

Gender and Developmental Differences in Adolescents' Conceptions of Moral Reasoning

Kathleen M. Galotti,¹ Steven F. Kozberg,² and Maria C. Farmer³

Received December 29, 1989; accepted August 3, 1990

Eighth graders, 11th graders, and college sophomores wrote open-ended essays that characterized their own approaches to moral reasoning. Their essays were scored for the presence of various themes. Students also responded to three dilemmas from the Defining Issues Test (Rest, 1979). Students then rated their own use of justice and care orientations in moral reasoning, using a previously developed instrument (Ford and Lowery, 1986; Lyons, 1983). Characterizations of moral reasoning became more multifaceted and thorough with grade, and correlated with many traditional measures of moral reasoning. Males and females showed some, but relatively few, differences in characterizing their own moral reasoning, although females tended to respond more thoroughly to the open-ended task. Gender differences, when they occurred, were mostly found on measures that assessed so-called feminine issues or concerns, but not on traditional measures of moral reasoning. Characterizations of moral reasoning clustered into five dimensions, raising questions about the utility of the constructs of justice and care orientations.

The first and third authors received financial support from a Ford-Mellon grant made to Carleton College. Results of this study were presented at the 1990 meeting of the Society for Research in Adolescence.

¹Associate Professor of Psychology at Carleton College. Received Ph.D. in Psychology and M.S.E. in Computer and Information Sciences at the University of Pennsylvania, and has research interests in the development of reasoning skills, decision making, and conceptual development. To whom reprint requests should be addressed at Department of Psychology, Carleton College, Northfield, MN 55057-4025.

²Counseling Psychologist and Lecturer in Psychology at Carleton College. Received Ph.D. in Counseling and Guidance from the University of Wisconsin, Madison, and has research interests in adolescent development and adolescent psychopathology.

³Received B.A. in Psychology from Carleton College in 1990. Has interests in adolescence and chemical dependence.

INTRODUCTION

Current controversies in the area of moral development center around gender differences, both in performance on moral reasoning tasks and in overall orientation toward moral reasoning. Questions about the appropriate measures of moral reasoning, moral development, and moral "style," are energetically debated. At the same time, the development of moral reasoning abilities is being seen in the context of broader intellectual and epistemological developmental phenomena. As a result of these discussions, the study of moral reasoning has broadened considerably in the last decade.

Our work investigates adolescents' characterizations of their own moral reasoning. In particular, we examine gender and developmental differences in self-reported, open-ended descriptions of how moral concerns are identified. Relationships between these descriptions and other assessments of moral reasoning and moral orientation are also explored. The purpose of this research is to see how well adolescents' conceptions of their own reasoning correlate with actual episodes, and to assess how these conceptions differ as a function of gender and development.

Traditional research on moral reasoning characterized moral development as a hierarchical progression from egocentrism, focusing on rewards and punishments, to a more autonomous, reflective, and principled stance in which competing claims are weighed and integrated (see Rest, 1983, for a review). This developmental scheme is most often associated with Kohlberg (1976; Colby and Kohlberg, 1987; see Colby *et al.*, 1983, and Kohlberg, 1984, for an overview of recent developments in the theory). Adolescents are described as becoming increasingly sophisticated about and concerned with issues of justice, fairness, and rights with age and with developmental changes in cognitive organization (see Rest, 1983, for a more extensive discussion).

Gilligan (1977, 1982, 1988a) challenged this framework as a description of (1) all moral development, and (2) the assumptions and perspectives of all adolescents. In particular, Gilligan argued for gender-related differences that "signify differences in moral orientation" (1988b, p. 8). Gilligan seemed to agree with Kohlberg on some aspects of his theory, e.g., the stage-like nature of moral development (1982; Gilligan and Belenky, 1980). At the same time, she asserted that women impose a "distinctive construction on moral problems, seeing moral dilemmas in terms of conflicting responsibilities" (1982, p. 105), rather than in terms of abstract principles, rules, or competing rights, as she characterized Kohlberg's description. She believed women's moral reasoning is "contextual and narrative" (1982, p.

19), and is oriented more toward issues of caring, responsibility for others, and the connection of self to others.

One of Gilligan's (1977, 1982) initial claims, that traditional Kohlbergian measures of moral development were biased against women, has been extensively disputed. Walker (1984, 1986b, 1989; Walker *et al.*, 1987), conducting both meta-analyses of previous studies, and original investigations, provided significant evidence to refute Gilligan's claim (but see Baumrind, 1986, and Walker, 1986a, for a debate over the compellingness of the argument). Thoma (1986), conducting meta-analyses and secondary analyses of reasoning using Rest's (1979) Defining Issues Test (DIT) measure (described below), did in fact discover a gender effect, with females at all ages and educational levels scoring *higher* than males, although the differences were small.

Another of Gilligan's claims, that women's and men's moral reasoning center around different concerns and issues, has received more support. Researchers studying aspects of moral reasoning and identity development have reported findings that provide direct and indirect evidence for gender differences in the types of issues and concerns typically focused upon (Bussey and Maughan, 1982; Galotti *et al.*, 1991; Gilligan *et al.*, 1988; Hanson and Mullis, 1985; Lyons, 1983; Scheidel and Marcia, 1985; Yussen, 1977).

Belenky *et al.* (1986) advanced a claim even broader in scope than those of Gilligan (1977, 1982). They argued that women's and men's reasoning about a variety of topics differed in fundamental ways. Women were held to seek connectedness and understanding, and men, rigor and proof. In sum, the emphases, goals, and standards of evaluation in reasoning were argued to differ for men and women, although the difference was not claimed to be an absolute one.

