one that is fast and operates outside of conscious control, and another that works more slowly and is available to conscious control. Dual-process theories have a long and diverse history (Epstein, 1994; Evans, 1984; Hammond, 1996; Johnson-Laird, 1983; Klein, 1998; Levinson, 1995; Pollock, 1991; Posner & Synder, 1975; Reber, 1993; Shiffrin & Schneider, 1977; Sloman, 1996), and, in fact, these dual-process models have become the orthodox view in cognitive psychology (Bargh, 1994; Chaiken et al., 1989; Chaiken & Trope, 1999; Devine, 1989; Petty & Cacioppo, 1986; Trope, 1986). Although these two systems of processing have been variously conceived, there seems to be general agreement concerning their essential features. Stanovich and West (2000) have proposed the nomenclature of System 1 and System 2 to capture the family resemblance that exists among dual-processing theories.
System 1 refers to information processing that is fast, automatic, and associative. It is thought to be the product of biological acquisition, exposure, and personal experience. System 2 processing is relatively slow, controlled, and rule-based. It is more demanding of cognitive resources, and is considered to be the product of cultural and formal acquisition (Stanovich & West, 2000).

As it relates to the moral domain, the received view is that moral development in general, and moral judgments in particular are driven by System 2 processing. Yet, as research in cognitive science illustrates, much of human behavior is governed by cognitive processes that are not deliberate, effortful, and conscious, but are instead tacit, implicit, and automatic. It is not likely that cognitive automaticity should govern so many aspects of human functioning, but leave the moral domain exempt. Indeed, until recently, the nearly universal notion that moral judgments are the product of System 2 information processing has served to narrow the field of inquiry in this domain. It is encouraging, therefore, to point out the emerging research supporting a variety of System 1 accounts of moral judgments; these include: social intuitions (Haidt, 2001), heuristics (Baron, 1993; Gigerenzer, 2008; Sunstein, 2005), chronic accessibility (Lapsley & Narvaez, 2004; Narvaez, Lapsley, Hagele, & Lasky, 2006), moral expertise (Dreyfus & Dreyfus, 1991; Narvaez & Lapsley, 2005), and a moral faculty (Cushman, Young, & Hauser, 2006; Hauser, 2006; Mikhail, 2000).¹ The present study is particularly concerned with Hauser’s (2006) conception of the moral faculty.

¹ The list provided above is not a comprehensive review of the different System 1 theories of morality. Additionally, though the theories listed above do have some general similarities, they are also quite different from one another in many ways. For a detailed account of the distinction between System 1 and System 2 processing as they relate to the moral domain, as well as a description of several different forms of System 1 theories, see Lapsley and Hill (2008).
The Moral Faculty

**Rawls’ Linguistic Analogy.** Hauser’s thesis is that humans possess a certain moral faculty which enables us to produce moral judgments based upon intentions, causes, and consequences of actions (Hauser, Young, & Cushman, 2008). His theory relies heavily on the normative ethical position of the political philosopher John Rawls (1971), who articulated a theory of justice that was grounded on the notion of fairness. Indeed for Rawls, justice and fairness are identical concepts. In addition, Rawls also compared the rules of justice to the rules of language, and the acquisition of moral rules to the acquisition of the rules of language as understood by Chomsky (e.g., Chomsky 1986, 1988, 2000).

One of the central claims of Chomsky’s views regarding the acquisition of language is that humans possess a language faculty; that is, an “organ” allowing us to perceive, learn, and produce languages (Chomsky, 1986, 1991a, 1991b). In a general sense, this organ can be thought of as an instinct to acquire language (Pinker, 1994); more specifically, it can be thought of as a set of principles that facilitate language development in humans (Hauser et al., 2008). These principles work outside of conscious awareness, and, therefore, are not available to reflection. One example Hauser provides is related to the verb contraction of “is”. It is clear that we are able to use the contracted form of *is* in sentences such as: “Frank’s more foolish than Joe is,” we cannot, however, use the contracted form of the verb in a sentence such as: “Frank is more foolish than Joe’s.” This is not a rule that is taught in schools, and yet this is also not a mistake that is made by children. Hauser explains that work in linguistics (Anderson & Lightfoot, 2000) seems to suggest there is an operative principle that says something like “The unit of
sound ‘s is too small to be alone, and when we use the contracted form, it must be followed by another word”.

Consequently, one of the fundamental contributions of modern linguistics is the understanding that knowledge about language is in many cases operative but not expressed (Hauser et al., 2008). The operative principles seem to work in the background and without our awareness, while the expressed principles are those that we can identify and explicitly describe. This is important to keep in mind, and it illustrates a crucial point—one that will soon be revisited as it relates to the moral faculty. Essentially, it seems that in the case of language we know more than we think we know (Hauser et al., 2008). Hauser makes the argument that this is similarly true for morality: that we know more than we think we know (Hauser, 2006, Hauser et al., 2008).

Hauser’s claim is that there exists a moral faculty that is similar in nature to Chomsky’s conception of the language faculty. Therefore, whereas Chomsky presented the ideas of a universal grammar, and a language acquisition device, Hauser argues that there is a universal moral grammar and a moral acquisition device. These are components of a moral faculty. Moreover, an important feature of this moral faculty—as with Chomsky’s language faculty—is that there are certain operative principles that seem to work outside of conscious awareness.

Evidence for the Moral Faculty

In a 2007 study Hauser and colleagues tested one such principle. In particular, this study tested the principle of double effect (also referred to as the doctrine of double effect in the philosophy literature). This principle has a long history in philosophical discourse,

---

2 It should be noted that the notion of universal moral grammar was first presented by Mikhail, Sorrentino, and Spelke (1998), and was further refined in Mikhail (2000).
and dates back to St. Thomas Aquinas (13th c/1988). The principle of double effect states that in some cases it may be permissible to cause harm to an individual in order to bring about a greater good, provided the harm is not the necessary means to the greater good but, rather, merely a foreseen side effect (Fischer & Ravizza, 1992; Kamm, 1998b; Mikhail, 2000; Thomson, 1970, 1976). Put differently, in general, we are more willing to allow harm that produces a greater good as long as the harm does not play a causal role in bringing about the good. In their 2007 study, Hauser and colleagues examined this principle using different forms of trolley problems (Fischer & Ravizza, 1992; Foot, 1967; Kamm, 1992; Kamm, 1998b; Thomson, 1970).

Phillipa Foot (1967) presented the original trolley problem in a paper discussing abortion and the doctrine of double effect.

A trolley is running out of control down a track. In its path are five people who have been tied to the track by a mad philosopher. Fortunately, you could flip a switch which will lead the trolley down a different track to safety. Unfortunately, there is a single person tied to that track. Should you flip the switch? (Foot, 1967, pp. 8–9).

Since its inception, the trolley problem has been reformulated and modified many times in order to examine different philosophical questions—most notably: abortion, euthanasia, and the distinction between killing and letting die (Thompson, 1970, 1976; Singer, 2000; Unger, 1996). The result has been an entire family of trolley problems capable of testing a variety of different philosophical questions. Although the term “trolley problem” will be used throughout this paper, this phrase will refer more generally to a class of problems in which there exists a moral dilemma that pits the harming of one person versus the harming of several people—as in Foot’s (1967) original formulation.
Thus, there can be many trolley problems that, in fact, do not make reference to trolley cars.

Hauser and colleagues (2007) use several different forms of trolley problems to examine permissible harm in general, and the doctrine of double effect in particular. Specifically, their study investigated whether or not individuals make moral judgments in accordance with the principle of double effect, and, furthermore, if they are able to provide sufficient justification for the judgments they produce. Participants in Hauser’s study (n = 2612) took part in an online study in which they were presented with several short dilemmas (i.e., trolley problems) and asked whether or not the actions described within the dilemmas were permissible (Yes/No response). The scenarios were designed specifically so that they would differ only according the principle under examination: the principle of double effect. Here is a sample trolley problem taken from Hauser et al. (2007).

Scenario 1: Ned is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Ned sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and they will not be able to get off the tracks in time. Fortunately, Ned is standing next to a switch, which he can throw, that will temporarily turn the train onto a side track. There is a heavy object on the side track. If the train hits the object, the object will slow the train down, thereby giving the men time to escape. Unfortunately, the heavy object is a man, standing on the side track with his back turned. Ned can throw the switch, preventing the train from killing the men, but killing the man. Or he can refrain from doing this, letting the five die. Is it morally permissible for Ned to throw the switch? (see Figure 1; Hauser et al., 2007, pp. 18–19)
Hauser et al. (2007) reported that approximately 56% of participants responded that it was, in fact, permissible for Ned to throw the switch that would send the trolley onto the side track killing the lone bystander, but saving the five men on the main track. Consider a second scenario also used by Hauser and colleagues (2007).

Scenario 2: Oscar is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Oscar sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Fortunately, Oscar is standing next to a switch, which he can throw, that will temporarily turn the train onto a side track. There is a heavy object on the side track. If the train hits the object, the object will slow the train down, thereby giving the men time to escape. Unfortunately, there is a man standing on the side track in front of the heavy object, with his back turned. Oscar can throw the switch, preventing the train from killing the men, but killing the man. Or he can refrain from doing this, letting the five die. Is it morally permissible for Oscar to throw the switch? (see Figure 2; Hauser et al., 2007, p. 19)
As illustrated above, Scenarios 1 and 2 are very similar. For this reason, it is particularly interesting that, in the case of Scenario 2, 72% of participants thought it permissible for Oscar to throw the switch. This raises the question: Why would it be permissible for Oscar to throw the switch but not Ned? In comparing the two scenarios, we find the only relevant difference is the presence of a large weight on the side track in Scenario 2. As it turns out, the presence of the large weight on the side track in Scenario 2 is quite important.

