Conscience in Childhood: Old Questions, New Answers

Nazan Aksan and Grazyna Kochanska
University of Iowa

Although conscience has been the focus of reflection for centuries, fundamental questions regarding its organization have not been fully answered. To address those questions, the authors applied structural equation modeling techniques to longitudinal data comprising multiple behavioral measures of children’s conscience, obtained in parallel fashion at 33 and 45 months. The measures encompassed moral emotion (guilt and empathic distress) and rule-compatible conduct (internalization of maternal prohibitions and requests and of another adult’s rules). Confirmatory factor analyses supported a differentiated view of conscience with 2 latent factors at both ages: Moral Emotion and Rule-Compatible Conduct. The structure of conscience was remarkably stable over time. The coherence between Moral Emotion and Rule-Compatible Conduct factors increased as children grew older.

The construct of conscience has a long history in many human intellectual endeavors, including philosophy, religion, literature, and the arts. Developmental psychologists have long pondered when and how children come to experience moral emotions and to behave according to rules and values with no need for external monitoring (Grusec, 1997; Kagan, in press). As with many developmental topics (Maccoby, 1984), two prominent approaches to children’s conscience have emerged over the years.

One approach is rooted in cognitive development and emphasizes universals in developmental course (Kohlberg, 1969; Piaget, 1932). Researchers who adopt this framework seek to uncover uniform processes and factors that characterize the developmental changes in children’s judgments and reasoning in morally relevant contexts. This historical tradition continues in the work of prominent contemporary social–cognitive theorists (Nucci & Turiel, 1978; Smetana, 1997; Turiel, 1998; Turiel, Killen, & Helwig, 1997). Though expanded to include observations of children’s interactions with peers and parents, this approach preserves the focus on cognitive representations of moral rules. Acquisition of values in arbitrary conventional versus universal moral domains is among the central questions social–cognitive theorists study.

The other approach is rooted in the socialization tradition (Maccoby, 1984; Sears, Rau, & Alpert, 1965). It emphasizes consistency and stability of observed variation in emotion, behavior, and cognition across contexts and time as well as variation in children’s individual trajectories. Researchers who adopt this framework seek to understand the processes and factors that account for variation in children’s increasingly complex emotional, behavioral, and cognitive repertoires in morally relevant contexts.

Although conscience has been the focus of reflection for centuries, several questions remain unanswered. To address those questions, the authors applied structural equation modeling techniques to longitudinal data comprising multiple behavioral measures of children’s conscience, obtained in parallel fashion at 33 and 45 months. The measures encompassed moral emotion (guilt and empathic distress) and rule-compatible conduct (internalization of maternal prohibitions and requests and of another adult’s rules). Confirmatory factor analyses supported a differentiated view of conscience with 2 latent factors at both ages: Moral Emotion and Rule-Compatible Conduct. The structure of conscience was remarkably stable over time. The coherence between Moral Emotion and Rule-Compatible Conduct factors increased as children grew older.

In this study, we adopt an individual differences perspective to address fundamental questions concerning the structure and stability of conscience during the early preschool years. Several emotional and behavioral milestones relevant to children’s conscience emerge during that period. For example, toddlers and young preschoolers acquire the capacity to regulate their conduct and emotion in contextually appropriate ways. Although most of these skills are acquired and honed during daily parent–child interactions in the conventional domain (e.g., “Don’t play with the VCR”), some concern the moral domain (e.g., “Don’t hit your brother”) (Kopp, 1989; Smetana, 1995). Variability in these rudimentary emotional and behavioral capacities, whether in the conventional or moral domain, is both commonly found and critical for future adjustment in morally relevant contexts (Frick et al., 2003).

There is a growing consensus that those emerging components of conscience need to be studied more comprehensively, integrating emotion, conduct, and cognition (Blasi, 1980; Burton, 1984; Dienstbier, Hillman, Lehnhoff, Hilman, & Valkenaar, 1975; Grusec & Goodnow, 1994; Grusec & Kuczynski, 1997; Hoffman, 1970, 1983; Kagan, in press; Radke-Yarrow, Zahn-Waxler, & Chapman, 1983; Thompson, 1998; Turiel, 1998). However, researchers’ understanding of the extent of coherence among early manifestations of conscience remains limited.

