Meta-Analyses of Age and Sex Differences in Children’s and Adolescents’ Prosocial Behavior

Richard A. Fabes and Nancy Eisenberg
Arizona State University

Part of this work was supported by a grant from the National Science Foundation (DBS-9208375) to both authors and by a Research Scientist Development Award from the National Institute of Mental Health (K02 MH00903) to Nancy Eisenberg. The authors wish to thank Pat Maszk and Sherri Souza for their help in obtaining and coding the materials for this research. Portions of this paper appear in; Eisenberg, N., & Fabes, R. A. (1998). Prosocial development. In W. Damon (Ed.), Handbook of Child Psychology, Fifth Edition (Vol 3: Social, Emotional, and Personality Development, N. Eisenberg [Ed.]). Address correspondence to Richard Fabes, Dept. of Family Resources & Human Development, Arizona State University, Tempe, AZ. 85287-2502, email: rfabes@asu.edu. © 1998 by the authors. All rights reserved.

Sample of studies. Several procedures were used to obtain studies used in the meta-analysis. First, we used a computer-based information search of PsycINFO (Psychological Abstracts) from 1974-1994. The key words in the search included altruism, prosocial behavior, helping behavior, sharing, caring, donating, and comforting. We also searched the reference lists in the journal articles, books, and chapters in our sample of studies. Finally, we examined the studies published in recent volumes of developmental journals (i.e., Developmental Psychology, Child Development, Merrill-Palmer Quarterly, Journal of Adolescent Research, Journal of Early Adolescence, Applied Developmental Psychology).

The criteria for including studies in the prosocial analyses were (a) the dependent variable had to meet our definition of prosocial behavior, (b) participants were not sampled from a clinical population (handicapped children, delinquents), (c) participants were children and/or adolescents not over the age of 21, (d) the article was published, (e) the data reported were for individual children rather than data at the dyadic, triadic, or class levels, (f) the index of prosocial behavior was for real rather than pretend or play prosocial behavior, and (g) the article contained results that were sufficient to either calculate an effect size for age differences or reported that the age difference was nonsignificant (without accompanying data; in this case a 0 was given for the effect size). Studies of allocation behavior and studies of competition versus cooperation (e.g., Kagan & Knight, 1979, Knight & Kagan, 1977a,b) were not included for two reasons: (a) these measures vary in terms of the degree to which they can be interpreted as prosocial, and (b) allocation measures differ somewhat in form from all other measures and, due to the number of studies containing such measures, they would overly influence the meta-analysis. The resulting total sample consisted of 155 studies yielding 478 effect sizes ($M$ number of effect sizes per study = 2.67, $SD = 3.95$).  

Variables coded from each study. The following data were recorded from each study: (a) sample size for older and younger children (when no specific sizes were reported we assumed equal numbers of older and younger children; $M_s = 38.58$ and $39.44$, $SD_s = 39.28$ and $53.35$, respectively), (b) type of index (instrumental helping, sharing/donating, comforting, aggregated index of prosocial behavior), (c) method (observation, self-report, other-report [peers, parents, teachers]), (d) design (correlational/naturalistic, structured/experimental), (e) target of prosocial response (children, adults, unknown/unspecified), (f) mean age of sample (when grade was reported, it was translated into years; $M$ age $= 7.83$ years, $SD = 3.03$), (g) mean age of oldest and youngest children in sample ($M_s = 9.31$ and $7.83$ years, $SD_s = 3.51$ and $2.80$, respectively), and (h) year of publication (articles that were currently in press were given a 1995 publication date). These variables were coded by a single rater. A second coder independently coded 44 of the studies in these lists. The two raters agreed on 86% of the codings.

Methods of computation. Initially, the effect size $g$ (difference between the means of two groups, divided by the pooled standard deviation) was computed for each variable (using Johnson’s [1993] software). The estimator $g$ was then corrected for bias and the unbiased effect size $d$ was used in the analyses (Hedges & Olkin, 1985). Positive values for $d$ represent higher levels of a prosocial behavior for older than for younger children. Computation of effect size was based on (a) 61 reports of means and standard deviation, (b) 73 reports of correlations, (c) 155 $F$ tests, (d) 8 $t$ tests, (e) 76 reports of proportions, (f) 28 reports of significance levels, (g) 6 reports of chi square, and (h) 71 cases in which authors reported no significant differences but did not report any data (a $d$ of 0 was given in these cases to indicate no difference).

Potential outliers were identified in both samples by computing a schematic box plot and determining which effect sizes were outside of the inner fence (Glass, McGaw, & Smith, 1981). Based on these criteria, 12 effect sizes were eliminated from the sample.

We took two approaches to our analysis of age differences in prosocial behavior. First, we were interested in examining effect sizes as a function of the specific age group comparison. Because some of the studies utilized correlational analyses to examine the relation of the age to prosocial behavior ($n = 55$), specific comparisons between age-related groups could not be calculated. Thus, for the analysis of prosocial behavior as it relates to specific age comparisons we relied on those studies in which specific age groups were contrasted ($n = 125$).

Secondly, we were interested in the predictors of the magnitude of effect sizes in age differences in prosocial behavior. In this analysis, we used all available studies. We computed least square weighted univariate regression analyses (in which each individual predictor was examined separately) and multivariate regressions (in which each predictor was examined while simultaneously controlling for the other predictors in the regression equation).

Age group analyses. From the 125 studies in which there were age-related comparisons reported, 378 effect
sizes (excluding outliers) were computed (with an average of 3.02 effect sizes per study). Although there is not a single convention to handle the potential problems of nonindependent effect sizes, we elected to be conservative in our approach. Thus, in the analysis of age differences in prosocial behavior we did the following: (a) within any study, we included only one effect size for each unique combination of samples used in an age group comparison and (b) when more than one effect size within a study was calculated for a specific age group comparison, we randomly selected one of these to include in our sample of effect sizes.

To examine specific age-related changes in children’s prosocial behavior, we categorized the children in each specific age comparison into one of the following age groups (based on the mean age of the children in a particular age group): (a) infants (less than 3 years of age), (b) preschool (3-6 years of age), (c) childhood (7-12 years of age), (d) adolescent (13-17 years of age), and (e) young adults (18-21 years of age). Because young adults were included in only four effect size comparisons, we excluded this group from our analyses. Comparisons were made both across and within age groups (e.g., a comparison of 8 and 10-year-olds is labeled a childhood/childhood comparison). Thus, the final sample of effect sizes included 265 effects from the 125 studies (2.12 effect sizes per study; see Table 1). Mean ages of samples broken down by age comparison group are presented in Table 1.

### Table 1. Summary of Sample Qualities for Meta-Analyses of Age Differences in Prosocial Behavior.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
</table>
| All Effect Sizes in Data Base (n = 478)
  Unweighted Effect Size | 0.37 |
  95% Confidence Interval (lower/upper) | 0.32/0.41 |
| Total Effect Sizes in which Age Groups were Compared (n = 378)
  Unweighted Effect Size | 0.35 |
  95% Confidence Interval (lower/upper) | 0.31/0.40 |
| Effect Sizes Used in Age Group Analyses (n = 265)
  Unweighted Effect Size | 0.38 |
  95% Confidence Interval (lower/upper) | 0.33/0.43 |
| Total Effect Sizes in which Effect Sizes were Calculated from Correlations (n = 88)
  Unweighted Effect Size | 0.23 |
  95% Confidence Interval (lower/upper) | 0.14/0.32 |
| Effect Sizes from Correlational Data Used in Regression Analyses (n = 50)
  Unweighted Effect Size | 0.21 |
  95% Confidence Interval (lower/upper) | 0.11/0.32 |
| Total Effect Sizes Used in Regression Analyses (n = 315)
  Unweighted Effect Size | 0.36 |
  95% Confidence Interval (lower/upper) | 0.31/0.40 |