If Hauser is correct, the presence of the weight accounts for the differences between the permissibility ratings of Scenario 1 versus Scenario 2. Though it seems to be a subtle difference, the presence of the large object on the side track substantially changes the nature of the lone bystander’s role in Scenario 2. In Scenario 1 the lone bystander is actually playing a causal role in stopping the runaway trolley; were it not for his presence on the side track, the redirected trolley car would merely continue around the loop in the
side track and then proceed to collide with the five men on the main track. In Scenario 2, however, the presence of the bystander on the side track has no bearing on whether or not the five men will be saved. Using counterfactuals we are able to determine that even if the bystander were not present in Scenario 2, the five men would still be saved since the large weight will stop the trolley. Thus, the role of the bystander is quite different in Scenario 1 versus Scenario 2; while his presence is incidental in Scenario 2, it is essential in Scenario 1 (assuming we aim to save the five men). By now, the difference between these two scenarios, and how they relate to the principle of double effect should be coming into sharper focus.

Recall that the principle of double effect suggests that in some cases it may be permissible to cause harm in order to bring about a greater good provided the harm is not the necessary means to the greater good but merely a foreseen side effect. It seems clear that in Scenario 1 the harm that is being caused to the innocent bystander is not merely a foreseen side effect; instead, causing harm to the bystander is the means by which the five men are saved. In the case of Scenario 2, however, it does seem that causing harm to the bystander is, in fact, merely a side effect. Therefore, according to the principle of double effect, it is reasonable to believe that there is a moral difference between the actions of Ned versus those of Oscar. Specifically, it seems that harm that is intended as the means to an end (as in the case of Ned) is morally worse than harm that is incidental—though foreseen (as in the case of Oscar). This was reflected in the pattern of responses: participants were more likely to rate the action of Oscar—but not Ned—as permissible. Thus, it would appear that people tend to conform their judgments in accordance with the principle of double effect (Hauser et al., 2007).
In addition to asking participants whether or not the actions of the agent in the dilemma were permissible, Hauser and colleagues (2007) also asked participants to provide justifications for their ratings. More specifically, participants were asked to provide justifications when they gave divergent responses to scenarios that differed according to the principle of double effect (e.g., Scenario 1 and Scenario 2). For instance, if a participant responded that the actions of Oscar were permissible, but those of Ned were not, that participant would be asked to explain the reason for the difference.

The participants’ justifications were coded into three categories. Category 1 represented a sufficient justification, which points out a factual difference between the two scenarios and claims this as the reason for the moral judgments. For example, it is a sufficient justification to note that the death of the bystander is a necessary means to saving the five workers in Scenario 1, but not in Scenario 2. Justifications in Category 2 represented those that were insufficient. These justifications failed to identify factual differences between the two scenarios, and in some cases demonstrated uncertainty about their reasons for their decisions. An example of an insufficient justification would be “It was a gut feeling” (Hauser et al., 2007, p. 13). The third and final category of justifications were those in which the participant made reference to their own additional assumptions. An example of this would be to suggest that the five men on the main track would hear the trolley in time to get out of its path.

An analysis of the justifications indicated that a very small number of participants were able to provide sufficient justifications. Hauser and colleagues (2007) conclude that participants seemed to operate in accordance with the principle of double effect, yet they are generally unaware that they are doing so. Thus, it seems that people are more willing
to allow harm to an individual in order to bring about a greater good, but only if the harm is a foreseen by-product and not a necessary means toward achieving that greater good; however, this is not a distinction that individuals are able to articulate. Hauser and colleagues (2007) take this to be evidence for the notion that the principle of double effect is an example of knowledge that is operative but not expressed—at least in most cases.

**More Evidence for the Moral Faculty**

In a separate study by the same team of researchers (Cushman et al., 2006) two additional principles of harm were investigated in order to determine their potential influence on moral judgments. In addition to examining the principle of double effect, Cushman and colleagues also tested what they referred to as the *contact principle*, and the *action principle*. The contact principle states that harm that is caused through physical contact is morally worse than equivalent harm that is caused without physical contact (Cushman et al., 2006). For example, according to this principle, it would be considered morally worse to push someone down a flight of stairs than to pull a lever that drops them down the same flight of stairs. The action principle simply states that harm that is caused through action is morally worse than equivalent harm that results from the omission of action (Cushman et al., 2006). According to this, it would be morally worse to poison someone than it would be to allow a victim of poisoning to die by not providing them with an available antidote. The action principle has been studied in the cognitive literature, where it is generally referred to as *omission bias* (Baron & Ritov, 1994, 2004; Spranca, Minsk, & Baron, 1991).
The methodology used to test these principles of harm is essentially the same as that used by Hauser and colleagues (2007). Participants were presented with a number of dilemmas and asked to rate the permissibility of the suggested action in the dilemmas. One minor difference between the two studies is that the study by Hauser and colleagues (2007) utilized a dichotomous response format for participants’ judgments of permissibility, whereas participants in Cushman’s study (2006) provided ratings on a Likert scale that ranged from 1 (Forbidden) to 10 (Obligatory). After completing permissibility ratings for a number of dilemmas, participants were asked to provide justifications for their ratings. More specifically, participants were presented with pairs of scenarios that they had previously rated, and asked to explain the divergent permissibility ratings (assuming they had not rated the scenarios as equally permissible). As in the study by Hauser and colleagues (2007), scenario pairs were organized in accordance with the principles being examined—just as Scenario 1 and Scenario 2 above differ according to the principle of double effect.

The justifications were coded in much the same way as Hauser et al. (2007). The justifications could be categorized into one of five groups: (1) sufficiency, (2) failure, (3) uncertainty, (4) denial, and (5) alternative explanation. A sufficient justification is one in which the participant identified a factual difference between the two scenarios, and claimed (or implied) that this was the basis for the judgments. The failure code was assigned for justifications in which the participant suggested an alternative principle but not one that could account for the actual pattern of judgments. The uncertainty code was assigned to justifications in which the participant explicitly referenced their own uncertainty with regard to the difference in judgments, or they directly stated that they
were unable to provide a justification. The denial code was assigned to justifications in which the participant stated that they did not believe there was a moral difference between the actions described in the two scenarios. Finally, the alternative explanation code was assigned for justifications in which the participant made reference to facts that were not present in the actual text of the scenario, or else they claimed to have made an error when making the permissibility judgments (e.g., clicking the wrong number).

Cushman and colleagues (2006) concluded that participants tended to make moral judgments in accordance with all three principles being tested. More specifically, (1) participants judged harm that was caused through physical contact as morally worse than harm without contact, (2) they judged harm caused through action as worse than harm resulting from omission, and (3) they judged harm that was used as the means to an end as morally worse than harm that was foreseen but unintended. In addition, Cushman and colleagues (2006) report an interesting pattern with respect to participants’ abilities to provide sufficient justifications for their judgments. In particular, Cushman’s findings indicate that there is a graded level of difficulty in the ability to provide sufficient justifications. Of the three principles of harm being examined, the action principle had the highest percentage of sufficient justifications (81%), the contact principle had the second-highest percentage of sufficient justifications (60%), and finally the principle of double effect had the lowest percentage of sufficient justifications (32%).

On the basis of these findings, Cushman and colleagues (2006) conclude that some principles of harm are available to conscious reflection (e.g., action principle, and contact principle), while others are not (e.g., principle of double effect). Some principles lend themselves more to conscious reflection, while other principles seem to be the
product of non-conscious processing and, as a result, they are not as readily explained via conscious reflection.

Taken together, the results of Cushman and colleagues (2006), as well as Hauser and colleagues (2007), seem to indicate that (1) there exist certain principles of harm according to which individuals make moral judgments, and (2) these principles of harm are in some cases operative but not expressed. Hauser concludes that these findings are evidence for a moral faculty that all individuals possess. This faculty effortlessly guides our moral judgments through the application of implicit principles of harm.

Development of the Moral Faculty

One of the interesting issues raised by this line of research concerns the etiology of the moral faculty. A recent study by Pellizzoni, Siegal, and Surian (2010) attempted to address this issue, and concluded that children as young as 3 and 4 years of age not only understand the kinds of moral dilemmas used by Hauser, but also that they generally tend to make moral judgments that are similar to those of adults; specifically, they make judgments in accordance with the contact principle. The findings of Pellizzoni and colleagues (2010) should be interpreted with caution, however, insofar as their methodology conflates the contact principle and the principle of double effect. Nonetheless, the studies described by Pellizzoni and colleagues (2010) are quite encouraging in the sense that seem to support the notion that children—even very young children—are able to understand several versions of the classic trolley problem.

To date, research examining the moral faculty has not provided a satisfactory account of its development. Moreover, it is not clear from this line of research how the distinction between operative and expressed principles is best understood. For instance, it
would appear that even the principles that generally seem to be operative but not expressed (e.g., the principle of double effect) can, in fact, be expressed by certain individuals. This seems to indicate that the division between operative and expressed principles is not a “hard” distinction but, rather, that these are general categories with at least some degree of overlap. This raises at least two further questions. First, what determines the ability to provide sufficient justifications for operative principles? Second, does this ability show development—perhaps as a result of greater expertise with the ethical dilemmas of everyday life?

**The Present Study**

Borrowing from the methodology of Cushman and colleagues (2006), the present study attempts to shed light on some of the issues raised in the preceding paragraphs. More specifically, the present study had two primary goals: (1) to investigate whether or not there are age-related differences in the degree to which individuals produce moral judgments in accordance with principles of harm, and (2) to test whether or not there are age-related differences with respect to the ability to provide sufficient justifications for moral judgments. This study investigated these issues using a sample of adolescents from grades 9 and 12.

