European Journal of Developmental Psychology

ISSN: 1740-5629 (Print) 1740-5610 (Online) Journal homepage: http://www.tandfonline.com/loi/pedp20

Moral judgement in adolescents: Age differences in applying and justifying three principles of harm

Paul C. Stey , Daniel Lapsley & Mary O. McKeever

To cite this article: Paul C. Stey , Daniel Lapsley & Mary O. McKeever (2013) Moral judgement in adolescents: Age differences in applying and justifying three principles of harm, European Journal of Developmental Psychology, 10:2, 206-220, DOI: 10.1080/17405629.2013.765798

To link to this article: http://dx.doi.org/10.1080/17405629.2013.765798

Published online: 07 Mar 2013.

EUROPEAN JOURNAL OF DEVELOPMENTAL PSYCHOLOGY, 2013
Vol. 10, No. 2, 206–220, http://dx.doi.org/10.1080/17405629.2013.765798

Routledge
Taylor & Francis Group

Moral judgement in adolescents: Age differences in applying and justifying three principles of harm

Paul C. Stey, Daniel Lapsley, and Mary O. McKeever

Department of Psychology, University of Notre Dame, Notre Dame, IN, USA

This study investigated the application and justification of three principles of harm in a cross-sectional sample of adolescents in order to test recent theories concerning the source of intuitive moral judgements. Participants were 46 early ($M_{age} = 14.8$ years) and 40 late adolescents ($M_{age} = 17.8$ years). Participants rated the permissibility of various ethical dilemmas, and provided justifications for their judgements. Results indicated participants aligned their judgements with the three principles of harm, but had difficulty explaining their reasoning. Furthermore, although age groups were consistent in the application of the principles of harm, age differences emerged in their justifications. These differences were partly explained by differences in language ability. Additionally, participants who used emotional language in their justifications demonstrated a characteristically deontological pattern of moral judgement on certain dilemmas. We conclude adolescents in this age range apply the principles of harm but that the ability to explain their judgements is still developing.

Keywords: Moral development; Moral cognition; Trolley problems; Adolescent; Judgement.

One of the central questions in moral psychology concerns how individuals arrive at moral judgements. Kohlberg's (1969) cognitive developmental theory famously argued that moral judgements are the result of conscious reasoning about justice that undergoes ontogenetic development. Recently there is growing interest in alternative accounts emphasizing non-conscious, intuitive cognitive processes as the source of moral judgements (e.g., Cushman, Young, & Hauser, 2006; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Haidt, 2001).

For instance, there has been a great deal of interest in the role emotions play in producing moral judgements (e.g., Blair, 1995; Greene et al., 2001, Haidt, 2001). Blair (1995), in particular, argued that deficiencies in certain moral emotions may

Correspondence should be addressed to Paul C. Stey, Department of Psychology, University of Notre Dame, 118 Haggar Hall, Notre Dame, IN 46556, USA. E-mail: pstey@nd.edu

explain the association between psychopathy and a failure to distinguish moral rule violations from conventional violations. Additionally, Greene and colleagues (2001) have proposed a dual-process model whereby deontological judgements are the product of non-conscious emotional processes, and utilitarian judgements are the product of conscious reasoning. Similarly, Haidt's social intuitionist model (2001) asserts that moral judgements are generated by intuitions that make a "sudden appearance" into consciousness, and that moral judgements automatically generated by intuitions lead to post hoc moral reasoning. Still others point to implicit deontological principles as the basis for decisions of right and wrong (Mikhail, 2000). According to this account, moral judgements result from a native mechanism and its sensitivity to features of a situation (e.g., intentionality, causality) and non-conscious application of deontological rules.

A recent study by Cushman and colleagues (2006) investigated the possibility that moral judgements are the product of certain implicit principles that guide judgements of permissible harm. This line of research draws on the work of Mikhail (2000), who suggested that humans possess a specialized moral faculty analogous to Chomsky's language faculty. Three principles were investigated: the contact principle, the action principle, and the principle of double effect ("intention principle" in their paper). According to the contact principle, harm caused through physical contact is morally worse than equivalent harm caused without physical contact. The action principle views harm that is caused through action as morally worse than equivalent harm resulting from the omission of action. The principle of double effect asserts that 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.