Clearly, the development of moral standards and values presupposes the ability to construct, test, and refine one's assumptions and methods of evaluation of claims and arguments. Researchers studying epistemological development (Belenky *et al.*, 1986; Clinchy and Zimmerman, 1982; Perry, 1970) generally characterize it as involving increasing complexity, internality, and personal responsibility for one's worldview, reasoning, and values. Gender differences have also been noted (Belenky *et al.*, 1986; Clinchy and Zimmerman, 1982). In this context, therefore, the finding of gender differences in moral reasoning is an expected one.

However, few empirical studies that directly examine adolescents' epistemological orientations have been carried out. Moreover, few studies of moral reasoning have looked at adolescents' own self-described goals and emphases in their reasoning. Lyons (1983) developed a coding scheme

to articulate the two orientations described by Gilligan (1977, 1982). One aspect of this scheme involves an overall orientation to moral reasoning. These two orientations are called the "justice" orientation and the "care" orientation (Gilligan, 1988a; Lyons, 1983; Ford and Lowery, 1986). Lyons found that women tended to articulate a care orientation and men, a justice orientation, although the gender difference was not absolute, nor were the two orientations mutually exclusive. Similar findings have been reported by Donenberg and Hoffman (1988), Pratt *et al.* (1988), and Rothbart *et al.* (1986). Gilligan and Attanucci (1988) later reported that 53 of their 80 subjects showed either a care focus (i.e., showed a predominant orientation to a care orientation) or a justice focus. The remaining 27 subjects used both orientations, with neither predominating. Orientation was significantly related to gender: Only 1 male (of 46) showed a care focus, and 30 showed a justice focus, while 12 of the 34 females showed a care focus, and 10 showed a justice focus.

Walker (1986b, 1989), using different scoring criteria, did not find evidence for gender differences in moral orientation among elementary through high school students, although he did among their parents. Walker *et al.* (1987) found evidence that people typically use a combination of both orientations. Ford and Lowery (1986), who developed an objective self-report measure of orientation, found only small gender differences in their sample of undergraduate students, and the differences disappeared when the importance and difficulty of reported dilemmas were held constant. Galotti (1989) also found few gender differences in the way male and female undergraduates characterized the issues they concerned themselves with in moral reasoning.

Walker (1989) examined a different typology of orientations to moral reasoning. He scored the responses of children and adults (age range: 5–63 years) for the presence of any of the following four orientations originally described by Kohlberg (1976; the descriptions are paraphrased from Walker, 1989, p. 158): *normative*, emphasizing duty and adherence to prescribed rules and roles; *fairness*, emphasizing liberty, equity, equality, reciprocity, and contract; *utilitarianism*, emphasizing the self's or other's welfare or happiness as a consequence of moral actions; and *perfectionism*, emphasizing the attainment of good conscience, harmony of self and others, dignity, and autonomy. The themes were treated as mutually exclusive in the scoring.

Walker (1989) found no gender differences in the use of these four orientations, despite his prediction (based on Gilligan's 1977 and 1982 writings) that males would use the first two orientations more than

females, with the reverse pattern occurring on the latter two. However, developmental differences in the use of the four orientations were strong. With age, the use of normative and utilitarianism orientations declined, and the use of fairness and perfectionism orientations increased.

In the present study, we examine a similar issue, namely, characterizing the responses to open-ended moral questions, in new ways. There are two major differences between our work and previous work. First, we ask our respondents to *describe* their own moral reasoning, in addition to responding to supplied hypothetical dilemmas. Previous work has not yet addressed people's conceptions of their own reasoning, and how this differs as a function of age or gender. Second, we use a different typology to describe these characterizations. Specifically, the themes we score are not mutually exclusive, so the scoring allows for none, some, or all themes to be counted for a response. The taxonomy, adapted from Galotti (1989), includes 11 themes, more specific in character than either those presented by Lyons (1983) or Kohlberg (1976). Characterizations of moral reasoning are compared to other performance measures, such as the DIT P score (Rest, 1979), and the self-reported use of justice and care orientations (Ford and Lowery, 1986). We also explore the degree to which these characterizations can be classified along the dimensions of "justice" and "care," and to what degree other descriptions of orientation prove necessary.

METHOD

Subjects

The sample consisted of 186 students: 61 eighth graders (27 males, 34 females), 73 eleventh graders (34 males, 39 females), and 52 college sophomores (27 males, 25 females). Eighth graders were recruited from a southeastern Minnesota public middle school and a southeastern Minnesota public junior high school; 11th graders were recruited from the corresponding public senior high schools in the same communities. One or two of the authors visited study halls at these schools, and administered the surveys to groups—4 to 30 students. A stratified random sample of college students at two private liberal arts colleges in southeastern Minnesota were also invited to participate by letter. Those who accepted the invitation were sent materials through campus mail. All subjects participated voluntarily, and without compensation.

Materials

Each subject received a large envelope containing three surveys, a set of instructions outlining the sequence in which they were to be filled out, and a consent form.⁴

The first instrument was the short form of the DIT (Rest, 1979), consisting of three moral dilemmas (Heinz, Escaped Prisoner, and Newspaper). For each dilemma, there was a set of standard issues that subjects were asked to rate the importance of, and then rank.

The second survey consisted of a single sheet of paper, containing the question, "When faced with a moral dilemma, what issues or concerns influence your decision?" Written instructions that preceded the question emphasized the open-ended nature of the question, and defined "moral dilemma" as "a situation in which different things that you value come into conflict."