Two classic studies were the first to address questions concerning the organization of conscience by examining coherence among its components (Hartshorne & May, 1928–1930; Sears et al., 1965). They asked the following: Are different manifestations of conscience, or character, interrelated? Is there an inner core of
one’s character that is relatively stable and that accounts for one’s relatively coherent behavior and affect in morally relevant contexts? Hartshorne and May (1928–1930) concluded that various behaviors and affects were basically unrelated and thus unpredictable from one context to another in a large group of school-age children. Sears et al. (1965) attempted to integrate the psychoanalytic focus on emotions of guilt with social-learning focus on moral conduct in a small group of preschool-aged children, observed in the context of transgressions. They concluded that children’s emotion and behavior were only modestly related.

Two consecutive investigations have revisited Hartshorne and May’s (1928–1930) original data to address questions of the structure of conscience (Burton, 1963; Rushton, Brainerd, & Pressley, 1983). Burton (1963) used principal-components analysis, and Rushton et al. (1983) applied rules of aggregation. Findings suggested that the early claims of the relative lack of coherence among different aspects of character had been overstated and that in similar contexts, individuals did show a modest degree of consistency. In our own research, we have also often found that moral emotion and conduct are interrelated (Kochanska, Padavich, & Koenig, 1996). Several other studies have yielded variable results, from absent or low consistency to substantial (for reviews, see Eisenberg & Fabes, 1998; Grusec & Lytton, 1988; Radke-Yarrow et al., 1983). The first goal of this study was to examine the factor structure of preschoolers’ emotionality and behavior relevant to conscience.

Another classic question of individual differences concerns the stability of young children’s conscience. Several empirical studies, including our own, suggest significant stability for the specific components of conscience (Grusec & Lytton, 1988; Kochanska, Aksan, & Nichols, 2003; Kochanska, Coy, & Murray, 2001; Kochanska, Gross, Lin, & Nichols, 2002). Moreover, research on origins of conduct problems has convincingly shown that deficient conscience in childhood—callousness, impaired moral emotions, and disrupted internalization of rules of behavior—launches a clear trajectory leading to antisocial personality well into adolescence and adulthood (Frick et al., 2003; Frick & Ellis, 1999; Lykken, 1995; Shaw & Wimslow, 1997). Some prosocial dispositions also have been found to be relatively stable from the preschool age to young adulthood (Eisenberg et al., 1999, 2003).

Collectively, those studies speak to rank-order stability, on a variable-by-variable basis, either in specific components of conscience that are measured similarly over short periods of time (e.g., Kochanska et al., 2001) or in functional continuity of early differences over long periods of time (e.g., Frick et al., 2003). As Caspi (1998) has argued, however, stability in individual differences can be studied at the level of latent structures.

Structural stability refers to the degree of invariance that governs the relation between the observable, fallible measures and the unobservable latent constructs over time. For example, scores on various facets of personality inventories typically reflect the familiar latent Big Five factors from late adolescence into adulthood (Costa & McCrae, 1992). Stability in factor structure, sometimes referred to as measurement equivalence (Baltes, Reese, & Nesselroade, 1977), is routinely incorporated into research on adolescent and adult personality as well as other aging research (e.g., Byrne, Shavelson, & Mutheu, 1989; Hertzog & Schaie, 1986; Marsh & Grayson, 1994). With the exception of maternally reported temperament (Pedlow,anson, Prior, & Oberklaid, 1993), measurement equivalence has very rarely been examined in observational studies with young children. The second goal of this study was to examine structural stability in young preschoolers’ conscience.

The Current Study

We examined the latent structure of conscience and its stability in a group of preschoolers at 33 and 45 months, using observational measures. At 33 and 45 months, we conducted two laboratory sessions that encompassed naturalistic yet carefully scripted paradigms. At both ages, we assessed the children’s guilt following transgressions and their empathy to an adult who simulated distress. In addition, we assessed children’s rule-compatible conduct in the absence of surveillance, including internalization of maternal prohibitions, maternal requests, and another adult’s rules.

Theoretically driven questions pertinent to the factor structure of fallible, observed measures and stability in factor structure over time are best addressed with structural equations modeling (SEM) techniques (Bollen, 1989; Marsh & Grayson, 1994). Unlike the traditional correlational strategies that rely heavily on arbitrary criteria, SEM techniques can formally test the viability of different models of conscience organization as well as its stability over time.

We applied modern SEM techniques to test two alternative views of conscience organization and structure. The simplest model is consistent with the psychoanalytic tradition in that it advocates a monolithic internal structure like superego that organizes emotion and conduct. We refer to this model as the uniform model. The alternative model is consistent with the social-learning tradition in that it advocates contextual specificity (Mischel, 1968) or the distinctiveness of emotion and conduct. We refer to this model as the differentiated model. Those alternative views are depicted in the panels of Figure 1. Figure 1A depicts the uniform model, and Figure 1B depicts the differentiated model of conscience. Those two models were tested separately at 33 and 45 months.