Cushman and colleagues (2006) presented adults with ethical dilemmas where the methods of causing harm (i.e., contact, action, intention) varied systematically in order to test the three principles of harm. Participants were asked to rate the permissibility of the actions in each dilemma. Ratings between dilemmas that make up a pair were compared to test the extent to which participants made ratings that aligned with the three principles of harm. For example, one vignette in the pair involves the protagonist causing harm through action while the other vignette in the pair describes a protagonist causing harm through the omission of action. If participants align their judgements with the action principle, they would rate the vignette in which the harm results from omission as more permissible than the vignette in which the harm results from action. This is what Cushman and colleagues found: participants rated harm caused by contact as worse than harm caused without contact; they rated harm caused by commission as worse than harm caused by omission; and they rated intended harm as worse than harm that was a foreseen side effect.

Cushman and colleagues (2006) also tested the extent to which participants could explain their divergent ratings across vignettes that made up a pair. For

example, participants were asked why they rated the action of the protagonist as more acceptable in one vignette (e.g., harm-by-omission) than another (e.g., harm-by-commission). Results showed that although participants made judgements aligning with the principles of harm, they often had difficulty providing a satisfactory account of their reasoning. This was especially true for the principle of double effect, where only 30% of justifications were successful (compared to 60% and 81% for the contact and action principles, respectively). The authors concluded that moral judgements are sometimes driven by non-conscious processes, or else align with ethical principles of which participants are unaware or struggle to articulate. Moreover, some moral principles are more accessible to conscious reasoning than others (e.g., action principle).

This line of research has implications for understanding the implicit features of moral judgement, but with few exceptions (e.g., Pellizzoni, Siegal, & Surian, 2010), there has been little research from a developmental perspective. The extant research has been limited to samples of adults and hence little is known about age-related variability in the application and justification of these principles. One reason there has been little interest in developmental variability and mechanisms driving these moral judgements is that much of the research conducted to date has adopted a nativist perspective (e.g., Mikhail, 2000). However, we might expect to find considerable variability in the implicit features of moral judgement (e.g., accessibility of moral heuristics). Adolescence is, after all, a period of expanding intellectual abilities including advances in formal operations, problem solving and decision making, and also exposure to more challenging academic content. If the source of moral judgements is a developmental achievement or experience dependent, then one might expect age differences between younger and older adolescents in how they reason about principles of harm.

However, if few age differences obtain, and if adolescents perform similarly to the pattern of young adults, then doubts are raised concerning the degree of articulacy required for making moral judgements. Such an outcome would inform an important debate. Taylor (1989) argued that *strong evaluation* "is essential to being a functional moral agent", and this requires the capacity to make transparent, deliberative judgements concerning what is higher and lower, worthy and unworthy. In contrast, Flanagan (1990) asserted that strong evaluation overstates the degree to which "moral decency" requires articulate self-reflection. In his view one can acknowledge and make judgements regarding the standards "without ever having linguistically formulated the standard and without ever possessing the ability to do so when pressed" (Flanagan, 1990, p. 53).

In addition to examining intuitive moral judgements with younger adolescent samples, we also analyse the data in an innovative way to reveal the work of affective mechanisms in reaching moral judgements. For example, Greene and colleagues (2001) asserted that the contact principle is driven by an affective response to the thought of causing direct, physical harm. However, to what extent—if any—are other principles of harm influenced by affective responses?

We will examine the degree to which affective considerations influence intuitive moral judgements in an adolescent sample.

