Older and More Moral?—Age-related Changes in Performance on Piagetian Moral Reasoning Tasks

LORRAINE McDONALD, IAN STUART-HAMILTON

Summary

Three age groups (teenage; 50–65 and 65+) completed a written version of Piaget's moral reasoning tasks to determine the general level of moral reasoning and performance on eight separate categories. Significant age effects were found. Though the nature of the age group difference was mixed, younger subjects consistently achieved lower levels of performance. Results indicate that Piaget overestimated all subjects' level of performance on his tasks and that, although there are differences in different areas of moral reasoning, the relationship with age is not uniform in its direction.

Introduction

Piaget's *The Moral Judgement of the Child* [1] contains the bulk of his research concerning the moral development of children. Piaget proposed that the peak of moral development (representative of the Formal Operational Stage) was reached in mid-adolescence (circa 15 years). Moral reasoning in adults is not discussed. The interpretation of this omission is that level of performance will remain stable throughout life, as stated in relation to other areas [2]. However, it has been suggested that further research into the area of moral reasoning would be likely to produce illuminating results. This was proposed by Hooper, Fitzgerald and Papalia [3], but has hitherto remained uninvestigated, perhaps because Kohlberg's work [4], derived from Piaget's theories but addressing adult moral reasoning, has deflected attention. This is a surprising omission. Given that so much of the research on moral reasoning stems from Piaget's, either in support or in reaction, it is difficult to conceive that our 'knowledge' of moral development in older adults is based upon extrapolation. Principally, it seems to be assumed that because older children and adolescents perform 'correctly' on Piaget's moral tasks, older people will do the same. However, the same argument might be applied to Piagetian cognitive tasks, and several studies have demonstrated that older adults perform less well on such tasks as conservation [5], classification and seriation [6], and animism [7]. Generally, older people are less adept at 'traditional' tests of childhood reasoning [8, 9]. Again, following the suggestion of Hooper et al. [3] that Piaget's tasks may be unsuitable for adults because of 'motivational salience and intrinsic attractiveness', it may be assumed that the tests are too simplistic for older test subjects. No evidence is provided to support this argument, which in any case ignores the generally cryptic nature of psychological studies and the acquiescence of volunteer subjects.

The purpose of this study was to investigate whether older people perform Piagetian moral development tasks in the manner in which their author suggests. If they do, then the tacitly held supposition inherent in the research literature is justified. If not, then one has to question the salience of models based on 'sophisticated' moral reasoning problems such as Kohlberg's, when there are 'flaws' in reasoning on apparently simpler tasks.

Method

Subjects: One hundred and ten subjects (age range 14–78 years, mean 44.518) were divided into three age groups for comparative purposes: 'teenage' (n = 44, mean 14.7, SD 0.451); the 'middle-aged' (n = 37, mean 57.7, SD 5.19); and the 'older' (n = 29, mean 71.1, SD 4.16). The young subjects were volunteers from a mixed (non-selective) comprehensive school in the North West of England. Older subjects were recruited from a senior citizens' activity group or had volunteered after reading a newspaper article. All were resident in the Midlands and were living independently in their own homes. Mean years of education were 12.5 years (SD 3.20) for middle-aged subjects and 11.9 years (SD 3.05) for older subjects. All subjects were unpaid, were free from marked auditory and visual impairment and were capable of reading and writing. Sample size was determined primarily by response to recruitment measures. Rates of volunteering in age groups were judged to be satisfactory.

Procedure and test materials: Subjects were tested in groups
on written forms of Piaget's stories of moral reasoning, an example of which is shown in the Appendix. The test was one of a battery of tests with no time limitations for completion. Subjects were presented with a pamphlet containing 35 stories of moral reasoning and an answer sheet with one question corresponding to each story. A choice of responses was offered after each question (for 28 of the stories two, but on the remaining seven stories, three). Instructions about how to fill in the answer sheet were printed on a cover sheet and were read aloud to the entire group by the experimenter. Anyone unsure about how to proceed was invited to ask for further advice.