The final survey, also on a single page, was the instrument developed by Ford and Lowery (1986). It described justice and care orientations to moral reasoning, and asked subjects to rate on a 7-point scale the degree to which each perspective was a part of their own thinking about moral dilemmas. The justice orientation was described as follows:

One way to approach thinking about a moral conflict is to see yourself involved in a *process of judging*. This includes *standing back* from the problem at hand to consider the most *fair* way to resolve the dilemma. It often means thinking of certain rules or laws which you believe should govern the way people behave. The problem, as you perceive it, involves *competing sets of values*. You seek what is *most just* by considering the rights of all involved.

The care/responsibilities orientation was described as follows:

One way to approach thinking about a moral conflict is to see yourself involved in *relationship* with others. This means you have certain *responsibilities* to be *concerned for others* and to consider how what you do will *help or hurt* those involved. Conflicts are considered in the *context* of their effect on relationships between people, and whether those *relationships will be maintained, restored or damaged*. You seek a way to *respond* that will minimize the hurt to all involved.

After presentation of each orientation, a 7-point scale was presented. One end of the scale (1) was labeled, "Not at all a part; I never think of things such as this." The midpoint (4) was labeled, "Sometimes a part; I sometimes think of things such as this," and the other end point (7) was labeled, "A very great part; I always think of things such as this." Intermediate points on the scale were not labeled. Subjects were asked to circle one of the 7 integers on each scale.

⁴Copies of the instruments are available on request.

Table I. Mean P-Scores, Justice and Care Orientation Ratings by Grade and Gender^a

Measure	Grade and gender					
	8th— male	8th— female	11th— male	11th— female	College— male	College— female
P score	18.00 (10.54)	21.76 (10.94)	32.00 (13.88)	24.72 (11.00)	48.71 (17.62)	49.08 (16.45)
Justice rating	5.07 (1.21)	4.68 (1.32)	5.24 (0.89)	5.36 (1.16)	5.33 (1.36)	5.44 (1.04)
Care rating	4.93 (1.69)	5.53 (1.21)	4.97 (1.10)	6.10 (1.35)	5.59 (1.09)	5.55 (1.19)

^aNumbers in parentheses are standard deviations.

Procedure

The three surveys, together with an instruction sheet and a consent form, were enclosed in envelopes. Students in Grades 8 and 11 participated during classes or study halls, in groups of 4–30. One or two authors attended each session. After briefly explaining the nature of the study, they reviewed the instructions for each task with the group of students. They then remained available to answer individual questions. When students completed the three surveys, they put all materials back in the envelope and handed it in. The average time for students to complete all tasks was approximately 35 minutes.

Surveys were sent through campus mail to college students in the initial sample who indicated a willingness to participate. They completed all surveys at their own convenience, but were asked to do so during one session. Anecdotal evidence suggests they complied, and also took approximately 35 minutes to complete all of the tasks.

RESULTS

Ratings

P scores, reflecting the proportion of “principled” moral reasoning, were calculated from the DIT rankings after scanning all DIT responses for consistency (see Rest, 1979). Self-reported ratings of justice and care orientations, coming from the third survey, were also recorded. Table I presents the means for all three measures, by gender and grade. A 2 (gender) × 3 (grade) analysis of variance (ANOVA) performed on P scores revealed a main effect for grade ($F[2,163] = 64.52, p < .001$). Mean P scores were 20.17, 27.70, and 48.90 for 8th graders, 11th graders, and col-

lege sophomores, respectively. Tukey tests showed that all means differed ($p < .05$). The interaction between gender and grade was only marginally significant ($F[2, 163] = 2.61, p < .08$), and there was no main effect for gender.

A 2 (type of rating: justice or care) \times 2 (gender) \times 3 (grade) mixed ANOVA, with repeated measures on the first factor, was run on the self-reported ratings. It yielded a main effect for type of rating, with care ratings ($M = 5.48$) significantly higher than justice ratings ($M = 5.18; F[1, 179] = 4.86, p < .05$). Overall, males reported significantly lower ratings than females (M [combined] = 5.17 vs. 5.46, $F[1, 179] = 4.26, p < .05$). There was also a main effect for grade, with 8th graders, 11th graders, and college sophomores reporting mean combined ratings of 5.06, 5.43, and 5.49, respectively ($F[2, 179] = 3.91, p < .05$). Tukey tests revealed that 8th graders differed from college students ($p < .05$), but no other differences were reliable.

One of the most important findings from this analysis was a statistically significant interaction between gender and rating ($F[1, 179] = 7.07, p < .01$). Tukey's tests showed that males' justice and care ratings ($M = 5.21$ and 5.14 , respectively) did not differ, although the corresponding ratings for females ($M = 5.14$ and 5.79 , respectively), were significantly different ($p < .05$). Correspondingly, males and females did not differ significantly in their justice ratings, but did in their use of care ratings ($p < .05$). No other interactions approached statistical significance.

Correlations were also computed between justice and care ratings and P scores. The Pearson product-moment correlations with P scores were $r(185) = .02, ns$, and $.08, ns$, for justice and care ratings, respectively. The correlation between justice and care ratings was $r(185) = .11, p < .10$.

Open-Ended Essays

Subjects' responses to the open-ended essay question ("When faced with a moral dilemma, what issues or concerns influence your decision?") were typed and identified only by subject number. They were scored independently by the three authors, who were blind to a subject's gender, grade, or school. Each essay was scored for the presence/absence of several themes. The set of themes is an extension of those used by Galotti (1989). The 11 themes, along with the scoring criteria, are shown in Table II. Interrater reliabilities over the three raters, computed using coefficient alpha, ranged from .72 to .97, with a median of .85.