The present study

This study is the first to examine the application and justification of intuitive moral principles in early and late adolescents. Following the methodology of Cushman and colleagues (2006), participants were asked to rate and justify the permissibility of harm in ethical dilemmas. Regarding the application of principles of harm, we expected our adolescent sample to mirror adults' pattern of judgements (e.g., Cushman et al., 2006; Mikhail, 2000). We expected adolescents would: rate harm caused by physical contact as morally worse than harm caused without contact, rate harm-by-commission as worse than harm-by-omission, and rate intended harm as worse than harm that is simply a foreseen side effect. We expected age differences, however, in the justification of moral judgements—with older adolescents more readily and more accurately articulating their reasoning. We anticipated this based on prior research demonstrating the importance of tacit knowledge and the extent to which higher-order cognitions are inaccessible to reflection (e.g., Nisbett & Wilson, 1977). We believe that older participants, having more developed cognitive abilities, and greater experience articulating their reasoning, would outperform younger participants in this respect. Evidence from the Piagetian and Kohlbergian traditions also demonstrate the positive relationship between age and explicit moral knowledge. Finally, based on Greene et al. (2001), we expected that justifications for harm involving physical contact would more likely invoke affective language than would justifications for harm without contact.

METHOD

Participants

Participants were 86 adolescents recruited from a public school in the Midwestern United States from two grade levels: grade 9 ($n = 46$, 28 female, $M_{age} = 14.8$ years, $SD = 0.50$) and grade 12 ($n = 40$, 20 female, $M_{age} = 17.8$ years, $SD = 0.48$). Students completed the study on a computer and participated in a brief structured interview. They were compensated $5 for participating.

Materials

Ethical dilemmas. Participants were presented with ethical dilemmas testing the three principles of harm: the contact principle, the action principle, and the principle of double effect. Participants rated the permissibility of the actions described in the vignettes on a scale of 1 (*Completely wrong*) to 8 (*Completely*

TABLE 1
Attributes of test vignettes

Vignette	*Contact*	*Action*	*Intended harm*	*Flesch-Kincaid Index*	*Source*
Ned (*Loop* Case)	No	Yes	Yes	3.7	Mikhail (2000)
Oscar (*Loop Weight* Case)	No	Yes	No	5.0	Mikhail (2000)
Frank (*Footbridge* Case)	Yes	Yes	Yes	5.5	Mikhail (2000)
Joe	No	No	No	5.9	Cushman et al. (2006)
Casey	No	No	Yes	6.0	Cushman et al. (2006)
Dan	No	No	Yes	5.8	Thomson (1976)
Chris	No	Yes	Yes	5.2	Thomson (1976)
Zach	Yes	Yes	Yes	4.1	Thomson (1976)
Bob	No	Yes	Yes	5.9	Thomson (1976)
Mark	No	No	Yes	5.7	Thomson (1976)

right). Black-and-white illustrations accompanied the vignettes. All vignettes described an agent choosing a course of action resulting in harming either one or five individuals. The dilemmas were versions of the classic trolley problem (Foot, 1967), dilemmas adapted from prior research (e.g., Cushman et al., 2006; Mikhail, 2000), and dilemmas adapted from philosophical research (Thompson, 1976; see the appendix for sample vignettes). Dilemmas included only male protagonists and victims in order to limit gender-by-gender interactions of participant and protagonist or victim. Each principle of harm was tested using three pairs of vignettes. To test these principles of harm, all 10 of the test vignettes possessed a combination of harm caused by (a) contact/no contact, (b) commission/omission, and (c) intention/no intention. Table 1 illustrates the attributes of all test vignettes.

To minimize the possibility that reading ability affected judgements, vignettes had readability indices of 6.0 or lower according to the Flesch–Kincaid readability index. This readability level was three full grade levels below the younger students.

Standardized test scores. Since verbal ability may play a role in judgements or justifications, participants' standardized test scores were obtained from school records—specifically, their English/Language Arts composite scores on the Indiana Statewide Testing for Educational Progress–Plus (ISTEP +).

Procedure

Data collection was conducted at the participants' school. Participants were removed from class one at a time and directed to a quiet room in the school where they completed the study on a computer. The study was divided into two parts:

(1) rating the permissibility of actions in ethical dilemmas; and (2) providing justifications for the ratings during a structured interview.