We also applied SEM techniques to examine the extent of stability in the factor structure of conscience from 33 to 45 months. When the configuration of various components of conscience is similar over time, we can proceed to even more interesting developmental questions. For example, we can ask whether the relations of various component scales to their respective underlying latent construct remain invariant over time. Longitudinal confirmatory factor analyses are optimally suited to address this question of equality in factor loadings (Caspi, 1998). Equality in factor loadings over time increases our confidence that the purported latent construct(s) retains its meaning over time (Kaplan, 2000; Marsh & Grayson, 1994; Meredith, 1964). When the hypothesis of equality in factor loadings cannot be rejected, we can further proceed to ask questions of invariance in unique variances of component scales. This strategy essentially involves conducting graded tests of stability in factor structure, often referred to as metric invariance. In summary, we addressed two questions. First, we examined the organization of two aspects of conscience, emotion and conduct, by testing its factor structure at 33 and 45 months. Second, we examined the stability of that factor structure over time.
**Method**

**Participants**

One hundred twelve families with infants (mostly White; 97% of mothers, 92% of fathers) volunteered for the study when the infants were 9 months old. Distribution of education, for mothers and fathers, respectively, was as follows: high school education (26% and 32%), associate degrees (15% and 10%), college degrees (14% and 36%), and some postgraduate education (18% and 22%). Annual family income distribution was as follows: 40% made less than $40,000, 23% made between $40,000 and $60,000, and 25% made more than $60,000 (see Kochanska, Coy, Tjebkes, & Husarek, 1998).

![Diagram of Uniform and Differentiated Models of Conscience](image)

*Figure 1.* The uniform (A) and the differentiated (B) models of the latent organization of conscience. The scales of the latent variables were defined by fixing their variances to the value of 1. M = mother; E = experimenter.

**Overview of Measures and Data Reduction**

In this study, we use data from assessments at 33 ($N = 104$; 52 girls) and 45 ($N = 101$; 49 girls) months, when parallel measures of children’s emotion and conduct were collected. At both ages, children and mothers were observed in two 2–3-hr-long laboratory sessions conducted by a young female experimenter.

The measures encompassed moral emotions (guilt and empathic distress) and rule-compatible conduct (internalization of maternal prohibitions and requests and of the experimenter’s rules). All measures were behavioral and coded from videotapes. Reliability of coding was typically established on 15% of cases. Multiple teams of coders were involved. Coders periodically realigned to prevent drift. The value of .61 for kappas (a substantial value; Landis & Koch, 1977) was adopted as minimal. Coding systems are...
summarized briefly, and readers are referred to earlier publications for greater detail.

We adopted different data reduction procedures than in earlier publications. Because multivariate normality is a critical assumption of SEM techniques, it is important to screen component scales for elevations in skewness and/or kurtosis (Bollen, 1989). Whenever raw skewness or kurtosis values were 2.0 or above, we applied square root transformations. In addition, we used linear transformations (divisions by appropriate constants) to bring component scales to a common metric (ranging from 0 to 1) to weigh each equally in paradigm-specific composites. Both the linear and nonlinear transformations were carried out in parallel at both ages to render metric invariance tests meaningful. Those transformations are explicitly described.

**Measures**

**Guilt Following Transgressions**

**Procedure.** Extensive empirical details are in Kochanska et al. (2002). At each age, we administered carefully scripted mishap paradigms, in which the child was led to believe that he or she had damaged the experimenter’s special, very valued objects. Afterward, the experimenter reassured the child that no harm was done by producing a fixed object (an exact but whole replica).

**Coding.** Gaze avoidance (brief or long) and various indicators of bodily tension (squirming, backing away, head down, hunched shoulders, hugging self, covering face with hands) were coded on a presence–absence basis in 5-s segments. Additionally, negative and positive affect (strong and mild) were coded in 60-s intervals. More than one category could be credited in any segment. Kappas ranged from .84 to .93 for gaze avoidance, from .62 to .77 for bodily signs of tension, and from .60 to .77 for affect.