Table II. Description of Essay Themes Scoring Criteria and Interrater Reliabilities

Interrater ^a reliability	Title and description of theme ^b
.86	What Others Would Think/Feel: Whether subject (S) consults others, or imagines others in a similar situation to make a decision.
.91	Effect on Others: Whether S considers the possible benefit or harm to other people in making the decision.
.83	Situation Specifics: Whether S makes explicit reference to the decision "depending on the situation."
.85	Effect on Self: Whether S considers the possible benefit or harm to him/herself in making the decision, including personal guilt, social reaction.
.81	What I Feel: Whether S mentions "intuition" or "gut feeling" in making a decision, or mentions "how I would feel." Emphasis on feeling.
.97	Religious Teachings: Whether S considers formal religious teachings (e.g., the Bible, Talmud) or reasons from experience in an organized religion (to be distinguished from "conscience" or "personal sense of right and wrong").
.85	Greater Societal Good: Whether S refers explicitly to the concerns of society in general or to the "greatest good."
.96	Legal Issues: Whether S refers to the local, state, or federal laws.
.72	Personal Code of Ethics: Whether S makes explicit references to a personal set of moral values or ethics, or a personal sense of right and wrong or to specific principles that are valued most highly; the values of the issues to the S; the principles by which S was raised.
.85	Systematic Reasoning: Whether S describes trying to reason logically, systematically, without being affected by mood or emotion. If S uses some of the following phrases "weigh the pros and cons," "gathered all the facts," "think it through" "look at all the possibilities."
.84	Rights of Others: Whether S explicitly mentions the (legal, personal, or moral) rights of other people (to be distinguished from the benefits or harm to people; see "Effect on Others," above).

^aComputed over three authors, using coefficient alpha. Median interrater reliability = .85.

^bCoding descriptions used by authors, who rated independently, and were blind to gender and grade.

Table III. Mean Proportion Usage of Themes in Open-Ended Essay by Grade and Gender^a

Theme	Grade and gender					
	8th— male	8th— female	11th— male	11th— female	College— male	College— female
What Others Think	.08 (.27)	.26 (.45)	.37 (.49)	.45 (.50)	.15 (.36)	.44 (.51)
Effect on Others	.50 (.51)	.53 (.51)	.40 (.50)	.58 (.50)	.59 (.50)	.80 (.41)
Situation Specifics	.08 (.27)	.15 (.36)	.03 (.18)	.15 (.36)	.30 (.47)	.12 (.33)
Effect on Self	.46 (.51)	.68 (.47)	.47 (.51)	.65 (.48)	.30 (.47)	.56 (.51)
What I feel	.04 (.20)	.09 (.29)	.10 (.31)	.13 (.33)	.26 (.45)	.36 (.49)
Religious Teachings	.00 (.00)	.00 (.00)	.03 (.18)	.15 (.36)	.15 (.36)	.24 (.44)
Greater Societal Good	.08 (.27)	.00 (.00)	.00 (.00)	.10 (.30)	.26 (.45)	.32 (.48)
Legal Issues	.08 (.27)	.12 (.33)	.07 (.25)	.08 (.27)	.44 (.51)	.48 (.51)
Personal Code Ethics	.15 (.27)	.09 (.29)	.13 (.35)	.30 (.46)	.48 (.51)	.64 (.49)
Systematic Reasoning	.08 (.27)	.09 (.29)	.17 (.38)	.18 (.38)	.37 (.49)	.40 (.50)
Rights of Others	.08 (.27)	.00 (.00)	.03 (.18)	.05 (.22)	.11 (.32)	.20 (.41)
Total number of themes scored	1.62 (1.06)	2.00 (1.33)	1.80 (1.03)	2.80 (1.40)	3.41 (1.25)	4.56 (1.61)

^aNumbers in parentheses are standard deviations.

The mean proportion use of themes by gender and grade is shown in Table III. Hierarchical loglinear analyses (see Norusis, 1985) were run for each theme to discover relationships among use as a function of gender, grade, or their interaction. These analyses showed an association between use and gender for two themes: "What Others Think" (overall likelihood ratio [L.R.] chi-square [4] = 3.95; $p = .41$) and "Effect on Self" (overall L.R. chi-square [8] = 6.91; $p = .55$); in both cases, females' use was higher. The loglinear analyses showed an association between use and grade for the following themes: "What Others Think" (overall L.R. chi-square [4] = 3.95; $p = .41$); "What I Feel" (overall L.R. chi-square [6] = 3.93; $p = .69$); "Religious teachings" (overall L.R. chi-square [6] = 6.21; $p = .40$); "Legal Issues" (overall L.R. chi-square [6] = 2.94; $p = .81$); "Personal Code of Ethics" (overall L.R. chi-square [6] = 7.34; $p = .29$); "Systematic Reasoning" (overall L.R. chi-square [6] = 2.66; $p = .85$) and "Rights of Others" (overall L.R. chi-square [6] = 6.93; $p = .33$). In all cases but the first, use rose with grade; use of the first theme appeared more curvilinear (see Table III). These analyses revealed an interaction between grade, gender, and use for the following two themes: "Situation Specifics," and "Greater Societal Good" (overall L.R. chi-square [0] = 0.00; $p = 1.00$ for both). Finally, Table III also shows the total number of themes scored in essays (an essay could contain none, one, or any number of themes). A 2 (gender) \times 3 (grade) ANOVA on this measure revealed a main effect for gender, with males and females using an average of 2.27 and 2.97 themes, respectively ($F[1,176] = 18.91$, $p < .001$). There was also a main effect for grade ($F[2,176] = 42.89$, $p < .001$). Eighth graders, 11th graders, and college sophomores used a mean number of 1.83, 2.37, and 3.96 themes, respectively. Tukey tests showed that 8th and 11th graders both differed from college sophomores ($p < .01$), but not from each other. There was no significant interaction between gender and grade.