Judgements of permissibility. Participants were presented with 10 test vignettes (and 1 control) in two blocks. Recall that a critical feature for investigating participants' adherence to the principles of harm is that each vignette is part of at least one pair. Vignettes that make up a pair are nearly identical, but differ according to one of the principles of harm. As shown in Figure 1, vignettes making up a pair were in opposing blocks to minimize the likelihood that participants would see two vignettes from a test pair one after another. This helped mitigate the impact of order effects of earlier judgements on later judgements. The order of blocks was counterbalanced, and vignettes within blocks were randomized.

The text and illustrations of the vignettes were presented on the computer, which also played an audio narration to avoid undue influence of differences in reading ability. Additionally, animation was used to direct participants' attention to the relevant features of the vignette. For instance, when the narration read, "Ned is walking near train tracks when he sees a train that is out of control", a red circle briefly appeared around the train car. When the narration read, "Ahead on the track are five people", the circle briefly highlighted these five.

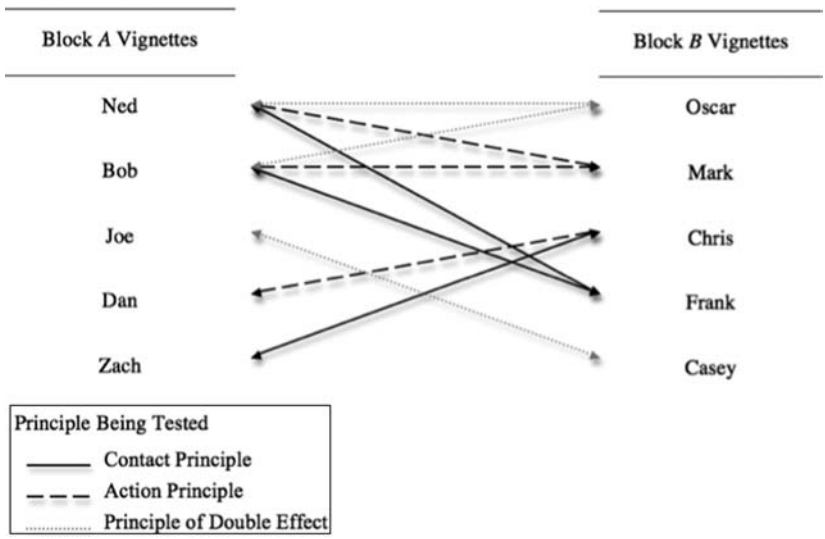

Figure 1. Summary of Block A and B vignettes, and how individual vignettes are paired in order to test the relevant principles of harm. Solid lines connecting two vignettes indicate that the pair of vignettes tests the contact principle. Dashed lines indicate that the pair tests the action principle. And dotted lines indicate that the pair tests the principle of double effect.

Justifications. Once participants completed the judgements they were asked to provide justifications for all vignette pairs that were given divergent judgements of permissibility. For example, a participant who gave different permissibility ratings on two members of a vignette pair (e.g., Ned and Frank) was asked to justify the divergent ratings. Thus, the number of justifications was dependent on the number of disparate judgements of permissibility on the vignette pairs. A maximum of five justifications was imposed to comply with time constraints at the school, but each participant was given no expressed time limit. Justifications were solicited via structured interview to control for differences due to writing ability. Interviews were conducted by a male experimenter.

For each justification, participants viewed a side-by-side display of the two vignettes (images and text) on which they provided divergent ratings, and were asked to explain their reason for the difference in ratings. A sequence of three probes was used in rare cases where participants had difficulty understanding what was required. The order of justification prompts was randomized. Interviews were audio recorded and transcribed.

Coding of justifications. Justifications were coded according to a modified version of Cushman and colleagues' (2006) coding system. Eight codes could be assigned to a justification: "mistest", sufficiency, failure, dumbfounding, denial, alternative explanation, emotion, and error. The first three are mutually exclusive; others may co-occur. A mistest was assigned when a participant: ranted or avoided the question, or misunderstood the comparison task or stimuli. Sufficient justifications identified a factual difference between the two vignettes, and claimed (or implied) that as the basis for the divergent judgements. The failure code was assigned to justifications suggesting an alternative principle that could not account for the pattern of judgements (e.g., "It's wrong to kill"). The dumbfounding code was used when the participant explicitly referenced his or her own uncertainty regarding the judgements, or directly stated their inability to provide a justification. The denial code was assigned when the participant denied there was a moral difference between the two dilemmas. The added assumption code was used when the participant referenced facts not present in the vignette. The emotion code was assigned when participants referenced their own emotional response to the vignette, or the emotional response they expected of the protagonist. Finally, the error code was assigned when participants reported an attitudinal shift or claimed to have made an error in responding (e.g., clicking incorrectly).