**Data reduction.** We tallied all instances of gaze avoidance, bodily tension, negative affect, and positive affect after weighing (i.e., doubling) occurrences of longer or stronger expressions. Square root functions succeeded in reducing skewness–kurtosis, and linear transformations were used to bring all component scales to range from 0 to 1. After reversing positive affect scores, we averaged all four measures into a Guilt composite at each age. The summary statistics were as follows: $M = .05, SD = .15$, and $\alpha = .72$ at 33 months and $M = .04, SD = .14$, and $\alpha = .68$ at 45 months.

**Empathic Distress**

**Procedure.** At both 33 and 45 months, we administered simulated distress paradigms adapted from Zahn-Waxler and colleagues (Zahn-Waxler, Robinson, & Emde, 1992). At both ages, the experimenter dropped a large box filled with neatly sorted objects (e.g., paper clips) on her foot and feigned both physical pain and distress over the spilled clutter for 30 s.

**Coding.** Gaze avoidance, various indicators of bodily tension (squirming, hunched shoulders), aggression toward the experimenter, discontinuing ongoing play, sad facial expressions, and neutral–happy facial expressions were coded on a presence–absence basis in 5-s intervals. More than one type of category could be credited in a 5-s segment. Kappas ranged from .94 to .95 at 33 months and from .61 to .81 at 45 months for all dimensions.

**Data reduction.** We first computed relative frequency scores for those categories. After reversing aggression and neutral–happy expressions, we summed all dimensions into an Empathic Distress composite at each age. There was no evidence of significant skewness–kurtosis at either age. The descriptive statistics were as follows: $M = .16, SD = .20$, and $\alpha = .72$ at 33 months and $M = -.02, SD = .16$, and $\alpha = .55$ at 45 months.

**Internalization of Maternal Prohibition**

**Procedures.** The paradigms were parallel at 33 and 45 months (details are in Kochanska et al., 2001). There was a low shelf with very attractive objects in the laboratory living room. The experimenter asked the mother to treat those as off limits for the child and to enforce the prohibition throughout the sessions. At the end of the second session, the child was observed alone for 8 min with those objects. The mother left to the adjoining room, having reminded the child about the prohibition and having asked the child to engage in a dull sorting activity, set up in front of the shelf.

**Coding.** Children’s sorting activity assigned by mother, deviation (full-blown play), and other permissible activity were coded on a presence–absence basis in 5-s intervals as part of a set of mutually exclusive and exhaustive codes (details are in Kochanska et al., 2001). Latency to deviate was also coded. Kappas were .96 and .92 at 33 and 45 months, respectively.

**Data reduction.** We computed relative frequency scores for each of those behaviors and successfully transformed them to reduce skewness–kurtosis. After using linear transformations on the latency to deviate measure, we averaged sorting activity, reversed deviation, and other activity measures into a composite of Internalization of Maternal Prohibition at each age. The summary statistics were as follows: $M = .47, SD = .20$, and $\alpha = .77$ at 33 months and $M = .52, SD = .19$, and $\alpha = .71$ at 45 months.

**Internalization of Maternal Request**

**Procedures.** After the mother–child toy cleanup during each laboratory session (two at 33 and 45 months each), the mother asked the child to finish the chore by himself–herself and left to the other room. The child was then alone for 5 min (with enough toys left to be cleaned up).

**Coding.** Children’s eager cleanup and playing with toys were coded on a presence–absence basis in 10-s intervals as part of a set of mutually exclusive and exhaustive codes. In addition, coders noted the number of segments that elapsed from the beginning of the task to completion of cleanup. Kappas were .85 at 33 months and .96 at 45 months.

**Data reduction.** We computed the relative frequency of eager cleanup and playing with toys and the percentage of segments that elapsed from beginning to end of cleanup (time to task completion). Only the latter scores displayed skewness–kurtosis. Those were successfully transformed with square root functions. We averaged all three measures into a composite of Internalization of Maternal Request at each age after reversing play scores. The summary statistics were as follows: $M = .03, SD = .22$, and $\alpha = .62$ at 33 months and $M = .06, SD = .23$, and $\alpha = .74$ at 45 months.

**Internalization of the Experimenter’s Rules**

**Procedures.** The same paradigm was used at both ages. At each time, the child was observed alone for 3 min, playing a game whose rules made it practically impossible to win. The experimenter introduced the rules during a friendly but serious talk with the child, asking the child not to cheat, explaining the importance of honesty, and making sure the child understood the rules. The child was then promised an attractive prize for winning. After 3 min, the experimenter returned to “discover” that she had used the wrong rules; she modified the rules and let the child play again to win a prize. The game involved guessing the animals hidden under a cloth, without looking, and using only a single finger to touch the cloth (Kochanska, Forman, Aksan, & Dunbar, 2005).

