

# Temperament and Parenting during the First Year of Life Predict Future Child Conduct Problems

Benjamin B. Lahey · Carol A. Van Hulle ·  
Kate Keenan · Paul J. Rathouz · Brian M. D’Onofrio ·  
Joseph Lee Rodgers · Irwin D. Waldman

Published online: 21 June 2008  
© Springer Science + Business Media, LLC 2008

**Abstract** Predictive associations between parenting and temperament during the first year of life and child conduct problems were assessed longitudinally in 1,863 offspring of a representative sample of women. Maternal ratings of infant fussiness, activity level, predictability, and positive affect each independently predicted maternal ratings of conduct problems during ages 4–13 years. Furthermore, a significant interaction indicated that infants who were both low in fussiness and high in predictability were at very low risk for future conduct problems. Fussiness was a stronger predictor of conduct problems in boys whereas fearfulness was a stronger predictor in girls. Conduct problems also were robustly predicted by low levels of early mother-report cognitive stimulation when infant temperament was controlled. Interviewer-rated maternal responsiveness was a robust predictor of conduct problems, but only among infants low in fearfulness. Spanking

during infancy predicted slightly more severe conduct problems, but the prediction was moderated by infant fussiness and positive affect. Thus, individual differences in risk for mother-rated conduct problems across childhood are already partly evident in maternal ratings of temperament during the first year of life and are predicted by early parenting and parenting-by-temperament interactions.

**Keywords** Infant temperament · Early parenting · Conduct problems

Many theorists have suggested that mothers who responsively meet the needs of their infants following birth provide a necessary foundation for the development of regulated and competent emotional and social behavior (Bornstein 2002; Bowlby 1958; Kochanska 1997a). The infant plays an important role in this potentially critical process, as responsive parenting involves reciprocal transactions between infants and mothers (Anderson et al 1986; Bates et al. 1998; Bell 1977; Chess and Thomas 1984; Lytton 1990; Maccoby 1992). Thus, it is likely that individual differences in maternal responsiveness reflect characteristics of the mother, dimensions of infant temperament, and interactions among these variables (Lerner et al. 1989; Thomas and Chess 1977).

The “truly early starter model (TESM) of antisocial behavior” was advanced to apply a transactional view of early temperament and parenting to the origins of conduct problems (Shaw et al. 2000). The TESM includes the hypothesis that the combination of infant fussiness and lack of maternal responsiveness creates aversive mother–infant interactions which set the stage for coercive interactions that foster conduct problems later in childhood (Patterson 1982). In addition, the TESM hypothesizes significant interactions (in the sense of greater than additive combinations) between infant temperament and early parenting. The specific

---

B. B. Lahey (✉) · C. A. Van Hulle · P. J. Rathouz  
Department of Health Studies (MC 2007), University of Chicago,  
5841 South Maryland Avenue,  
Chicago, IL 60637, USA  
e-mail: blahey@uchicago.edu

K. Keenan  
Department of Psychiatry, University of Chicago,  
5841 South Maryland Avenue,  
Chicago, IL 60637, USA

B. M. D’Onofrio  
Indiana University,  
Bloomington, IN, USA

J. L. Rodgers  
University of Oklahoma,  
Norman, OK, USA

I. D. Waldman  
Emory University,  
Atlanta, GA, USA

hypothesis offered is the combination of low maternal responsiveness and aversive infant behavior (e.g., fussy and persistent bids for maternal attention) is associated with particularly high risk for later conduct problems. The TESM also suggests that the predictive associations between infant factors and later conduct problems may be stronger in boys than girls (Shaw et al. 2000).

Based on extensive non-human animal research, Meaney and Szyf (2005) offered a very different model of effects of parenting during infancy. They hypothesized that variations in maternal care during infancy have lasting effects of the offspring's behavior through "environmental programming." That is, variations in early parenting can cause enduring alterations in gene expression that can be passed on to future generations. Regardless of the specific theoretical model, there is considerable reason to suspect that experiences during infancy play an important and perhaps critical role in the origins of psychopathology. Understanding that role will require an understanding of which dimensions of infant temperament and early parenting operate independently or interactively to create vulnerability for later psychopathology.

### Infant Temperament

Early temperament is viewed both as the earliest indicator of biologically based individual differences that directly presage later psychopathology and as early childhood characteristics that influence, and interact with, the social environment to foster the learning of psychopathology (Bates et al. 1998; Chess and Thomas 1984; Keenan 2000; Sanson et al. 1993; Shaw et al. 2000). Much remains to be learned about such hypothesized relations between infant temperament and later adjustment, however. As Caspi (2000) put it: "Behavioral differences among children are apparent very early in life... Are such behavioral differences, or temperamental styles, evanescent qualities or do they presage the life patterns to follow?" (p. 158).

A number of different models and measures of infant temperament have been advanced (Buss and Plomin 1975; Carey 1970; Rothbart 1981; Thomas and Chess 1977). Each distinguishes somewhat different dimensions of temperament, but they include individual differences in dimensions such as infant activity level, positive affect, fearfulness, fussiness, and predictable rhythmicity. Many previous longitudinal studies have examined the extent to which such dimensions of early temperament predict future conduct problems. Three reviews (Keenan 2000; Sanson et al. 2004; Shaw et al. 2000) document that maternal ratings of temperament during the infant and toddler years significantly predict future maternal ratings of childhood conduct problems. Most previous studies used small and

non-representative samples, but yielded generally consistent findings. Nonetheless, three basic questions remain largely unanswered: (1) how early in life does temperament predict future conduct problems, (2) which dimensions of temperament predict future conduct problems, and (3) over how long a period of future development does early temperament predict future conduct problems? Existing studies clearly suggest that temperament measured *after* the first year of life predicts behavior problems into middle childhood, but there is less evidence that temperament measured *during* the first year of life predicts future conduct problems beyond the preschool years. For example, Prior et al. (2001) found that ratings of infant temperament at 4–8 months predicted mother rated behavior problems at 3–4 years, but these early temperament ratings did not continue to predict behavior problems after the children reached 5 years. In contrast, Bates et al. (1998) found that temperamental resistance to control measured after the first year of infancy (at 13 and 24 months) did predict behavior problems into middle childhood. Similarly, Guerin et al. (1997) found that maternal ratings of difficult temperament at 18 months predicted parent-rated conduct problems through age 12 years of age. In addition, two prospective longitudinal studies indicate that individual differences in temperament-like behavioral characteristics at age 3 years predict aggression in late childhood (Raine et al. 1998) and predict antisocial behavior in early adulthood (Caspi et al. 1996).

Only four previous studies provide evidence relevant to the long-term prediction of later conduct problems from measures of temperament during the potentially important first year of life. A longitudinal study of approximately 120 infants found that maternal ratings of difficult temperament at 6 months predicted mother-rated conduct problems through age 17 years, but not teacher or youth reports of conduct problems (Olson et al. 2000). A longitudinal study of 100 Finnish infants found that maternal ratings of fussiness at 6 months significantly predicted both parent and youth composite reports of a broad range of emotional and behavior problems at ages 14–15 years, but infant activity, fearful response to novelty, and biological predictability did not (Teerikangas et al. 1998). In contrast, in a prospective study of 180 infants, maternal ratings of infant temperament at 3 and 6 months did not significantly predict maternal ratings of conduct problems during adolescence (Aguilar et al. 2000). In a sample of about 400 infants, Colder et al. (2002) found that infant fearfulness and activity level during the first year of life predicted conduct problems during 4–8 years in interaction (not individually), but did not test other dimensions of temperament as predictors. It is very important to continue to test the hypothesis that variations in temperament during the first year of life presage behavior problems into later childhood

and beyond using samples with sufficient power to detect meaningful associations and interactions. Because some models of temperament suggest that interactions among dimensions of temperament may be important in predicting future problem behavior (e.g., Rothbart and Bates 1998), it will be important to continue to test such interactions.

### Role of Parenting during Infancy

Because of the theoretical importance of parenting during infancy, a number of studies have prospectively linked measures of parenting during infancy to later conduct problems. Pettit and Bates (1989) conducted three observational assessments of 29 infant–parent dyads at three different ages and found that a pattern of parenting defined by affectionate contact (6 months), affectionate teaching (13 months), and verbal stimulation (24 months) inversely predicted behavior problems at 4 years. In a study of 125 low-income children, Shaw et al. (1998) found that high maternal responsiveness at 12 months was inversely related to problem behavior at 42 months. Similarly, maternal report of spanking across 0–23 months was found to predict conduct problems 4 years later, even after controlling for maternal report of difficult temperament during infancy and toddlerhood (Slade and Wissow 2004). Only two prospective studies have found links between parenting during the first year of life and conduct problems beyond the preschool period. In sample of 77 mostly substance-abusing women, observational ratings of maternal responsiveness at 4 and 12 months were found to predict child conduct problems at 10 years of age (Wakschlag and Hans 1999). In addition, in a sample of about 120 mothers, observations of warm maternal responsiveness when the infant was age 6 months of age predicted both mother-rated and youth reported conduct problems at age 17 years (Olson et al. 2000). In addition, low levels of cognitive stimulation and affection at 13 months predicted youth reported conduct problems at age 17 years in this study (Olson et al. 2000). In contrast, a study of 180 infants did not find that observed maternal responsiveness at 6 months and other early parenting dimensions predict maternal ratings of conduct problems during adolescence (Aguilar et al. 2000).

There is some evidence from studies of preschool and school age children of interactions among parenting variables on child outcomes. For example, the degree of association of spanking with child conduct problems is different at different levels of maternal emotional support (McLoyd and Smith 2002). Therefore, it is important to explore potential interactions among parenting behaviors in samples with sufficient power to detect interactions.

### Demographic Factors in Temperament and Parenting during Infancy

The development of conduct problems must be understood in the context of its complicated demography (Loeber et al. 1998). Much can be learned about these demographic differences in conduct problems by studying temperament and parenting in infancy. For example, there is clear evidence that males exhibit more conduct problems than females after age 4 years (Keenan and Shaw 1997; Lahey et al. 1999; Moffitt et al. 2001). It is interesting, then, that literature reviews have not found evidence of sex differences in temperament or parenting during infancy (Keenan and Shaw 1997, 2003). Subsequent studies have generally replicated these conclusions, although Gartstein and Rothbart (2003) found that 3–12 month infant girls were rated as significantly less active and more fearful than male infants. Similarly, a study of a large Finnish sample found that infant girls were rated significantly higher on both temperamental fearfulness and predictability (Martin et al. 1997). This conflicting evidence makes it important to test for sex differences in infant temperament in representative samples that are large enough to detect such differences. It also would be important to know if there are sex differences in the magnitudes of predictive associations between temperament and parenting and later conduct problems. If there are neither sex differences in levels of infant temperament and parenting, nor in the extent to which they predict later conduct problems, that would suggest that the later emergence of sex differences in conduct problems is due to factors that come into play after infancy. On the other hand, if there are sex differences in infant temperament or parenting, or in the magnitudes of predictions of later child conduct problems, that would imply that infant variables play a role in the origins of sex differences in later conduct problems.

There also is evidence that children living in lower socioeconomic environments are more likely to develop conduct problems than children living in higher socioeconomic environments (Lahey et al. 1999). Correspondingly, there is evidence that some aspects of the parenting of infants, such as cognitive stimulation, vary with socioeconomic status (Bornstein et al. 2003; Karrass et al. 2003). If socioeconomic differences in the aspects of early parenting that predict future conduct problems can be identified consistently in representative samples, that would suggest that early parenting could be a factor that contributes to socioeconomic differences in conduct problems (Bornstein et al. 2003). There is currently little evidence on race–ethnic differences in parenting during the first year of life from representative samples, but one study found that African American mothers report using spanking during preschool and elementary school years more often than

Hispanic and Non-Hispanic European American mothers (McLoyd and Smith 2002). In addition, some studies have found that spanking predicts higher levels of childhood conduct problems in non-Hispanic European American families but not in African American families (Deater-Deckard and Dodge 1997; Slade and Wissow 2004). Again, describing any race-ethnic differences in levels of parenting and temperament in infancy, and any differences in their predictive relations to later conduct problems, will inform theories of the origins of conduct problems in important ways.