In addition to examining the use of individual themes, we also examined the use of types of themes, classifying some as "stereotypically masculine," and some as "stereotypically feminine." Following Gilligan (1977, 1982, 1988), Belenky *et al.* (1986), Galotti (1989), and Lyons (1983), we classified as "feminine" those themes that emphasized feelings and emotions, effects on others or on self, contextual reasoning (e.g., "it depends on the specifics of the situation"). Thus, the first five themes shown in Tables II and III were classified as feminine. The remaining themes were classified as masculine because they emphasized noncontextual reasoning, reasoning that came from established authorities (religion, the law, the Golden Rule), and/or reasoning that attenuated aspects such as emotion

Table IV. Mean Proportion Usage of Possible Masculine and Possible Feminine^a Themes, by Gender and Grade^b

Type of theme	Grade and gender					
	8th— male	8th— female	11th— male	11th— female	College— male	College— female
Possible masculine	.08 (.11)	.05 (.09)	.07 (.14)	.14 (.14)	.30 (.20)	.38 (.18)
Possible feminine	.23 (.18)	.34 (.22)	.27 (.19)	.39 (.21)	.32 (.22)	.46 (.25)

^aComputed by counting the number of themes classified as masculine or feminine that were actually used in an essay.

^bNumbers in parentheses are standard deviations.

and feeling. Thus, the last six themes displayed in Tables II and III were classified as masculine.⁵

We next computed the proportion of possible masculine and feminine themes by counting the number of themes actually used in an essay relative to the total number possible (six and five, respectively). These measures were subjected to a 2 (theme type) \times 2 (gender) \times 3 (grade) mixed ANOVA, with repeated measures on the first factor. Table IV presents the mean use of each type of theme by gender and grade.

The ANOVA showed significant main effects for all variables. Over all subjects, the use of possible feminine themes ($M = .34$) was higher than the corresponding use of possible masculine themes ($M = .16$; $F[1, 176] = 65.59$, $p < .001$). Females showed a higher use of both types of themes than males (M [combined] = .28 and .21; $F[1, 176] = 19.33$, $p < .001$). Use of both types of themes also increased with grade, with the combined means for 8th graders, 11th graders, and college sophomores equal to .17, .23, and .36 ($F[2, 176] = 36.16$, $p < .001$). Tukey tests ($p < .05$) showed that all means differed.

In addition, there were two significant interactions. The first, between type of theme (masculine, feminine) and gender ($F[1, 176] = 4.05$, $p < .05$), showed a significant ($p < .01$ by a Tukey test) difference between males' and females' use of possible feminine themes ($M = .27$ and .39), but not in their use of possible masculine themes ($M = .15$ and .17). The second interaction was between type of theme and grade ($F[2, 176] = 7.98$, $p < .001$). Mean use of possible masculine themes was .06, .11, and .34 for 8th graders, 11th graders, and college sophomores, respectively. The corresponding means for use of possible feminine themes were .29, .34,

⁵One might question the classification of "Religious Teachings" and "Personal Code of Ethics" as masculine. Although it may be true that women are thought to be more religious than men, we performed this classification to take account of the fact that in both themes, an appeal to a noncontextual set of rules, standards, or prohibitions is being made. It should be noted, however, that neither theme revealed a gender difference in pattern of usage.

Table V. Rotated Factor Loadings of Theme Usage

Theme	Factor				
	External Prescriptions	Effects	Nonemotional Deciding	Contextual Factors	Personal Values
What Others Think	.40	.04	-.08	-.67	.20
Effect on Others	.03	.84	.17	.23	.09
Situation Specifics	.16	.04	-.10	.69	.19
Effect on Self	-.01	.79	-.23	-.26	-.04
What I feel	-.07	.11	.01	-.08	.74
Religious Teachings	.67	-.04	-.15	-.09	.30
Greater Societal Good	.13	-.02	.55	.07	.19
Legal Issues	.73	.11	.31	.01	-.13
Personal Code Ethics	.16	-.12	.27	.17	.57
Systematic Reasoning	-.03	-.01	.83	-.09	-.03
Rights of Others	.51	-.13	.13	.42	-.29
Eigenvalue	1.75	1.48	1.24	1.14	1.02
Percent variance accounted for	15.9	13.5	11.0	10.4	9.3

and .38. Tukey tests ($p < .01$) showed that the means for the two types of themes differed for 8th and 11th graders, but not for college sophomores. No other interactions approached significance.

A related analysis took as the dependent measure the proportion of all themes scored in an essay that were masculine (note that the corresponding measure for feminine themes is one minus this measure). A 2 (gender) \times 3 (grade) on this measure revealed only a main effect for grade, with means for 8th graders, 11th graders, and college sophomores equal to .19, .26, and .54 ($F[2, 163] = 19.11, p < .001$). Tukey tests revealed that college sophomores differed from the other two groups ($p < .01$), who did not differ.

One might question how legitimate it is to categorize these 11 themes into two categories, labeled "masculine" and "feminine." To address this question, we performed a principal components factor analysis on presence or absence of each theme (recall that each essay could contain any number of themes). Table V presents the rotated factor loadings for each theme. The five extracted factors were named after examination of the patterns of loadings. Only two of the five factors can readily be labeled "masculine" (nonemotional deciding) or "feminine" (effects) based upon examination of the factor loadings. The other three factors (external prescriptions, contextual factors, personal values) loaded upon both types of themes.