**Coding.** Children’s various legal behaviors (e.g., touching with finger tip) and illegal behaviors (e.g., peeking under cloth, using more than one finger) were coded on a presence–absence basis in 3-s intervals. More than one behavioral category could be credited in a 3-s segment. In addition, latencies to various illegal behaviors were noted. Kappas were between .92 and .90 at 33 months and .84 at 45 months.

**Data reduction.** The sum of various legal and reversed illegal behaviors and the average of latencies to various illegal behaviors displayed skewness and kurtosis. All three measures were successfully transformed with square root functions and transformed to range from 0 to 1. We then
averaged all three measures into a composite of Internalization of Experimenter’s Rule at each age. The summary statistics were as follows: \( M = .54, SD = .19 \), and \( \alpha = .71 \) at 33 months and \( M = .57, SD = .18 \), and \( \alpha = .77 \) at 45 months.

**Results**

**Overview**

We first examined the intercorrelations among various components of conscience. We then conducted SEM analyses to address the following two questions. First, we examined the factor structure of conscience components separately at 33 and 45 months. Second, we examined the degree of stability in that factor structure across time.

Specifically, we conducted confirmatory factor analysis to address the question of whether the various components of children’s rule-compatible conduct and emotion should be viewed as indicators of a broadly construed latent construct of conscience or as correlated but distinct aspects of children’s characteristics. Those models were examined separately at 33 and 45 months by comparing two nested models. Finally, we conducted graded tests of stability in the factor structure of conscience from 33 to 45 months.

All model-fitting analyses were conducted with LISREL 8 on variances–covariances (Jöreskog & Sörbom, 2001). Although transformations were successful in reducing or eliminating skewness and kurtosis, multivariate normality cannot be assumed, especially given the small sample size. Thus, we computed asymptotic covariance matrices and used Satorra–Bentler scaling in overall model chi-squares, denoted as \( \chi^2_{SB} \) (Satorra & Bentler, 2001). Furthermore, substantive inferences were not altered by the minimal differences observed in robust maximum-likelihood versus normal theory maximum-likelihood estimation.

All model comparisons involved nested models. Nested model comparisons constitute exact tests of the specific implied constraints in a target model relative to the comparison model. When the exact tests yield a significant difference in model chi-squares, the specific constraint(s) of the more restrictive model can be rejected (Bentler & Bonett, 1980). In addition, we relied on several fit indices. Goodness-of-fit indices together with the nonsignificance of the overall model chi-square allow us to evaluate whether the various assumptions and constraints of the substantive models provide an adequate explanation of the observed phenomena. Such indices allow researchers to gauge the adequacy of the substantive models in isolation.

There are several ways to classify various model fit indices. Here, we adopted Kaplan’s (2000) framework and chose one index from each of three classes of alternatives to model chi-square. Those indices were as follows: root-mean-square error of approximation (RMSEA) and its 90% confidence interval (CI), expected cross-validation index (ECVI) and its 90% CI, and the comparative fit index (CFI). When the model is acceptable in the population of interest, we expect RMSEA to range from .05 to .08 or less and CFI to take on values of .90 and higher. The values for the ECVI are evaluated relative to the value this index takes for the saturated model, which necessarily has a perfect fit. When the value of the ECVI in the target model is lower relative to its value for the saturated model, we have greater confidence that the results would hold in an independent sample of the same size.

**Organization of Conscience**

**Correlations Among Components of Conscience**

Table 1 shows the correlations among all composite measures of conscience. All correlations were meaningful and in the expected direction. Children’s guilt and empathic distress were positively correlated at both ages. For the most part, children’s internalized conduct was significantly and positively correlated across various contexts including maternal prohibitions, requests, and another adult’s rules. Furthermore, all components of conscience showed significant rank-order stability from 33 to 45 months.

