Developmental Psychology
2010, Vol. 46, No. 2, 379–390

0012-1649/10/$12.00 DOI: 10.1037/a0015203

Differential Susceptibility to Parenting and Quality Child Care

Michael Pluess and Jay Belsky
Birkbeck University of London

Research on differential susceptibility to rearing suggests that infants with difficult temperaments are disproportionately affected by parenting and child care quality, but a major U.S. child care study raises questions as to whether quality of care influences social adjustment. One thousand three hundred sixty-four American children from reasonably diverse backgrounds were followed from 1 month to 11 years with repeated observational assessments of parenting and child care quality, as well as teacher report and standardized assessments of children's cognitive-academic and social functioning, to determine whether those with histories of difficult temperament proved more susceptible to early rearing effects at ages 10 and 11. Evidence for such differential susceptibility emerges in the case of both parenting and child care quality and with respect to both cognitive-academic and social functioning. Differential susceptibility to parenting and child care quality extends to late middle childhood. J. Belsky, D. L. Vandell, et al.'s (2007) failure to consider such temperament-moderated rearing effects in their evaluation of long-term child care effects misestimates effects of child care quality on social adjustment.

Keywords: differential susceptibility, temperament, child care, parenting

How child characteristics of individuality and rearing experiences interact in shaping children's functioning is a topic of long-standing interest to students of child development (e.g., Rothbart & Bates, 2006; Wachs & Gandour, 1983). Recently, much work in this area, especially that concerning the interaction of parenting and temperament, has been recast in differential susceptibility terms (Belsky, 2005; Belsky, Bakermans-Kranenburg, & van IJzendoorn, 2007). Central to Belsky's (1997a, 1997b, 2005) evolutionary-inspired differential susceptibility hypothesis is the proposition that children vary in the extent to which they are affected—for better and for worse—by their rearing experiences, because in ancestral times (and even possibly to this day) this would have had reproductive fitness payoffs for parents, children, and their siblings. More specifically, in cases in which parental influence proved beneficial from a fitness standpoint, not only would the affected children have had their fitness enhanced but so would parents and even unaffected siblings as a result of shared genes (i.e., inclusive fitness). By the same token, in cases in which parental influence (inadvertently) undermined affected children's fitness, their unaffected siblings would have benefited from not being as penalized and, owing to shared genes, so would have the parents and (affected) siblings. In sum, natural selection should have shaped human development such that some children in a family would be more influenced than others by their rearing experiences (see also Boyce & Ellis, 2005).

After theoretically deriving the general hypothesis that children would differ in their susceptibility to parental influence on the basis of evolutionary first principles (see Belsky, 2005; Belsky & Pluess, in press), Belsky (1997a, 1997b) observed that highly negatively emotional infants and toddlers, as well as those scoring high on measures of "difficult" temperament, appeared disproportionately susceptible to rearing (for review, see Belsky, 2005)—and this was true whether infant negativity–difficulty measurements were based on behavioral assessments carried out in the laboratory or on maternal reports of temperament. More specifically, characteristics generally associated with difficult temperament (Chess & Thomas, 1989)—proneness to negative emotional expression, low adaptability, high activity, and low emotional regulation—appear to moderate effects of early experiences. For example, Kochanska and colleagues (Kochanska, 1993; Kochanska, Clark, & Goldman, 1997) found that the quality of maternal discipline—gentle guidance versus forceful control—accounted for substantially more variance in the self-control of infants and toddlers who scored high rather than low on a behavioral measure of fearfulness. In work explicitly testing the differential susceptibility hypothesis, Belsky, Hsieh, and Crnic (1998) observed that the combined power of mothering and fathering across the second and third year of life in predicting 36-month externalizing prob-

Michael Pluess and Jay Belsky, Institute for the Study of Children, Families and Social Issues, Birkbeck University of London, London, United Kingdom.

Preparation of this article was supported by a grant (PBBS1-120809) for doctoral studies of the Swiss National Science Foundation awarded to Michael Pluess. The research described herein was supported by a cooperative agreement (U10-HD25420) with the National Institute of Child Health and Human Development (NICHD). This article was the result of a collaboration of the two named authors using data collected under the direction of the NICHD network authors that has been placed in the public domain. The NICHD network of authors merits our appreciation for ensuring that these data were gathered and made available to all network authors and others who received permission to conduct scientific studies using them. The NICHD network authors, however, have no responsibility for how we have analyzed the data, the results we report, and the conclusions we draw. We would like to express our gratitude to Margaret Burchinal and Antero Malin for their helpful support in the statistical analyses.