## Goals and Hypotheses

The goals of the present paper are to determine if temperament and parenting during the first year of life, both singly and in interaction with one another, predict future conduct problems across 4–13 years of age. Tests will be conducted using prospective longitudinal data on the same children from infancy through early adolescence in a large and diverse population-based sample. Based on theory and previous studies, a number of specific hypotheses will be tested:

1. Dimensions of infant temperament that may reflect what is referred to generally as “difficult temperament” (i.e., high activity, high fearfulness, and low predictability) are hypothesized to predict future child conduct problems (Guerin et al. 1997; Olson et al. 2000). In particular, infant fussiness is predicted to be robustly related to future child conduct problems (Teerikangas et al. 1998; Shaw et al. 2000).
2. Maternal responsiveness during the first year of life is predicted to be inversely related to future child conduct problems (Kochanska 1997a; Olson et al. 2000; Shaw et al. 2000; Wakschlag and Hans 1999). In addition, cognitive stimulation will be inversely related to future child conduct problems (Olson et al. 2000).
3. A significant interaction will be found between maternal responsiveness and dimensions of infant temperament that the mother may find to be aversive (infant fussiness, fearfulness, and activity) in the prediction of future child conduct problems (Shaw et al. 2000).
4. Associations between infant fussiness and later conduct problems, and between maternal responsiveness and later conduct problems, will be stronger in boys than girls (Shaw et al. 2000).
5. Spanking will predict childhood conduct problems more strongly in non-Hispanic European American families than African American families (Deater-Deckard and Dodge 1997; Slade and Wissow 2004).

In addition, because we have only begun to study relations between factors in the first year of life and later

conduct problems, all main effects and all possible interactions among temperament dimensions, among parenting dimensions, and between temperament and parenting dimensions will be tested in an exploratory spirit. The statistical power provided by this large sample provides an unprecedented opportunity to test interactions. These systematic tests of all main effects and interactions will provide an empirical basis for future refinements of theoretical models of early factors in the origins of conduct problems. Demographic variables related to childhood conduct problems will be included in these analyses to describe any demographic differences in temperament and parenting, to control potential demographic confounds, and to test interactions between demographic factors and infant predictors.

## Materials and Methods

### Sample

#### *Mother-Generation Sample: The National Longitudinal Survey of Youth (NLSY79)*

The NLSY79 survey was funded by the Bureau of Labor Statistics to study the future US workforce. A nationally representative household sample of 6,111 14–22 year old male and female youth who were not in the military was selected for the NLSY79 using a complex survey design. An additional sample of 3,652 African American and Hispanic youth was selected for the NLSY79 mother-generation sample to oversample these groups. The NLSY79 sample available for the present analyses consisted of the 4,926 females (1,472 African American, 977 Hispanic, and 2,477 non-Hispanic European American and other groups) who had given birth to children who have participated in the assessments of offspring. The present analyses are based on data from the 1986–2004 assessments of the offspring. The response rate for the initial NLSY79 assessment was 90% of the eligible sample. Participants were re-interviewed annually from 1979 through 1994 and every 2 years since then. Retention rates for the NLSY79 during follow-up assessments were 90% or better during the first 16 waves and have stayed above 80% since.

#### *Offspring-Generation Sample: Children of the NLSY79 (CNLSY)*

Biennial assessments of the biological children of women in the NLSY79 began in 1986, with 95% of the offspring assessed (Chase-Lansdale et al. 1991). For budgetary reasons, however, a random 38% of the African American



and Hispanic oversamples was not recruited in 2000. The data missing from the 2000 assessment can be considered to be missing completely at random (Little and Rubin 1987). The average retention rate for the repeated assessments was above 90% through 2004. Although there were relatively little missing data due to nonparticipation a large amount of data was missing by design. Only a subset of the full sample of CNLSY sample could be used in the present analyses because of the ages at which they were assessed. The offspring whose data could be used in the present analyses participated in an assessment of temperament and parenting at 0–11 months of age. Nearly equal numbers of infants of each age in months participated, with 46.4% being less than 6 months in age ( $M=5.8$  months,  $SD=3.5$  months). In addition, the subset of the CNLSY sample was limited to participants whose childhood conduct problems had been rated by their mothers in at least two of the five biannual assessments across 4–13 years. Because different offspring entered the study in different calendar years following their births, different offspring could be assessed different numbers of times through the last assessment in 2004 for which data are available. The percent of participants with completed assessments of conduct problems at each age was: 4–5 years=93%; 6–7 years=92%; 8–9 years=82%; 10–11 years=70%; and 12–13 years=61%. Even at ages 12–13 years there were 1,131 assessments, providing substantial power to test predictive associations. Among the participants in the present analyses, 86% had three or more assessments of their childhood conduct problems, and 70% had four or more assessments.

Because children could enter the biennial sequence of assessments only in even numbered calendar years, only half of the offspring of the NLSY79 mothers were eligible for the assessment of infant temperament during the first year of life, with the remaining half not included in the present analyses. Other CNLSY offspring were not included in the present analyses because they were too old for the temperament assessment at the time of the first assessment in 1986 or have not yet been assessed twice during childhood. These missing data can be considered missing at random (Little and Rubin 1987), which is accommodated well by the longitudinal analyses. Nonetheless, the data missing by design are likely to influence its representativeness. Therefore, it is necessary to test for potential biases in the subsample used in the present analyses.

It should be noted that five papers cited in this paper also were based on varying subsets of observations from the CNLSY. Colder et al. (2002) predicted child outcomes at 8 years using data from the 1986–1994 assessments on a restricted subsample and Slade and Wissow (2004) used data from the 1986–1998 assessments. McLoyd and Smith (2002) assessed the association of parenting measures with change in a global measure of child adjustment across

1988–1994 by following CNLSY children who were 4–5 years of age in 1988. In addition, two papers on the validity of the Home Observation for Measurement of the Environment (HOME-SF; Bradley et al. 2001a, b) conducted time-varying correlational analyses using data from the 1986–1994 assessments.

The present analyses differ from these earlier CNLSY papers in three main ways. First, the present analyses are based on at least three additional biannual assessments of the CNLSY sample that were not included in the earlier papers (i.e., 2000, 2002, and 2004). This both allows coverage of a longer span of development and allows the analyses to be based on the offspring of NLSY79 mothers who were born over a broader range of maternal ages than in previous papers. The latter means that the data used in the present analyses are less biased by early maternal age and more representative of the US population. Second, unlike most previous papers, we limited our analyses to the predictor variables of temperament and parenting measured during the first year of life and limited our response variable to conduct problems, rather than the broader and less specific measures of child maladjustment (which conflated internalizing and externalizing problems) used in most previous studies of this sample. Third, the present analyses tested predictive associations of early parenting with future conduct problems while controlling for early temperament and vice-versa.

#### *Participants Available for Preliminary Factor Analyses*

Maternal ratings of infant temperament obtained at 0–11 months were available on 2,562 offspring of the CNLSY mothers. Respondents other than the child's biological mother completed temperament ratings in <1% of cases. Eighteen (0.7%) children were dropped because they had missing data on >50% of temperament items. After dropping these children, 94% of the 2,544 children had missing data on no items and 98% had missing data on less than two items.

#### *Participants Available for Longitudinal Analyses*

Infant temperament ratings obtained at 0–11 months were available on 2,040 children who also had at least two follow up assessments of child conduct problems over 4–13 years. Because 177 parents did not provide information on family income, which was an essential covariate in all longitudinal analyses, the analyses were based on the remaining 1,863 participants. Among these participants, there were  $\leq 0.5\%$  missing data on each infant temperament variable. There were fewer completed mother–infant interactions during which interviewers rated parenting behaviors ( $n=1,519$ ), however, because infants were not always

available at the time of the assessment of the mother. Within the group with completed mother–infant interactions, there was <0.5% missing data on each parenting measure. The characteristics of the participants in the present analyses are described in Table 1. Correlations among the demographic and infant predictor variables are presented in Table 2.

## Measures

Federal health and policy studies, such as the CNLSY, use a measurement strategy that is designed to assess multiple constructs without overburdening participants. Short forms of scales often are developed that maximize correlations with full scales. Therefore, repeated assessments of large sample sizes are used to detect reliable signals in the data.

### Measures of Infant Temperament

Mothers rated their 0–11 month old infants on 17 temperament items using a five-point scale: almost never;

less than half the time; half the time; more than half the time; almost always. These temperament items were based primarily on a subset of the items used in the Infant Behavior Questionnaire (IBQ; Rothbart 1981) that were selected to represent five dimensions of infant temperament: activity level, predictability of cycles and moods, positive affect, fearfulness, and fussiness (Mott et al. 1995).

There is disagreement regarding the factor structure of these CNLSY temperament items. Based on an exploratory factor analysis of a subset of the across 6–23 months, Baydar (1995) recommended distinguishing only two temperament dimensions. In contrast, the NLSY User's Guide recommends distinguishing five factors, but also provides alternative scorings for fewer factors. In order to choose among these recommendations, we conducted confirmatory factor analyses (CFAs) using Mplus 4.1 based on item ratings residualized on the infant's age in months at the time that temperament was rated, while taking clustering within CNLSY families into account. The five-factor temperament model was based on the scale structure originally assumed for the CNLSY infant temperament measure. The fit of this model was compared to nested

**Table 1** Unweighted Proportions, Means, and Standard Deviations (in parentheses) for Demographic Variables, Infant Temperament, Parenting during Infancy, and Child Conduct Problems in Female and Male Participants

	Girls N=878 (47.4%)	Boys N=976 (52.6%)
Infant's race–ethnic classification		
Hispanic	166 (18.9%)	184 (18.8%)
African American	242 (27.6%)	260 (26.6%)
Non-Hispanic European American	470 (53.5%)	532 (54.5%)
Total family income at age (1986 US dollars)	\$43,543 (106,928)	\$33,653 (59,251)
Mother's age at first birth	24.1 (5.0)	23.8 (5.2)
Mean age in months at time of assessment of temperament and parenting in first year of life	5.7 (3.5)	5.8 (3.4)
Numbers of assessments of conduct problems at each age in years	4 or 5: 848 6 or 7: 827 8 or 9: 746 10 or 11: 626 12 or 13: 544	4 or 5: 940 6 or 7: 919 8 or 9: 809 10 or 11: 701 12 or 13: 597
Maternal ratings of infant temperament ( <i>z</i> scores) <sup>a</sup>		
Activity	−0.25 (3.27)	0.16 (3.22)
Fearfulness	0.18 (3.27)	−0.21 (3.18)
Fussiness	−0.06 (2.02)	0.04 (1.99)
Positive affect	−0.02 (2.61)	−0.02 (2.52)
Predictability	0.06 (2.37)	−0.05 (2.39)
HOME-SF parenting measures		
Maternal responsiveness ( <i>z</i> scores) <sup>a</sup>	0.00 (0.24)	0.00 (0.24)
Cognitive stimulation ( <i>z</i> scores) <sup>a</sup>	−0.00 (0.92)	−0.02 (0.87)
Spanking infant <sup>a</sup>	65 (7.4%)	85 (8.8%)
Mean conduct problems over 4–13 years ( <i>z</i> scores)	−0.20 (0.85)	0.16 (1.07)

<sup>a</sup> Residualized on the infant's age in months at the time of assessment for this table. Standardized *z* scores for temperament, parenting, and conduct problems are presented here to facilitate comparisons of girls and boys, but raw scores were used in statistical analyses.