**Model-Fitting Analyses**

We next addressed the question of whether children’s moral emotionality (guilt following transgressions, empathic distress in simulated distress contexts) is best viewed as a distinct character-

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>At 33 months</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Guilt</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Empathic distress</td>
<td>.44**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Internalization of M prohibition</td>
<td>.10</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Internalization of M request</td>
<td>.16</td>
<td>.17</td>
<td>.32**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Internalization of E rule</td>
<td>.10</td>
<td>.15</td>
<td>.20**</td>
<td>.21*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>At 45 months</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Guilt</td>
<td>.37**</td>
<td>.42**</td>
<td>.24*</td>
<td>.26**</td>
<td>.24*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Empathic distress</td>
<td>.12</td>
<td>.28**</td>
<td>.32**</td>
<td>.15</td>
<td>.54**</td>
<td>.47**</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Internalization of M prohibition</td>
<td>.15</td>
<td>.20*</td>
<td>.54**</td>
<td>.29**</td>
<td>.21*</td>
<td>.27**</td>
<td>.31**</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Internalization of M request</td>
<td>.04</td>
<td>.13</td>
<td>.16</td>
<td>.26**</td>
<td>—.08</td>
<td>.10</td>
<td>.14</td>
<td>.18</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>10. Internalization of E rule</td>
<td>.20*</td>
<td>.24*</td>
<td>.27**</td>
<td>.12</td>
<td>.32**</td>
<td>.31**</td>
<td>.31**</td>
<td>.40**</td>
<td>.02</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note. N varies from 96 to 104. M = mother; E = experimenter.*

* * p < .05. ** p < .01 or better.
We estimated two nested models at each age. To that end, we considered a model that assumed a single, uniform factor of conscience, with all five components of conscience as its indicators. We then considered another model that assumed that children’s guilt and empathic distress formed a correlated but distinct latent dimension from children’s rule-compatible conduct (see Figure 1). These models were parameterized so that the variances of the latent factors were fixed to unity to define their respective scales. Table 2 presents the overall model chi-square, the 90% CI for RMSEA and ECVI, and the CFI values for each model.

**Table 2**

<table>
<thead>
<tr>
<th>Model</th>
<th>At 33 months</th>
<th>At 45 months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2_{SB}$</td>
<td>$df$</td>
</tr>
<tr>
<td>Uniform model</td>
<td>21.15</td>
<td>5</td>
</tr>
<tr>
<td>Differentiated model</td>
<td>4.25</td>
<td>4</td>
</tr>
<tr>
<td>Uniform model</td>
<td>8.83</td>
<td>5</td>
</tr>
<tr>
<td>Differentiated model</td>
<td>3.21</td>
<td>4</td>
</tr>
</tbody>
</table>

**Note.** Listwise $N = 100$ at 33 months and 99 at 45 months. CI = confidence interval; RMSEA = root-mean-square error of approximation; ECVI = expected cross-validation index.

We began examining the extent of stability in factor structure by fitting the differentiated model to 33- and 45-month data simultaneously. The purpose of these initial models was to obtain a baseline comparison model against which to evaluate the restrictions implied in tests of factorial stability over time. These models were parameterized so that the guilt measures served as the index indicator for the Moral Emotion latent factor, and the internalization of the maternal prohibition measures served as the index indicator for the Rule-Compatible Conduct factors.

The fit of the initial baseline model was adequate, $\chi^2_{SB}(29) = 34.93, p = .21, \text{CFI} = .98$; the point estimates were RMSEA = .06 and ECVI = .95. The 90% CIs were .00−.10 for RMSEA and .86−1.17 for the ECVI with a saturated model value of 1.17. Despite adequate overall fit, we examined this model for evidence of component misfit, which could arise from omitted correlations among the error terms of specific indicators over time, autocorrelations. Not surprisingly, examination of the modification indices revealed one significant autocorrelation for the internalization of maternal prohibition between 33 and 45 months.

Such autocorrelations are routinely encountered in longitudinal data, and their omission can compromise the estimates of factor correlations, factor loadings, and unique variances, in effect jeopardizing inferences on stability in factor structure (Marsh & Gray-White, 1987).
son, 1994; Sörbom, 1975). Therefore, we freed the autocorrelation for the internalization of maternal prohibition indicator, and the overall model fit improved, $\Delta \chi^2_{\text{SB}}(1) = 7.16, p < .05$. The maximum-likelihood estimates, associated standard errors, factor correlations, and squared correlations for each of the indicators along with the overall fit statistics for this model are provided in Figure 2.

**Factorial invariance.** We next proceeded with the graded tests of stability in factor structure over time. We first tested invariance in the factor loadings of the Moral Emotion and Rule-Compatible Conduct factors. In other words, we reestimated the baseline model depicted in Figure 2, with equality constraints on factor loadings over time. The hypothesis of equality in factor loadings could not be rejected, $\Delta \chi^2_{\text{SB}}(3) = 6.89, \text{ns}$, supporting invariance in factor loadings from 33 to 45 months for both Moral Emotion and Rule-Compatible Conduct factors. We next tested invariance in the unique variances of each indicator over time. The hypothesis of equality in both factor loadings and unique variances could not be rejected, $\Delta \chi^2_{\text{SB}}(5) = 7.47, \text{ns}$, indicating invariance in unique variances of the indicators over time.