Mean Age (and Standard Deviation) of Older/Younger Group by Age Group Comparison for Age Group Analyses

| Infant/Infant | 2.00 (.33)/1.40 (.33) |
| Preschool/Infant | 4.58 (1.12)/2.13 (.57) |
| Preschool/Preschool | 5.59 (.89)/3.90 (.77) |
| Childhood/Preschool | 9.03 (1.72)/5.43 (.62) |
| Childhood/Childhood | 10.40 (1.30)/7.59 (1.13) |
| Adolescent/Preschool | 12.50 (.63)/9.04 (1.73) |
| Adolescent/Childhood | 13.36 (1.64)/9.04 (1.73) |
| Adolescent/Adolescent | 15.71 (.63)/12.98 (.67) |

**Note.** Effect sizes are positive for differences favoring older children.  

- Total number of effect sizes across and within all studies used.  
- Total number of effect sizes calculated from studies that presented comparisons of children of specific ages (excluding outliers).  
- Only those effect sizes calculated from data comparison children of specific ages with only one effect size per study (excluding studies that included young adults and excluding the outliers).  
- Total number of effect sizes from studies in which a correlation between age and prosocial behavior was calculated (excluding outliers).  
- Only those effect sizes calculated from correlational data with only one effect size per study (excluding outliers).  
- Effect sizes of data set “C” combined with those from data set “D”.  


Table 1 also includes a descriptive summary of the data sets for the meta-analysis of age differences in prosocial behavior. Overall mean unweighted positive effect sizes were obtained, indicating a difference favoring older children.

Table 2 presents the overall effect size for the entire sample and the effect sizes broken down for each age group, as well as each categorical coding classification broken down by age group. These effect sizes are weighted by the reciprocal of its variance (d+) see Hedges & Olkin, 1985). Because we only used published studies, we computed a Fail Safe N (Cooper, 1979) to determine the number of additional studies that would be needed to reverse a conclusion that a significant relationship exists at least at the p < .001 (two-tailed) level. Based on these calculations, it would take 8,363 additional studies with effect sizes of 0 (or totaling 0) to change our significance from p < 0.001 for the overall effect size of age differences in prosocial behavior.

Although we can be confident that the overall age difference differed from 0.0 (the value indicating exactly no sex difference), the significant homogeneity statistic (Q) indicated that the effect sizes in this sample were not consistent across studies (see Table 2). Our next step was to disaggregate the effect sizes.

Table 2. Tests of Categorical Models for Age Differences in Prosocial Behavior Effect Sizes

<table>
<thead>
<tr>
<th>Variables and Class</th>
<th>Between-Class effect (Q_b)</th>
<th>n</th>
<th>Weighted effect size (d+)</th>
<th>95% CI for d+ (lower/upper)</th>
<th>Homogeneity (Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td>265</td>
<td>0.26*</td>
<td>0.23/0.29</td>
<td>749.56***</td>
</tr>
<tr>
<td>Age Group Comparison</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infant/Infant</td>
<td></td>
<td>10</td>
<td>0.26*</td>
<td>0.06/0.45</td>
<td>7.62</td>
</tr>
<tr>
<td>Preschool/Infant</td>
<td></td>
<td>11</td>
<td>0.15*</td>
<td>-0.01/0.30</td>
<td>13.36</td>
</tr>
<tr>
<td>Childhood/Preschool</td>
<td></td>
<td>30</td>
<td>0.24***</td>
<td>0.13/0.35</td>
<td>31.54</td>
</tr>
<tr>
<td>Childhood/Childhood</td>
<td></td>
<td>75</td>
<td>0.33***</td>
<td>0.28/0.39</td>
<td>189.94***</td>
</tr>
<tr>
<td>Adolescent/Preschool</td>
<td></td>
<td>6</td>
<td>0.68***</td>
<td>0.48/0.87</td>
<td>7.62</td>
</tr>
<tr>
<td>Adolescent/Childhood</td>
<td></td>
<td>37</td>
<td>0.13***</td>
<td>0.06/0.19</td>
<td>191.72***</td>
</tr>
<tr>
<td>Adolescent/Adolescent</td>
<td></td>
<td>11</td>
<td>0.06*</td>
<td>-0.03/0.16</td>
<td>34.41***</td>
</tr>
<tr>
<td>Type of Prosocial Behavior (by age group)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infant/Infant</td>
<td></td>
<td>4</td>
<td>0.34*</td>
<td>0.03/0.64</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.21*</td>
<td>-0.35/0.51</td>
<td>0.85</td>
</tr>
<tr>
<td>Preschool/Infant</td>
<td></td>
<td>3</td>
<td>0.06</td>
<td>-0.27/0.40</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.38*</td>
<td>0.01/0.75</td>
<td>2.70</td>
</tr>
<tr>
<td>Preschool/Preschool</td>
<td></td>
<td>2</td>
<td>0.21*</td>
<td>-0.35/0.51</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.00</td>
<td>-0.60/0.60</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.27*</td>
<td>-0.07/0.61</td>
<td>2.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.09</td>
<td>-0.12/0.31</td>
<td>9.05</td>
</tr>
<tr>
<td>Preschool/Preschool</td>
<td></td>
<td>1</td>
<td>0.08</td>
<td>-0.35/0.51</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.55*</td>
<td>0.05/1.01</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0.26**</td>
<td>0.07/0.45</td>
<td>10.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>0.22**</td>
<td>0.07/0.36</td>
<td>17.32</td>
</tr>
<tr>
<td>Childhood/Preschool</td>
<td></td>
<td>16</td>
<td>0.64*</td>
<td>0.50/0.79</td>
<td>34.60***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>0.28***</td>
<td>0.09/0.47</td>
<td>9.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>41</td>
<td>0.36*</td>
<td>0.28/0.43</td>
<td>97.88***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>0.16*</td>
<td>0.06/0.26</td>
<td>17.98</td>
</tr>
<tr>
<td>Children/Childhood</td>
<td></td>
<td>17</td>
<td>0.40*</td>
<td>0.29/0.52</td>
<td>21.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.18*</td>
<td>0.07/0.30</td>
<td>15.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>59</td>
<td>0.34*</td>
<td>0.28/0.40</td>
<td>139.86***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.09*</td>
<td>-0.07/0.24</td>
<td>6.00</td>
</tr>
<tr>
<td>Adolescent/Preschool</td>
<td></td>
<td>1</td>
<td>1.39***</td>
<td>0.42/2.37</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.54***</td>
<td>0.16/0.91</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.69***</td>
<td>0.46/0.92</td>
<td>4.98</td>
</tr>
<tr>
<td>Adolescent/Childhood</td>
<td></td>
<td>22</td>
<td>0.02*</td>
<td>-0.06/0.11</td>
<td>105.79***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.33*</td>
<td>-0.06/0.72</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>0.65*</td>
<td>-0.05/0.85</td>
<td>6.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.13*</td>
<td>-0.02/0.27</td>
<td>53.43***</td>
</tr>
<tr>
<td>Adolescent/Adolescent</td>
<td></td>
<td>9</td>
<td>0.01</td>
<td>-0.14/0.12</td>
<td>31.61***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.14*</td>
<td>0.01/0.27</td>
<td>0.22</td>
</tr>
<tr>
<td>Method (by age group)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infant/Infant</td>
<td></td>
<td>8</td>
<td>0.21*</td>
<td>0.01/0.43</td>
<td>4.79</td>
</tr>
<tr>
<td>Observation</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Report</td>
<td></td>
<td>2</td>
<td>0.47*</td>
<td>0.01/0.92</td>
<td>1.84</td>
</tr>
<tr>
<td>Preschool/Infant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other-Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preschool/Preschool</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All codings were from observational studies.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Design (by age group)