These factor loadings were used to compute five factor scores for each subject, by multiplying the factor loading for a particular theme by 1 if the theme was present in the essay and 0 otherwise, and summing over

Table VI. Mean Factor Scores (Superordinate Theme) by Grade and Gender^{a,b}

Factor	Grade and gender					
	8th— male	8th— female	11th— male	11th— female	College— male	College— female
External Prescription	.18 (.30)	.23 (.39)	.26 (.34)	.44 (.48)	.68 (.56)	.93 (.66)
Effects	.77 (.73)	1.01 (.70)	.72 (.64)	.99 (.64)	.74 (.68)	1.11 (.70)
Nonemotional Deciding	.15 (.29)	.03 (.27)	.12 (.45)	.19 (.44)	.70 (.63)	.78 (.53)
Contextual Factors	.05 (.34)	-.13 (.43)	-.24 (.48)	-.19 (.54)	.25 (.47)	-.04 (.64)
Personal Values	.15 (.26)	.20 (.30)	.23 (.34)	.44 (.48)	.59 (.50)	.79 (.50)

^aNumbers in parentheses are standard deviations.

^bSee Table V and text for a description of these factors and their computation.

all 11 themes. These five factor scores were subjected to 2 (gender) × 3 (grade) ANOVAs. Table VI presents the means for each of the five factors, by gender and grade.

None of the ANOVAs yielded a significant interaction between grade and gender. However, three of the ANOVAs yielded a main effect of gender, in each case with females having higher scores than males. These were for the factors “External Prescriptions” ($F[1,176] = 5.32; p < .05; M = .37$ for males and $.49$ for females), “Effects” ($F[1,176] = 8.08; p < .01; M = .74$ for males and 1.03 for females), and “Personal Values” ($F[1,176] = 6.36; p < .05; M = .32$ for males and $.45$ for females). Significant main effects for grade were found for four of the five factors: “External Prescriptions” ($F[2,176] = 25.41; p < .001; M = .21, .36,$ and $.80$ for 8th graders, 11th graders, and college students, respectively), “Nonemotional Deciding” ($F[2,176] = 35.80; p < .001; M = .08, .16,$ and $.74$ for 8th graders, 11th graders, and college students, respectively), “Contextual Factors” ($F[2,176] = 6.02; p < .01; M = -.05, -.21,$ and $.11$ for 8th graders, 11th graders, and college students, respectively); and “Personal Values” ($F[2,176] = 22.78; p < .001; M = .17, .35,$ and $.69$ for 8th graders, 11th graders, and college students, respectively). Tukey tests showed that the means for college students differed from the other two, which did not differ ($p < .05$) for all of the factors except “Contextual Factors,” where only the 8th graders and college students differed.

Finally, we examined associations between use of specific themes and other more global measures of performance. Specifically, the use of each individual theme was correlated with P scores and justice and care ratings. The purpose of this analysis was to see whether the subjects’ report of what issues or concerns they typically considered bore any relationship to their thinking about specific moral dilemmas (as indexed by the P score) or to other descriptions they provided of their own moral reasoning (the

Table VII. Correlations of Theme Usage with Other Measures

Theme	Measure		
	P-score	Justice rating	Care rating
What Others Think	-.03	.09	-.01
Effect on Others	.21 ^b	-.06	.17 ^a
Situation Specifics	.11	.02	.07
Effect on Self	-.08	-.07	.14 ^a
What I feel	.17 ^a	.07	.09
Religious Teachings	.12 ^a	.06	.11
Greater Societal Good	.37 ^c	.07	.14
Legal Issues	.29 ^c	.21 ^b	.07
Personal Code Ethics	.28 ^c	.03	.07
Systematic Reasoning	.27 ^c	.10	-.04
Rights of Others	.06	.14 ^a	-.02

^a*p* < .05.^b*p* < .01.^c*p* < .001.

justice and care ratings). Table VII presents these correlations. In general, P scores tended to correlate moderately with a number of themes typically thought to be masculine (see below for more discussion), such as "Greater Societal Good," "Legal Issues," and "Personal Code of Ethics." Justice and care ratings generally did not correlate with most themes, with some notable exceptions. Justice ratings correlated with two masculine themes ("Legal Issues" and "Rights of Others") and care ratings correlated with two feminine themes ("Effect on Others" and "Effect on Self"), although in all four cases, the correlations were small in magnitude.

DISCUSSION

Differences in conceptions of moral reasoning, as a function of grade and of gender, were evident. Older students wrote more complex essays, as indexed by the number of themes scored. Moreover, the types of issues students focused upon also changed as a function of grade. Older students were more likely to use the following themes in their essays: "What I Feel," "Religious Teachings," "Legal Issues," "Personal Code of Ethics," "Systematic Reasoning," and "Rights of Others," while a curvilinear relationship with grade existed for "What Others Think." This pattern of results suggests that, with age, there is an increasing view of moral reasoning as prescriptive, noncontextual, and autonomous.

One problem in interpreting many of the grade effects, however, is that of confounding of age/grade with educational setting. If our college students come from a more highly educated, motivated, or affluent popula-

tion than the younger students, the source of the grade effects becomes less clear. This issue will need to be addressed in future work.

Gender differences in conceptions of moral reasoning were also found. Females' essays contained more themes, suggesting either more complexity in their thinking, or more compliance with task instructions. Of the 11 themes scored, gender differences appeared for 2. Both of these, "What Others Think," and "Effect on Self," were predicted by previous work. Furthermore, females' essays were more likely to reflect the superordinate theme of "Effects" than were males' essays, again in keeping with similar findings from previous work.