**Summary.** Collectively, these analyses showed a remarkable degree of stability in the factor structure of conscience from 33 to 45 months. The squared multiple correlations generally ranged from fair to moderate for both factors at both ages, indicating significant sources of unique variance in each indicator not shared by their common factors.

**Discussion**

Although the questions addressed in this study are old ones, our approach is new. Our multitrait and multiassessment longitudinal study produced a robust behavioral database that included moral emotions and several forms of rule-compatible conduct. The contemporary data-analytic strategies allowed us to theorize about directly unobservable constructs that underlie those data. The study makes dual contributions: It advances researchers’ substantive knowledge of early conscience and illustrates the usefulness of new analytical approaches to this classic issue.

Over the last 2 decades, it has become broadly recognized that even toddlers and young preschoolers show a rich repertoire of...
moral emotions and a variety of behaviors that reflect fledgling internalization of rules and values (Emde, Biringen, Clyman, & Oppenheim, 1991; Kagan, 1981; Kochanska, 1994; Radke-Yarrow et al., 1983). A growing body of research, consistent with common sense parental wisdom, has indicated that those early forms of conscience have some degree of stability and that they are an important fundamental component of future adjustment.

**Organization of Conscience**

We revisited the long-standing question concerning the latent organization of conscience. Is its organization uniform? This implies that there exists a latent trait that determines a broad range of observed moral affects and forms of conduct. Or is its organization differentiated in nature? This implies that conscience is composed of distinct but related traits of moral emotion and rule-compatible conduct.

Our findings suggest that the differentiated view of conscience, which depicts moral emotion and rule-compatible conduct as related yet nevertheless distinct constructs, provided a better fit to the data during the preschool years. Those two respective latent factors, Moral Emotion and Rule-Compatible Conduct, were at least moderately coherent. This pattern of findings has implications for our ability to predict, for individual children, moral emotion from rule-compatible conduct and vice versa. Children who express guilt on transgressions and empathic distress in response to another’s distress are also likely to display internalized conduct in the absence of surveillance.

Note that the psychoanalytic theories adopted a model of morality that assumes the existence of an intrapsychic, higher-order entity consistent with the uniform model, whereas the social learning theories posited processes consistent with the differentiated model of conscience. Thus, our findings appear to support the predictions of the social-learning theories. However, we also obtained evidence of increasing coherence between children’s rule-compatible conduct and moral emotion from 33 to 45 months. In fact, the uniform view of conscience showed adequate fit to the data at 45 months. Evidence of greater coherence over time may suggest that a stable inner core of conscience may be emerging toward the latter half of the preschool years, consistent, in fact, with the view of superego in the psychoanalytic theory.

**Stability in the Organization of Conscience**

We also examined the degree of stability in the factor structure of conscience. As summarized earlier, in most developmental research with children, stability is often understood in terms of rank-order stability on a variable-by-variable basis. Examination of structural stability has been advocated as essential for progress in longitudinal research on personality development (Caspi, 1998).

We found evidence of a remarkable degree of stability in the latent organization of early conscience. For example, the magnitude of the relations of each specific indicator to their respective latent factors and the unique variances of specific components of conscience (e.g., internalization of the experimenter’s rule) not shared by their common factors remained invariant from 33 to 45 months. Such extensive stability in factor structure increases our confidence that various measures of children’s moral emotion and rule-compatible conduct are organized similarly during the preschool years.

An important caveat concerning structural stability is that we examined it when measures from all five paradigms were available. Some of those paradigms are not developmentally viable at younger ages (e.g., cheating games), whereas others show ceiling effects at later ages (e.g., internalization of maternal prohibition). In other words, the remarkable degree of stability in the factor structure applies to a relatively short developmental window. Therefore, we need to be cautious in generalizing those findings to a broader age range.

**Implications and Future Directions**

Our findings suggested that moral emotion and rule-compatible conduct become more coherent during the course of the preschool years. Future studies need to examine the factors and the mechanisms responsible for the growing coherence among early aspects of conscience. Greater coherence over time could simply reflect a learning process whereby repeated exposure to similar events builds a coalescing inertia in response styles. Those styles eventually become solidified and entrenched in behavioral patterns consistent with a uniform model of conscience. Alternatively, greater coherence over time could reflect the role of an emerging sense of self, an agency that may serve to actively organize the initially disparate components into a coherent system (Harter, 1998; Kochanska, 2002).