<table>
<thead>
<tr>
<th>Category</th>
<th>Observation</th>
<th>Self-Report</th>
<th>Other-Report</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infant/Infant</strong></td>
<td>28</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.23***</td>
<td>-0.09/0.65</td>
<td>31.39</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.28</td>
<td>-0.09/0.65</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>Preschool/Preschool</strong></td>
<td>64</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.41***</td>
<td>0.25/0.65</td>
<td>153.64***</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.45***</td>
<td>0.25/0.65</td>
<td>2.52</td>
</tr>
<tr>
<td><strong>Childhood/Preschool</strong></td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.08**</td>
<td>-0.03/0.19</td>
<td>6.38</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.34**</td>
<td>0.29/0.40</td>
<td>149.10**</td>
</tr>
<tr>
<td><strong>Adolescent/Preschool</strong></td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.71****</td>
<td>0.50/0.92</td>
<td>6.97</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.49**</td>
<td>-0.01/0.99</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Design (by age group)</strong></td>
<td>1.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infant/Infant</td>
<td>6</td>
<td>4</td>
<td>0.80</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.16</td>
<td>0.11/0.71</td>
<td>5.21</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.41**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preschool/Preschool</td>
<td>20</td>
<td>10</td>
<td>21.36</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.19***</td>
<td>0.15/0.55</td>
<td>8.43</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.35***</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Childhood/Preschool</strong></td>
<td>24.06***</td>
<td>62</td>
<td>147.46***</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.12**</td>
<td>0.35/0.48</td>
<td>18.42</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.42***</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Childhood/Childhood</strong></td>
<td>76</td>
<td>9</td>
<td>20.98</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.13**</td>
<td>0.04/0.22</td>
<td></td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.36****</td>
<td>0.30/0.41</td>
<td>161.76***</td>
</tr>
<tr>
<td><strong>Adolescent/Preschool</strong></td>
<td>1.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.00</td>
<td>-0.11/0.11</td>
<td>24.14***</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.29***</td>
<td>0.12/0.28</td>
<td>166.91***</td>
</tr>
<tr>
<td><strong>Adolescent/Adolescent</strong></td>
<td>6.09***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>8</td>
<td>3</td>
<td>2.39</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.17**</td>
<td>0.04/0.29</td>
<td>1.03</td>
</tr>
<tr>
<td><strong>Target (by age group)</strong></td>
<td>5.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infant/Infant</td>
<td>2</td>
<td>1</td>
<td>0.00</td>
</tr>
<tr>
<td>Child</td>
<td>-0.02</td>
<td>-0.37/0.60</td>
<td>0.10</td>
</tr>
<tr>
<td>Adult</td>
<td>0.33**</td>
<td>0.07/0.64</td>
<td>2.39</td>
</tr>
<tr>
<td>Unknown/Unspecified</td>
<td>0.68**</td>
<td>0.13/0.23</td>
<td>0.00</td>
</tr>
<tr>
<td>Preschool/Infant</td>
<td>1.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>9</td>
<td>1</td>
<td>5.19</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.19</td>
<td>0.02/0.37</td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td>0.03</td>
<td>-0.44/0.39</td>
<td>0.00</td>
</tr>
<tr>
<td>Unknown/Unspecified</td>
<td>0.00</td>
<td>-0.59/0.59</td>
<td>0.00</td>
</tr>
<tr>
<td>Preschool/Preschool</td>
<td>0.06</td>
<td>28</td>
<td>31.40</td>
</tr>
<tr>
<td>Child</td>
<td>0.23***</td>
<td>0.12/0.35</td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Unknown/Unspecified</td>
<td>2</td>
<td>0.28</td>
<td>0.08</td>
</tr>
<tr>
<td>Childhood/Preschool</td>
<td>36.75***</td>
<td>61</td>
<td>173.18**</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.39**</td>
<td>0.33/0.46</td>
<td></td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>1.16**</td>
<td>0.76/1.56</td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td>0.12**</td>
<td>0.01/0.22</td>
<td>11.26</td>
</tr>
<tr>
<td>Unknown/Unspecified</td>
<td>0.12**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Childhood/Childhood</td>
<td>13.93***</td>
<td>76</td>
<td>145.12**</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.33***</td>
<td>0.28/0.38</td>
<td></td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.67***</td>
<td>0.34/1.00</td>
<td>1.40</td>
</tr>
<tr>
<td>Adult</td>
<td>0.13**</td>
<td>0.02/0.25</td>
<td>6.71</td>
</tr>
<tr>
<td>Unknown/Unspecified</td>
<td>0.13**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adolescent/Preschool</td>
<td>7.89</td>
<td>5</td>
<td>7.54</td>
</tr>
<tr>
<td>Child</td>
<td>0.69***</td>
<td>0.48/0.89</td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Unknown/Unspecified</td>
<td>1</td>
<td>0.60***</td>
<td>0.00</td>
</tr>
<tr>
<td>Adolescent/Childhood</td>
<td>9.68**</td>
<td>26</td>
<td>129.64***</td>
</tr>
<tr>
<td>Naturalistic/Correlational</td>
<td>0.12**</td>
<td>0.01/0.24</td>
<td>59.72**</td>
</tr>
<tr>
<td>Experimental/Structured</td>
<td>0.12**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td>0.12**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown/Unspecified</td>
<td>0.12**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adolescent/Adolescent</td>
<td>10.71***</td>
<td>5</td>
<td>1.92</td>
</tr>
<tr>
<td>Child</td>
<td>0.18***</td>
<td>0.06/0.29</td>
<td>21.78**</td>
</tr>
<tr>
<td>Adult</td>
<td>-0.14*</td>
<td>-0.31/0.07</td>
<td></td>
</tr>
<tr>
<td>Unknown/Unspecified</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Effect sizes are positive for differences favoring older children. CI = Confidence Interval. *Significance indicates rejection of the hypothesis of homogeneity. Age Classifications: Infants = less than 3 years of age; Preschool = 3 - 6 years of age; Childhood = 7 - 12 years of age; Adolescent = 13 - 17 years of age. **p < .05, ***p < .01, ****p < .005, *****p < .001. Within each coding category for each age comparison grouping, effect sizes with similar alphabetic superscripts are significantly different (post hoc contrasts), ps at least < .05.
Table 2 includes tests of univariate categorical models (analogous to main effects in an analysis of variance), tests of the significance of between-class effects ($Q_h$), tests of the homogeneity of the effect sizes within each class ($Q_{sh}$), the mean weighted effect size for each categorical variable, and the 95% confidence interval. Because the between-class test was significant for age comparison groups (see Table 2), post-hoc comparisons among the mean effects sizes for the different age comparison groups were conducted. All age group comparisons resulted in positive weighted effect sizes (indicating that the older age group evidenced greater prosocial behavior). Further, all comparison effect sizes were significant ($p$s at least $< .05$) except for those in the preschool/infant and adolescent/adult age groups. The greatest effect size was found in comparisons of adolescents/preschoolers, with moderate effects sizes found for comparisons of infant/infant, childhood/preschool, and childhood/childhood groups, and small effect sizes found for preschool/infant, adolescent/childhood, and adolescent/adolescent comparisons.

We next examined how age differences in prosocial behavior varied as a function of study quality variables. Data for the study quality variables broken down by appropriate age groups are presented in Table 2. Variation in the magnitude of age differences in prosocial behavior as a function of the type of prosocial behavior was not significant for the three youngest age group comparisons. Across the remaining age group comparisons, the magnitudes of age differences were relatively constant in size when the type of prosocial behavior was aggregated, sharing, or comforting. In contrast, the magnitude of effect sizes in instrumental helping varied more across these age groups. The magnitudes of effect sizes were relatively high when the type of prosocial behavior was instrumental help for childhood/preschool and childhood/childhood comparisons and relatively low for the adolescent/childhood and adolescent/adolescent comparisons.

The magnitude of the effect size differed significantly by the method of collection (e.g., observation, self-report, other-report) only for childhood/preschool, childhood/childhood, and adolescent/childhood comparison groups (note that effect sizes could not be contrasted for the preschool/infant comparison group because all effect sizes were from observational studies). Specifically, for both the childhood/preschool and childhood/childhood age groups, effect sizes for age differences were significantly higher when measured by observation or self-report than when measured by reports obtained from others (e.g., parents, peers, and/or teachers). For the adolescent/childhood comparisons, effect sizes were significantly smaller when measured by self-report methods than by observational or other-report methods (see Table 2).