However, on many of the masculine measures (i.e., the DIT P score, justice ratings, all 6 of the masculine themes for essays, and the mean proportion of use of the possible masculine themes), males' and females' performances were not distinguishable. Gender differences that emerged were usually on feminine measures (i.e., 2 of the 5 feminine themes, care ratings, mean proportion use of the possible feminine themes) in each case, with females scoring higher than males. These results support the predictions of Belenky *et al.* (1986), Gilligan (1977, 1982, 1988b), and Lyons (1983), and are consistent with similar findings reported by Donenberg and Hoffman (1988) and Rothbart *et al.* (1986). On two superordinate themes ("Personal Values" and "External Prescriptions"), where males might have been predicted to show higher scores than females, the opposite was in fact true.

At the same time, the gender differences were not apparent everywhere, as might have been expected. Nine of the 11 essay themes did not reveal gender differences (although 7 of the 11 showed grade differences, suggesting that the measures were sensitive enough to detect differences). No gender differences in P scores were found, nor were justice ratings different. No gender differences emerged for two of the five superordinate themes ("Nonemotional Deciding" and "Contextual Factors"). The gender similarities support conclusions of researchers studying similar or related tasks (Archer and Waterman, 1988; Ford and Lowery, 1986; Galotti, 1989; Galotti and Kozberg, 1987; Walker, 1986b, 1989; Walker *et al.*, 1987), who also report few to no gender differences.

One finding relevant to Gilligan's predictions was the pattern of increasing masculinity with grade: All of the themes that show increased use with grade are classified as masculine, the mean use of possible masculine themes rose with age (in contrast to the mean use of feminine themes), and the proportion of all themes scored in an essay rose dramatically (from .2 to .5) over the three age groups. Moreover, of the 4 superordinate themes that showed grade effects, 3 drew more heavily from masculine than from feminine themes, while only one showed the opposite pattern. This

pattern of findings is consistent with Donenberg and Hoffman's (1988) finding, that subjects become more justice oriented with age.

This pattern of results reinforces the idea that mature moral reasoning is associated with noncontextual, nonemotional reasoning, at least in the middle-class sample of subjects studied here. However, two things should be noted. The first is that the trends just mentioned held for both males and females. The second is that the use of feminine themes, especially "What Others Think," "Effect on Others," and "Effect on Self," did not decrease with age, and indeed remained at fairly high levels for all ages studied. Moreover, the use of possible feminine themes was twice as high as the use of possible masculine themes, replicating the findings of Galotti (1989). This in turn reinforces the idea that moral concerns typically classified as feminine are important to all subjects over a wide range of ages (and presumably, moral levels of development; Galotti, 1989).

The factor analysis on theme use suggests that the classification of moral reasoning into two categories—justice/rights vs. care/responsibility—may be too simplistic. That five factors were needed to account for the patterns of responses, and that only two factors drew exclusively from either masculine or feminine themes, suggests that moral reasoning of students across a wide range of developmental and educational levels is more multidimensional than binary. It is thus perhaps unsurprising that overall justice and care ratings were, for the most part, uncorrelated with other performance measures, such as P score, use of most individual themes, or with each other.

The correlations between P scores and theme use yielded significant associations for 7 of the 11 themes, most of them masculine. Again, this speaks to the face validity of the essay responses, and is predicted by the DIT's orientation to issues of competing rights, justice, and fairness.

One potential confound in the data merits discussion. This is the fact that subjects wrote their open-ended responses *after* having completed the DIT. It might therefore be argued that the responses were contaminated by the immediately previous exposure to the three specific DIT dilemmas. This possibility cannot be ruled out. However, it is worth noting that the results from the open-ended essays are consistent with previously reported results where order effects were not an issue (e.g., Donenberg and Hoffman, 1988; Galotti, 1989; Rothbart *et al.*, 1986). These consistencies argue against the view that order effects account for all, or even most, of the effects reported here.

Three directions for future work are suggested. One concerns the effects of education on conceptions of moral reasoning. It is not clear from the present study how many of the grade effects are accounted for by differences in years or types of education. Future investigation should examine

more closely the kinds of experiences that elicit changes in conceptions of moral thinking.

A second issue concerns reasoning about actual, rather than hypothetical, moral dilemmas. The present results suggest a link between conceptions of moral thinking and the latter. It will be informative to chart the relationships between conceptions of moral reasoning and actual confrontations with moral issues.

Finally, the issue of how one's sense of self as a moral reasoner develops must be investigated in a broader framework. It seems evident that moral reasoning is a part of, and possibly an influence on, reasoning about identity, and reasoning about the nature of knowledge and action. The degree to which moral, cognitive, epistemological, and identity development interrelate is a topic worthy of detailed investigation. We hope future work will explore how moral reasoning influences, and is in turn influenced by, these other types of reasoning.

ACKNOWLEDGMENT

We thank Al Johnson, Jan Mitchell, John Metzke, Cynthia Trapp, and Alice Thomas for their help in recruiting subjects, and Paula Erickson and Valerie Sims for their help with data analyses and coding, and Jon Baron for comments on an earlier version of the manuscript.