The purpose of sampling participants in these age groups is fourfold. First, since the majority of the research examining these principles of harm has been conducted on college students and adults, and since there is interest in establishing the origin of the moral faculty, extending the current age range downward will allow us to test the stability of this phenomenon in a younger sample. Second, a sample of this age range can reasonably be expected to understand scenarios similar to those found in the Hauser and
Cushman studies described above. That is, this sample will not be so drastically different from the samples used in previous studies so as to necessitate different (i.e., less complex) stimuli. Third, a sample of students from these age groups will enable us to examine the possibility of developmental differences, while still allowing us to utilize the same measures and methodology for the youngest and the oldest participants. Thus, there is an advantage to having a relatively narrow age range in the sense that it will allow us to use the same measures across all age groups. Finally, research suggests that the analytic system of processing that guides judgment and decision-making seems to be well developed by middle adolescence (Klaczynski, 2001).

**Hypotheses**

**Hypothesis 1.** Based on prior research with adults, it is expected that participants will conform their judgments to the three principles of harm being tested, but that they will, in many cases, be unable to provide an account of the reasoning behind their judgments. That is, they will tend to rate harm caused by physical contact to be worse than harm caused without contact; they will rate harm caused by action as worse than harm brought about through the omission of action; and they will rate harm that is intended as a means to an end as morally worse than harm that is merely a foreseen side-effect. However, when asked to explain their judgments, it is expected that participants will often be unable to provide an explicit account of their reasoning process.

**Hypothesis 2.** It is expected that participants at both grade levels will demonstrate a similar pattern of conforming their judgments to the three principles of harm under investigation. Meaning, we do not anticipate grade differences on the pattern of judgments of permissibility. This would be in line with the theoretical framework of
Hauser (2006) and Mikhail (2000), and it would suggest that the moral faculty—if it is a developmental construct at all—is well developed by this time. Alternatively, if it is the case that we find age-related differences with respect to individuals’ adherence to certain principles of harm, this would seem to indicate that the moral faculty may develop over time.

**Hypothesis 3.** It is expected, however, that there will be age-related differences in the ability to provide sufficient justifications for moral judgments. Specifically, it is expected that younger individuals will tend to provide fewer sufficient justifications than older individuals. This may indicate that as individuals age, the principles of harm that were once operative but not expressed become available to conscious reflection and, therefore, can be articulated. This would follow from the wealth of theoretical and empirical work of Kohlberg (1969, 1971, 1981), which has shown that explicit knowledge of morality develops over time. Conversely, if we find that there are no age-related differences in the ability to provide sufficient justifications, this would seem to indicate that this ability is either (a) not a developmental construct, or (b) well developed by the time individuals reach the age of our youngest participants.
METHOD

Participants

Participants were 86 adolescents from grades 9 ($n = 46$, 28 female, $M_{age} = 14.8$, $SD = 0.50$) and 12 ($n = 40$, 20 female, $M_{age} = 17.8$, $SD = 0.48$). Participants were recruited from a local high school in the South Bend Community School Corporation. Students were offered compensation in the amount of $5 for their participation in the study. An experimenter visited 9th- and 12th-grade classrooms to distribute consent forms and briefly explain the study. The experimenter then returned to the classrooms one week after the original visit in order to begin data collection. Participants were removed from class one at a time and directed to a quiet room in the school’s library. Once there, participants completed an assent form before being directed to the study on a laptop computer.

Materials

Trolley problems. As in the Cushman and colleagues (2006) study described above, participants in this study were presented with a number of trolley problems and asked to rate the permissibility of the actions described in the scenarios. These scenarios all have a similar structure in the sense that they describe an agent who must choose a course of action that will result in either the harming of one individual, or the harming of several individuals. Again, in keeping with the methodology used by Cushman and colleagues (2006), participants rated the permissibility of the actions described in the
scenarios on a scale of 1 (*Completely Wrong*) to 8 (*Completely Right*).\(^3\) The scenarios were accompanied by a schematic diagram that depicted the details of the dilemma (see Appendix A). Furthermore, in order to limit the possibility that judgments of permissibility would be affected by differences in reading ability or comprehension, the text of all scenarios had a readability index of 6.0 or lower—as measured by Flesch-Kincaid readability index. As indicated in Table 1 below, the average readability index across all 11 scenarios is 5.1, with lowest at 3.7 and the highest at 6.0. Thus, even the most complex passage still has a readability level three full grade levels below the students in the younger age group.

### TABLE 1

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Contact</th>
<th>Action</th>
<th>Intended Harm</th>
<th>Flesch-Kincaid Index</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Ned”</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>3.7</td>
<td>Mikhail, 2000</td>
</tr>
<tr>
<td>“Oscar”</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>5.0</td>
<td>Mikhail, 2000</td>
</tr>
<tr>
<td>“Frank”</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>5.5</td>
<td>Hauser, 2006</td>
</tr>
<tr>
<td>“Joe”</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>5.9</td>
<td>Cushman et al., 2006</td>
</tr>
<tr>
<td>“Casey”</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>6.0</td>
<td>Cushman et al., 2006</td>
</tr>
<tr>
<td>“Dan”</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>5.8</td>
<td>Original</td>
</tr>
<tr>
<td>“Chris”</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>5.2</td>
<td>Original</td>
</tr>
<tr>
<td>“Zach”</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>4.1</td>
<td>Original</td>
</tr>
<tr>
<td>“Bob”</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>5.9</td>
<td>Original</td>
</tr>
<tr>
<td>“Mark”</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>5.7</td>
<td>Original</td>
</tr>
</tbody>
</table>

**ATTRIBUTES OF TEST SCENARIOS**

All participants were presented with 11 scenarios (10 test scenarios, and 1 control

---

\(^3\) The study described in Cushman et al. (2006) used a rating scale with “forbidden” and “obligatory” as the anchors; however, since the present study used a considerably younger sample, it seems prudent not to use terminology with which participants might be unfamiliar. Moreover, a subsequent study by Cushman (2008) investigated the possible differences that may result from using the language of “wrong” and “right” as opposed to the “forbidden” and “obligatory”, and concluded that the difference in word choice did not affect participants’ judgments of permissibility.
The 10 test scenarios were similar in form to the classic trolley problems first conceptualized by Foot (1967). That is, they involve an agent faced with a moral dilemma, and depending on the actions of the agent in the scenario, there will either be five lives lost, or one life lost. These 10 test scenarios made up 9 test pairs. It is possible for a scenario to be part of multiple test pairs (see Table 2). A test pair is any pair of two scenarios that differ in accordance with only one specific principle. Take, for instance, the “Ned” and “Oscar” scenarios described earlier. These two scenarios constitute a test pair, and they differ according to the principle of double effect.

### TABLE 2

<table>
<thead>
<tr>
<th>SCENARIO PAIRS CORRESPONDING TO PRINCIPLES OF HARM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact Principle</td>
</tr>
<tr>
<td>Test Pair 1</td>
</tr>
<tr>
<td>Test Pair 2</td>
</tr>
<tr>
<td>Test Pair 3</td>
</tr>
</tbody>
</table>

A control scenario was created in order to ensure that participants were providing judgments that considered the details of the scenarios (see the “Ben” scenario in Appendix A). This control scenario seems similar in nature to the test scenarios, but instead of offering a choice between two courses of action that both result in a number of lives lost, the control scenario offers a choice of two actions, one of which involves lives lost and another which involves no loss of life. The use of a control scenario was helpful in identifying participants who responded without regard for the specific details of the dilemmas. The data from these participants was excluded from analyses.
Additionally, the amount of time that participants spent on each scenario was automatically recorded. This served two purposes. First, it was used as an exclusion criterion. If, for instance, a participant completed a scenario in only a matter of seconds, it is likely that they did not make a judgment based on the details of the scenario. Second, recording the amount of time spent on each scenario provided us with additional data that could prove informative in subsequent analyses. For example, it could be that differences in time spent on test scenarios differentially predict participants’ judgments of permissibility.

As in the study by Cushman and colleagues (2006) described above, the present study investigated the effect of three principles of harm: the contact principle, the action principle, and the principle of double effect. Each of these three principles of harm was tested using three pairs of scenarios (see Table 2). In order to test these principles of harm, all 10 of the test scenarios possess a combination of three of the following six attributes: (1a) harm caused by contact, (1b) harm caused without contact, (2a) harm caused by action, (2b) harm resulting from the omission of action, (3a) harm intended as the means to an end, (3b) harm that was merely a foreseen by-product. Each scenario possesses only combinations of three of these attributes that do not logically contradict one another. It is difficult to imagine, for instance, that there could be a scenario in which someone causes harm through physical contact and with the omission of action. An example of a possible combination of attributes is: 1b, 2a, 3a. This combination of attributes constitutes a scenario in which there is harm that is intended as the means to an end and is caused by action but without contact. Incidentally, this combination of
attributes describes those of the Ned scenario discussed above. The complete text of these scenarios can be found in Appendix A.

**Standardized test scores.** Although the test scenarios all have a readability index that is three full grade levels below the younger age group, it is possible that verbal ability may play a role in judgments of permissibility or in the ability to provide justifications for judgments. In order to control for the possible impact of language ability on judgments and justifications, participants’ standardized test scores were obtained from their school records—in particular, their English/Language Arts composite scores on the Indiana Statewide Testing for Educational Progress-Plus (ISTEP).

**Procedure**

Data collection for this study was conducted at the participants’ school during the normal school day. Thus, students who were willing—and whose parents had consented—were removed from their classrooms (providing their teachers had also approved this in advance) in order to complete the study. The experimenter ran participants one at a time in a quiet room at the school. The study was administered on a laptop computer brought to the school by the experimenter. The study was divided into two parts. The first portion involved rating the permissibility of actions described in different scenarios (i.e., the trolley problems described above). The second portion involved participants providing justifications for the ratings they provided.