**Table 2** Unweighted Pearson Correlations Among Demographic and Predictor Variables

	Predictability	Fearfulness	Positive affect	Fussiness	Spanking	Maternal responsiveness	Cognitive stimulation
Activity	0.02	0.16***	0.13***	0.19***	0.03	-0.05	-0.01
	1,850	1,844	1,846	1,849	1,846	1,511	1,502
Predictability		-0.11***	0.23***	-0.22***	-0.10***	0.10***	0.16***
		1,843	1,842	1,845	1,842	1,508	1,499
Fearfulness			-0.01	0.34***	0.06*	-0.10***	-0.09**
			1,840	1,839	1,836	1,502	1,493
Positive affect				-0.12***	-0.05*	0.06**	0.14***
				1,843	1,838	1,504	1,495
Fussiness					0.11***	-0.12***	-0.16***
					1,841	1,507	1,498
Spanking						-0.04	-0.03
						1,510	1,501
Maternal responsiveness							0.29***
							1,502

Ns are below each correlation. Ns for interviewer-rated maternal responsiveness and cognitive stimulation are smaller because infants were not always available for mother–child interactions.

\* $p < 0.05$ , \*\* $p < 0.001$ , \*\*\* $p < 0.0001$

models in which scales were combined to create broader scales, as suggested by Mott et al. (1995) and consistent with other models of temperament (e.g., Bates et al. 1979). In the four-factor model, fearfulness and fussiness were combined to define negative affectivity. In the three-factor model, fearfulness, and fussiness, and the inverse of positive affect were combined to define emotionality. In the two-factor model predictability, fearfulness, fussiness, and the inverse of positive affect were combined to define difficult temperament. All temperament dimensions were combined in the one-factor model.

The recommended five-factor model of CNLSY temperament items fit significantly better than the next best-fitting four-factor model (Satorra–Bentler scaled-difference  $\chi^2 = 280.13$ ,  $df = 4$ ,  $p < 0.0001$ ). All other models fit significantly less well. Fit indices for the five-factor model indicated a close fit (comparative fit index = 0.94; root mean square error of approximation = 0.04). Therefore, the five infant temperament scales used in the present analyses were: *activity level* (squirms and kicks during feeding; waves arms during feeding; moves around in the crib during sleep), *predictability* (sleepy about the same time each evening; hungry about the same time each evening; wakes up in the same mood each morning), *fearfulness* (cries or turns away from strangers; cries or turns away from unfamiliar dog or cat; cries when left alone in a room; cries or turns away from a doctor, dentist, or nurse); *positive affect* (smiles or laughs when you play with him or her; smiles or laughs when plays alone; smiles or laughs in the bath), and *fussiness* (often fussy or irritable; trouble soothing infant when crying or upset; often cries or fusses

compared to most babies; cries or becomes upset in response to noise). The item with the lowest loading on any factor was crying in response to noises. We tested an alternative model in which this item loaded on the fearfulness factor with other crying items, but the loading was still  $< 0.30$  and the fit of the alternative model was not better than the original five-factor model. On the assumption that crying in response to noises is more related to sensory modulation than fearfulness (Dunn 2002; Goldsmith et al. 2006), we dropped this item. Cronbach’s alpha coefficients for the raw temperament scales were modest: activity = 0.66; predictability = 0.59; positive affectivity = 0.71; fearfulness = 0.61; and fussiness = 0.60.

*Measure of Parenting During Infancy*

Parenting at 0–11 months was measured using the CNLSY Infant/Toddler Short Form of the widely used HOME-SF (Caldwell and Bradley 1984). The HOME was constructed rationally and through factor analysis (Bradley and Caldwell 1984) and has been validated by correlations with observational measures of parenting (Bradley et al. 2003). The Infant/Toddler HOME-SF is composed of 8 ratings of the mother’s parenting and the physical home environment made by the interviewer following the assessment of the mother and infant, and ten maternal-report items on her parenting behaviors and the home environment. Although there are mean differences in HOME-SF items across US race–ethnic groups in the CNLSY, differences related to family socioeconomic status are greater and account for most race–ethnic differences

when controlled. Moreover, differences in HOME-SF scores related to poverty are proportional across European American, African American, and Hispanic American groups (Bradley et al. 2001a).

Three dimensions of parenting are distinguished in the Infant/Toddler HOME-SF: maternal responsiveness, cognitive stimulation, and spanking/restraint. In a study of CNLSY data from the 1986–1994 assessments, Bradley et al. (2001a) treated the three HOME-SF dimensions as time-varying covariates across 4–13 years. They found that cognitive stimulation and maternal responsiveness were concurrently associated with the development of receptive vocabulary and cognitive stimulation and spanking/restraint was concurrently associated with the global total score on the mother-completed Behavior Problem Index (Mott et al. 1995). Similarly, Linver et al. (2002) found cognitive stimulation to be concurrently correlated with a global behavior problem score during the preschool years.

We conducted CFAs to test the three-dimensional structure of the HOME-SF parenting items, partly because we dropped four items from the HOME-SF to focus more on maternal parenting than on the home environment and perhaps to make the scale more applicable across socioeconomic levels and cultures: number of soft/role play toys child has; number of push/pull toys child has; child sees father figure daily; child often eats with mother and father. The recommended three-factor model of age-adjusted ratings of the Infant/Toddler HOME-SF parenting items fit significantly better than the next best-fitting two-factor model (Satorra–Bentler scaled-difference  $\chi^2=1,222$ ,  $df=2$ ,  $p<0.0001$ ). Fit indices for the three-factor model indicated a close fit (comparative fit index=0.90; root mean square error of approximation=0.04). These three parenting dimensions were *maternal responsiveness* (interviewer rating of the mother speaking to child two or more times; mother responding verbally to child's speech; mother kissing, hugging, or caressing the child; mother providing interesting toys or activities; mother keeping child in view; and the play environment in the home being safe), *cognitive stimulation* (maternal reports that the child often gets out of house; child has a number of books; mother often reads to child; child is often taken to grocery store; mother often talks to child while working or doing chores), and *spanking/restraint* (mother report that she spanked the child at least once in the last week; interviewer report that the mother slapped or spanked child at least once during assessment; interviewer report that the mother physically restricted child at least three times during the observation).

Cronbach's alpha coefficients for the three parenting dimensions were modest: cognitive stimulation=0.58; maternal responsiveness=0.68; and spanking/restraint=0.28. Because the internal consistency of spanking/restraint was unacceptable, we created a dichotomous indicator of

spanking based on either the mother reporting spanking the infant during the last week or the interviewer observing spanking during the assessment. The prevalence of mother-reported spanking was 7.3% and the prevalence of observed spanking was 1.7%, with 8.1% of mothers being reported to spank by one or both reports. The two reports of spanking co-occurred at greater than chance levels (odds ratio=5.30; 95% confidence interval=2.2–12.4), providing validation of each report of spanking.

#### *Measure of Child Conduct Problems*

Across 4–13 years of age, mother's rated their children's adjustment using the Behavior Problem Index (BPI; Peterson and Zill 1986). The BPI was created for the CNLSY by selecting the items from the Child Behavior Checklist (Achenbach 1978) that had the strongest correlations with the corresponding CBCL factor scores (Peterson and Zill 1986). Mothers rated each of their children in each assessment wave using a three-point scale for each item: "often true"=2, "sometimes true"=1, or "not true"=0. Based on similarities to items used in previous studies (Lahey et al. 2006), mean ratings of seven BPI items were selected before data analysis to define child *conduct problems*: cheats or tells lies; has trouble getting along with teachers; disobedient at home; disobedient at school; bullies or is cruel or mean to others; breaks things on purpose or deliberately destroys his/her own or another's things; and does not seem to feel sorry after misbehaving. To test this selection of items, we conducted a confirmatory factor analysis of the 13 "externalizing" items in the BPI and found that the best-fitting model distinguished a factor comprised of these seven conduct problem items from other items that measured hyperactive and oppositional behavior (D'Onofrio et al. 2008). Across ages 4–13 years, the range of Cronbach's alpha coefficients for the conduct problems measure ranged across age from 0.68–0.80, median=0.74 (Lahey et al. 2006). Importantly, the CNLSY measure of child conduct problems is valid in the sense of robustly predicting convictions for nontrivial offenses during adolescence in the CNLSY (Lahey et al. 2006).

#### *Methods of Longitudinal Data Analysis*

Longitudinal tests of infant predictors of future childhood conduct problems over 4–13 years used longitudinal Poisson regression models estimated in generalized estimating equations (GEE; Zeger and Liang 1986), specifying autoregressive correlation structures in SAS GENMOD. These analyses were weighted using the mother's 1986 sampling weight to assign equal weights to siblings within families. GEE models the average value of the response variable for each subset of individuals who share the same



value of the predictor variable. Because GEE is based on estimates of averages rather than the entire distribution of values, GEE is less restricted by distributional assumptions than other approaches to longitudinal analysis. All GEE models were conducted specifying Poisson working distributions, but unlike other longitudinal models, GEE can be used when distributions of values of response variables do not conform exactly to a particular distribution (normal, Poisson, etc.).

All statistical tests in the present GEE analyses used robust (“empirical”) standard errors because an adjustment for overdispersion from the Poisson distribution is automatically included and its use reduces concern about correct specification of the within-subject covariance structure. Like the other cited studies of the CNLSY sample, the clustering of the sample within sampling units and families was not formally modeled in these analyses. Fortunately, the robust standard errors used in GEE also minimize the effect of any incorrect specification of standard errors due to such clustering.

In all cases, predictive associations between infancy variables and later child conduct problems were tested in simultaneous longitudinal regression analyses in which each infant predictor of subsequent child conduct problems was tested while controlling for all other predictors and the demographic control variables (the child’s age in years at the time of each assessment of child conduct problems, the infant’s sex (coded 1=males; 0=females), maternal age at first birth, race–ethnicity, and total family income). In the log-linear models conducted in GEE, the regression coefficient ( $\beta$ ) represents the log relative mean count of conduct problems for a one-unit difference in the predictor variable. For example, in the model predicting conduct problems (range 0–11 problems) across 4–13 years from age-adjusted units of infant fussiness (range –3 through 9), adjusting for time, demographic variables, and the other dimensions of temperament, the coefficient of  $\beta=0.05$  indicates each of the 12 single-unit differences in fussiness was independently associated with 5% greater conduct problems in each wave because  $\exp(0.05)=1.05$ . Interpreting  $\beta$  as an estimate of “effect size” in this way in GEE, however, requires recognition that the range of scores for each predictor variable differs. As a result, a smaller  $\beta$  for a predictor with a larger range of scores could indicate a stronger association than a larger  $\beta$  for a predictor with a smaller range of scores. Therefore, more straightforward supplemental estimates of effect size also were provided for key findings.

After testing “main effects” for all predictors, all predictor-by-age (of the child at each assessment of conduct problems across 4–13 years) interactions were tested. A significant predictor-by-age interaction would indicate that magnitude of the prediction of child conduct problems

changed over 4–13 years. No predictor-by-age interactions were significant, however.

Controlling the demographic covariates, we tested the interaction of each temperament measure (before residualizing for age in months) with the infant’s age in months when temperament was measured to determine if these temperament measures predict future conduct problems differentially when temperament is measured at different ages during the first year of life. None of the interactions was significant at  $p<0.05$ .