A very consistent pattern of findings was obtained when the effect sizes were broken down by the type of design (effect sizes could not be contrasted for preschool/infant and adolescent/preschool comparison groups because all effect sizes were from naturalistic/correlational studies or structured/experimental studies, respectively). Across all remaining age comparison groups, effect sizes were greater in experimental/structured designs than in naturalistic/correlational designs. Finally, the magnitude of the effect size differed significantly by the target of the prosocial behavior, but this was true only for childhood/preschool, childhood/childhood, and adolescent/adolescent groups. In the first two age comparison groups, effect sizes were larger when the target was an adult and lowest when the target was an adult. In contrast, for the adolescent/adolescent comparison, the effect size was greater when the target was a child in comparison to an adult.

There also were age-related differences in study characteristics. For example, instrumental help was relatively unlikely to be used as a measure of prosocial behavior for children under 7 years of age. Moreover, naturalistic/correlational designs were relatively likely to be used with younger children whereas experimental/structured designs were likely to be used with older children. Additionally, adults were likely to be used as targets of children's prosocial behavior at the youngest and oldest age group comparisons whereas children were likely to targets for groups not at the extremes (see Table 2). Thus, age-related differences in prosocial behavior may be a function of differences in study characteristics that vary with the age of the sample. Our next step was to examine this possibility.

Regression analyses. To examine the prediction of age differences in prosocial behavior while controlling for study qualities, we computed multivariate least squares weighted regression analyses (see Table 3). For comparative purposes, we also computed univariate weighted regression analyses (when a predictor was examined in relation to effect size without controlling for other study qualities).

The sample of studies in these analyses is the larger sample that includes studies that compared specific age groups and studies that used correlational analyses ($n = 315$). In these analysis, we used a weighted regression procedure (each effect size was weighted by the inverse of its variance) and included the five continuous predictors (mean age of sample, age of youngest in sample, age of oldest in sample, sample size, and year of publication) and $u - 1$ contrasts (where $u$ is the number of predictors in a category) for each categorical variable (the categorical predictors were dummy coded, see Table 3). Thus, one code (e.g., comforting for type of behavior) is not presented in Table 3 (but was used in the contrast).
Table 3. Tests of Models for Age Differences in Prosocial Behavior.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariate Models</th>
<th>Multivariate Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta</td>
<td>Beta</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>****</td>
</tr>
<tr>
<td>Continuous Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Age of Sample</td>
<td>-.116</td>
<td>-.849</td>
</tr>
<tr>
<td>Age of Youngest in Sample</td>
<td>-.150</td>
<td>-.235</td>
</tr>
<tr>
<td>Age of Oldest in Sample</td>
<td>-.040</td>
<td>.513</td>
</tr>
<tr>
<td>Sample Size</td>
<td>-.240</td>
<td>-.226</td>
</tr>
<tr>
<td>Year of Publication</td>
<td>-.147</td>
<td>-.116</td>
</tr>
<tr>
<td>Categorical Variables (Dummy Coded Contrasts)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of Prosocial Behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrumental Help</td>
<td>-.047</td>
<td>-.050</td>
</tr>
<tr>
<td>Sharing</td>
<td>.192</td>
<td>-.008</td>
</tr>
<tr>
<td>Aggregated Index</td>
<td>-.194</td>
<td>-.132</td>
</tr>
<tr>
<td>Method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Report</td>
<td>-.042</td>
<td>.127</td>
</tr>
<tr>
<td>Other-Report</td>
<td>-.118</td>
<td>.158</td>
</tr>
<tr>
<td>Design</td>
<td>-.292</td>
<td>-.278</td>
</tr>
<tr>
<td>Target of Prosocial Behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child</td>
<td>.140</td>
<td>-.071</td>
</tr>
<tr>
<td>Adult</td>
<td>-.069</td>
<td>.048</td>
</tr>
<tr>
<td>Data Type</td>
<td>.085</td>
<td>.139</td>
</tr>
<tr>
<td>Multiple R</td>
<td>.45</td>
<td>****</td>
</tr>
</tbody>
</table>

Note. Models are weighted least squares regressions calculated with weights equal to the reciprocal of the variance for each effect size. Beta = standardized regression coefficient. n = 315. \( Q_0 = 736.04, p < .0001 \). \( Q_R = 185.74, p < .0001 \). Embryonic dummy coded variables reflect contrasts comparing this variable (coded 1) to all other categories (coded 0). Embryonic Naturalistic/Correlational designs = 1, Structured/Experimental designs = 0. Embryonic Cross-Section Age Comparisons = 1, Correlations with Age = 0. \( p < .05 \), \( p < .01 \), \( p < .005 \), \( p < .001 \).

The univariate tests indicated that four of the continuous variables were significantly related to age differences in prosocial behavior. Effect sizes for age differences were smaller as the mean age of sample increased, as the age of the youngest children in the sample increased, as sample size increased, and in studies published later. For the categorical variables, effect sizes indicative of change with age were larger when the measure of prosocial behavior was sharing and smaller when it was an aggregated index. Additionally, effect sizes were smaller when the data were obtained from other reporters (parents, teachers, peers), when the design of the study was naturalistic/correlational, when the target of the prosocial act was an adult, and when the type of age comparison was correlational rather than specific age group contrasts (see Table 3). Finally, when the target of prosocial behavior was a child or when the type of comparison was a specific age group rather than correlational, effect sizes tended to be larger.

In the multivariate regression analyses, important differences relative to the findings of the univariate analyses sometimes were found. After controlling for study qualities, all four of the continuous variables that were significant in the univariate models remained significant in the multivariate model. In addition, age of oldest children in the sample also was significant; as the age of the oldest children in the sample increased, the magnitude of effect size increased as well (see Table 3).

Although we found significant univariate predictions for sharing and aggregated indexes of prosocial behavior, effect size was not predicted by type of prosocial measure after partialling out other study and sample qualities. In contrast to the univariate analyses, effect sizes were now significantly predicted by self reports (i.e., they were larger when measured by self-reports), but the significant inverse
univariate prediction of effect sizes by other-reports was now a significant positive predictor in the multivariate analysis (a suppression effect which is difficult to interpret, Cohen & Cohen, 1975; see Table 3). Finally, as in the univariate analyses, when the design was naturalistic/correlational, or when the data were from correlational analyses, effect sizes were relatively small.

The findings of our meta-analysis suggest that age differences in prosocial behavior are complex: they differed in magnitude as a function of the specific age comparison made, the way in which prosocial behavior was studied, and the type of age-related analysis reported. However, across all studies and study qualities, we found a significant, positive effect size for age differences in prosocial behavior. Thus, our data support the conclusion that as children get older, prosocial behaviors generally are more likely to occur. This pattern was found for all specific age group comparisons, although there was considerable variation in the magnitude of effect sizes across different age group comparison (see Table 2).

Moreover, a potential source of discrepancy across studies is the fact that preschool children's prosocial activity often occurs in the context of dramatic play (e.g., playing the role of a parent). Such prosocial behavior does not involve responses to real needs of others. Because dramatic play declines during the preschool period (Singer, 1973), there may appear to be a drop in prosocial action during this period. For example, Bar-Tal et al. (1982) found that younger children evidenced greater frequencies of total prosocial behavior when pretend and real prosocial actions were combined. However, older children exhibited more real prosocial behavior than did younger children. Because we focused our meta-analysis on real prosocial behavior only, our data would not be influenced by changes in children's tendencies to role-play prosocial behavior in the context of pretend play.

According to the multivariate regression analysis, the magnitude of the age differences increased as the age span between the youngest and oldest children in the samples increased (as age of the youngest decreased and as the age of the oldest increased, effect sizes tended to get larger; see Table 3). We now turn to a discussion of the possible developmental processes that may contribute to these differences.