REFERENCES

- Archer, S. L., and Waterman, A. S. (1988). Psychological individualism: Gender differences or gender neutrality? *Human Develop.* 31: 65-81.
- Baumrind, D. (1986). Sex differences in moral reasoning: Response to Walker's (1984) conclusion that there are none. *Child Develop.* 57: 511-521.
- Belenky, M. F., Clinchy, B. M., Goldberger, N. R., and Tarule, J. M. (1986). *Women's Ways of Knowing*. Basic Books, New York.
- Bussey, K., and Maughan, B. (1982). Gender differences in moral reasoning. *J. Personal. Social Psychol.* 42: 701-706.
- Clinchy, B., and Zimmerman, C. (1982). Epistemology and agency in the development of undergraduate women. In Perun, P. (ed.), *The Undergraduate Woman: Issues in Educational Equity*. Lexington Books, Lexington, MA.
- Colby, A., and Kohlberg, L. (1987). *The Measurement of Moral Judgment (Vol. 1)*. Cambridge University Press, Cambridge.
- Colby, A., Kohlberg, L., Gibbs, J., and Liberman, M. (1983). A longitudinal study of moral judgment. *Monographs of the Society for Research in Child Development* 48 (1, Serial No. 200).
- Donenberg, G. R., and Hoffman, L. W. (1988). Gender differences in moral development. *Sex Roles* 18: 701-717.
- Ford, M. R., and Lowery, C. R. (1986). Gender differences in moral reasoning: A comparison of the justice and care orientations. *J. Personal. Social Psychol.* 50, 777-783.
- Galotti, K. M. (1989). Gender differences in self-reported moral reasoning: A review and new evidence. *J. Youth Adoles.* 18: 475-488.

- Galotti, K. M., and Kozberg, S. F. (1987). Older adolescents' thinking about academic/vocational and interpersonal commitments. *J. Youth Adoles.* 16: 313-330.
- Galotti, K. M., Kozberg, S. F., and Appleman, D. (1991). Younger and older adolescents' thinking about commitments. *J. Exp. Child Psychol.* 50: 324-339.
- Gilligan, C. (1977). In a different voice. *Harvard Educat. Rev.* 47: 481-517.
- Gilligan, C. (1982). *In a Different Voice*. Harvard University Press, Cambridge, MA.
- Gilligan, C. (1988a). Preface. In Gilligan, C., Ward, J. V., Taylor, J. M., and Bardige, B. (eds.), *Mapping the Moral Domain*. Harvard University Press, Cambridge, MA.
- Gilligan, C. (1988b). Remapping the moral domain: New images of self in relationship. In Gilligan, C., Ward, J. V., Taylor, J. M., and Bardige, B. (eds.), *Mapping the Moral Domain*. Harvard University Press, Cambridge, MA.
- Gilligan, C., and Attanucci, J. (1988). Two moral orientations. In Gilligan, C., Ward, J. V., Taylor, J. M., and Bardige, B. (eds.), *Mapping the Moral Domain*. Harvard University Press, Cambridge, MA.
- Gilligan, C., and Belenky, M. F. (1980). A naturalistic student of abortion decisions. In Selman, R. L., and Yando, R. (eds.), *New Directions for Child Development: Clinical Developmental Psychology* (No. 7). Jossey-Bass, San Francisco.
- Gilligan, C., Ward, J. V., Taylor, J. M., and Bardige, B. (eds.) (1988). *Mapping the Moral Domain*. Harvard University Press, Cambridge, MA.
- Hanson, R. A., and Mullis, R. L. (1985). Age and gender differences in empathy and moral reasoning among adolescents. *Child Study J.* 15: 181-188.
- Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In Lickona T. (ed.), *Moral Development and Behavior: Theory, Research and Social Issues*. Holt, Rinehart & Winston, New York.
- Kohlberg, L. (1984). *Essays in Moral Development. Volume II: The Psychology of Moral Development*. Harper & Row, San Francisco.
- Kohlberg, L., Levine, C., and Hewer, A. (1983). *Moral Stages: A Current Formulation and a Response to Critics*. S. Karger, Basel.
- Lyons, N. P. (1983). Two perspectives: On self, relationships, and morality. *Harvard Educat. Rev.* 53: 125-145.
- Norusis, M. J. (1985). *SPSSx Advanced Statistics Guide*. McGraw-Hill, New York.
- Perry, W. G., Jr. (1970). *Forms of Intellectual and Ethical Development in the College Years: A Scheme*. Holt, Rinehart & Winston, New York.
- Pratt, M. W., Golding, G., Hunter, W., and Sampson, R. (1988). Sex differences in moral orientation. *J. Personal.* 56: 373-391.
- Rest, J. R. (1979). *Revised Manual for the Defining Issues Test*. Minnesota Moral Research Projects, Minneapolis.
- Rest, J. R. (1983). Morality. In Flavell, J. H., and Markman, E. M. (eds.), Mussen, P. H. (series ed.), *Handbook of Child Psychology: Vol. 3. Cognitive Development*. Wiley, New York.
- Rothbart, M. K., Hanley, D., and Albert, M. (1986). Gender differences in moral reasoning. *Sex Roles* 15: 645-653.
- Scheidel, D. G., and Marcia, J. E. (1985). Ego identity, intimacy, sex role orientation, and gender. *Develop. Psychol.* 21: 149-160.
- Thoma, S. J. (1986). Estimating gender differences in the comprehension and preference of moral issues. *Develop. Rev.* 6: 165-180.
- Walker, L. J. (1984). Sex differences in the development of moral reasoning: A critical review. *Child Develop.* 55: 677-691.
- Walker, L. J. (1986a). Sex differences in the development of moral reasoning: A rejoinder to Baumrind. *Child Develop.* 57: 522-526.
- Walker, L. J. (1986b). Experiential and cognitive sources of moral development in adulthood. *Human Develop.* 29: 113-124.
- Walker, L. J. (1989). A longitudinal study of moral reasoning. *Child Develop.* 60: 157-166.
- Walker, L. J., de Vries, B., and Trevethan, S. D. (1987). Moral stages and moral orientations in real-life and hypothetical dilemmas. *Child Develop.* 58: 842-858.
- Yussen, S. R. (1977). Characteristics of moral dilemmas written by adolescents. *Develop. Psychol.* 13: 162-163.