The source of the stories was Piaget's *The Moral Judgement of the Child*. The stories were classified into eight categories: carelessness, lying, collective responsibility, stealing, punishment, immanent justice, equality and authority, and justice between children. The stories of carelessness, lying and stealing focused upon appreciation of varying motives that preceded and outcomes that followed incidents involving damage (stories of carelessness) or a wrongdoing (lying and stealing). Those of equality and authority and justice between children involved issues concerning attitudes to authority. Stories concerning punishment examined the type of punishment the subject deemed appropriate for transgressions. Stories concerning collective responsibility required the subject to express a choice, in varying circumstances, of whether a group should be punished for the misdemeanour of an individual. Immanent justice stories described misfortune occurring to a person who had done something wrong but in an unconnected area; the subject was asked whether the misfortune would have occurred if the subject had behaved properly. Each story was an exact copy from the English translation of *The Moral Judgement of the Child* but the order of presentation was randomized to avoid categorical responses. The questions were taken from transcripts of Piaget's interviews with children concerning each story (as recorded in *The Moral Judgement of the Child*). The response choices were classified as representing either a higher or lower level of moral reasoning, as indicated by Piaget in his assessment of children's responses to stories.

For every question the subject was awarded either 1 or 2 points. The higher level answer of the two earned the most points. These points were accumulated to give an overall score for each category, and an overall total score. To ease comparison, each score was re-expressed as being out of a total of 10. Thus the higher the score, the higher the level of moral reasoning indicated.

### Results

The results are displayed in Table I. These were analysed using a two-way mixed ANOVA, the results of which are shown in Table II. There was a significant age group difference and differences between the repeated measures. The interaction was not significant. *Post-hoc* analysis, using the Fisher PLSD test (a multiple comparison test), revealed differences between the teenage and the middle-aged groups on the 'total carelessness' and 'stealing' scales (PLSD = 0.44 and 0.65 respectively; both significant at <0.05) and between the teenage and the middle-aged and teenage and older on the 'justice between' scale (PLSD = 0.56 and 0.59 respectively; both significant at <0.05). No other comparisons were significant.

### Discussion

Overall, significant differences in performance were recorded by subjects in each age group. This result contrasts with Piaget's theory which suggests that uniform responses would be given by subjects across the age range included in the groups in this study. The pattern of subject responses across categories is, however, not uniform in direction. This suggests that moral reasoning may not be a unitary concept; that what Piaget would have described as a 'high' level response on a question in one category does not guarantee a 'high' level response on another category. It may be argued that this is due to a scaling problem associated with test design, but it is remarkable that there are more than a handful of 'mistakes' on a test which should, by accepted wisdom, be passed correctly by all subjects older than mid-adolescence.

<table>
<thead>
<tr>
<th>Age group</th>
<th>All subjects</th>
<th>Teenage</th>
<th>Middle-aged</th>
<th>Older</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carelessness</td>
<td>9.5 (0.52)</td>
<td>9.2 (0.39)</td>
<td>9.8 (1.13)</td>
<td>9.4 (1.54)</td>
</tr>
<tr>
<td>Lying</td>
<td>8.8 (0.48)</td>
<td>8.6 (1.13)</td>
<td>9.0 (1.53)</td>
<td>8.9 (1.60)</td>
</tr>
<tr>
<td>Stealing</td>
<td>8.8 (1.47)</td>
<td>8.4 (1.53)</td>
<td>9.3 (1.57)</td>
<td>8.9 (1.60)</td>
</tr>
<tr>
<td>Punishment</td>
<td>8.52 (1.21)</td>
<td>8.1 (1.04)</td>
<td>8.5 (0.94)</td>
<td>8.36 (1.05)</td>
</tr>
<tr>
<td>Collective responsibility</td>
<td>8.1 (0.92)</td>
<td>8.0 (0.97)</td>
<td>8.2 (0.94)</td>
<td>8.1 (0.80)</td>
</tr>
<tr>
<td>Immanent justice</td>
<td>7.9 (1.37)</td>
<td>7.5 (1.55)</td>
<td>8.0 (1.47)</td>
<td>8.1 (1.71)</td>
</tr>
<tr>
<td>Equality &amp; authority</td>
<td>7.6 (1.88)</td>
<td>7.6 (1.19)</td>
<td>7.5 (0.49)</td>
<td>7.7 (0.67)</td>
</tr>
<tr>
<td>Justice between children</td>
<td>8.7 (1.29)</td>
<td>8.1 (1.15)</td>
<td>8.9 (0.95)</td>
<td>9.1 (1.56)</td>
</tr>
<tr>
<td>Total score</td>
<td>8.4 (1.48)</td>
<td>8.4 (0.52)</td>
<td>8.5 (0.39)</td>
<td>8.6 (0.37)</td>
</tr>
</tbody>
</table>