Processes Potentially Related to Changes with Age in Prosocial Responding

For some theorists, the primary source of the increase in prosocial and altruistic behavior across age is socio-cognitive development (e.g., Burleson, 1994), including attentional processes (attending to the needs of others), (b) evaluative processes (evaluating behaviors and situations in terms of moral standards), and (c) planning processes (Krebs & Van Hesteren, 1994). In our view, these processes include encompass more than purely socio-cognitive development, and other aspects of responding (e.g., moral emotions, regulatory capacities) may partially account for age-related changes in prosocial behavior.

As noted by Krebs and Van Hesteren (1994) and Hoffman (1982), attention to the needs of others transforms egoistic affect to other-oriented affect, rendering it increasingly altruistic. Throughout infancy and childhood, children develop an increasingly refined understanding of others' emotional states and cognition processes, and are better able to decode other people's emotional cues (Barnett, Darcie, Holland, & Kobasigawa, 1981; see Eisenberg, Murphy, & Shepard, 1997). As is discussed shortly, such perspective taking and related socio-cognitive skills are associated with prosocial responding. Moreover, with age, children are more likely to have the social experience necessary to perceive need in social contexts in which overt cues of distress are ambiguous or subtle (see Pearl, 1985).

Children's abilities to evaluate situational factors and behavioral options also become more complex and probably more accurate with age. For example, children's abilities to evaluate the potential costs and benefits for prosocial behavior become more sophisticated with age (Black, Weinstein, & Tanur, 1980). Younger children appear to weigh costs to the self more than do older children when deciding whether or not to assist others (see Eisenberg, 1986) and are less attuned to the benefits of prosocial behavior (Lourenço, 1990, 1993; Perry, Perry, & Weiss, 1986).

Bar-Tal, Raviv, & Leiser (1980) proposed that children's helping behavior develop in six stages. The first three stages involve helping behaviors that are compliant and in which the child anticipates the gain of material rewards (or the avoidance of punishment). The next two stages involve compliance with social demands and generalized reciprocity. The final stage represents true altruism. Young children are viewed as helping only under certain circumstances, mostly in compliance with external authority.

At these early stages, children evaluate helping primarily in regard to the material and physical costs and benefits. The intermediate stages involve evaluative processes related to shared costs and benefits, and the most advanced stages involve evaluations that focus on the functioning of social systems and psychological integrity (of themselves and others; Krebs & Van Hesteren, 1994).

Bar-Tal and colleagues have found some support for their hypothesized developmental changes in children's motives for helping: older children tend to assist more often than do younger children in contexts in which the effects of compliance and rewards or costs are minimized (Bar-Tal et al., 1980; Raviv, Bar-Tal, & Lewis-Levin, 1980; see Bar-Tal, 1982; Eisenberg, 1986). Although Bar-Tal and colleagues sought to delineate a developmental sequence in prosocial motivation, the data concerning this issue are inconclusive (i.e., it is not clear whether all of their proposed stages actually emerge in the specified order; see Eisenberg, 1986). Moreover, children's reported motives for their prosocial behavior change in ways that generally are consistent with Bar-Tal's stages. Although even preschoolers sometimes give simple other-oriented and pragmatic reasons for their peer-directed prosocial actions (Eisenberg, Lundy, Shell, & Roth, 1985; Eisenberg, Pasternak, Cameron, & Tryon, 1984; Eisenberg-Berg & Neal, 1979), researchers generally have found a decrease with age in self-oriented, hedonistic reasons for helping and an increase in other-oriented, internalized and altruistic motives and reasons for prosocial behavior (e.g., Bar-Tal & Nissim, 1984; Bar-Tal et al., 1980; Ugurel-Semin, 1952; see Bar-Tal, 1982; Eisenberg, 1986). Findings vary with contextual variables (e.g., if an adult is present; see Eisenberg, 1986) and are not always consistent (Hertz-Lazarowitz, 1983; Boehnke et al., 1989); nonetheless, overall the evidence of developmental change in children's motives for assisting...
Age-related changes in children's evaluative processes (as well as personal goals; see Eisenberg, 1986) also are reflected in children's prosocial moral reasoning (i.e., reasoning about moral dilemmas in which one person's needs or wants conflict with those of others in a context in which the role of authorities, laws, rules, punishment, and formal obligations is minimal). Typically individuals are presented with hypothetical moral conflicts (e.g., about helping an injured child rather than going to a social event) and their reasoning about the conflict is elicited. Based on both cross-sectional and longitudinal research, Eisenberg and her colleagues have identified an age-related sequence of children's prosocial reasoning. Preschool and early elementary school children tend to use primarily hedonistic reasoning or needs-oriented (primitive empathic) prosocial reasoning. Hedonistic reasoning decreases sharply in elementary school and increase slightly in adolescence; needs-oriented reasoning increases until mid-childhood and then levels of in use. In elementary school, children's reasoning begins to reflect concern with others' approval and enhancing interpersonal relationships, as well as the desire to behave in stereotypically "good" ways. However, such reasoning (particularly approval reasoning) appears to decline somewhat in high school. Beginning in late elementary school or thereafter, children begin to express reasoning reflecting abstract principles, internalized affective reactions (e.g., guilt or positive affect about the consequences of one's behavior for others or living up to internalized principles and values) and self-reflective sympathy and perspective taking. Thus, although children and adolescents alike sometimes verbalize immature modes of reasoning, children's moral reasoning becomes more abstract, somewhat less self-oriented, and increasingly based on values, moral principles, and moral emotions with age (Carlo, Eisenberg, & Knight, 1992; Carlo, Koller, Eisenberg, Pacheco, & Loguercio, in press; Eisenberg, 1986; Eisenberg, Carlo, Murphy, & Van Court, 1995; Eisenberg, Miller, Shell, McNalley, & Shea, 1991; Eisenberg-Berg, 1979a). As is discussed later, these age-related changes are linked to prosocial behavior; thus, the processes reflected in children's moral reasoning likely play some role in the age-related increase in quantity and quality of prosocial behavior. However, these processes may include age-related changes in goals and values, as well as in sociocognitive skills required for high level moral reasoning (see Eisenberg, 1986; Staub, 1978, 1992).

Although sociocognitive processes may underlie the development of children's prosocial behaviors, engaging in these processes does not ensure that prosocial actions be enacted. Affective and regulatory developments in childhood also may account for age-related changes in prosocial behavior. Eisenberg and Fabes (1992) proposed a model of social behavior based on individuals' tendencies to react emotionally and to regulate emotional and behavioral reactions. Spontaneous prosocial behavior is viewed as most likely to occur in individuals who are optimally regulated (i.e., moderate use of inhibitory control, flexible coping, etc.).

Although Eisenberg and Fabes (1992) considered these qualities to reflect, in part, temperamental differences in individuals, it is also clear that these qualities change with development. For example, with increasing age, children generally are better able to control emotional and behavioral impulses (Block & Block, 1980; Mishel, Shoda, & Rodriguez, 1989). As such, one would expect older children, relative to younger ones, to be more likely to respond optimally in emotionally evocative situations—that is, to respond sympathetically and with prosocial behavior. Support for the hypothesized relations between children's prosocial tendencies and their behavioral and emotional regulation is discussed later.

Developmental changes in both children's emotion regulation and in their sociocognitive skills (e.g., Hoffman, 1982) would be expected to contribute to developmental changes in prosocial behavior by influencing children's tendencies to respond empathically or sympathetically (see Eisenberg, Fabes, Murphy et al., 1994). Lennon and Eisenberg (1987b), in a review of the literature, found that age differences in empathy varied with the specific index of empathy used. In general, self-report of empathy was positively associated with age in preschool and elementary school years. However, findings were incongruent with older children and adolescents. Facial/gestural indices appeared to be either inversely related or unrelated to age in the early school years, perhaps due to increases with age in children's ability to mask their emotions. To further examine this issue, Fabes and Eisenberg (1996) conducted a meta-analysis of age differences in empathy in studies since 1983 and found an overall unweighted effect size of .24 (favoring older children). Moreover, Fabes and Eisenberg found that effect sizes varied significantly by method; they were greater for observational and self-report indices than for facial/gestural or other-report measures (p values at least < .05). These findings are consistent with the view age-related changes in vicarious emotional responding may contribute to changes with age in prosocial responding, particularly in contexts involving distress.