Correspondence concerning this article should be addressed to Jay Belsky, Institute for the Study of Children, Families and Social Issues, Birkbeck University of London, 7 Bedford Square, London WC1B 3RA, United Kingdom. E-mail: j.belsky@psychology.bbk.ac.uk

lems and inhibited behavior was substantially greater for children who manifested high levels of distress in the Strange Situation at 12 months than for those who did not. Early et al. (2002) replicated the inhibition results in an independent study of the transition to school: Whereas higher levels of maternal sensitivity forecast lower levels of inhibition during the first month of school in the case of children classified as wary, inhibited, and negative in infancy—after controlling for early wariness—no such parenting–school adjustment linkage emerged in the case of other children. As it turns out, it is not just studies using observational measures of infant negative emotionality (e.g., fear, inhibition, or wariness) that provide evidence of differential susceptibility but also studies relying on maternal reports of infant temperament. For example, Feldman, Greenbaum, and Yirmiya (1999) linked synchrony in mother–infant interaction during face-to-face exchanges when infants were 9 months of age with self-control at 2 years of age, measured by compliance with maternal cleanup requests and capacity to refrain from eating candy until granted permission. In accordance with the differential susceptibility hypothesis, mutually synchronous mother–infant interaction predicted greater self-control more strongly in the case of children who, according to maternal reports, had had more difficult temperaments in infancy than those who did not.

Evidence of differential susceptibility is not restricted to investigations of parenting, as several recent studies show that effects of early child care experience also seem to vary as a function of child temperament. Blair (2002) reported, for example, that the beneficial effects of an early intervention program for low-birth-weight, preterm babies involving, among other things, high-quality, center-based child care was restricted to or most pronounced in the case of those scoring high in negativity as infants, as reported by mothers. More recently, a reanalysis of 54-month and first-grade data from the large-scale National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD) revealed that effects of quality of care across the first 4.5 years of life on problem behavior and social skills were evident exclusively in the case of children scoring high on difficult temperament in the first 6 months of life (Pluess & Belsky, in press). Interestingly, no such temperamentally moderated effects of child care emerged when quantity and type of care were considered, leading to the decision to restrict investigation of temperamentally moderated child care effects in this inquiry to quality of care.

It is also of interest that Bradley and Corwyn (2008) discerned similar evidence of differential susceptibility vis-à-vis (mother-reported) temperamental difficulty using NICHD SECCYD data when maternal sensitivity measured across the infant, toddler, and preschool years was the rearing predictor and problem behavior in first grade was the outcome to be explained. Relatedly, Stright, Gallagher, and Kelley (2008) found much the same when first-grade academic competence, social skills, student–teacher relationship, and peer status were the focus of inquiry.

Although all the just-cited evidence is consistent with the evolutionary-inspired theoretical prediction that children should vary in their susceptibility to environmental influences in a for-better-and-for-worse manner (Belsky, 1997a, 1997b, 2005; Belsky & Pluess, in press), the fact that it was only empirical evidence that highlighted the role of negative emotionality and difficult temperament in moderating rearing influences raises the question as to why this should be the case. Several nonmutually exclusive explanations have been advanced with regard to this fundamental issue. Drawing on primate evidence, Suomi (1995, 1997) suggested that highly fearful, inhibited, "up-tight" rhesus macaques spend more time than other young monkeys observing the world around them, thereby learning more than others about how to function effectively in their social world. Kochanska (1993) drew explicitly on Dienstbier's (1985) thinking on anxiety to account for her results, arguing that more negatively emotional, fearful, inhibited infants have lower thresholds for anxiety, thereby making them more easily aroused by discipline and thus responsive to it. Not unrelatedly, Belsky (2005) contended that a negatively emotional, difficult temperament reflects a highly sensitive nervous system on which experience—of both the positive and the negative variety—registers especially strongly (see also Aron & Aron, 1997). Finally, Boyce and Ellis (2005) hypothesized that a sensitive stress response system makes physiologically reactive children more biologically sensitive to context.