**Judgments of permissibility.** During the first portion, participants were presented with 11 scenarios. These scenarios were presented in two blocks. Recall that a critical feature for investigating participants’ justifications of the three principles of harm is that each scenario is part of a pair. The two scenarios that make up a pair are nearly
identical, but differ with respect to one of the principles of harm. For the sake of simplicity, let us designate that each pair of dilemmas contains an $A$ scenario, and a $B$ scenario. Thus, of the 10 total test scenarios, there are 5 $A$ scenarios and 5 $B$ scenarios (see Figure 3 below for a list of the $A$ and $B$ scenarios and the principles tested by the pairs).

![Figure 3. Diagram of $A$ and $B$ Scenarios Illustrating the 9 Test Pairs.](image)

During the first block participants were randomly presented with a total of 6 scenarios: either all of the $A$ scenarios and the control scenario, or all of the $B$ scenarios and the control scenario. In the second block participants were presented with the 5 scenarios from the remaining group in random order. To clarify, participants began with either the $A$ scenarios, or the $B$ scenarios (counterbalancing based on random
assignment). In the second block they were presented with whichever group of scenarios they had not yet seen. For instance, if they were presented with the A scenarios in the first block, then they saw the B scenarios in the second block. Thus, the order of the scenario groups was counterbalanced, and the scenarios within the groups were randomized. The A scenarios and the B scenarios were presented in separate blocks in order to minimize the likelihood that participants saw two scenarios from the same test pair one after another or in close temporal proximity. This was intended to help mitigate the impact that participants’ judgments on earlier scenarios have on their later judgments.

Scenarios were presented to the participant on the computer. A schematic illustration of the scenario appeared along with the text of the scenario below the image. In order to control for differences in reading ability, the computer also played an audio narration of the text of the scenario—essentially, reading the scenario to the participant. Additionally, animation was used to highlight the relevant features of the scenario as the narration progressed. For instance, in the “Ned” scenario, when the narration reads “Ned is walking near train tracks when he sees a train that is out of control”, a large red circle appears around the train car. When the narration reads “Ahead on the track are five people”, the circle shifts so as to highlight these five individuals. We believe this system helps to hold the participant’s attention, and that it minimizes the possibility of confusion.

**Justifications.** Once the participants completed the 11 scenarios (10 test and 1 control) they moved on to the justifications portion of the study. Recall that this portion of the study is concerned with participants’ justifications of the judgments of permissibility they provided in the previous section. Recall also that the 10 test scenarios make up 9 test pairs containing an A scenario and a B scenario. Finally, recall that the A
and B scenarios of a test pair differ with respect to one of the principles of harm under investigation. The second portion of this study, then, is interested in the cases in which participants provided divergent judgments of permissibility within a pair of scenarios. For example, if a participant rated the permissibility of the actions described in the “Ned” scenario as a 4, and the permissibility of the actions described in “Oscar” scenario as a 6 (where higher scores indicate more permissibility or less “wrongness”), then this participant would be asked to provide a justification for the divergent ratings.

During the justifications portion of this study participants were asked to provide justifications for up to 5 pairs of dilemmas. The number of justifications that the participants were asked to provide was contingent upon their judgments of permissibility on the 10 test scenarios. Since participants were only asked to provide justifications when they gave divergent ratings on two scenarios that make up a test pair, there could be between 0 and 9 scenarios pairs for which justifications would be needed: 0 if they gave no divergent ratings within test pairs, and 9 if they gave divergent ratings across all test pairs. However, the ceiling of a maximum of 5 requested justifications was imposed in order to comply with anticipated time constraints. The request for justifications was in the form of a very brief structured interview conducted by the experimenter. These brief interviews were audio taped and transcribed for later coding. Previous studies investigating justifications of moral judgments (e.g., Cushman et al., 2006; Hauser et al., 2007) have used a written response format for the participants’ justifications; however, due to the age characteristics of the sample in the present study, the use of brief interviews was meant to help control for any possible difference in justifications due to verbal or writing ability.
Participants were presented with a side-by-side view of the two scenarios (schematic diagrams along with text) making up the test pair on which they provided divergent ratings of permissibility. The experimenter then asked the participant “Could you explain why you rated this one [experimenter indicates the scenario that was rated as less permissible] as worse than this one [experimenter indicates the scenario rated as more permissible]”. If the participant seemed confused or did not respond, the experimenter used the following sequence of 3 probes: (1) “Do you understand what I’m asking? Why is what’s happening here [experimenter indicates the scenario rated as less permissible] worse than what’s happening here [experimenter indicates the scenario rated as more permissible].” (2) “Is there anything about this one [experimenter indicates the scenario rated as less permissible] that makes it more wrong than this one [experimenter indicates the scenario rated as more permissible].” (3) “Do you see any differences between these two that might make one worse than the other?” If after this the participant had not attempted to provide a justification, the experimenter gave instructions to move on to the next justification. The participant was then presented with the next test pair for which they were asked to provide a justification and the above sequence was repeated. This process continued until they were presented with a maximum of five test pairs for which they were asked to provide justifications—less, if they provided divergent ratings across 4 or fewer test pairs. The order of the test pair justification prompts was randomized. After this, participants were directed to the screen inquiring about demographic information.

**Coding of justifications.** The participants’ justifications were coded according to the coding scheme described in Cushman and colleagues (2006). There are seven codes
that can be assigned to an individual justification: (1) “mistest” (2) sufficiency, (3) failure, (4) uncertainty, (5) denial, (6) alternative explanation, and (7) error. The first three codes (i.e., mistest, sufficient, and failed) are mutually exclusive; the last four codes can co-occur with any other code.

The mistest code is assigned for several different reasons including: when a participant rants or avoids the questions (e.g., “These are so unrealistic!”), when a participant misunderstands that a comparison between the two scenarios is the relevant task, or—most commonly—when a participant clearly misunderstands the stimuli (e.g., “I think it would be wrong for Mark to step on the man’s hand.”). A sufficient justification is one in which the participant identifies a factual difference between the two scenarios, and claims (or implies) that this was the basis for the difference in judgments. The failure code is assigned when the participant suggests an alternative principle but not one that could account for the actual pattern of judgments. The uncertainty code is used for justifications in which the participant explicitly references their own uncertainty regarding the difference in judgments, or they directly state that they are unable to provide a justification. The denial code is assigned to justifications in which the participant denies that there is a moral difference between the actions described in the two scenarios. The added assumption code is used for justifications in which the participant references facts that are not present in the actual text of the scenario. Finally, the error code is assigned when participants claim to have made an error when making the judgments of permissibility (e.g., clicking the wrong number), or—more commonly—when they report an attitudinal shift.
Once the participant completed the demographics portion, the study was concluded. The participant was debriefed, thanked, and compensated ($5) for their participation. If the participant had any questions, they were answered at this time. If the participant did not provide divergent judgments of permissibility within any of the 9 test pairs of scenarios, then they were not required to take part in the justifications portion, and the study was complete after the demographics portion of the study. Completion of the study generally took between 15 and 20 minutes depending on the speed with which the participant completed the judgments portion, and on the number of justifications that the participant was asked to provide.
RESULTS

Plan of Analysis

Of the 94 individuals who completed the study, 4 were excluded for failing the control scenario (see Ben scenario in Appendix A) and 4 were excluded for response times below the 4-second threshold deemed necessary for comprehension of the scenarios (Cushman, et al., 2006). Participants were asked to complete two tasks in this study. In the first, they provided judgments of permissibility on 10 test scenarios. These 10 scenarios are paired in nine ways (a scenario may be part of multiple pairs); each of these pairs tests one principle of harm. There are three scenario pairs testing each principle. Paired sample t-tests were performed on each of the nine scenario pairs to determine whether participants rated one scenario as morally worse (i.e., less permissible) than the other scenario in the pair. Since all comparisons were clearly planned and guided by a priori hypotheses, familywise error correction was deemed unnecessary (Keppel & Wickens, 2004; Kirk, 1995).

The second task required participants to provide justifications for their judgments of permissibility. The request for justifications came in the form of a structured interview. Responses were transcribed and coded by independent raters. The coding scheme and procedure for checking inter-rater agreement are described in detail below.

Preliminary analyses. Preliminary analyses revealed a subtle effect of gender with respect to permissibility ratings. Across all scenarios, males tended to provide higher
permissibility ratings. However, this effect only achieved statistical significance in the case of the “Casey” and “Joe” scenarios, $F(1,84) = 4.60, p = .04$, and $F(1,84) = 4.00, p = .05$, respectively.

There were also slight differences with respect to ISTEP English/Language Arts scores. In particular, preliminary analyses revealed that for 12th graders—but not 9th graders—ISTEP scores are related to the proportion of sufficient justifications on both contact principle pairs, and the action principle pairs, $r = .49, p = .01$, and $r = .44, p = .01$, respectively (see Appendix B for summary of ISTEP scores).

**Judgments of Permissibility**

Participants rated the permissibility of the actions described in each scenario. Ratings ranged from 1 (Completely Wrong) to 8 (Completely Right). The descriptive statistics for the 9th- and 12th-grade groups on all test scenarios are reported in Table 3 and Table 4, respectively. Both skewness and kurtosis were within the acceptable range across all scenarios.