An unadjusted alpha level of  $p<0.05$  was used based on the assumption that Type II errors are more serious than Type I errors during the early stages of research on a phenomenon (Cohen 1982). That is, Type I errors are likely to be detected by future failures to replicate, whereas Type II errors may discourage future research (Cohen 1994). If one were to apply a Bonferroni correction to the alpha level for the primary regression model (predicting childhood conduct problems using the five infant temperament scales, three dimensions of parenting during infancy, five demographic control variables, and testing the repeated assessment effect of age, the corrected alpha level would be  $p<0.0033$ . As reported below, some infant predictors would remain statistically significant (i.e., infant fussiness and maternal cognitive stimulation) following such an adjustment, but others would not. Moreover, our emphasis is not on significance tests but on estimates of effect sizes for the predictive relations from infancy into childhood (Cohen 1994).

#### Graphing Conventions and Supplemental Estimates of Effect Size

As detailed below, no longitudinal prediction analysis found a significant interaction between any infant predictor and the child’s age in the repeated assessments of child conduct problems conducted over 4–13 years of age. That is, there was no evidence of significant differences in the slopes of conduct problems across increasing age associated with any predictor variable. Therefore, to simplify graphic presentations of the many findings, median numbers of conduct problems across all of the repeated assessments across 4–13 years are presented in all figures. Medians are presented instead of means because the distributions of conduct problems were skewed.

To supplement the interpretation of  $\beta$  coefficients, estimates of effect sizes for the magnitude of predictive associations between the continuous temperament and parenting predictors and future conduct problems also were generated by comparing the level of conduct problem scores during 4–13 years of infants who were in the top 25% of the unweighted sample distribution of each dimension of infant temperament and parenting to

those of infants who were in the lower 25% of the distributions of these predictors. Because the distribution of conduct problem scores was skewed, the magnitudes of these differences between the two ends of the distributions were quantified using Rosnow and Rosenthal's (1996) effect-size correlation ( $\rho_{ES}$ ) based on Spearman's rank correlations.

## Results

### Tests of Sample Bias

Because a selected subset of the CNLSY offspring was used in the present analyses, tests were conducted to determine how the participants in the analyses differed from the remainder of the sample on key demographic variables. These cross-sectional comparisons were conducted using log-linear regression based on robust variance estimators in which the demographic variables were all entered simultaneously. A total of 2,027 offspring had both sufficient data on infant temperament and childhood conduct problems to be included in the factor analysis, compared to the 8,879 offspring who were not included in the factor analyses. There was a significantly greater proportion of male children (53.1%) included in the analyses than excluded (50.6%),  $\chi^2=4.20$ ,  $df=1$ ,  $p<0.05$ . There was not a significant difference in the proportions of children of African American and Hispanic mothers. The families of children who were included had higher weighted total family incomes when the mother was age 30 years (\$46,026, expressed in 1986 dollars), than those of children who were excluded (\$36,439),  $z=3.35$ ,  $p<0.001$ . Similarly, children who were included had mothers with older ages at the birth of their first child (25.0 years), than children who were excluded (22.4 years),  $z=3.35$ ,  $p<0.001$ . Therefore, to more accurately estimate unbiased population parameters, all regression analyses controlled for the infant's sex, total adjusted family income when the mother was age 30, and the mother's age at first birth. Because of variations in income and maternal age across race-ethnic groups in the USA, binary contrasts for maternal classifications of race-ethnicity as African American and Hispanic were also included as covariates in all cross-sectional and longitudinal analyses. In the CNLSY, the African American and Hispanic classifications were mutually exclusive.

### Demographic Differences in Temperament and Parenting during Infancy

We conducted an initial series of analyses to describe any differences in infant temperament and parenting variables

that are related to demographic factors.<sup>1</sup> Cross-sectional regression models were estimated separately for each infant temperament and parenting scale (residualized on age in months of the infant at the time of the assessment), with child sex, race-ethnicity, family income, and maternal age at first birth entered simultaneously as the predictor variables. Consistent with Gartstein and Rothbart (2003) and Martin et al. (1997), there was a small, but significant sex difference in fearfulness, with males rated as less fearful than female infants,  $\beta=-0.42$ ,  $z=-2.63$ ,  $p<0.01$ . Unlike these previous studies, there were not significant sex differences in infant fussiness or activity level. In addition, there were no significant differences in the levels of maternal responsiveness or cognitive stimulation, or in the prevalence of spanking, received by female and male infants.

Other demographic factors also were found to be associated with infant temperament. Mothers with higher incomes rated their infants as less fearful ( $\beta=-0.00$ ,  $z=-4.81$ ,  $p<0.0001$ ), but family income was not significantly related to the other dimensions of temperament. African American mothers rated their infants as fussier ( $\beta=0.89$ ,  $z=7.16$ ,  $p<0.0001$ ), more active ( $\beta=0.97$ ,  $z=4.90$ ,  $p<0.0001$ ), more fearful ( $\beta=1.12$ ,  $z=5.46$ ,  $p<0.0001$ ), and less predictable ( $\beta=-0.95$ ,  $z=-6.62$ ,  $p<0.0001$ ) than non-Hispanic European American mothers. Hispanic mothers rated their infants as more fearful ( $\beta=0.46$ ,  $z=2.16$ ,  $p<0.04$ ) and less predictable ( $\beta=-0.66$ ,  $z=-4.07$ ,  $p<0.0001$ ) than non-Hispanic European American mothers. Women who first gave birth at older ages rated their infants as less fearful ( $\beta=-0.05$ ,  $z=-2.66$ ,  $p<0.01$ ).

### Demographic Differences in Parenting during Infancy

There were no significant differences in parenting related to child's sex. African American mothers ( $\beta=-0.27$ ,  $z=-4.62$ ,  $p<0.0001$ ) and Hispanic mothers ( $\beta=-0.33$ ,  $z=-5.00$ ,  $p<0.0001$ ) each reported providing lower levels of cognitive stimulation than did non-Hispanic European American mothers. Interviewers rated African American mothers as providing lower levels of maternal responsiveness ( $\beta=-0.05$ ,  $z=-2.92$ ,  $p<0.005$ ) than non-Hispanic European American mothers; Hispanic mothers did not differ from non-Hispanic European American mothers on this parenting variable.

In addition, African American mothers were more likely to spank their infants (13.0%) than non-Hispanic European American mothers (5.9%). Hispanic mothers (7.8%) did not spank their infants significantly more often than non-Hispanic European American mothers.

<sup>1</sup> Complete tabled results of all cross-sectional and longitudinal analyses are available from the authors.

Women with higher family incomes reported providing greater cognitive stimulation of their infants ( $\beta=0.00$ ,  $z=2.66$ ,  $p<0.01$ ), were rated higher in maternal responsiveness ( $\beta=0.00$ ,  $z=2.08$ ,  $p<0.05$ ), and were more likely to spank their infants ( $\beta=0.00$ ,  $z=2.13$ ,  $p<0.05$ ). Women who first gave birth at older ages received higher scores on cognitive stimulation ( $\beta=0.05$ ,  $z=10.38$ ,  $p<0.0001$ ) and maternal responsiveness ( $\beta=0.01$ ,  $z=6.10$ ,  $p<0.0001$ ), and were less likely to spank their infants ( $\beta=-0.08$ ,  $z=-3.83$ ,  $p<0.0001$ ).

## Results of Longitudinal Prediction Analyses

### *Demographic Predictors of Child Conduct Problems*

In a longitudinal model simultaneously testing the extent to which each demographic factor predicts repeated measures of child conduct problems across ages 4–13 years, male sex of the infant ( $\beta=0.30$ ,  $z=6.59$ ,  $p<0.0001$ ), family income ( $\beta=-0.00$ ,  $z=-2.01$ ,  $p<0.05$ ), African American race–ethnicity ( $\beta=0.13$ ,  $z=2.69$ ,  $p<0.01$ ), and maternal age at first birth ( $\beta=-0.03$ ,  $z=-5.06$ ,  $p<0.0001$ ) predicted mean levels of childhood conduct problems across 4–13 years, but Hispanic race–ethnicity did not ( $\beta=0.01$ ,  $z=0.12$ ,  $p=0.90$ ) relative to non-Hispanic European American families. With the linear term for age across ages 4–13 years in the model, the quadratic term for age also was significant ( $\beta=0.01$ ,  $z=4.87$ ,  $p<0.0001$ ), indicating a small but statistically significant decrease and then increase in mean conduct problems across this age span, as previously reported by Lahey et al. (2006) in an earlier analysis based on the CNLSY. Therefore, all of these demographic variables were included as covariates in all subsequent longitudinal analyses, including the contrast between Hispanic and non-Hispanic European American families, to reduce confounding of the infant predictors with demographic factors related to infant temperament, parenting during infancy, and childhood conduct problems.

### *Infant Temperament as a Predictor of Child Conduct Problems*

In a longitudinal regression model with all demographic variables and each of the five temperament dimensions entered simultaneously as predictors, maternal ratings of conduct problems across 4–13 years were predicted by maternal ratings of infant activity level ( $\beta=0.02$ ,  $z=3.54$ ,  $p<0.0005$ ), infant predictability ( $\beta=-0.02$ ,  $z=-2.73$ ,  $p<0.01$ ), infant fussiness ( $\beta=0.05$ ,  $z=4.53$ ,  $p<0.0001$ ), and the infant's positive affect ( $\beta=-0.02$ ,  $z=-2.11$ ,  $p<0.04$ ). When all interactions between each dimension of infant

temperament and the child's age in each repeated assessment were added to the model, no interaction was significant. Furthermore, when all interactions with the quadratic term for age also were added to the model, none was significant. These findings are illustrated in Fig. 1 using the graphing conventions described above.

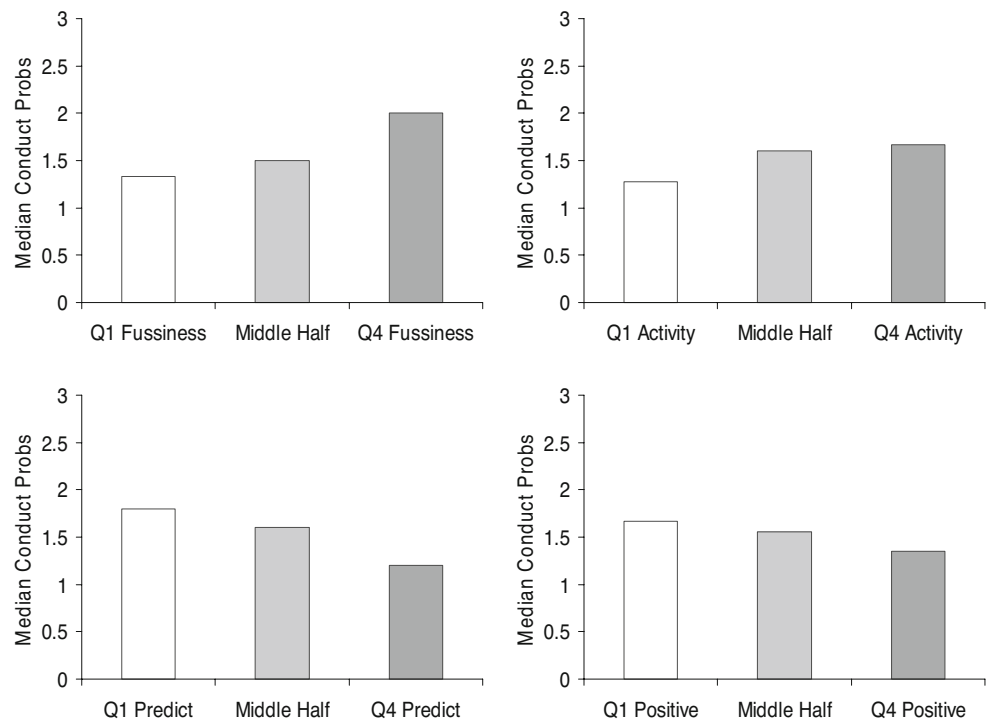
*Temperament-by-Infant's Sex Interactions* When interactions between the infant's sex and each temperament dimension were added to the main effects model, the sex-by-fearfulness interaction was significant,  $\beta=-0.03$ ,  $z=-2.26$ ,  $p<0.03$ . This reflected a stronger positive predictive association between fearfulness and future conduct problems in girls. In addition, consistent with the prediction of Shaw et al. (2000), the sex-by-fussiness interaction was significant,  $\beta=0.05$ ,  $z=2.09$ ,  $p<0.04$ , indicating a stronger positive predictive association between fussiness and future conduct problems in boys (Fig. 2).