RESULTS

Judgements of permissibility

The contact principle suggests that harm caused through physical contact is morally worse than harm caused without contact. To test this, participants rated

several vignettes that pitted harm caused by contact (e.g., Frank vignette) against harm caused without contact (e.g., Ned vignette). Descriptive statistics appear in Table 2. If participants conformed to the contact principle, their ratings of permissibility on the Frank vignette—for example—would be lower than their ratings on the Ned vignette. As Table 3 indicates, in two of three comparisons, both 9th- and 12th-grade students reported that harm caused by contact was less permissible than harm caused without contact.

The action principle suggests that harm caused by action is morally worse than harm resulting from omission. To test this, participants rated vignettes that pitted harm caused by action (e.g., Chris vignette) against harm caused by omission (e.g., Dan vignette). If subjects conformed to the action principle, their permissibility ratings on the Chris vignette—for instance—would be lower than their ratings on the Dan vignette. As shown in Table 3, in one of three comparisons, both 9th- and 12th-grade students reported that harm caused by action was less morally permissible than harm caused by omission.

The principle of double effect suggests that harm that is intended as the means to an end is morally worse than harm that is only a foreseen side effect. To test this principle, participants rated vignettes that pitted harm caused intentionally (e.g., Ned vignette) against harm that was a foreseen side effect (e.g., Oscar vignette). If subjects conformed their judgements to the principle of double effect, their permissibility ratings would be lower on the Ned vignette—for example—compared to the Oscar vignette. As Table 3 illustrates, this was observed in two vignette pairs testing the principle of double effect. This indicates that participants rated dilemmas in which the harm was intended as the means to an end as morally worse than dilemmas in which harm was a foreseen side effect. Again, just as with the contact principle and the action principle, the pattern of judgements was essentially identical for the 9th- and 12th-grade groups.

TABLE 2
Summary of descriptive statistics

	Grade 9 group Mean (SD)	*Grade 12 group Mean (SD)*
Ned	3.87 (1.96)	4.45 (1.91)
Bob	3.76 (2.10)	3.63 (1.97)
Joe	5.02 (1.94)	5.40 (1.95)
Dan	4.07 (1.91)	2.95 (1.85)
Zach	3.13 (1.64)	2.73 (2.11)
Frank	3.28 (1.64)	2.95 (2.14)
Oscar	4.72 (1.81)	4.95 (1.97)
Mark	3.61 (1.75)	4.02 (2.19)
Chris	3.57 (1.96)	2.50 (1.92)
Casey	5.17 (1.54)	5.30 (2.03)

TABLE 3
Summary of paired-sample *t*-tests on vignette pairs within grade

		Grade 9 group				*Grade 12 group*			
		M_{Diff} *(SD)*	*t(45)*	*p (two-tailed)*	*Effect size (d)*	M_{Diff} *(SD)*	*t(39)*	*p (two-tailed)*	*Effect size (d)*
Contact principle									
Contact	*No contact*								
Frank	Ned	−0.59 (1.53)	−2.60	.01	0.39	−1.50 (1.73)	−5.50	<.001	0.87
Frank	Bob	−0.49 (1.66)	−1.96	.06	0.30	−0.68 (2.06)	−2.08	.04	0.33
Zach	Chris	−0.44 (1.96)	−1.50	.14	0.23	0.23 (1.78)	0.80	.43	0.13
Action principle									
Commission	*Omission*								
Ned	Mark	0.26 (1.48)	1.19	.24	0.18	0.43 (1.81)	1.49	.15	0.23
Bob	Mark	0.15 (1.75)	0.59	.56	0.09	−0.40 (2.31)	−1.10	.28	0.17
Chris	Dan	−0.50 (1.41)	−2.41	.02	0.36	−0.45 (1.21)	−2.33	.03	0.37
Principle of double effect									
Intention	*No intention*								
Ned	Oscar	−0.85 (2.04)	−2.81	.01	0.42	−0.50 (1.50)	−2.11	.04	0.33
Bob	Oscar	−0.96 (2.09)	−3.11	<.01	0.46	−1.33 (1.62)	−5.16	<.001	0.81
Casey	Joe	0.15 (1.25)	0.83	.41	0.13	−0.10 (1.60)	−0.40	.70	0.06