**TABLE 3**

DESCRIPTIVE STATISTICS FOR 9TH-GRADE GROUP ON TEST SCENARIOS

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness Statistic</th>
<th>Std. Error</th>
<th>Kurtosis Statistic</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ned</td>
<td>46</td>
<td>1</td>
<td>8</td>
<td>3.87</td>
<td>1.96</td>
<td>0.14</td>
<td>0.35</td>
<td>–0.91</td>
<td>0.69</td>
</tr>
<tr>
<td>Bob</td>
<td>46</td>
<td>1</td>
<td>8</td>
<td>3.76</td>
<td>2.10</td>
<td>0.32</td>
<td>0.35</td>
<td>–0.92</td>
<td>0.69</td>
</tr>
<tr>
<td>Joe</td>
<td>46</td>
<td>1</td>
<td>8</td>
<td>5.02</td>
<td>1.94</td>
<td>–0.34</td>
<td>0.35</td>
<td>–0.73</td>
<td>0.69</td>
</tr>
<tr>
<td>Dan</td>
<td>46</td>
<td>1</td>
<td>8</td>
<td>4.07</td>
<td>1.91</td>
<td>–0.16</td>
<td>0.35</td>
<td>–1.00</td>
<td>0.69</td>
</tr>
<tr>
<td>Zach</td>
<td>46</td>
<td>1</td>
<td>7</td>
<td>3.13</td>
<td>1.64</td>
<td>0.19</td>
<td>0.35</td>
<td>–0.75</td>
<td>0.69</td>
</tr>
<tr>
<td>Frank</td>
<td>46</td>
<td>1</td>
<td>7</td>
<td>3.28</td>
<td>1.64</td>
<td>0.31</td>
<td>0.35</td>
<td>–0.40</td>
<td>0.69</td>
</tr>
<tr>
<td>Oscar</td>
<td>46</td>
<td>1</td>
<td>8</td>
<td>4.72</td>
<td>1.81</td>
<td>–0.36</td>
<td>0.35</td>
<td>–0.42</td>
<td>0.69</td>
</tr>
<tr>
<td>Mark</td>
<td>46</td>
<td>1</td>
<td>7</td>
<td>3.61</td>
<td>1.75</td>
<td>0.24</td>
<td>0.35</td>
<td>–0.79</td>
<td>0.69</td>
</tr>
<tr>
<td>Chris</td>
<td>46</td>
<td>1</td>
<td>8</td>
<td>3.57</td>
<td>1.96</td>
<td>0.22</td>
<td>0.35</td>
<td>–0.74</td>
<td>0.69</td>
</tr>
<tr>
<td>Casey</td>
<td>46</td>
<td>1</td>
<td>8</td>
<td>5.17</td>
<td>1.54</td>
<td>–0.34</td>
<td>0.35</td>
<td>0.45</td>
<td>0.69</td>
</tr>
</tbody>
</table>
TABLE 4

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Statistic</th>
<th>Std. Error</th>
<th>Statistic</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ned</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>4.45</td>
<td>1.91</td>
<td>−0.27</td>
<td>0.37</td>
<td>−0.58</td>
<td>0.73</td>
</tr>
<tr>
<td>Bob</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>3.63</td>
<td>1.97</td>
<td>0.22</td>
<td>0.37</td>
<td>−0.77</td>
<td>0.73</td>
</tr>
<tr>
<td>Joe</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>5.40</td>
<td>1.95</td>
<td>−0.73</td>
<td>0.37</td>
<td>0.00</td>
<td>0.73</td>
</tr>
<tr>
<td>Dan</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>2.95</td>
<td>1.85</td>
<td>0.66</td>
<td>0.37</td>
<td>−0.61</td>
<td>0.73</td>
</tr>
<tr>
<td>Zach</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>2.73</td>
<td>2.11</td>
<td>0.90</td>
<td>0.37</td>
<td>−0.63</td>
<td>0.73</td>
</tr>
<tr>
<td>Frank</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>2.95</td>
<td>2.14</td>
<td>0.92</td>
<td>0.37</td>
<td>−0.40</td>
<td>0.73</td>
</tr>
<tr>
<td>Oscar</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>4.95</td>
<td>1.97</td>
<td>−0.45</td>
<td>0.37</td>
<td>−0.71</td>
<td>0.73</td>
</tr>
<tr>
<td>Mark</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>4.02</td>
<td>2.19</td>
<td>0.24</td>
<td>0.37</td>
<td>−0.92</td>
<td>0.73</td>
</tr>
<tr>
<td>Chris</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>2.50</td>
<td>1.92</td>
<td>1.43</td>
<td>0.37</td>
<td>1.29</td>
<td>0.73</td>
</tr>
<tr>
<td>Casey</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>5.30</td>
<td>2.03</td>
<td>−0.43</td>
<td>0.37</td>
<td>−0.70</td>
<td>0.73</td>
</tr>
</tbody>
</table>

DESCRIPTIVE STATISTICS FOR 12TH-GRADE GROUP ON TEST SCENARIOS

Contact principle. Figure 3 illustrates the scenario pairs, and the principle of harm being tested by each pair. Table 5 below summarizes the paired sample t-tests by group along with measures of effect size. As indicated in Table 5, of the three scenario pairs testing the contact principle, two achieved significance or marginal significance in the expected direction: the Frank and Ned pair, as well as the Frank and Bob pair.
**TABLE 5.**
SUMMARY OF PAIRED SAMPLE T-TESTS ON SCENARIO PAIRS

<table>
<thead>
<tr>
<th></th>
<th>Grade 9</th>
<th></th>
<th></th>
<th>Grade 12</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Difference</td>
<td>SD</td>
<td>t(45)</td>
<td>p  (two-tailed)</td>
<td>Effect Size (d)</td>
<td>Mean Difference</td>
</tr>
<tr>
<td><strong>Contact Principle</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frank – Ned</td>
<td>−0.59</td>
<td>1.53</td>
<td>−2.60</td>
<td>.01</td>
<td>.39</td>
<td>−1.50</td>
</tr>
<tr>
<td>Frank – Bob</td>
<td>−0.49</td>
<td>1.66</td>
<td>−1.96</td>
<td>.06</td>
<td>.30</td>
<td>−0.68</td>
</tr>
<tr>
<td>Zach – Chris</td>
<td>−0.44</td>
<td>1.96</td>
<td>−1.50</td>
<td>.14</td>
<td>.23</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Action Principle</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ned – Mark</td>
<td>0.26</td>
<td>1.48</td>
<td>1.19</td>
<td>.24</td>
<td>.18</td>
<td>0.43</td>
</tr>
<tr>
<td>Bob – Mark</td>
<td>0.15</td>
<td>1.75</td>
<td>0.59</td>
<td>.56</td>
<td>.09</td>
<td>−0.40</td>
</tr>
<tr>
<td>Chris – Dan</td>
<td>−0.50</td>
<td>1.41</td>
<td>−2.41</td>
<td>.02</td>
<td>.36</td>
<td>−0.45</td>
</tr>
<tr>
<td><strong>Principle of Double Effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ned – Oscar</td>
<td>−0.85</td>
<td>2.04</td>
<td>−2.81</td>
<td>.01</td>
<td>.42</td>
<td>−0.50</td>
</tr>
<tr>
<td>Bob – Oscar</td>
<td>−0.96</td>
<td>2.09</td>
<td>−3.11</td>
<td>&lt;.01</td>
<td>.46</td>
<td>−1.33</td>
</tr>
<tr>
<td>Casey – Joe</td>
<td>0.15</td>
<td>1.25</td>
<td>0.83</td>
<td>.41</td>
<td>.13</td>
<td>−0.10</td>
</tr>
</tbody>
</table>
The “Frank” and “Ned” scenarios constitute a test pair that investigates the effect—if any—of the contact principle. According to this principle, it was expected participants would rate the actions of Frank as morally worse than the actions of Ned, because Frank is causing harm through physical contact, whereas the harm that Ned causes is mediated by a mechanical device and, thus, less direct. As expected, both the 9th-grade group, and the 12th-grade group rated Frank’s actions as morally worse than Ned’s, $t(1,45) = -2.60, p = .01$, and $t(1,39) = -5.50, p < .001$, respectively.

This same effect was demonstrated in testing the difference between the mean permissibility ratings on the “Bob” and Frank scenarios. Just as with Ned, Bob is causing harm in an indirect manner, without physical contact, while Frank is causing harm through direct physical contact. Both the 9th- and 12th-grade groups tended to rate the actions of Frank as morally worse than those of Bob, $t(1,45) = -1.96, p = .06$, and $t(1,39) = -2.08, p = .04$, respectively. Note that in the case of the 9th-grade group this effect did not quite reach statistical significance, however, the trend is in the expected direction, and the effect size (see Table 5) is comparable to that of the 12th-grade group.

**Action principle.** Of the three scenario pairs testing the action principle, only one achieved significance in the expected direction. The “Chris” scenario and the “Dan” scenario constitute a test pair that examined the effect of the action principle. Recall that, according to the action principle, harm that is caused through action is considered morally worse than harm resulting from the omission of action. In accordance with this, it was expected that participants would rate the actions of Chris to be morally worse than the actions of Dan, because Chris is causing harm through action. This is exactly the
pattern that was observed; both the 9th- and 12th-grade groups rated the actions of Chris as morally worse than the actions of Dan, $t(1,45) = -2.41, p = .02$, and $t(1,39) = -2.33, p = .03$, respectively.

**Principle of double effect.** Of the three scenario pairs testing the principle of double effect, two achieved significance in the expected direction. According to the principle of double effect, it was expected that permissibility ratings on the Ned scenario would be significantly lower than ratings on the Oscar scenario. This pattern was anticipated because Ned’s actions involve causing harm that is intended as the means to an end; whereas, in the case of Oscar, the harm that results from his actions is merely a foreseen side-effect of his actions—the harm has no causal role in bringing about the greater good in Oscar’s case. As predicted, this effect was observed. Participants in both the 9th-grade group and the 12th-grade group rated the actions of Ned as morally worse than those of Oscar, $t(1,45) = -2.81, p = .01$, and $t(1,39) = -2.11, p = .04$, respectively.

The same effect was discovered when examining the difference in scores on the Bob and Oscar scenarios. Again, Bob’s actions involve causing harm that is intended as the means to an end, whereas the harm that Oscar causes is merely a foreseen by-product. As expected, participants in both the 9th-grade group and the 12th-grade group rated the actions of Bob as worse than those of Oscar, $t(1,45) = -3.11, p < .01$, and $t(1,39) = -5.16, p < .001$, respectively.