*Temperament-by-Race–Ethnicity Interactions* When all interactions between each temperament dimension and each race–ethnic group were tested, only the interaction between infant fussiness Hispanic ethnicity was significant,  $\beta=-0.06$ ,  $z=-1.96$ ,  $p<0.05$ . This indicated a stronger predictive relation between infant fussiness and future conduct problems in non-Hispanic European American families than in Hispanic families.

*Temperament-by-Temperament Interactions* When all two-way interactions among the five temperament dimensions were added to the main effects model, only the interaction between fussiness and predictability was significant,  $\beta=0.01$ ,  $z=2.22$ ,  $p<0.03$ . This appears to reflect a ceiling effect in which there was a stronger inverse relation between predictability and conduct problems among infants who are not already at risk due to high fussiness and vice-versa. As shown in Fig. 3, this suggests that infants with both low fussiness and high predictability are at very low risk for future conduct problems.

*Supplemental Effect Size Estimates for Infant Temperament* To facilitate interpretation, effect size estimates for the main effects of infant temperament were calculated by comparing levels of future conduct problems in infants in the top and bottom quartiles of the distributions of infant temperament scores. Effect–size correlations were  $\rho_{ES}=0.21$  for infant fussiness and  $\rho_{ES}=-0.20$  for predictability, which were in the “medium effect size” range of 0.15–0.24 (Rosnow and Rosenthal 1996). Effect–size correlations were  $\rho_{ES}=0.12$  for infant activity level and  $\rho_{ES}=-0.10$  for positive affect which were in the “small effect size” range of  $<0.15$  (Rosnow and Rosenthal 1996). When the

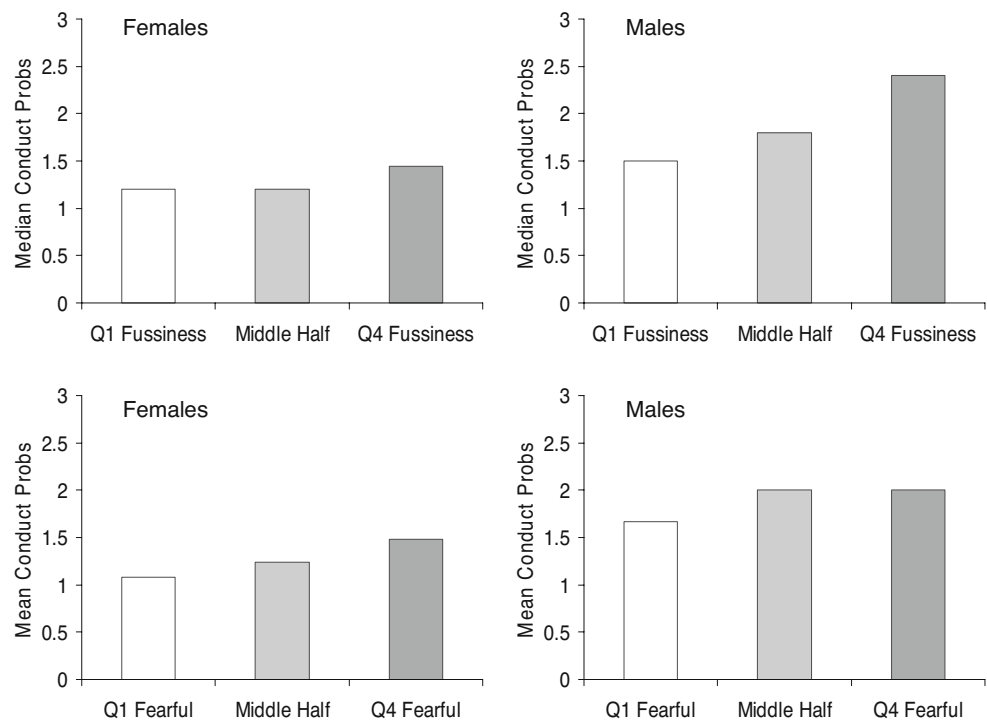
**Fig. 1** Weighted medians of the average maternal rating of child conduct problems across 4–13 years of age among children with maternal ratings in the top 25%, middle 50%, or bottom 25% of the sample distributions of fussiness (*upper left*), activity level (*upper right*), predictability of rhythms and mood (*lower left*), and positive affect (*lower right*) during infancy. Note that prospective associations of these infant predictors with later conduct problems were tested using repeated measures of conduct problems across 4–13 years in longitudinal analyses; the median level of conduct problems across these ages is used here only to facilitate graphic presentation



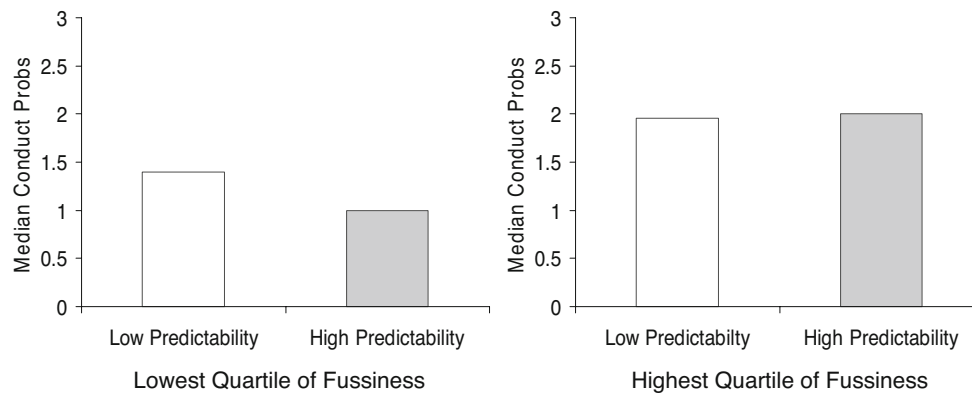
temperament-by-temperament interaction was considered, the effect size for predictability was  $\rho_{ES}=-0.04$  (“small effect”) among infants in the highest quartile of fussiness and  $\rho_{ES}=-0.17$  (“medium effect”) among infants in the lowest quartile of fussiness. When the temperament-by-sex

interaction was considered, the effect size for fussiness was  $\rho_{ES}=0.27$  (“large effect”) for males and  $\rho_{ES}=0.15$  (“medium effect”) for girls. Similarly, the effect size for fearfulness was  $\rho_{ES}=0.11$  (“small effect”) for males and  $\rho_{ES}=0.15$  (“medium effect”) for girls.

**Fig. 2** Weighted medians of the average maternal rating of child conduct problems across 4–13 years of age among infants with maternal ratings in the top 25%, middle 50%, or bottom 25% of the sample distributions of the two dimensions of infant temperament, presented separately for girls and boys for predictors for which there were significant interactions with the infant’s sex: fussiness (*upper left and right*) and fearfulness (*lower left and right*). Note that prospective associations of these infant predictors with later conduct problems were tested using repeated measures of conduct problems across 4–13 years in longitudinal analyses; the median level of conduct problems across these ages is used here only to facilitate graphic presentation







**Fig. 3** Weighted medians of the average maternal rating of child conduct problems across 4–13 years of age among infants with maternal ratings in the top or bottom quartile of the sample distributions of infant predictability, presented separately for infants with maternal ratings in the top quartile (*left panel*) or bottom quartile (*right panel*) of the sample distributions of infant fussiness to

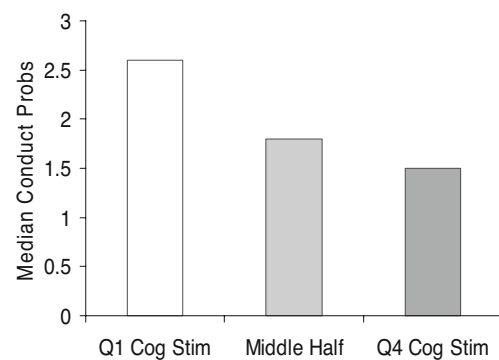
illustrated the significant fussiness-by-predictability interaction. Note that prospective associations of these infant predictors with later conduct problems across 4–13 years in longitudinal analyses; the median level of conduct problems across these ages is used here only to facilitate graphic presentation

*Parenting During Infancy as a Predictor of Future Child Conduct Problems*

With all demographic variables and each of the three dimensions of parenting during infancy entered simultaneously as predictors in the full sample, the hypothesis that childhood conduct problems across 4–13 years would be predicted by cognitive stimulation was supported,  $\beta=-0.13, z=-4.35, p<0.0001$ , but the hypothesis that childhood conduct problems would be predicted by maternal responsiveness was not supported,  $\beta=-0.15, z=-1.41, p=0.16$ . In addition, the predictive association between spanking during infancy and childhood conduct problems did reach the conventional 0.05 level of statistical significance,  $\beta=0.14, z=1.86, p=0.06$ .

In order to at least partially control for potential child effects on parenting (Bell 1977), the five dimensions of temperament were added to all subsequent longitudinal models assessing parenting. As illustrated in Fig. 4, when all demographic variables, all dimensions of parenting, and all dimensions of infant temperament were controlled, childhood conduct problems were still robustly predicted by cognitive stimulation,  $\beta=-0.11, z=-3.76, p<0.0005$ , but not by spanking,  $\beta=0.10, z=1.34, p=0.18$ , or maternal responsiveness,  $\beta=-0.14, z=-1.34, p=0.18$ . When all interactions between each dimension of parenting during infancy and the child’s linear and quadratic terms for age in each repeated assessment in these longitudinal analyses were added to the model, no interaction was significant.

interactions between race–ethnicity and either cognitive stimulation or maternal responsiveness when temperament and other demographic variables were controlled. The hypothesis that spanking during infancy would predict childhood conduct problems more strongly among non-Hispanic European American families than African American families was not supported, as interaction did not reach the 0.05 level of significance ( $\beta=-0.24, z=-1.85, p=0.06$ ). When African American and Hispanic families were compared on the strength of the predictive association between spanking and future conduct problems, the interaction term also did not reach the 0.05 level of significance ( $\beta=-0.31, z=-1.90, p=0.06$ ). In contrast, there was a significant interaction between spanking and the comparison between Hispanic and non-Hispanic European



**Fig. 4** Weighted medians of the average maternal rating of child conduct problems across 4–13 years of age among children with maternal reports in the top 25%, middle 50%, or bottom 25% of the sample distribution of cognitive stimulation of the infant. Note that the prospective association of cognitive stimulation during infancy with later conduct problems was tested using repeated measures of conduct problems across 4–13 years in longitudinal analyses; the median level of conduct problems across these ages is used here only to facilitate graphic presentation

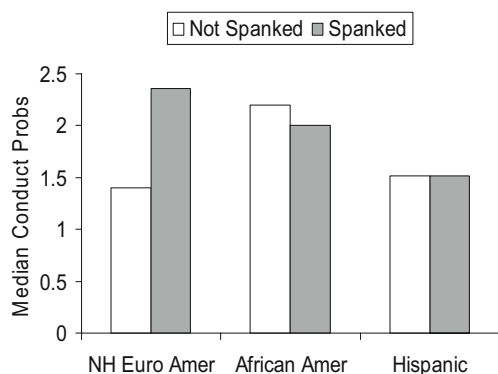
*Parenting-by-Demographic Interactions* When each two-way interaction between the three parenting measures and the infant’s sex were added to the main effects model, with temperament and other demographic variables controlled, none was significant. There also were no significant

American ethnicity ( $\beta=-0.56$ ,  $z=-3.24$ ,  $p<0.005$ ). This indicated a positive predictive association between early spanking and later conduct problems in non-Hispanic European American families, but not among Hispanic families (Fig. 5).