It was important that the latent factors generally explained less than half of the systematic variance in their respective indicators. If those variances were high, we would have expected antecedents of various moral emotions and various forms of conduct to overlap. Because those variances were low, however, we would only expect partial overlap among their respective predictors. Future studies need to document both the distinct and common sources of influence among various indicators of conduct and moral emotion.

Findings also showed that the correlation between internalization of maternal request in the cleanup context and other aspects of rule-compatible conduct diminished from 33 to 45 months. In fact, internalization of maternal request did not load significantly on the rule-compatible conduct factor at 45 months. This pattern of diminishing coherence over time for children’s behavior in the cleanup context has been repeatedly observed in other studies (Kochanska, Aksan, & Koenig, 1995; Kochanska, Coy, & Murray, 2001).

We believe this pattern may reflect differences in task demands of the cleanup context, which requires the maintenance of a boring activity, and the task demands of prohibition paradigms (e.g., prohibited toys and cheating games). The cleanup tasks have dual demands. They require both restraint from engaging in alternative pleasant activities and a sustained engagement in a boring, mundane activity. In contrast, the prohibition paradigms require only restraint from certain activities. Future studies need to be designed to better understand the interplay of task demands in relation to acquisition of internalized conduct across various contexts.

Future research on emerging morality would greatly benefit from expanded application of SEM techniques. For example, latent growth curve modeling techniques (e.g., Burchinal & Appelbaum, 1991; Hussong, Curran, & Chassin, 1998) or methods that permit
the profiling of children’s rank order across components of conscience both within and across time (e.g., Cairns, Bergman, & Kagan, 1998; Cate, Huston, & Nesselroade, 1986; Nagin, 1999) could elucidate patterns of developmental change not evident from the analyses we conducted in this study. Those techniques may be particularly suitable for understanding the nature of events that impact adaptive and maladaptive pathways critical to the acquisition of societal values and norms (Cairns et al., 1998; Caspi, 1998).

Limitations

This study has several limitations that are a source of caution. First, we did not address the third classic component of conscience, moral cognition. Second, all our measures were behavioral. Future studies need to use both questionnaire and observed measures to permit a fully multitrait–multimethod approach to conscience organization. Third, it is possible, though not likely, that our measures of empathic distress reflect personal distress rather than other-oriented sympathy. For example, children’s sad facial expressions and askance gazes were generally directed either at the experimenter or the spilled box contents, consistent with other-oriented sympathy, rather than directed away from the situation indicating withdrawal, consistent with personal distress. However, we cannot claim with certainty that our measures did not tap personal distress at all.

Fourth, our sample size is small for SEM methods (Quintana & Maxwell, 1999). Social developmentalists face a difficult choice between a wish to heed appeals to increasingly use SEM techniques, using rich observational data (Collins, Maccoby, Steinberg, Hetherington, & Bornstein, 2000; McArdle, 1991), and a wish to observe rigorous sample size requirements those techniques impose (Quintana & Maxwell, 1999). However, the confidence intervals for the RMSEA were generally tight for the model in Figure 2, and we did not encounter any minimization difficulties or other indications of empirical underidentification. Thus, we believe our basic inferences concerning the factor structure of conscience were not compromised. Finally, our measures tapped rule-compatible conduct more consistent with norms in the conventional than moral domain (Smetana, 1997; Turiel, 1998). Nevertheless, we believe the findings speak to the organization of morality early in life. Several findings are central to that thesis.

Children acquire the capacity to regulate their conduct and emotions in contextually appropriate ways throughout the 2nd and 3rd years of life (Emde et al., 1991; Hoffman, 1983; Kopp, 1989; Thompson, 1998). Evidence shows that those fledgling capacities predict disordered conduct in the school years (e.g., Frick et al., 2003; Kochanska et al., 2003). For example, measures of rule-compatible conduct used in this study, sampled from the conventional domain, predicted teacher ratings of children’s aggressive, prosocial behaviors in the expected direction in early school years (Kochanska et al., 2003).

Few questions in social development are as important as how young children gradually internalize values and norms of their families and societies and how they move toward internally regulated conduct. The contemporary analytic techniques provide us with new tools to revisit those classic questions.

References


Received March 22, 2004
Revision received December 15, 2004
Accepted January 30, 2005

---

**Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted**

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to Demarie Jackson at the address below. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In your letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Demarie Jackson, Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.