Developmental changes in children's experience-based competencies also affect their ability to engage in prosocial behavior. Peterson (1983a,b) found that when children were specially trained on relevant tasks, age-related increases in helping evaporated. The data in our meta-analysis also suggest that experience-based developmental competencies may contribute to age-related differences in prosocial behavior. For example, age differences in prosocial behavior were relatively pronounced when the index of prosocial behavior was instrumental helping. Older children may provide more direct, instrumental assistance because they are possess greater physical and social competence than do younger children.

In summary, changes in prosocial behavior are complex and are influenced by a variety of methodological factors. Moreover, the precise developmental mechanisms involved in producing these changes are not yet fully explicated and likely involve cognitive, motivational, social, and physical processes. The next wave of research should include studies devoted to identifying when and how age-related changes in the sociocognitive, emotional, and regulatory capabilities change with age and jointly affect prosocial responding.

**Sex Differences in Children's Prosocial Behavior**

Based on stereotypic gender roles, females generally are expected and believed to be more responsive, empathic, and prosocial than are males whereas males are expected to be relatively independent and achievement oriented (Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Parsons & Bales, 1955; Spence,
Helmreich, & Stapp, 1974). Further, cross-cultural work has verified that gender differences in prosocial responding are not limited to only a few cultures and may develop with age. For example, Whiting and Edwards (1973) found that helpfulness and support giving generally were greater for girls than boys across six different cultures, although these differences were significant for older but not younger children.

Despite the prevailing view that females are more prosocial than males, the empirical evidence is equivocal (see Moore & Eisenberg, 1984; Radke-Yarrow, Zahn-Waxler, & Chapman, 1983, for reviews). In fact, Eagly and Crowley (1986) conducted a meta-analysis of sex differences in adults' helping behavior and found that men helped more than women, particularly in situations involving instrumental and chivalrous assistance. Importantly, sex differences in helping were extremely inconsistent across studies and were successfully predicted by various attributes of the studies. A similar meta-analysis has not been conducted with studies that include children. Thus, we now present such a meta-analysis, followed by a more general review and discussion of relevant literature.

A Meta-Analysis of Sex Differences and Children’s Prosocial Behavior

Sample of studies. The same procedures and criteria as those used in the meta-analysis on age differences were used in this meta-analysis. The resulting sample consisted of 259 studies yielding a total of 450 effect sizes (M number of effect sizes per study = 1.74, SD = 1.32).

Variables coded from each study. The following information was recorded from each study: (a) sample size for female and male subjects (Ms = 52.03 and 51.98, SDs = 61.61 and 59.91, respectively; when sample sizes were not broken down by sex of child we assumed equal numbers of males and females), (b) type of index (instrumental helping, sharing/donating, comforting, aggregated prosocial behavior, being kind/considerate), (c) method (observation in natural or experimental contexts, self-report, other-report [peers, parents, teachers]), (d) design (correlational/naturalistic, structured/experimental), (e) target of prosocial behavior (children, adults, unknown/unspecified), (f) mean age of sample (M age = 7.93 years, SD = 3.58, range = 1.2 to 19.5 years), (g) age span of children (age of oldest - age of youngest; M = 2.53 years, SD = 2.17), and (h) year of publication.

Computation and analysis of effect sizes. Once again, the effect size g was computed for each variable and then corrected for bias. The unbiased effect size estimator d was used in the analyses. Positive values for d represent higher levels of a behavior for females than for males. For analysis of prosocial behavior, computation of effect size was based on (a) 64 reports of means and standard deviations, (b) 36 reports of correlations, (c) 72 F tests, (d) 19 t tests, (e) 19 reports of proportions, (f) 29 reports of significance levels, (g) 14 reports of chi square, and (h) 197 cases where the authors reported no significant differences but did not report any data (a d of 0 was given in these cases).

We again elected to be conservative in dealing with nonindependent effect sizes. Thus, in the analysis of sex differences in prosocial behavior we did the following: (a) when more than one effect size was calculated for similar variables we chose to use the median value for this variable and (b) when more than one effect size was presented for different variables from the same study for the same sample we randomly selected one of these.

We also identified potential outliers in both samples by computing a schematic box plot and determining which effect sizes were outside of the inner fence (Glass et al., 1981). Based on these criteria, 12 effect sizes were eliminated from the sample. Thus, final number of effect sizes for the meta-analysis of sex differences in prosocial behavior was 272.

Results. Across all effects, the mean unweighted effect size was positive (indicating a difference favoring females; see Table 4). We now turn to the meta-analytic computations using only the sample of 227 effect sizes (i.e., only one effect size per sample minus outliers).

Table 4. Descriptive Summary of Sex Differences in Prosocial Behavior.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Effect Sizes in Data Base (n = 450)&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>M Unweighted Effect Size</td>
<td>0.18</td>
</tr>
<tr>
<td>95% Confidence Interval (lower/upper)</td>
<td>0.14/0.22</td>
</tr>
<tr>
<td>Median Unweighted Effect Size</td>
<td>0.00</td>
</tr>
<tr>
<td>Effect Sizes Used in Analyses (n = 272)&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>M Unweighted Effect Size</td>
<td>0.18</td>
</tr>
<tr>
<td>95% Confidence Interval (lower/upper)</td>
<td>0.14/0.21</td>
</tr>
<tr>
<td>Median Unweighted Effect Size</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. Positive effect sizes indicate larger difference favoring girls. <sup>a</sup>Total number of effect sizes across and within all studies used. <sup>b</sup>Independent effect sizes excluding outliers.
Table 5 includes the overall effect size for all samples in the meta-analysis and the effect sizes broken down by each categorical coding classification. These effect sizes are weighted by the reciprocal of its variance (d+) and a Fail Safe N (Cooper, 1979) was computed to determine the number of additional studies that would be needed to reverse a conclusion that a significant relation exists at the p < .001 (two-tailed) level. For the overall analysis, it would take 3,271 additional studies with effect sizes of 0 (or totaling 0) to change the significance to p > .001. Even though it is clear that the overall sex difference presented in Table 5 differed from 0 (the value indicating exactly no sex difference), its meaning can be questioned because the significant homogeneity statistic, Q, indicated that the effect sizes in this sample were not consistent across studies. The next step was to disaggregate the effect sizes.

Table 5. Univariate Tests of Categorical Models for Sex Differences in Prosocial Behavior Effect Sizes

<table>
<thead>
<tr>
<th>Variables and Class</th>
<th>Between-Class effect (Qh)</th>
<th>Weighted effect size (d+)</th>
<th>95% CI for d+ (lower/upper)</th>
<th>Homogeneity (QW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>272</td>
<td>0.20****</td>
<td>0.18/0.22</td>
<td>473.78****</td>
</tr>
<tr>
<td>Type of Prosocial Behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrumental Help</td>
<td></td>
<td>62</td>
<td>0.14ab</td>
<td>63.08</td>
</tr>
<tr>
<td>Being Kind/Considerate</td>
<td></td>
<td>9</td>
<td>0.42a,c</td>
<td>16.15</td>
</tr>
<tr>
<td>Comforting</td>
<td></td>
<td>20</td>
<td>0.17c,d</td>
<td>36.88</td>
</tr>
<tr>
<td>Sharing/Donating</td>
<td></td>
<td>117</td>
<td>0.13c,d</td>
<td>146.95****</td>
</tr>
<tr>
<td>Aggregated Index</td>
<td></td>
<td>64</td>
<td>0.31b,c,d</td>
<td>156.22****</td>
</tr>
<tr>
<td>Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td></td>
<td>196</td>
<td>0.13ab</td>
<td>247.67**</td>
</tr>
<tr>
<td>Self-Report</td>
<td></td>
<td>36</td>
<td>0.28a</td>
<td>98.20****</td>
</tr>
<tr>
<td>Other-Report</td>
<td></td>
<td>40</td>
<td>0.33ab</td>
<td>95.30****</td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlational/Naturalistic</td>
<td></td>
<td>120</td>
<td>0.26a</td>
<td>239.77****</td>
</tr>
<tr>
<td>Structured/Experimental</td>
<td></td>
<td>152</td>
<td>0.14a</td>
<td>206.35***</td>
</tr>
<tr>
<td>Target of Prosocial Behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child</td>
<td></td>
<td>190</td>
<td>0.15ab</td>
<td>251.98****</td>
</tr>
<tr>
<td>Adult</td>
<td></td>
<td>23</td>
<td>0.28a</td>
<td>60.69****</td>
</tr>
<tr>
<td>Unknown/Unspecified</td>
<td></td>
<td>59</td>
<td>0.28b</td>
<td>133.84****</td>
</tr>
</tbody>
</table>