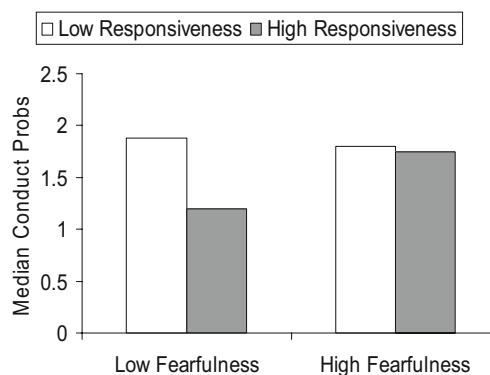
**Parenting-by-Parenting Interactions** When all two-way interactions among the parenting dimensions were added to the main effects model, there were no significant interactions.

**Parenting-by-Temperament Interactions** We hypothesized significant interactions between maternal responsiveness and dimensions of infant temperament that the mother may find to be aversive (infant fussiness, fearfulness, and activity) in the prediction of future child conduct problems (Shaw et al. 2000). In analyses controlling all demographic variables, all temperament dimensions, and each parenting dimension, the hypothesized interaction of maternal responsiveness and temperament fearfulness was significant ( $\beta=0.06$ ,  $z=2.03$ ,  $p<0.05$ ). As illustrated in Fig. 6, this interaction indicated that maternal responsiveness was a robust inverse predictor of childhood conduct problems only among infants low in fearfulness. The other hypothesized interactions between maternal responsiveness and fussiness and activity level were not statistically significant at  $p<0.05$ .

In addition, exploratory analyses revealed significant interactions between spanking and infant fussiness ( $\beta=-0.07$ ,  $z=-2.23$ ,  $p<0.03$ ) and spanking and infant positive affect ( $\beta=-0.05$ ,  $z=-2.14$ ,  $p<0.04$ ). These interactions with spanking suggest ceiling effects, with the



**Fig. 5.** Weighted medians of the average maternal rating of child conduct problems across 4–13 years of age among children with who were or were not spanked during the first year of life, separately for non-Hispanic European American, African American, and Hispanic infants. Note that the prospective associations of these infant predictors with later conduct problems were tested using repeated measures of conduct problems across 4–13 years in longitudinal analyses; the median level of conduct problems across these ages is used here only to facilitate graphic presentation



**Fig. 6** Weighted medians of the average maternal rating of child conduct problems across 4–13 years of age among children whose mothers were rated by interviewers as being in the top or bottom quartiles of maternal responsiveness during the first year of life, presented separately for infants in the lowest quartile (*left panel*) or in the highest quartile of the distribution of infant fearfulness scores (*right panel*). Note that the prospective associations of these infant predictors with later conduct problems were tested using repeated measures of conduct problems across 4–13 years in longitudinal analyses; the median level of conduct problems across these ages is used here only to facilitate graphic presentation

modest prediction of conduct problems from spanking being less strong among more fussy infants and among infants with more positive affect.

**Supplemental Effect Size Estimates for Early Parenting** A rank-order effect size correlation was calculated for the only significant main effect for early parenting by comparing levels of future conduct problems in infants in the top and bottom quartiles of cognitive stimulation scores. This estimate was in the “large effect” range for cognitive stimulation,  $\rho_{ES}=0.26$  (Rosnow and Rosenthal 1996). When the interaction of race–ethnicity and spanking was considered, the effect–size correlation compared children who were spanked during infancy to those who were not in different groups. This was  $\rho_{ES}=0.11$  (“small effect”) among non-Hispanic European American families,  $\rho_{ES}=0.05$  among African American families, and  $\rho_{ES}=-0.01$  among Hispanic families.

When parenting-by-temperament interactions were considered, the effect size for predicting childhood conduct problems from greater maternal responsiveness during infancy was  $\rho_{ES}=-0.25$  (“large effect”) among infants in the lowest quartile of the distribution of fearfulness scores and  $\rho_{ES}=-0.09$  (“small effect”) among infants in the highest quartile of fearfulness. The effect size for spanking during infancy was  $\rho_{ES}=0.10$  (“small effect”) among infants in the lowest quartile of the distribution of fussiness and  $\rho_{ES}=-0.00$  among infants in the highest quartile of fussiness. The effect size for spanking during infancy was  $\rho_{ES}=0.10$  (“small effect”) among infants in the lowest quartile of the distribution of fussiness and  $\rho_{ES}=-0.03$  among infants in the highest quartile of fussiness.

## Discussion

The results of the present longitudinal analyses support the idea that it is possible to predict conduct problems across childhood from temperament measured during the first year of life. In addition, dimensions of parenting during the first 12 months of life were found to predict future conduct problems when temperament was controlled. Because no interaction between the measures of temperament and parenting during infancy and the ages during childhood when conduct problems were measured (4–13 years) was significant, there was no evidence that the prediction of conduct problems was attenuated by length of the follow-up period.

To advance understanding of the role of infant temperament and parenting in the development of conduct problems, it is important to expose specific hypotheses to “severe risk of refutation” (Popper 1963). Hypotheses that consistently are not supported are refuted and must be revised or replaced (Platt 1964; Popper 1963). In the present analyses, we tested a number of specific hypotheses based on theory and previous studies regarding infant temperament, early parenting, and their interactions. In addition, because much basic information remains to be learned in this area, systematic exploratory tests of main effects and interactions were conducted to induce additional hypotheses to be tested in future studies.

### Infant Temperament

The present findings supported the specific hypothesis that maternal ratings of infant fussiness during the first year of life predict future maternal ratings of child conduct problems (Shaw et al. 2000; Teerikangas et al. 1998). In addition, these findings supported the more general hypothesis that two other temperament dimensions (infant activity level and predictability) that also may contribute to what has been referred to as “difficult temperament” (Guerin et al. 1997; Olson et al. 2000) independently predict future conduct problems. No hypothesis was offered regarding infant positive affect, but it also was found to independently predict future conduct problems inversely. The supplemental effect size estimates for the dimensions of infant temperament illustrated in Fig. 1 reached the “medium effect” size range for infant fussiness and predictability, but were in the “small effect” size range for fearfulness and positive affect (Rosnow and Rosenthal 1996). The prediction of future conduct problems from infant temperament in the first year of life is generally consistent with the broad hypothesis that early individual differences in challenging temperament foster transactions with the environment that increase risk for early-onset conduct problems (Lahey and Waldman 2003; Moffitt

1993; Shaw et al. 2000). In addition, the dimensions of fussiness and fearfulness refer to high levels of crying, irritability, and difficulty being soothed, which may reflect deficits in early emotion regulation (Kopp 1989; Keenan 2000; Keenan and Shaw 2003). Thus, these findings also are consistent with the hypothesis that deficits in early childhood emotion regulation foster later conduct problems (Keenan and Shaw 2003).

All interactions among the five temperament dimensions were tested without specific hypotheses, but the only significant interaction was between infant fussiness and predictability. This indicated that infants who were “easy” in the sense of being both low in fussiness and high in predictability were found to be at particularly low risk for childhood conduct problems.

There were few significant differences among race–ethnic groups in the extent to which infant temperament predicted future conduct problems. The one exception was that infant fussiness predicted child conduct problems better in non-Hispanic European American families than among Hispanic families. If this finding is replicated, it will be important to explore its meaning in studies of the interface between child characteristics, culture, and other factors correlated with culture, such as income and neighborhood characteristics.

Consistent with the hypothesis of Keenan and Shaw (1997), we found few significant sex differences in mean levels of maternal ratings of temperament during the first year of life. Nonetheless, like Gartstein and Rothbart (2003) and Martin et al. (1997), we found that infant girls were rated slightly but significantly higher on fearfulness. In addition, we found two significant sex differences in the extent to which infant temperament predicted future childhood conduct problems that had not been reported previously. First, as hypothesized (Shaw et al. 2000), infant fussiness predicted future conduct problems more strongly in males than in females. Indeed, the effect size for the prediction of child conduct problems from infant fussiness in males ( $\rho_{ES}=0.27$ ; “large effect”) was nearly twice as large as for girls ( $\rho_{ES}=0.15$ ; “medium effect”).

It is not currently clear why fussiness would be a stronger predictor of future conduct problems in boys than girls. Some developmental models suggest that female and male infants are at similar risk for future conduct problems based on their infant characteristics, but sex differences in socialization beginning in toddlerhood reduce risks for female children more than for male children (Crick and Zahn-Waxler 2003; Keenan and Shaw 1997). Because there is evidence that girls may have a higher threshold for the same causal influences on antisocial behavior as boys (Van Hulle et al. 2007), understanding the reasons for such sex differences should be a priority for studies of the early origins of conduct problems.

The second interaction between a dimension of temperament and the infant's sex was not predicted. Infant fearfulness accounted for slightly but significantly more variance in the prediction of child conduct problems among female than male infants. Note that fearfulness was a positive predictor of future conduct problems in both sexes. This suggests that the disposition among infants to cry or turn away in response to some kinds of changes in the environment (e.g., presence of strangers, animals, or the caretaker leaving the room) does not index an inhibitory process in the same way that ratings of fearfulness do at age 3 years (e.g., Raine et al. 1998). Some support for this possibility can be seen in the almost complete lack of stability of maternal ratings of social fearfulness from 18 months to 4 years (Goldsmith 1996). That is, it is possible that what is measured as "fearfulness" changes from infancy through the preschool period. Therefore, fearfulness during the first year of life may reflect irritability rather than an inhibitory process. The present finding that fearfulness and fussiness are positively correlated at  $r=0.34$  (Table 2) is consistent with this interpretation.

#### Parenting During the First Year of Life

**Cognitive Stimulation** The present findings supported the hypothesis based on Olson et al. (2000) that greater mother-reported cognitive stimulation of the infant protects infants from future conduct problems (Fig. 4). Indeed, the supplemental effect size estimate was in the "large effect" range for cognitive stimulation (Rosnow and Rosenthal 1996). Why would cognitive stimulation during the first year of life robustly predict lower levels of future childhood conduct problems? One possibility is that the HOME-SF measure of cognitive stimulation broadly reflects affectionate and caring parenting (Pettit and Bates 1989) rather than only cognitive stimulation. A second possibility is that cognitive stimulation in infancy may facilitate language development, which is consistent with the hypothesis that individual differences in early language development play an important role in socialization (Keenan and Shaw 1997, 2003). That is, young children with better developed language may be better able to understand adult commands and express their needs to adults.

A third possibility is that language development may facilitate the development of emotion regulation (Kopp 1989, 1992). Young children with better-developed language understand concepts of emotion better and display better emotion regulation than children with less advanced language skills (Eisenberg et al. 2005; Kopp 1992). In the context of the latter two hypotheses, it is important that Bradley et al. (2001a) found that HOME-SF cognitive stimulation was correlated with the development of recep-

tive vocabulary in the CNLSY. This is consistent with the possibility that maternal cognitive stimulation operates partly by stimulating language development, which facilitates both the socialization of behavior and the development of emotion regulation. Fourth, because early parental cognitive stimulation may foster school readiness (Brooks-Gunn and Markman 2005) and because poor achievement in reading may increase risk for conduct problems (Trzesniewski et al. 2006), it is possible that early cognitive stimulation indirectly influences risk through academic achievement.

**Maternal Responsiveness** In tests of main effects, we did not support the hypothesis that interviewer-rated maternal responsiveness during infancy inversely predicts childhood conduct problems (Kochanska 1997a; Olson et al. 2000; Shaw et al. 2000; Wakschlag and Hans 1999). As discussed below, however, there was a significant interaction between maternal responsiveness and infant fearfulness.