Age differences. As indicated in Table 3, 9th and 12th graders demonstrated a nearly identical pattern of ratings. However, to further investigate the possibility of group differences, repeated measures analyses of variance (ANOVAs) were conducted on each vignette pair. For each analysis, the within-subject factor was Vignette Pair, and Grade was the between-subject factor. Of the nine total vignette pairs, only one contact principle pair (Frank and Ned), demonstrated a significant Grade by Dilemma interaction, $F(1, 83) = 6.51$, $p = .01$.

Justifications

In the second portion of the study participants provided justifications for rating the actions of the protagonist in one vignette as morally worse than the actions of the protagonist in another vignette. For example, in a contact principle pair, if participants rated the actions of a protagonist that causes harm by contact (e.g., Frank) as morally worse than those of a protagonist causing equivalent harm without contact (e.g., Ned) or vice versa, then they were asked to explain their reason for different ratings.

The first author coded all justifications ($N = 367$). A second coder, blind to the hypotheses of the study, also coded 25% of the justifications ($n = 92$). Inter-rater agreement on this subset revealed a Cohen's kappa of .75, which is considered substantial (Landis & Koch, 1977).

Overall pattern of justifications. We were interested in whether participants could provide sufficient justifications for their judgements. Both 9th and 12th graders had difficulty providing sufficient justifications for their judgements. Additionally, the pattern of sufficient justifications by principle replicates the pattern reported by Cushman and colleagues (2006) in that participants were most successful providing sufficient justifications for their judgements of action principle pairs, and least successful with principle of double effect pairs. Ninth graders provided sufficient justifications for 54.6% of action principle pairs, 44.7% of contact principle pairs, and 21.5% of principle of double effect pairs. Twelfth graders provided sufficient justifications for 68.1% of action principle pairs, 53.3% of contact principle pairs, and 31.3% of principle of double effect pairs.

Overall grade differences. As noted above, there were grade-level differences with respect to sufficient justifications. Twelfth-grade participants provided significantly more sufficient justifications for their judgements than did 9th graders, $F(1, 85) = 7.53$, $p < .01$. Grade differences for other justification codes were also analysed. Two differences emerged: 9th graders were significantly more likely to deny there was a moral difference between vignettes, $F(1, 85) = 5.20$, $p = .03$, and to report an error in responding,

$F(1, 85) = 4.22$, $p = .04$—generally indicating that subjects were no longer certain which was the morally worse vignette.

Grade differences by principle. We next analysed grade differences in sufficient justifications on each principle of harm. The proportion of sufficient justifications by principle was computed for each participant. For the contact principle, 12th graders provided a higher average proportion of sufficient justifications than did 9th graders, $F(1, 73) = 5.17$, $p = .03$. There was a similar trend for the action principle, with 12th graders again providing a higher proportion of sufficient justifications than 9th graders, $F(1, 70) = 3.15$, $p = .08$. There were no age differences on the principle of double effect. Thus, on two of the three principles, older adolescents offered a greater proportion of sufficient justifications.

In the next analysis ISTEP + scores were added as a covariate to determine whether the ability to provide sufficient justifications was related to language and reading skills. When ISTEP + scores were added as a covariate, the age differences were no longer apparent for the contact principle, $F(1, 63) = 2.65$, *ns*, and the action principle, $F(1, 60) = 2.45$, *ns*. Additionally, ISTEP + scores were negatively correlated with failed justifications, $r(77) = -.38$, $p = .001$. These results indicate that greater language ability is positively related to sufficient justifications, and negatively related to failed justifications.