**Grade differences.** In the previous analyses, tests of the effect of scenario were analyzed within grade. As indicated in Table 5, 9th-grade students and 12th-grade students demonstrated a nearly identical pattern of ratings with respect to scenario pairs. However, in order to further investigate the possibility of group differences, repeated measures
analysis of variance (RM-ANOVAs) were conducted on each scenario pair. For each analysis, the within-subject factor was the scenario pair and year-in-school was the between-subject factor. There were some minor gender differences with respect to permissibility ratings on the scenarios; thus, gender was also included as a covariate in these analyses. Levene’s test of the equality of error variances verified that the assumption of homogeneity of variance was satisfied. Of the five scenario pairs that proved significant in the analyses described above, only the Frank and Ned pair demonstrated a statistically significant group by dilemma interaction, $F(1, 83) = 6.51, p = .01$. Upon examining the means, it appears that the 9th-grade group rated the actions of Frank as more permissible compared to the 12th-grade group’s ratings, and the actions of Ned as less permissible compared to the 12th-grade group’s ratings.

**Summary of judgments.** Of the nine test pairs, five were statistically significant in the expected direction among the 9th graders and among 12th graders. Overall, participants tended to conform their judgments to the three principles of harm under investigation. That is, they tended to rate harm caused through physical contact as morally worse than harm caused without physical contact; they rated harm caused by action as worse than harm brought about by omission; and they rated harm that was intended as the means to an end as morally worse than harm that was merely a foreseen side-effect. Moreover, apart from the one instance described above, there were no major, systematic differences in the pattern of ratings between the 9th- and 12th-grade participants.
Justifications

In the second portion of the study participants were asked to provide justifications explaining why they rated the actions of the agent in one scenario as morally worse than the actions of the agent in another scenario. For instance, if the participant rated the actions of Ned as morally worse than those of Oscar, then in the justifications portion of the study they may be asked to explain the reason for the difference in ratings.

Inter-rater agreement. Justifications were coded according to the coding scheme used in Cushman and colleagues (2006), which is described above. The primary coder was responsible for coding every justification ($N = 404$). A second coder, who was blind to the hypotheses of the study, also coded 25% of the justifications ($n = 101$). Inter-rater agreement was calculated on this subset of justifications revealing a Cohen’s kappa value of .75, which indicates substantial inter-rater agreement (Landis & Koch, 1977). Disagreements between raters on this subset were resolved through discussion.

Overall pattern of justifications. In general, both the 9th and the 12th graders had difficulty providing sufficient justifications for their judgments (see Figure 4). It is also clear from Figure 4 that the general pattern of sufficient justifications by principle replicates the pattern reported by Cushman and colleagues (2006). Meaning, participants were most successful in providing sufficient justification for their judgments of action principle pairs, and least successful with doctrine of double effect pairs. The contact principle pairs fell somewhere in between. The 9th graders sufficiently justified 54.6% of the action principle pairs, 44.7% of the contact principle pairs, and 21.5% of the doctrine of double effect pairs. The 12th graders sufficiently justified 68.1% of the action principle
pairs, 53.3% of the contact principle pairs, and 31.3% of the doctrine of double effect pairs.

**Figure 4.** Overall Percentage of Sufficient Justifications by Principle for 9th and 12th Graders.

**Overall grade differences.** As indicated from the percentages reported above, there were grade-level differences. Overall, 12th-grade students provided significantly more sufficient justifications for their judgments than did 9th-grade students, $F(1, 85) = 7.53, p < .01$. Moreover, the 9th graders provided significantly more justifications in which they denied there was a moral difference between scenarios, $F(1, 85) = 5.20, p = .03$, and significantly more justifications in which they reported an error in responding, $F(1, 85) = 4.22, p = .04$. Again, almost invariably, the error code meant the participant expressed an attitudinal shift regarding which scenario they rated as morally worse.

**Grade differences by principle.** In order to untangle the overall grade differences in justifications, a proportion of sufficient justifications by principle was
computed for each participant. Thus, for each of the three principles of harm, participants had a score representing the number of sufficient justifications on scenario pairs differing by a certain principle divided by the number of possible instances for a justification on that principle. With respect to the contact principle, there does appear to be a difference between 9th and 12th graders with respect to the proportion of sufficient justifications. Specifically, 9th graders provided a lower average proportion of sufficient justifications than did 12th graders, $F(1, 73) = 5.17, p = .03$. However, as previously mentioned, preliminary analyses revealed 12th graders’ ISTEP English/Language Arts scores were related to the proportion of sufficient justifications on both the contact principle, and the action principle. Thus, the analysis above was repeated with ISTEP scores as a covariate. The result—with respect to the contact principle—was that the difference between 9th and 12th graders on the proportion of sufficient justifications was no longer statistically significant after controlling for ISTEP scores, $F(1, 63) = 2.65, p = .11$. A similar pattern emerged with respect to the action principle; when testing group difference on the proportion of sufficient justifications, there does appear to be a marginal effect of group, with 12th graders having a greater proportion of sufficient justifications than 9th graders, $F(1, 70) = 3.15, p = .08$. However, after controlling for ISTEP English scores, this difference becomes non-significant, $F(1, 60) = 2.45, p = .12$.

**Grade differences by scenario pair.** In order to further investigate the source of the difference in sufficient justifications, chi-square tests were conducted on individual scenario pairs. For the purposes of these analyses, the frequency of sufficient codes on each scenario pair was compared against the frequency of all insufficient codes combined (i.e., failed, uncertainty, denial, added assumption, mistest, and error codes collapsed). Of
the nine scenario pairs being tested, there were four on which 9th and 12th graders showed at least marginally significant differences in the pattern of sufficient versus insufficient justifications. On the Frank and Bob pair, 12th graders provided more sufficient justifications, and fewer insufficient justifications than 9th graders, $\chi^2(1, N = 42) = 5.30, p = .02$ (Figure 5). Similarly, on the Frank and Ned pair, 12th graders also provided more sufficient justifications, and fewer insufficient justifications than the 9th graders, $\chi^2(1, N = 40) = 3.90, p = .05$ (Figure 6). A similar—though not quite statistically significant—pattern emerged for the Ned and Oscar pair, $\chi^2(1, N = 41) = 3.45, p = .06$ (see Figure 7), as well as the Chris and Dan pair, $\chi^2(1, N = 34) = 2.99, p = .08$ (see Figure 8). Graphs of these four pairs appear below in the figures below. It is worth noting that each of the three principles of harm is represented in these four pairs—two correspond to the contact principle, and one each to the action principle and principle of double effect.

![Figure 5. Number of Sufficient and Insufficient Justifications for the Frank and Bob Scenario Pair for 9th and 12th Graders.](image-url)
Figure 6. Number of Sufficient and Insufficient Justifications for the Frank and Ned Scenario Pair for 9th and 12th Graders.

Figure 7. Number of Sufficient and Insufficient Justifications for the Ned and Oscar Scenario Pair for 9th and 12th Graders.
Figure 8. Number of Sufficient and Insufficient Justifications for the Chris and Dan Scenario Pair for 9th and 12th Graders.

**Strength of judgments and justifications.** In order to ensure that sufficient justifications are not merely the product of “stronger” moral judgments (i.e., more divergent ratings across the two scenarios making up the pair), difference scores for each scenario pair were computed. These difference scores allow an examination of whether or not participants who gave sufficient justifications on a given scenario pair might have also given more divergent ratings on the two scenarios making up the pair in question. The difference scores are essentially a measure of the “strength” or conviction of the participant’s moral judgments on that pair. For example, perhaps someone who gave a sufficient justification for the Ned and Oscar pair rated the Ned scenario as a 2 and the Oscar scenario as a 7 (where higher scores indicate more permissibility). But perhaps
someone who failed to provide a sufficient justification rated the Ned scenario as a 3 and the Oscar scenario as a 5. The first participant would have a difference score with an absolute value of 5; the second would have a difference score with an absolute value of 2. This second participant might reasonably be described as feeling less strongly about the difference between the two scenarios.4

The absolute value of the difference scores for each scenario pair were included as the within-subject factor in a one-way ANOVA with justification of the scenario pair (i.e., sufficient, or insufficient) as the between-subjects variable. This analysis was repeated for each of the nine scenario pairs. Of the nine pairs, only the Frank and Ned pair demonstrated a significant effect of group, with participants who provided sufficient justifications on this pair also having more divergent values on the Frank and Ned scenarios, $F(1, 38) = 9.53, p < .01$. This is an interesting finding, but the fact that only one of the nine scenario pairs demonstrated this effect suggests that the ability to justify one’s judgments is rather independent of the strength of those judgments.

**Unexpected Findings**

Over the course of data collection, it became clear during the interview portion of the study that several of the stimuli were being interpreted in unexpected and problematic ways. This was particularly true of the “Mark” scenario, but was also true of the “Zach” scenario.

**Problem with Mark.** The Mark scenario is part of two scenario pairs designed to test the action principle. The dilemma describes a large man on a footbridge. The man has slipped and is barely hanging on to the edge of the footbridge. He will fall to his

---

4 I am very grateful to the members of my thesis committee, especially Nicole McNeil, for having pointed out this possibility.
death if Mark does not help him back up. However, if this man were to fall to his death his body would slow the runaway train car sufficiently so as to allow the five people further down the track time to escape. Thus, if Mark refrains from acting, then five lives will be saved. If Mark helps the large man up, then the five people will be killed. It was expected that participants would rate the actions of Mark as more permissible than the actions of Ned and the actions of Bob, since both Ned and Bob are causing harm through direct action. However, on average, this was not the case.