*Note. Effect sizes are positive for differences favoring females. All effect sizes were significantly different from zero, all ps < .001. CI = Confidence Interval. Within each category, mean effect sizes with similar alphabetic superscripts are significantly different (posthoc ps at least < .05, see text). *Significance indicates rejection of the hypothesis of homogeneity. **p < .05, ***p < .01, ****p < .005, *****p <.001.

Table 5 contains tests of univariate categorical models and includes the mean weighted effect size for each categorical variable and the 95% confidence interval. Because the between-class tests were significant for each categorical variable, post-hoc comparisons among the mean effect sizes for the classes of variables within each category were computed. For type of prosocial behavior studied, the sex difference was more significant for aggregated indexes or indexes reflecting kindness/considerateness than for indexes reflecting instrumental help, comforting, or sharing. Moreover, sex differences were significantly greater when prosocial responding was measured with self-reports or reports from others than with observational methods. The magnitude of the sex difference was greater in correlational/naturalistic studies than in structured/experimental studies. Finally, sex differences in prosocial behavior were significantly greater when the target was an adult or was unspecified than when the target was another child (see Table 5).
prosocial behavior was sharing/donating and when the target was a child (see Table 6).
Table 6. Univariate and Multivariate Tests of Sex Differences in Prosocial Behavior.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariate Models Beta</th>
<th>Multivariate Model Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Continuous Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Age</td>
<td>.104*</td>
<td>-.011</td>
</tr>
<tr>
<td>Age Span</td>
<td>-.178****</td>
<td>-.123**</td>
</tr>
<tr>
<td>Year of Publication</td>
<td>.213*****</td>
<td>.093</td>
</tr>
<tr>
<td>Sample Size</td>
<td>.221*****</td>
<td>1.288**</td>
</tr>
<tr>
<td><strong>Categorical Variables (Dummy Coded Contrasts)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of Prosocial Behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrumental Help(^a)</td>
<td>-.159****</td>
<td>-.148*</td>
</tr>
<tr>
<td>Being Kind (^a)</td>
<td>.171****</td>
<td>.129</td>
</tr>
<tr>
<td>Sharing (^a)</td>
<td>-.222*****</td>
<td>-.027</td>
</tr>
<tr>
<td>Aggregated Index(^a)</td>
<td>.310****</td>
<td>.128</td>
</tr>
<tr>
<td>Method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other-Report (^a)</td>
<td>.237*****</td>
<td>.201**</td>
</tr>
<tr>
<td>Self-Report (^a)</td>
<td>.169****</td>
<td>.148*</td>
</tr>
<tr>
<td>Design (^b)</td>
<td>.284*****</td>
<td>.031</td>
</tr>
<tr>
<td>Target of Prosocial Behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child (^a)</td>
<td>-.251*****</td>
<td>.087</td>
</tr>
<tr>
<td>Adult (^a)</td>
<td>.044</td>
<td>.117**</td>
</tr>
<tr>
<td>Multiple R</td>
<td>.489****</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Models are weighted least squares regressions calculated with weights equal to the reciprocal of the variance for each effect size. Beta = standardized regression coefficient. \(n = 272.\) \(Q_e = 338.35, p < .001.\) \(Q_e = 124.48, p < .001.\) Dummy coded variables reflect contrasts comparing this variable (coded 1) to all other categories (coded 0). \(^a\)Naturalistic/Correlational designs = 1, Structured/Experimental designs = 0. \(^* p < .05, \** p < .01, \*** p < .005, \**** p < .001.\)

To examine the simultaneous impact of both the continuous and categorical variables, we utilized a multivariate regression procedure in which the categorical variables were dummy-coded (see Table 6). In this regression, sample size (positively) and age span (negatively), but not mean age of sample or year of publication, continued to be significant predictors of sex differences in prosocial behavior when the other variables were controlled.

As in the meta-analysis for age differences in prosocial behavior, contrasts were computed to compare the categorical dummy-coded variables (with \(u - 1\) contrasts within each category) with other codings in the same category while controlling for all other study variables. When the effects of other study qualities were controlled, the category of instrumental help was significantly less predictive of sex differences in prosocial behavior than were other types of prosocial indexes. Being kind, sharing, and the aggregated index of prosocial behavior did not differ from other types of prosocial behaviors once study characteristic were controlled. Sex differences continued to be greater when measured by self- or other-reports. There was no effect of design once the other study qualities were controlled. Finally, when the targets of the prosocial behavior were children or adults, effect sizes were greater. Child targets no longer predicted effect sizes once other study qualities were controlled (see Table 6).

Although the multivariate model accounted for a significant amount of the variance in the magnitude of effect size for sex differences in prosocial behavior (\(R^2 = .24\), the test of model specification (\(Q_e\)) indicated that there was still a significant amount of the variance unexplained (see Table 6). This was due in part to the fact that a zero was entered as the effect size for a large number of studies in which investigators reported only that the sex difference in the prosocial measure was not significant but did not provide specific data. When we repeated the analysis controlling for whether we calculated an effect or not, the amount of variance explained increased to .50.

**Discussion**

The results of our meta-analysis support Eagly and Crowley’s (1986) conclusion that sex differences in adults’ prosocial behavior are inconsistent across studies and vary
as a function of the qualities of the studies. In contrast to Eagly and Crowley, our data indicate that girls tend to be more prosocial than boys. The homogeneity statistic in our meta-analysis indicated that there was a significant amount of the variance unexplained. We now examine review the literature with an eye toward understanding what factors may account for the unexplained variance.

In our meta-analysis, sex differences in prosocial behavior varied with the type of prosocial behavior. Sex differences (favoring girls) were larger for indexes of kindness/considerateness and for the aggregated indexes than for help or sharing in the univariate analyses. However, when study characteristics were controlled, the sex difference was significantly smaller for instrumental helping than for other measures and only marginally greater for kindness/consideration. Aggregated measures, which were a strong predictor in the univariate analyses, did not differ from other measures of prosocial behavior when study characteristics were controlled. The finding that the sex difference was greatest for self-report and other-report data than for observational data and that kindness/consideration and aggregated indexes did not differ significantly from other types of prosocial behavior in the degree of sex difference once study characteristics (including whether self-report, other-report, or observational measures were used) were controlled. Berman (1980) noted that sex differences in children's responsiveness to young children were greatest when responses were unspecified/unknown individuals. Moreover, sex differences tended to be larger when measured in naturalistic/correlational contexts than in structured/experimental contexts. Again, this may have to due with the fact that self- or other-reports are likely to be used in correlational designs. The multivariate analysis failed to reveal a significant effect of design once the other study qualities were controlled.

Because the kindness/consideration and the aggregated indexes often have been measured with self- or other-reports whereas helping and sharing have tended to be measured with observational procedures, the relatively large sex differences in the former global indexes may be partially a function of methodology. Recall that in the meta-analysis, sex differences were greater for self-report and other-report data than for observational data and that kindness/consideration and aggregated indexes did not differ significantly from other types of prosocial behavior in the degree of sex difference once study characteristics were controlled. The finding that the sex difference was particularly interesting because many of the studies in the adult literature in which males help more assess instrumental helping (Eagly & Crowley, 1986).