Emotions in justifications. Finally, we attempted something that has not yet been done: we examined instances of emotion words in participants' justifications. Interestingly, over 70% of participants had no emotion codes assigned to their justifications. Furthermore, by examining individual justifications as the unit of analysis, it became apparent that there were differences by principle with regard to the frequency of emotional language, $\chi^2(2, N = 367) = 6.67$, $p = .036$. Emotional language occurred most frequently in contact principle pairs and action principle pairs (no gender differences emerged). Looking at the frequency of emotion codes by vignette pair, it appeared that pairs that included the Mark vignette or the Frank vignette had a greater frequency of emotional language. The Frank vignette was the classic footbridge case (see appendix). Based on the work of Greene and colleagues (2001), it might be expected to have a higher frequency of emotional language. The Mark vignette is a modified footbridge case designed to test the action principle; the protagonist saves the five on the tracks by *allowing* the man on the footbridge to fall onto the tracks (see appendix). It was expected that participants would rate the actions of Mark as more permissible than the actions of protagonists in the action principle pairs. However, on average, this was not the case.

Results of paired *t*-tests revealed no statistically significant differences between the action principle pairs involving the Mark vignette. However, an interesting pattern emerged when examining justifications with emotion language. Of the participants who *conformed* to the action principle by rating

the actions of Mark to be more permissible than its action principle pair counterparts (i.e., Ned or Bob), 6% of their justifications referenced emotions. However, of the participants who did *not* conform to the action principle by rating the actions of Mark as morally worse than those of either Ned or Bob, 32% of the justifications involved emotional content.

This suggests the interesting possibility that individuals with emotion codes assigned to their justifications made moral judgements in a different manner than those without. As it turns out, the former group—call them "emoters"—were *less* likely to conform to the action principle on pairs involving the modified footbridge case (i.e., Mark vignette) including: the Bob and Mark pair, $\chi^2(1, N = 55) = 9.55$, $p = .002$, as well as a marginal effect on the Ned and Mark vignette pair, $\chi^2(1, N = 58) = 2.93$, $p = .087$. Interestingly, the use of emotion words in justifications also had a significant effect on conformity to the contact principle in the vignette pairs involving the classic footbridge case (i.e., Frank vignette). Emoters were more likely to conform to the contact principle on vignette pairs that contained the classic footbridge case: including the Frank and Ned vignette pair, $\chi^2(1, N = 62) = 4.70$, $p = .030$, as well as marginal effect on the Frank and Bob pair, $\chi^2(1, N = 57) = 3.28$, $p = .070$. Thus, on vignette pairs involving either the modified footbridge or the classic footbridge case, emotion codes in justifications were related to deontological judgements.

One possible explanation for this pattern of findings is that visual imagery has been shown to be preferentially related to characteristically deontological judgement (Amit & Greene, 2012). In vignette pairs involving the modified footbridge case (i.e., Mark vignette), this would mean *not conforming* to the action principle. Thus, perhaps the illustrations that accompanied the vignettes affected participants' judgements on the Mark vignette.[1]

DISCUSSION

The purpose of this study was to determine whether principles of harm are operative and available for reflection in adolescents as they are in adults. To this end we examined possible developmental differences in the application and justification of principles of harm. Additionally, we tested whether proficiency in justifications was mediated by language skills. Finally, we examined an innovative question concerning the role of affective mechanisms in accounting for certain moral judgements.

Regarding moral judgements, our data show no age differences between younger and older adolescents in the application of three principles of harm. Both

[1]This hypothesis was tested in a follow-up study on an adult sample using Amazon.com's Mechanical Turk. Results indicated participants ($N = 72$, $M_{age} = 27.9$ years, 34 females) gave significantly lower permissibility ratings to the modified footbridge vignette (i.e., Mark vignette) when it included an illustration compared to when there was none, $t(70) = 2.20$, $p = .031$, $d = 0.52$.

groups judged that harm caused by physical contact was morally worse than harm caused without contact, that harm caused by commission was worse than harm caused by omission, and that harm that was intended as the means to an end was morally worse than harm that was only a foreseen side effect. The ability of adolescents to reach these judgements on such hypothetical dilemmas does not appear to hinge on the acquisition of other cognitive development abilities that emerge during this period, and seems to provide evidence of greater moral competence among adolescents than would be expected, for example, from the Kohlberg tradition. Age differences may only be apparent on moral dilemma tasks that require participants to explicitly reason about action choices.