The results of the \( t \)-tests revealed no statistically significant differences between permissibility ratings on Ned and Mark, nor were there differences on ratings of Bob and Mark. It seems from the interviews that the likely explanation is that participants found the thought of letting someone fall to their death right before one’s eyes to be far too disturbing to allow. In the initial reading of the justifications it became clear that a considerable number of participants used emotion words when describing their justifications for the Ned and Mark scenario pair as well as the Bob and Mark scenario pair. Therefore, when coding the justifications, instances of emotional language were noted in all scenarios. This included any occasion in which the participant referenced their expectations of Mark’s emotions, the victim’s emotions, or their own emotions (e.g., Participant 55: “I feel like, if I were put into this situation, um, not helping the man up, I feel, I would feel something of guilt”).

An interesting pattern emerged when examining instances of justifications involving emotion language. Of the participants who rated the actions of Mark to be more permissible than the actions of either Ned or Bob (i.e., those who did conform to the action principle), 6% of the justifications referenced emotions; however, of the
participants who rated the actions of Mark as morally worse than those of either Ned or Bob (i.e., those who did not conform to the action principle), 32% of the justifications involved emotional content. Thus, participants who did not conform their judgments to the action principle on these two scenario pairs used more emotional language than those who did conform to the action principle.

This suggests an interesting possibility. Perhaps those individuals who did not conform to the action principle on either the Ned and Mark pair, or the Bob and Mark pair were more likely to use emotional language across all their justifications. This is, in fact, the pattern that was observed. Using participant’s judgments on the Ned and Mark pair as a grouping variable, it was observed that those who did not conform their judgments to the action principle (i.e., rated Mark’s actions as worse than Ned’s) used more emotional language throughout their justifications compared to those that who did adhere to the action principle in this case, $F(2, 83) = 8.23, p = .001$. The same pattern was observed with the Bob and Mark scenario pair; individuals who did not conform to the principle on this pair were more likely to use emotional language across all of their justifications, $F(2, 83) = 7.88, p = .001$. This pattern of findings may suggest that those participants not applying the action principle on the Mark scenario may be different in some relevant way compared to those who did apply this principle; specifically, those not applying the action principle may have been guided by affective responses, while those who applied the principle were not.

**Problem with Zach.** An interesting effect was also observed with respect to the Zach scenario. During the course of data collection, it became clear after several interviews that the text of the scenario contained a miscalculation that seemed to have an
unexpected and powerful impact on participant’s judgments of the permissibility of Zach’s actions. The Zach scenario describes a doctor (i.e., Zach) with five patients, all of whom need organ transplants. Zach also has one patient who is in a coma. The scenario asks the participant to rate how right or wrong it would be for Zach to cover his coma patient’s mouth, killing him, and then use his organs to save Zach’s five other patients. This scenario was designed to test the contact principle. It was expected that participants would rate the actions of Zach as worse than those described in the “Chris” scenario—in which the harm is caused without physical contact. However, this was not the case.

Paired sample $t$-tests revealed no statistically significant difference between participants’ rating on the Chris and Zach scenarios.Judging from the interview portion of the study, a possible explanation for this is that the Zach scenario describes a coma patient who “is barely breathing and is near death”, whereas the Chris scenario made no mention of this. It is worth noting that the description of the patient’s poor health was included so that the participant could reasonably be expected to imagine a situation in which simply covering someone’s mouth would lead to their death; as opposed to imagining a situation in which there is a violent struggle as a doctor intentionally smothers his patient. Describing a patient in poor health seems to allow for the possibility of a less graphic death. However, this obviously had an unintended effect. Several participants mentioned the coma patient’s poor health as the reason for their divergent ratings on the Zach and Chris scenarios. It seems that, in some cases, participants were more willing to allow the killing of a patient who is already near death, even if that killing involved physical contact.
This study sought to investigate age differences with respect to the application and justifications of three principles of harm. These principles were tested in a sample of adolescents in grades 9 and 12. Participants rated the moral permissibility of actions described in several ethical dilemmas, and were then asked to provide explanations for their judgments. The three principles under investigation included: the contact principle, the action principle, and the principle of double effect. The contact principle says that participants would consider harm caused by contact as worse than harm caused indirectly without physical contact. The action principle states that participants would rate harm caused by action as worse than harm brought about through the omission of action. Finally, the principle of double effect says participants would rate harm that is intended as the means to an end as worse than harm that is merely a foreseen side-effect.

Hypotheses Revisited

**Hypothesis 1: Understanding principles of harm.** Recall that the present study aimed to test three specific hypotheses. Hypothesis 1 (H1) posited that participants would demonstrate an operative understanding of the three principles of harm being tested. That is, they would conform their judgments of permissibility to the principles, yet be unable to articulate sufficient justifications for their judgments. Participants’ tendencies to conform—or *not* conform—their judgments to these three principles of harm was tested by examining their pattern of ratings across different scenario pairs.
Each principle of harm was tested using three scenario pairs. With respect to the contact principle, two of the three pairs found statistically significant effects of principle. In the case of the action principle, only one of the three pairs indicated a statistically significant effect of principle. Finally, in the case of the principle of double effect, again, two of the three pairs revealed a statistically significant effect of principle.

Overall, participants had difficulty justifying their ratings of permissibility. The pattern of justifications suggests a dissociation between judgments and justifications. That is, individuals conformed their judgments to these principles, yet they were often unable to provide an account of how they arrived at their judgments. This was particularly true of the principle of double effect. This dissociation between judgments and justifications has been thought to indicate that these principles of harm may operate outside of conscious awareness (Cushman et al., 2006; Hauser et al., 2007; Mikhail, 2007).

At the minimum, this pattern of findings provides partial support for H1. It does seem a bit disconcerting, however, that of the nine scenario pairs being tested, only five demonstrated statistically significant effects of principle. But this can likely be explained by the unexpected findings with the three scenarios involving either Mark or Zach.

The partial support of H1 does provide some evidence for Hauser (2006) and Mikhail’s (2000) understanding of the moral faculty. That is, it seems to be the case that participants make judgments about the moral permissibility of harmful actions according to a set of specific principles, which, in many cases, are not available to conscious reflection. These principles of harm take into consideration causal and intentional
features of actions, as well as the directness of the harm, and use this information to guide judgments of permissibility.

**Hypothesis 2: Grade differences in judgments.** Recall that Hypothesis 2 (H2) proposed that both 9th and 12th graders would demonstrate a similar pattern of conforming their judgments to the three principles of harm under investigation. Thus, no grade differences were expected in the application of the three principles. This hypothesis was largely supported. The 9th graders and the 12th graders demonstrated a nearly identical pattern of permissibility ratings; the scenario pairs that demonstrated a statistically significant effect of principle among the 9th graders also proved significant for the 12th graders. Furthermore, only one of the nine scenario pairs revealed a statistically significant group by ratings interaction. This suggests that the moral faculty, if it is a developmental construct, is well developed by the age of 14.

The support of H2 in the present study does align with the theoretical model of both Hauser (2006) and Mikhail (2000). That is, both 9th and 12th graders conformed their judgments to the principles of harm under investigation. However, the results from this study do not go so far as to support Hauser and Mikhail’s suggestion that elements of the moral faculty may be the product of some native systems. Further research—especially with a younger sample—would be required to substantiate this claim.

**Hypothesis 3: Age differences in justifications.** Finally, recall that Hypothesis 3 (H3) posited that there would be age-related differences in the ability to provide sufficient justifications for moral judgments. It was expected that younger individuals would tend to provide fewer sufficient justifications than older individuals. This may indicate that with age, the principles of harm that were once operative but not expressed,
become available to conscious reflection and can, therefore, be articulated. Again, this hypothesis was largely supported. As expected, 12th graders provided more sufficient justifications. It is interesting to note, however, that this difference was only apparent with respect to the contact principle and the action principle. There was no effect of group on the ability of participants to provide sufficient justification for scenario pair that differed according to the principle of double effect.

It is also interesting to note that verbal ability played a role in the ability to provide sufficient justifications for moral judgments. Specifically, verbal ability—as measured by ISTEP English/Language Arts scores—was related with the 12th graders’ proportion of sufficient justifications on both contact principle pairs, and action principle pairs. Additionally, after controlling for participants’ ISTEP scores, the difference between 9th and 12th graders in the ability to provide sufficient justifications is diminished. These are interesting and informative findings that have not previously appeared in the literature. And while this relationship should not come as a surprise, it is nonetheless interesting to point out that verbal ability seems to influence explicit knowledge of morality, which has in the past been viewed as the standard by which moral reasoning is measured (e.g., Kohlberg, 1969, 1971).

Taken together, these findings indicate that although there are grade differences with respect to the ability to provide justifications for one’s judgments, these differences are—at least in part—related to differences in verbal ability. One major implication of this finding is that the division between operative moral principles (i.e., those driving judgments), and expressed moral principles (i.e., those driving justifications) may merely be a function of differences on a certain cognitive capacities such as verbal ability.
Another implication of this finding for future research is to exercise caution whenever verbal reports are used in assessing moral knowledge. Indeed, this warning is perhaps relevant any time a verbal response format is used.

**Mark Revisited**

**Lessons learned.** The obvious advantage of using a verbal response format, however, is the wealth of information that qualitative data can provide. This was especially true in the case of the Mark scenario. This scenario was originally thought to be one on which participants would be quite willing to allow the harm to occur—since it would occur by way of an omission. This was not the case. Participants were quite bothered by the thought of allowing a man to plummet to his death before one’s very eyes. And it became apparent from the participants’ justifications that those individuals who did not conform to the action principle on scenario pairs involving Mark seemed to be more likely to cite emotions in their justifications compared to participants who did conform to the action principle. Essentially, those participants who found Mark’s actions morally worse than either Ned or Bob’s appear to have been at least partly influenced by an affective response.