**Spanking** The present findings did not confirm at the  $p < 0.05$  level the prediction that spanking during infancy would predict future childhood conduct problems better among non-Hispanic European American than African American families (Deater-Deckard and Dodge 1997). The unpredicted finding that spanking inversely predicts future conduct problems more strongly among non-Hispanic European American infants than Hispanic infants had not been reported previously (Fig. 5). It will be important to continue to investigate possible cultural differences in the effects of spanking during infancy, but the effect sizes for spanking appear to be small even among non-Hispanic European American families.

#### Parenting-by-Temperament Interactions

The present findings did not support the hypothesis of an interaction between maternal responsiveness and infant fussiness (Bates et al. 1998; Keenan 2000; Sanson et al. 1993; Shaw et al. 2000). We did find, however, a significant interaction between early maternal responsiveness and another aspect of temperament. As illustrated in Fig. 6, this indicated that maternal responsiveness was a robust predictor of future conduct problems among infants low in fearfulness (with an effect size in the "large" range), but was a weak predictor among fearful infants. It should be noted, however, that this interaction reflects the reverse of an earlier finding with older children. Kochanska (1997b) found that maternal responsiveness at age 2–3 years predicted less cheating in games by 4-year-old children, but did so more strongly in children who were highly fearful at 2–3 years. Again, this suggests that the construct



measured as fearfulness in infancy, which is positively related to conduct problems, may be quite different from fearfulness measured at 2–3 years. Unexpected interactions between spanking and fussiness and between spanking and positive affect also were significant, but the differences in effect sizes for spanking at different levels of these temperament were quite small.

### Prevention Trials

The current findings are consistent with the hypothesis that interventions focusing on parenting during the first year of life would be beneficial in preventing future child conduct problems. Indeed, two controlled trials already have demonstrated that interventions designed to teach mothers of high-risk infants to be more responsive (and to engage in some parenting behaviors related to cognitive stimulation) yield both changes in parenting and concomitant improvements in the infant's cognitive development, emotional behavior, and cooperation through the preschool years (Landry et al. 2006; van den Boom 1995). The results of these studies using randomized assignment to groups support the current finding that maternal parenting early in life predicts later conduct problems. When studies with different measurement strategies and different threats to their validity yield converging findings they increase confidence in those findings (Shadish et al. 2002). Conversely, the present findings of long-term prediction of childhood conduct problems from infant parenting strongly support longer-term evaluations of these programs to determine if they reduce conduct problems during later childhood. The present findings also suggest that greater emphasis be placed on increasing maternal cognitive stimulation in such early intervention programs and suggests that interactions with infant temperament should be considered in their design and evaluation.

### Limitations

The high cost of longitudinal studies of large and representative samples such as the CNLSY can be justified only when the studies serve multiple scientific and policy purposes. As a result, it is rarely possible to use lengthy questionnaires and the kinds of time-intensive observational measures of temperament and parenting that are feasible in smaller studies (e.g., Bates et al. 1998). Each type of study has its important advantages, however. Large studies of nationally representative samples have considerable statistical power (e.g., for detecting statistical interactions) and have high external validity because their results can be generalized to entire population (Shadish et al. 2002). In contrast, the results of smaller studies have less statistical

power and their samples typically are not fully representative of any population of reference. Furthermore, small samples usually do allow comparisons of subgroups (e.g., sexes or race-ethnic groups), which are often possible in larger studies. Nonetheless, the strong measurement possible in smaller samples affords them considerable internal validity. Smaller studies are ideal for conducting detailed examinations of constructs measured more briefly in larger studies. When the results of the two kinds of studies converge, their complementary advantages support strong conclusions (Shadish et al. 2002).

The measures of parenting used in the present analyses were largely based on maternal reports and concerns have been raised that maternal reports of parenting could be biased by recall requirements, social desirability, and parental mood (Morsbach and Prinz 2006). Although such biases may exist, maternal reports of parenting are substantially correlated with fathers' independent reports of the mother's parenting and by moderate to substantial correlations between maternal reports and observational measures of the same parenting behaviors (Lovejoy et al. 1999; Morsbach and Prinz 2006). These validating correlations are more impressive when one considers that observational measures of parenting are subject to biases themselves, including reactivity to the presence of observers, being limited to brief samples of parenting behavior, and a reduced ability to observe low-rate but important behaviors such as spanking. Nonetheless, the CNLSY did not include the kinds of independent observational assessments of mother-infant interactions that are free of maternal report biases (Keenan 2000; Tronick 1989). As a result, it is entirely possible that the predictive association between maternal reports of parenting and maternal ratings of later child conduct problems was affected by common method variance. Indeed, the major limitation of the present analyses is that both were based on maternal reports. It will be very important, therefore, to examine relations between infant temperament and reports of conduct problems from other informants in the future to rule out common method variance as an explanation.

Although the rate of spanking (i.e., 8.1%) was low in the present sample, it is well within the expected range for this age group. Nonetheless, the low rate of spanking raises the possibility of floor effects. That is, a less strict and discrete measure of negative discipline might have yielded a stronger predictive effect because of less restriction of range in the predictor. Nonetheless, there is evidence from a study of preschool children that "normative" levels of physical punishment at that age mostly reflect child effects on parental discipline rather than causal effects of physical punishment on child conduct problems (Jaffee et al. 2004). This raises the possibility that it may actually be desirable to use a strict definition of spanking during infancy.

Precisely because it is uncommon, spanking an infant under 1 year of age may be extreme enough to function causally like harsher physical maltreatment does later in later childhood (Jaffee et al. 2004).

Because the items used to operationalize infant temperament and parenting were selected from broader pools of available items, it is possible that other sets of items might have provided even stronger predictions of future conduct problems. Mother-rated infant fussiness was the strongest predictor in both the present study and in Teerikangas et al. (1998). Nonetheless, because not all items from the IBQ (Rothbart 1981) were used, it is possible that other unmeasured dimensions of temperament might be even more predictive than fussiness.

### Implications for Future Research

Keeping the limitations of the present study in mind, it is striking that temperament and parenting during the first year of life predicted future conduct problems through age 13 years. Because the antisocial behavior with the greatest social impact does not occur until after puberty, however, it will be important to follow up the present analyses after sufficient data on adolescent outcomes in the CNLSY are released in late 2008. This also will have the advantage of allowing tests of predictions based on maternal reports of infant temperament and parenting using antisocial behavior assessed by youth self report as the outcome variable. In addition, these data will allow infant temperament and parenting to be linked to developmental trajectories of antisocial behavior across childhood and adolescence. Moffitt (1993) hypothesized that difficult infant temperament and early maladaptive parenting predicts only early-onset, stable antisocial behavior, but few data are yet available to test this hypothesis.

Perhaps the most important questions for future research is *why* early temperament and parenting predict child conduct problems. The CNLSY is a genetically informative sample that includes twins, full siblings, half siblings, and cousin pairs. This will allow us to use quasi-experimental analyses to test hypothesized causal links between infant factors and future conduct problems. It is clear that both temperament and conduct problems are moderately heritable (Goldsmith et al. 1997; Rhee and Waldman 2002; Saudino 2005). There also is evidence that maternal ratings of temperamental emotionality at 14–36 months predict maternal ratings of externalizing problems at age 4 years largely because of shared genetic influences (Schmitz et al. 1999). Much remains to be learned about the mechanisms through which infant temperament, parenting during infancy, and later conduct problems are related, however.

**Acknowledgements** This research was supported by grant R01 MH070025. The research was approved by the Institutional Review Board of each author's institution.

### References

- Achenbach, T. M. (1978). The child behavior profile: I. Boys aged 6–11. *Journal of Consulting and Clinical Psychology, 46*, 478–488.
- Aguilar, B., Sroufe, L. A., Egeland, B., & Carlson, E. (2000). Distinguishing the early-onset/persistent and adolescence-onset antisocial behavior types: From birth to 16 years. *Development and Psychopathology, 12*, 109–132.
- Anderson, K. E., Lytton, H., & Romney, D. M. (1986). Mothers' interaction with normal and conduct-disordered boys: Who affects whom? *Developmental Psychology, 22*, 604–609.
- Bates, J. E., Freeland, C. A., & Lounsbury, M. L. (1979). Measurement of infant difficultness. *Child Development, 50*, 794–803.
- Bates, J. E., Pettit, G. S., Dodge, K. A., & Ridge, B. (1998). Interaction of temperamental resistance to control and restrictive parenting in the development of externalizing behavior. *Developmental Psychology, 34*, 982–995.
- Baydar, N. (1995). Reliability and validity of temperament scales of the NLSY child assessments. *Journal of Applied Developmental Psychology, 16*, 339–370.
- Bell, R. Q. (1977). Socialization findings re-examined. In R. Q. Bell, & R. V. Harper (Eds.), *Child effects on adults* (pp. 53–84). Hillsdale, NJ: Erlbaum.
- Bornstein, M. H. (2002). Parenting infants. In M. H. Bornstein (Ed.), *Handbook of parenting: Children and parenting* (vol. 1, 2nd ed., pp. 3–43). Mahwah, NJ: Lawrence Erlbaum.
- Bornstein, M. H., Hahn, C. S., Suwalsky, J. T. D., & Haynes, O. M. (2003). Socioeconomic status, parenting, and child development: The Hollingshead Four-Factor Index of Social Status and the Socioeconomic Index of Occupations. In M. H. Bornstein, & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 29–82). Mahwah, NJ: Lawrence Erlbaum.
- Bowlby, J. (1958). The nature of the child's tie to his mother. *International Journal of Psychoanalysis, 39*, 350–373.
- Bradley, R. H., & Caldwell, B. M. (1984). The HOME Inventory and family demographics. *Developmental Psychology, 20*, 315–320.
- Bradley, R. H., Caldwell, B. M., & Corwyn, R. F. (2003). The Child Care HOME Inventories: Assessing the quality of family child care homes. *Early Childhood Research Quarterly, 18*, 294–309.
- Bradley, R. H., Corwyn, R. F., Burchinal, M., Pipes McAdoo, H., & García Coll, C. (2001a). The home environments of children in the United States Part II: Relations with behavioral development through age thirteen. *Child Development, 72*, 1868–1886.
- Bradley, R. H., Corwyn, R. F., Pipes McAdoo, H., & García Coll, C. (2001b). The home environments of children in the United States Part I: Variations by age, ethnicity, and poverty status. *Child Development, 72*, 1844–1867.
- Brooks-Gunn, J., & Markman, L. B. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *Future of Children, 15*, 139–168.
- Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development*. New York: Wiley.
- Caldwell, B. M., & Bradley, R. H. (1984). *Home observation for measurement of the environment*. Little Rock, AR: University of Arkansas at Little Rock.
- Carey, W. (1970). A simplified method of measuring infant temperament. *Journal of Pediatrics, 77*, 188–194.