Sex differences in self- and other-reported prosocial behavior may reflect people's conceptions of what boys and girls are supposed to be like rather than how they actually behave. Parents emphasize prosocial behaviors and politeness more with their daughters than their sons (Power & Parke, 1986; Power & Shanks, 1989). Moreover, peers, parents, and teachers perceive girls as more prosocial than behavioral or self-reported data indicate is actually the case (Bernzweig, Eisenberg, & Fabes, 1993; Bond & Phillips, 1971; Shigetomi, Hartmann, & Gelfand, 1981). Further, parents even attribute girls' actions to inborn factors significantly more often than boys' actions whereas boys' prosocial actions are more likely to be viewed as due to environmental factors (Gretarsson & Gelfand, 1988). These findings are consistent with the view that girls' reputations for prosocial behavior are greater than the actual sex difference.

Sex differences in the literature may also be due, in part, to biases in measures of prosocial behavior. Zarbatany, Hartmann, Gelfand, and Vinciguerra (1985) argued that measures used to evaluate children's prosocial tendencies include a disproportionate number of sex-biased items favoring girls (items pertaining to feminine activities). They found that masculine items (e.g., helping get a cat out of a tree) elicited endorsements for boys and feminine-related and neutral items elicited endorsements for girls. These data support the notion that the sex differences in prosocial behavior are due in part to the items included on measures of prosocial behavior. Consistent with the masculine role and findings for adults (Eagley & Crowley, 1986), boys often may help as much or more than girls in situations in which there is some risk or need for certain types of instrumental activities.

The conditions under which prosocial action is measured also may influence the degree to which sex differences in prosocial behavior are found. In our univariate meta-analyses, sex differences in favoring girls tended to be larger when measured in naturalistic/correlational contexts than in structured/experimental contexts. Again, this may have to due with the fact that self- or other-reports are likely to be used in correlational designs. The multivariate analysis failed to reveal a significant effect of design once the other study qualities were controlled.

Findings in regard to sex differences in empathy and sympathy, like those for prosocial behavior, vary with the method used to assess empathy-related responding. As mentioned previously, Eisenberg and Lennon (1983; Lennon & Eisenberg, 1987a), in a meta-analytic review, found large differences favoring females for self-report measures of empathy, especially questionnaire indices. No gender differences were found when the measure of empathy was either physiological or unobtrusive observations of nonverbal behavior. In more recent work in which sympathy and personal distress were differentiated, similar results were obtained, similar findings, although they occasionally have found weak sex differences in facial reactions (generally favoring females) (see Eisenberg, Martin, & Fabes, 1996: Eisenberg, Fabes, & Miller, et al., 1989). Eisenberg and Lennon suggested that the general pattern of results was due to differences among measures in the degree to which both the intent of the measure was obvious and people could control their responses. Sex differences were greatest when demand characteristics were high (i.e., it was clear what was being assessed) and individuals had conscious control over their responses (i.e., self-report indices were used); gender differences were virtually nonexistent when demand characteristics were subtle and study participants were unlikely to exercise much conscious control over their responding (i.e., physiological indices). Thus, when gender-related stereotypes are activated and people can easily control their responses, they may try to project a socially desirable image to others or to themselves.

Fabes and Eisenberg (1996) conducted a follow-up meta-analysis of empathy data published since Eisenberg and Lennon's (1983) first review and found an overall unweighted effect size (favoring females) of .34. Relatively large effect sizes were found in self-report studies (significantly larger than in the studies involving other methods) and in studies where the targets of the empathic response were unspecified/unknown individuals. Moreover, sex differences tended to be larger with samples of older children. When sex differences were
examined by method, significant sex differences favoring girls were obtained self-report indexes (weighted effect size of .60, \( p < .001 \)) and observational measures (in which a combination of behavioral and facial reactions usually were used, .29, \( p < .001 \)). No sex differences were obtained for nonverbal (facial and physiological measures). Further, the sex difference in self-reported empathy/sympathy increased with mean age of the sample, beta = .24, \( p < .005 \). Sex differences in reported empathy may increase as children become more aware of, and perhaps are more likely to internalize in their self-image, sex-role stereotypes and expectations.

Of related interest, whereas there are no sex differences in prosocial moral reasoning with age, in later elementary school and beyond, girls use more of some relatively sophisticated types of prosocial moral reasoning than do boys whereas boys sometimes verbalize more of less mature types of reasoning (Carlo, Eisenberg, & Knight, 1992; Eisenberg, Carlo et al., 1995; Eisenberg, Miller et al., 1991; Eisenberg, Shell et al., 1987). Moreover, in adolescence, femininity is positively related to internalized prosocial moral reasoning (but also related to hedonistic reasoning for males; Carlo et al., in press). At this point in time, it is unclear the degree to which these sex differences, which generally are relatively weak, are due to real differences in moral reasoning or to differences in the ways in which adolescent males and females view themselves and desire to be viewed by others. However, the finding that children's moral reasoning frequently is related to their prosocial behavior is consistent with the view that children's prosocial moral reasoning does not merely reflect children's desire to reason in a socially acceptable manner.

It should also be noted that sex differences in prosocial behavior were smaller as the age span of the sample increased. Thus, it appears that older and younger children vary in the degree in which they may wish to present themselves (or are perceived to present themselves) in stereotypic ways regarding their prosocial behaviors. With increasing age, sex differences tended to get larger, although this effect was eliminated once other study qualities were controlled (see Table 6). Recall also that Eagly and Crowley (1986) found that men were more prosocial than were women. As such, it appears that sex differences in prosocial behavior may be moderated by developmental processes.

In summary, although girls appear to be more prosocial than boys on prosocial behavior, the issue of sex differences in prosocial responding and their origins is far from resolved. At this time it is difficult to discern the degree to which any difference reflects a difference in moral orientation versus other factors (e.g., self-presentation) and if the sex difference emerges with age (although age was related to effect size in the univariate analysis, there was no effect of age when study characteristics were controlled). In the future, there is a need to better assess the developmental trajectory of any sex differences, to investigate the origins of sex differences in prosocial behavior, and to examine factors that account for individual differences in prosocial responding within boys and girls.

Footnotes

1 A complete list of the studies, codings, and effect sizes is included in Appendix A.
References for Studies Included in Age Differences in Prosocial Behavior Meta-Analysis
(study ID in parentheses at end of reference)


Carlo, G., Koller, S. H., Eisenberg, N., Da Silva, M. S., & Frohlich, C. B. (in press). A cross-national study on the relations among prosocial moral reasoning, gender...
role orientations, and prosocial behaviors. Developmental Psychology. (475).


Eisenberg, N., Fabes, R. A., Karbon, M., Murphy, B. C., Carlo, G., & Wosinski, M. (in press). Relations of school children’s comforting behavior to empathy-related reactions and shyness. Social Development. (180)


Eisenberg, N., & Shell, R. (1986). Prosocial moral judgment and behavior in children: The mediating role of
cost. Personality and Social Psychology Bulletin, 12, 426-433. (37)


Lowe, R., & Ritchey, G. (1973). Relation of altruism to age, social class, and ethnic identity. Psychological Reports, 33, 567-572. (258)


References for Studies Included in Sex Differences in Prosocial Behavior Meta-Analysis (study ID in parentheses at end of reference)


Eisenberg, N., Fabes, R. A., Karbon, M., Murphy, B. C., Carlo, G., & Wosinski, M. (in press). Relations of school children’s comforting behavior to empathy-related reactions and shyness. Social Development. (180)


space to prosocial, antisocial, and social behaviors. 
Journal of Genetic Psychology, 138, 111-121. (366)


Wills, J. B., Feldman, N. S., & Ruble, D. N. (1977). Children’s generosity as influenced by deservedness of
reward and type of recipient. Journal of Educational Psychology, 69, 33-35. (159)