Though unanticipated, and not explained by Hauser’s (2006) model, this finding is quite intriguing, and suggests several avenues for future research. In particular, it would be interesting to further investigate the instances in which individuals tend to allow affective responses to guide their judgments of permissible harm. Although there is a large literature on the role of emotion in moral judgment (e.g., Haidt, 2001; Greene, 2007; Greene & Haidt, 2002; Pizzaro, 2000; Prinz, 2006), a more complete account of this relationship is beyond the scope of this discussion.
Implications and Future Directions

The findings suggest that while there is a great deal of similarity concerning the application of these principles of harm, there do appear to be age differences regarding the ability to provide justifications for these judgments. However, it also seems that these age differences in justifications are in part driven by differences in verbal ability.

One apparent avenue for future research would be a more thorough exploration of the developmental nature of the adherence to these principles of harm when making moral judgments. Though the present study found only minor age differences regarding the application of these principles, there is certainly still the possibility that this is a developmental phenomenon and that the age range of the present sample was simply unable to capture the relevant period of change. Future research may attempt to investigate these principles of harm in younger sample, or perhaps using a longitudinal design to better capture change.

Conclusion

The findings of this study indicate that adolescents in the 9th and 12th grades solve moral dilemmas in a very comparable fashion; specifically, it appears that their judgments are guided by the same set of principles. It also appears that both groups had difficulty providing sufficient justifications for their judgments. This suggests that, although individuals may apply this set of principles when making moral judgments, they appear to have difficulty providing an account of their reasoning. This point is not easily accounted for by Kohlberg’s (1969) Stage Theory of moral development, since Kohlberg requires that moral judgments be the product of explicit reasoning.
The age differences with respect to justifications are interesting, and initially seem to suggest that this might be a developmental phenomenon. However, the fact that the effect is somewhat diminished after controlling for verbal ability might indicate that the division between expressed moral knowledge (i.e., justifications) and operative moral knowledge (i.e., judgments) might simply be a product of differences in cognitive abilities—in this case: language. This is an interesting, and novel finding.

Finally, the pattern of findings on the Mark scenario is not easily explained using Hauser’s (2006) model. Indeed, the model makes no room for emotions in the production of moral judgments. This will require further investigation, since it does seem clear that a subset of the participants in the present study used emotion-laden justifications of their moral judgments.

In sum, the findings of this study provide at least partial support for Hauser (2006) and Mikhail’s (2000) conception of the moral faculty. However, it is apparent that there is still more work to be done in order to address the questions that remain unanswered and fully substantiate the model.


APPENDIX A

TEXT AND IMAGES OF ALL SCENARIOS

Scenario 1.

Ned is walking near train tracks when he sees a train that is out of control. Farther down the track there are five people who will not hear the train and will not be able to get off the track in time. Ned is standing next to a switch, which he can flip to turn the train on to a side track. If the train hits the object on the side track, the object will slow the train down, giving the men time to escape. The heavy object is 1 man, standing on the side track. Ned can flip the switch, preventing the train from killing the 5 people, but killing the 1 man. Or he can not flip the switch, letting the 5 die. How right or wrong would it be for Ned to flip the switch?5

5 Several of these scenarios and the images that accompany them were adapted from Hauser and colleagues (2007) and from Cushman and colleagues (2006).
Scenario 2.

Oscar is walking near the train tracks when he notices a train approaching out of control. Up ahead on the track are 5 people. Oscar is standing next to a switch, which he can flip to turn the train on to a side track. If the train hits the object on the side track, the object will slow the train down, giving the 5 people time to escape. There is 1 man standing on the side track in front of the heavy object. Oscar can flip the switch, preventing the train from killing the 5 people, but killing the 1 man. Or he can not flip the switch, letting the 5 die. How right or wrong is it for Oscar to flip the switch?
Scenario 3.

Frank is on a bridge that passes over train tracks. He sees an out-of-control train coming down the track. Farther down the track there are five people who will not hear the train and will not be able to get off the track in time. The only way to stop an out-of-control train is to drop a very heavy weight into its path. But the only available, heavy weight is a very large man wearing a backpack who is also on the bridge. Frank can push the man with the backpack onto the track in the path of the train, killing him; or he can not push the man, letting the five die. How right or wrong would it be for Frank to push the man onto the track?
Scenario 4.

Joe is in his motorboat at the mouth of a river when he notices a shark approaching. Farther down the river he sees five swimmers. If Joe stays where he is, his boat will block the river and prevent the shark from attacking the five swimmers. Joe also sees another swimmer drowning in the distance. Joe can save the one swimmer, but that will leave the river open to the shark. If Joe moves towards the one swimmer in his motorboat, the swimmer will live but the five swimmers will be eaten by the shark. If Joe stays where he is, the one swimmer will drown but the five swimmers will be safe. How right or wrong would it be for Joe to stay where he is?
Scenario 5.

Casey is in his motorboat near the mouth of a river when he notices a shark approaching. Farther down the river he sees five swimmers. One other swimmer is at the mouth of the river where the shark will attack him first. This will give the five swimmers time to escape. Casey can save the one swimmer by moving toward him in the boat, but saving the one swimmer would leave the river open to the shark. If Casey moves towards the one swimmer, the swimmer will live but the five swimmers will be eaten. If Casey stays where he is, the shark will eat the one swimmer but the five swimmers will be safe. How right or wrong would it be for Casey to stay where he is?
Scenario 6.

Dan is a doctor in a hospital. Dan has five patients who need organ transplants. One patient needs a heart, one needs liver, one needs a lung, and two need kidneys. Without transplants, they will die. One day, Dan is running tests on Tony, another one of his patients. Tony is in a coma. Dan notices that Tony has the same blood type as the patients who need transplants. As Dan is leaving Tony’s room, he notices that someone has accidentally unplugged the respirator that is keeping Tony alive. Without the respirator, Tony will die. Dan knows that Tony is an organ donor, and that if he died his organs could be used to save Dan’s other patients. If Dan leaves the respirator unplugged, Tony will die, but the five other patients will be saved. If Dan plugs the respirator back in, then Tony will live, but the five other patients will die. How right or wrong would it be for Dan to leave the respirator unplugged?
Scenario 7.

Chris is a doctor in a hospital. Chris has five patients who need organ transplants. One patient needs a kidney, one needs a heart, one needs a liver, and two need lungs. Without transplants, Chris’s patients will die. One day, Chris is running tests on Pat, another one of his patients. Pat is in a coma. Chris sees that Pat has the same blood type as the patients who need transplants. As Chris is leaving Pat’s room, he notices that the electrical cord for the respirator that is keeping Pat alive is barely plugged in. Kicking the cord could easily unplug it. Without the respirator, Pat will die. Chris knows that Pat is an organ donor, and that if he died his organs could be used to save Chris’s other patients. If Chris kicks the cord and unplugs the respirator, Pat will die, but the five other patients will be saved. If Chris does not kick the cord and unplug the respirator, then Pat will live, but the five other patients will die. How right or wrong would it be for Chris to kick the cord and unplug the respirator?
Zach is a doctor in a hospital. Zach has five patients who need organ transplants. One patient needs a liver, one needs heart, one needs a kidney, and two need lungs. Without transplants, Zach’s patients will die. One day, Zach is running tests on Tom, another one of the patients. Tom is in a coma. Zach sees that Tom has the same blood type as the patients who need transplants. As Zach is leaving Tom’s room, he notices that Tom is barely breathing and is near death. Covering Tom’s face would easily stop his breathing. Zach knows that Tom is an organ donor, and that if he died his organs could be used to save Zach’s other patients. If Zach covers Tom’s face, Tom will die, but the five other patients will be saved. If Zach does not cover Tom’s face, then Tom will live, but the five other patients will die. How right or wrong would it be for Zach to cover Tom’s face?
Scenario 9.

Bob is standing near a bridge that passes over train tracks. He sees an out-of-control train coming down the track. Farther down the track there are five people who will not hear the train and will not be able to get off the track in time. The only way to stop an out-of-control train is to drop a very heavy weight into its path. But the only available, heavy weight is a very large man wearing a backpack who is on the bridge. Bob can flip a switch that will open a trap door and drop the man with the backpack onto the track in the path of the train, killing him; or he can not flip the switch, letting the five die. How right or wrong would it be for Bob to flip the switch?
Scenario 10.

Mark is on a bridge that passes over train tracks. He sees an out-of-control train coming down the track. Farther down the track there are five people who will not hear the train and will not be able to get off the track in time. The only way to stop an out-of-control train is to drop a very heavy weight into its path. But the only available, heavy weight is a very large man wearing a backpack who is also on the bridge. Mark notices the large man stumble over the railing on the bridge. The large man is barely hanging on. He will fall into the path of the train if Mark does not help him. If he falls, his body will slow the train down letting the five people on the track escape. Mark can move toward the man and help him back up, saving him but letting the five people die, or he can stay where he is and allow the man to fall into the path of the train. How right or wrong would it be for Mark stay where he is, and allow the large man to fall?
**Scenario 11.**

Ben sees that an empty runaway train is headed down a track. There is nobody on the track. There is one person working on a side track. If Ben flips the switch, the train will turn onto the side track and hit the one person. If Ben does not flip the switch, the train will continue on the track, and nobody will be hit. How right or wrong would it be for Ben to flip the switch?
APPENDIX B

TABLE 6.
SUMMARY OF PARTICIPANTS’ ISTEP ENGLISH/LANGUAGE ARTS SCORES

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Range</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 9</td>
<td>38</td>
<td>580.05</td>
<td>431</td>
<td>685</td>
<td>254</td>
<td>63.19</td>
</tr>
<tr>
<td>Grade 12</td>
<td>39</td>
<td>605.79</td>
<td>552</td>
<td>734</td>
<td>182</td>
<td>33.70</td>
</tr>
</tbody>
</table>