- Caspi, A. (2000). The child is father of the man: Personality continuities from childhood to adulthood. *Journal of Personality and Social Psychology*, 78, 158–172.
- Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1996). Behavioral observations at age 3 predict adult psychiatric disorders: Longitudinal evidence from a birth cohort. *Archives of General Psychiatry*, 53, 1033–1039.
- Chase-Lansdale, P. L., Mott, F. L., Brooks-Gunn, J., & Phillips, D. A. (1991). Children of the National Longitudinal Survey of Youth: A unique research opportunity. *Developmental Psychology*, 27, 918–931.
- Chess, S., & Thomas, A. (1984). *Origins and evolution of behavior disorders: From infancy to early adult life*. Cambridge, MA: Harvard University Press.
- Cohen, P. (1982). To be or not to be: Control and balancing of Type I and Type II errors. *Evaluation and Program Planning*, 5, 247–253.
- Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, 49, 997–1003.
- Colder, C. R., Mott, J. A., & Berman, A. S. (2002). The interactive effects of infant activity level and fear on growth trajectories of early childhood behavior problems. *Development and Psychopathology*, 14, 1–24.
- Crick, N. R., & Zahn-Waxler, C. (2003). The development of psychopathology in females and males: Current progress and future challenges. *Development and Psychopathology*, 15, 719–742.
- Deater-Deckard, K., & Dodge, K. A. (1997). Externalizing behavior problems and discipline revisited: Nonlinear effects and variation by culture, context, and gender. *Psychological Inquiry*, 8, 161–175.
- D’Onofrio, B. M., Van Hulle, C. A., Waldman, I. D., Rodgers, J. L., Harden, K. P., Rathouz, P. J., et al. (2008). Smoking during pregnancy and offspring externalizing problems: An exploration of genetic and environmental confounds. *Development and Psychopathology*, 20, 139–164.
- Dunn, W. (2002). *The infant/toddler sensory profile manual*. San Antonio, TX: The Psychological.
- Eisenberg, N., Sadovsky, A., & Spinrad, T. L. (2005). Associations of emotion-related regulation with language skills, emotion knowledge, and academic outcomes. *New Directions in Child and Adolescent Development*, 109, 109–118.
- Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant Behavior and Development*, 26, 64–86.
- Goldsmith, H. H. (1996). Studying temperament via construction of the Toddler Behavior Assessment Questionnaire. *Child Development*, 67, 218–235.
- Goldsmith, H. H., Buss, K. A., & Lemery, K. S. (1997). Toddler and childhood temperament; expanded content, stronger genetic evidence, new evidence for the importance of environment. *Developmental Psychology*, 33, 891–905.
- Goldsmith, H. H., Van Hulle, C. A., Arneson, C. L., Schreiber, J. E., & Gernsbacher, M. A. (2006). A population based twin study of parentally reported tactile and auditory defensiveness in young children. *Journal of Abnormal Child Psychology*, 34, 393–407.
- Guerin, D. W., Gottfried, A. W., & Thomas, C. W. (1997). Difficult temperament and behaviour problems: A longitudinal study from 1.5 to 12 years. *International Journal of Behavioral Development*, 21, 71–90.
- Jaffee, S. R., Caspi, A., Moffitt, T. E., Polo-Tomas, M., Price, T. S., & Taylor, A. (2004). The limits of child effects: Evidence for genetically mediated child effects on corporal punishment but not on physical maltreatment. *Developmental Psychology*, 40, 1047–1058.
- Karrass, J., VanDeventer, M. C., & Braungart-Rieker, J. M. (2003). Predicting shared parent–child book reading in infancy. *Journal of Family Psychology*, 17, 134–146.
- Keenan, K. (2000). Emotion dysregulation as a risk factor for child psychopathology. *Clinical Psychology: Science and Practice*, 7, 418–434.
- Keenan, K., & Shaw, D. (1997). Developmental and social influences on young girls’ early problem behavior. *Psychological Bulletin*, 121, 95–113.
- Keenan, K., & Shaw, D. S. (2003). Starting at the beginning: Exploring the etiology of antisocial behavior in the first years of life. In B. B. Lahey, T. E. Moffitt, & A. Caspi (Eds.), *Causes of conduct disorder and juvenile delinquency* (pp. 153–181). New York: Guilford.
- Kochanska, G. (1997a). Mutually responsive orientation between mothers and their young children: Implications for early socialization. *Child Development*, 68, 94–112.
- Kochanska, G. (1997b). Multiple pathways to conscience for children with different temperaments: From toddlerhood to age 5. *Developmental Psychology*, 33, 228–240.
- Kopp, C. B. (1989). Regulation of distress and negative emotions: A developmental view. *Developmental Psychology*, 25, 343–354.
- Kopp, C. (1992). Emotional distress and control in young children. In N. Eisenberg & R. A. Fabes (Eds.), *Emotion and its regulation in early development: New directions in child development* (vol. 55, pp. 41–56). San Francisco: Jossey-Bass.
- Lahey, B. B., & Waldman, I. D. (2003). A developmental propensity model of the origins of conduct problems during childhood and adolescence. In B. B. Lahey, T. E. Moffitt, & A. Caspi (Eds.), *Causes of conduct disorder and juvenile delinquency* (pp. 76–117). New York: Guilford Press.
- Lahey, B. B., Miller, T. L., Gordon, R. A., & Riley, A. (1999). Developmental epidemiology of the disruptive behavior disorders. In H. Quay, & A. Hogan (Eds.), *Handbook of the disruptive behavior disorders* (pp. 23–48). New York: Plenum.
- Lahey, B. B., Van Hulle, C. A., Waldman, I. D., Rodgers, J. L., D’Onofrio, B. M., Pedlow, S., et al. (2006). Testing descriptive hypotheses regarding sex differences in the development of conduct problems and delinquency. *Journal of Abnormal Child Psychology*, 34, 737–755.
- Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills. *Developmental Psychology*, 42, 627–642.
- Lerner, J. V., Nitz, K., Talwar, R., & Lerner, R. M. (1989). On the functional significance of temperamental individuality: A developmental contextual view of the concept of goodness of fit. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 509–522). West Sussex: Wiley.
- Linver, M. R., Brooks-Gunn, J., & Kohen, D. E. (2002). Family processes as pathways from income to young children’s development. *Developmental Psychology*, 38, 719–734.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & Van Kammen, W. (1998). *Antisocial behavior and mental health problems: Explanatory factors in childhood and adolescence*. Mahwah, NJ: Lawrence Erlbaum.
- Lovejoy, M. C., Weis, R., O’Hare, E., & Rubin, E. C. (1999). Development and initial validation of the Parent Behavior Inventory. *Psychological Assessment*, 11, 534–545.
- Lytton, H. (1990). Child and parent effects in boys’ conduct disorder: A reinterpretation. *Developmental Psychology*, 26, 683–697.
- Maccoby, E. E. (1992). The role of parents in the socialization of children: An historical overview. *Developmental Psychology*, 28, 1006–1017.
- Martin, R., Wisenbaker, J., Baker, J., & Huttunen, M. (1997). Gender differences in temperament at six months and five years. *Infant Behavior and Development*, 20, 339–347.



- McLoyd, V. C., & Smith, J. (2002). Physical discipline and behavior problems in African American, European American, and Hispanic children: Emotional support as a moderator. *Journal of Marriage and Family*, *64*, 40–53.
- Meaney, M. J., & Szyf, M. (2005). Maternal care as a model for experience-dependent chromatin plasticity? *Trends in Neuroscience*, *28*, 456–463.
- Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*, 674–701.
- Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. (2001). *Sex differences in antisocial behaviour*. Cambridge, UK: Cambridge University Press.
- Morsbach, S. K., & Prinz, R. J. (2006). Understanding and improving the validity of self-report of parenting. *Clinical Child and Family Psychology Review*, *9*, 1–21.
- Mott, F. L., Baker, P. C., Ball, D. E., Keck, C. K., & Lenhart, S. M. (1995). *The NLSY children 1992: Description and evaluation*. Columbus, OH: Ohio State University Center for Human Resource Research.
- Olson, S. L., Bates, J. E., Sandy, J. M., & Lanthier, R. (2000). Early developmental precursors of externalizing behavior in middle childhood and adolescence. *Journal of Abnormal Child Psychology*, *28*, 119–133.
- Patterson, G. R. (1982). *Coercive family processes*. Eugene, OR: Castalia.
- Peterson, J. L., & Zill, N. (1986). Marital disruption, parent-child relationships, and behavior problems in children. *Journal of Marriage and the Family*, *48*, 295–307.
- Pettit, G. S., & Bates, J. E. (1989). Family interaction patterns and children's behavior problems from infancy to 4 years. *Developmental Psychology*, *25*, 413–420.
- Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353.
- Popper, K. (1963). *Conjectures and refutations*. London: Routledge.
- Prior, M., Smart, D., Sanson, A., & Oberklaid, F. (2001). Longitudinal predictors of behavioural adjustment in pre-adolescent children. *Australian and New Zealand Journal of Psychiatry*, *35*, 297–307.
- Raine, A., Reynolds, C., Venables, P. H., Mednick, S. A., & Farrington, D. P. (1998). Fearlessness, stimulation-seeking, and large body size at age 3 years as early predispositions to childhood aggression at age 11 years. *Archives of General Psychiatry*, *55*, 745–751.
- Rhee, S. H., & Waldman, I. D. (2002). Genetic and environmental influences on antisocial behavior: A meta-analysis of twin and adoption studies. *Psychological Bulletin*, *128*, 490–529.
- Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counterexamples on other people's published data: General procedures for research consumers. *Psychological Methods*, *1*, 331–340.
- Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development*, *2*, 569–578.
- Rothbart, M. K., & Bates, J. E. (1998). Temperament. In N. Eisenberg (Ed.), *Handbook of child psychology: Social, emotional, and personality development* (vol. 3, 5th ed., pp. 105–176). New York: Wiley.
- Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. *Social Development*, *13*, 142–170.
- Sanson, A., Prior, M., Smart, D., & Oberklaid, F. (1993). Gender differences in aggression in childhood: Implications for a peaceful world. *Australian Psychologist*, *28*, 86–92.
- Saudino, K. S. (2005). Behavioral genetics and child temperament. *Developmental and Behavioral Pediatrics*, *26*, 214–223.
- Schmitz, S. S., Fulker, D. W., Plomin, R., Zahn-Waxler, C., Emde, R., & DeFries, J. C. (1999). Temperament and problem behavior during early childhood. *International Journal of Behavioral Development*, *23*, 333–355.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Shaw, D. S., Bell, R. Q., & Gilliom, M. (2000). A truly early starter model of antisocial behavior revisited. *Clinical Child and Family Psychology Review*, *3*, 155–172.
- Shaw, D. S., Winslow, E. B., Owens, E. B., Vondra, J. I., Cohn, J. F., & Bell, R. Q. (1998). The development of early externalizing problems among children from low-income families: A transformational perspective. *Journal of Abnormal Child Psychology*, *26*, 95–107.
- Slade, E. P., & Wissow, L. S. (2004). Spanking in early childhood and later behavior problems: A prospective study of infants and young toddlers. *Pediatrics*, *113*, 1321–1330.
- Teerikangas, O. M., Aronen, E. T., Martin, R. P., & Huttunen, M. O. (1998). Effects of infant temperament and early intervention on the psychiatric symptoms of adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*, 1070–1076.
- Thomas, A., & Chess, S. (1977). *Temperament and development*. New York: Brunner/Mazel.
- Tronick, E. Z. (1989). Emotions and emotional communication in infants. *American Psychologist*, *44*, 112–119.
- Trzesniewski, K. H., Moffitt, T. E., Caspi, A., Taylor, A., & Maughan, B. (2006). Revisiting the association between reading achievement and antisocial behavior: New evidence of an environmental explanation from a twin study. *Child Development*, *77*, 72–88.
- van den Boom, D. C. (1995). Do first-year intervention effects endure? Follow-up during toddlerhood of a sample of Dutch irritable infants. *Child Development*, *66*, 1798–1816.
- Van Hulle, C. A., Rodgers, J. L., D'Onofrio, B. M., Waldman, I. D., & Lahey, B. B. (2007). Sex differences in the causes of self-reported adolescent delinquency. *Journal of Abnormal Psychology*, *116*, 236–248.
- Wakschlag, L. S., & Hans, S. L. (1999). Relation of maternal responsiveness during infancy to the development of behavior problems in high-risk youths. *Developmental Psychology*, *35*, 569–579.
- Zeger, S. L., & Liang, K.-Y. (1996). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*, 121–130.