### Table II. Two-way mixed ANOVA table of responses to each category and total score

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>F test</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teenage/middle-aged/older subjects within groups</td>
<td>2</td>
<td>42.49</td>
<td>21.24</td>
<td>11.27</td>
<td>0.0001</td>
</tr>
<tr>
<td>Repeated measure (B)</td>
<td>8</td>
<td>263.07</td>
<td>32.88</td>
<td>29.96</td>
<td>0.0001</td>
</tr>
<tr>
<td>AB</td>
<td>16</td>
<td>25.49</td>
<td>1.59</td>
<td>1.45</td>
<td>0.1111</td>
</tr>
<tr>
<td>B x subjects within groups</td>
<td>808</td>
<td>886.746</td>
<td>1.097</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Table I.** Mean scores and standard deviation of all subjects and groups on each category and total score.

---

**Table II.** Two-way mixed ANOVA table of responses to each category and total score.
Although Piaget predicted that all of these subjects would give answers indicative of higher levels of moral reasoning, no subject achieved 100% higher level answers. It would appear that the stage-wise deterioration suggested by some authors (e.g. [3]) has not occurred, as overall the older group (66 years plus) recorded more higher level answers than did the middle-aged (50—65 years) who in turn recorded a greater number of higher level responses than the teenage group.

How does one explain these results? The two most direct solutions are that either they reflect an age difference or a cohort difference. If it is an age difference, then the results provide evidence for moral development across the life span, reaching its apex in old age, facilitated by a lifetime of experience. A cohort-difference explanation portrays the younger generation as less moral than the older one, perhaps fulfilling the expectations of some sociological observers that modern society is increasingly violent, selfish and less religious. It is possible that older adults may have scored highly as in terms of age cohort they are closer to Piaget’s generation of subjects tested than are the younger subjects who completed this study. However, this still implies that moral values have changed in some way, hence the different responses recorded on this task. A third explanation is that the young are more pragmatic, whilst the old are more idealistic, even whimsical, in their judgements. The results of the three categories found to be statistically significant seem to justify this suggestion.

No methodological problems, as suggested by Hooper et al. [3], were noted. All subjects willingly completed the tests. Some reported that it was ‘interesting’, ‘thought-provoking’ and ‘entertaining’. There was no reason to suspect that any subject had given anything other than a genuine response.

In summary, there were two major findings in this study. The first was that, on some tasks, the teenagers have a lower level of moral reasoning (as defined by Piaget) than do elderly adults. The second finding was that all groups gave fewer ‘high’ responses than Piaget would have predicted. Little, if any, evidence was produced to suggest that a decline in moral reasoning occurs with increasing age. Differences appear to be of a more specific and subtle nature. The lack of complete uniformity in the results suggests that further more detailed work is required in this area to unravel the problems completely. However, this begs the question of the form this investigation might take. Because Piaget’s clinical method stresses the need for a relatively informal setting, how far can one move the original paradigm on to a statistically more rigorous footing without creating a radically different measure? Whilst there is arguably a justification for presenting some of Piaget’s tasks in a written format to older people, provided the questions remain ostensibly the same, if we start to alter the vehicle by which the questions are posed any further than this, or change the questions themselves, then we encounter the same criticisms which have dogged Piaget’s critics in other fields (e.g. cognitive development)—namely, that we are no longer giving the subjects the same test. What is interesting and important about the findings of this study is that people are manifestly not performing to the level which can be extrapolated from Piaget’s own work. Piaget was incorrect when he predicted that the developmental process attained a certain level and then never looked back. Of perhaps the greatest immediate interest is how widespread this phenomenon is and what aspects of the older person are most strongly linked with this change.

Acknowledgements
This research has been funded by Worcester College of Higher Education. Our gratitude is also extended to our volunteer panels.

References

Authors’ address
Address correspondence to Dr I. Stuart-Hamilton

Received in revised form 3 June 1996

Appendix

Example of Story used in Piagetian Moral Development Task
There was a little boy who disobeyed his mother. He took the scissors one day when he had been told not to. But he put them back in their place before his mother came home and she never noticed anything. The next day he went for a walk and crossed a stream on a little bridge. It gave way, and he fell into the water with a splash. If the boy had not taken the scissors would he still have fallen in the stream?