Support for this view is found in the second portion of our study where age differences were observed. Here, 12th-grade students provided significantly more sufficient justifications than 9th-grade students when asked to justify their judgements on both contact and action principle pairs. Interestingly, there were no differences in the ability to justify judgements on the more complex principle of double effect pairs—much like adults (Cushman et al., 2006), both groups struggled here.

One source of these age differences in justifications appears to be language skills. When ISTEP + scores were included as a covariate, age differences in the ability to provide sufficient justifications faded. ISTEP + scores were also negatively related to failed justifications, which indicates that individuals with greater language ability were less likely to provide justifications suggesting a principle that did not explain their pattern of judgements. This shows that the ability to articulate and justify moral judgements can be influenced by the ordinary general education curriculum and may not require an explicit moral education intervention.

One innovative feature of this study is our attempt to tie together two existing literatures: one showing that moral judgements are influenced by affective considerations (Greene et al., 2001), and another showing that judgements of permissible harm are the product of implicit principles (Cushman et al., 2006). To this end we examined the extent to which adolescents used emotion words in their justifications, and whether this was related to their application of three principles of harm. Our findings showed that participants were more likely to use emotion words in justifications of contact principle and action principle pairs, and that the use of emotion words in justifications was related to more deontological judgements in these pairs. This supports similar findings reported by Greene and colleagues (2001), who also showed that deontological judgements are related to affective responses in adults.

When the judgement and justification data are jointly considered, it is clear that many adolescents are making moral judgements that align with certain principles without the ability to articulate their reasoning. Cushman and colleagues (2006) also reported this pattern of findings in studies with adults. Hence, the data support Flanagan's (1990) claim that the ability to reach

competent moral judgements does not require a high degree of articulacy that proponents of strong evaluation suggest (e.g., Taylor, 1989).

Although we charted age-related variation, future research ought to examine age change in longitudinal designs. A second concern is the artificiality of the classic trolley problems. As research in the Kohlberg tradition has shown, moral judgements are sensitive to dilemma type. It may be the case that the application and justification of principles of harm would vary in contextually rich and more realistic vignettes. However, one advantage of the classic trolley problem is that it makes evident in very subtle ways the moral principles under consideration. Use of these dilemmas in the present study contributes to a growing literature where these dilemmas are an important methodological tool.

In sum, the present study showed there are no age differences in the ability of adolescents to make intuitive moral judgements. Older adolescents were better able to provide sufficient justification than were younger adolescents, but this advantage was partially explained by stronger language ability. Additionally, this study is the first to demonstrate the role of affective mechanisms in the justification of principles of harm.

REFERENCES

Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science*, *23*, 861–868.

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

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

Flanagan, O. (1990). Identity and strong and weak evaluation. In O. Flanagan & A. O. Rorty (Eds.), *Identity, character and morality: Essays in moral psychology* (pp. 37–66). Cambridge, MA: MIT Press.

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

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

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

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago, IL: Rand McNally.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Mikhail, J. M. (2000). *Rawls' linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in "A theory of justice"* (Unpublished doctoral dissertation). Cornell University, Ithaca, NY.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

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

Taylor, C. (1989). *Sources of the self: The making of modern identity*. Cambridge, MA: Harvard University Press.

Thomson, J. J. (1976). Killing, letting die and the trolley problem. *Monist*, *59*, 204–217.

APPENDIX

Frank Vignette (adapted from Mikhail, 2000)

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?

Joe Vignette (adapted from Cushman et al., 2006)

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?

Mark Vignette (adapted from Thomson, 1976)

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?