In view of all the theory and supporting evidence considered, the research reported herein sought to extend in several ways the investigation of differential susceptibility using data from the NICHD SECCYD. Most notably, we endeavored to determine whether temperament-moderated rearing effects detected through first grade remained evident when children were in fifth and sixth grade, implementing analytic models virtually identical to those employed by Belsky, Vandell, et al. (2007) and the NICHD Early Child Care Research Network (ECCRN) to evaluate child care effects through the end of middle childhood; the only change was the inclusion of interaction terms to test for differential susceptibility. Second, instead of investigating in separate inquiries differential susceptibility with respect to parenting and to child care, we sought to study both in a single inquiry. Finally, in addition to focusing on social functioning, we expanded the outcome set to include cognitive-academic achievement, using all the dependent variables analyzed by Belsky, Vandell, et al. (2007). Although most work highlighting moderating effects of early negativity or difficult temperament has focused upon social functioning (Belsky, 2005), this has not been the case exclusively (Blair, 2002), thereby giving rise to the possibility that maternal sensitivity and/or child care quality would disproportionately or exclusively prove related to the cognitive-academic functioning of children with histories of difficult temperament.

Important to appreciate, though, is that by also focusing upon social development, this study investigates the prospect that virtually all earlier analyses of child care quality effects using NICHD SECCYD data may have misestimated quality of care effects. That is, with the exception of the aforementioned differential susceptibility work by Pluess and Belsky (in press), prior investigations of child care effects by the NICHD ECCRN (1998, 2003b, 2005c, 2006; Belsky, Vandell, et al., 2007) dating back to when children were 3 years of age and through the age of 11 may have failed to detect quality-of-care effects on socioemotional adjustment because they did not evaluate differential susceptibility.

As in all prior NICHD SECCYD investigations of effects of child care, it is necessary in this inquiry to take into account selection effects. This is because it is widely appreciated and well established that variation in the quality, type, and quantity of care that children experience is not randomly assigned (NICHD

ECCRN, 2005a). By not then statistically controlling for important demographic and family covariates, it is possible that detected effects of child care could actually be effects of third variables that may influence family child care decision making and children's development; were this the case and selection effects not taken into account, effects of these demographic and family factors would masquerade as effects of child care. Although the inclusion of statistical controls cannot ensure that effects detected in a nonexperimental field study such as the present one are causal in nature, they do most certainly reduce this possibility. Because the present effort builds directly on the prior analysis of child care effects in the NICHD SECCYD through sixth grade and is specifically designed to determine whether results reported by Belsky, Vandell, et al. (2007) may have misestimated quality-of-care effects owing to the failure to take into account differential susceptibility, the covariates employed in this inquiry are the same as those used in that investigation; these are ones that have been found to be related to child care experience and children's development in prior research (NICHD ECCRN, 2005a).

Method

Participants

Families were recruited through hospital visits to mothers shortly after the birth of a child in 1991 in 10 locations in the United States. During selected 24-hr intervals, all women giving birth ($N = 8{,}986$) were screened for eligibility. From that group, 1,364 families completed a home interview when the infant was 1 month old and became the study participants. Details of the sampling plan can be found in NICHD ECCRN (2005a). In terms of demographic characteristics, 26% of the mothers had no more than a high school education at time of enrollment, 21% had incomes no greater than 200% of the poverty level at sixth grade, and 22% were minority (i.e., not non-Hispanic European American).

As with any longitudinal study, not all families participated in every wave of data collection. Relatively few families formally withdrew ($N = 291$, through sixth grade), but almost all had at least some missing data. Children with more complete data tended to be from families with more income, to have parents who provided more responsive and stimulating care, to experience more center care and spend more time in child care, and to show higher academic achievement scores over time than children with missing data. Children with and without missing data did not differ reliably on child care quality or teacher ratings. Imputation of missing data is described below.

Measures

Measures are described in terms of their roles in the analyses to be reported. A more detailed description of all included measures can be found in Belsky, Vandell, et al. (2007). Additional information about data collection procedures and measures is also documented in manuals of operation of the study, located at https://secc.rti.org. With the exception of the inclusion of the measurement of difficult temperament, all measures described are the same as those used in Belsky, Vandell, et al.'s evaluation of long-term child care effects.

Difficult temperament. Difficult temperament was assessed by maternal report at 6 months with an adapted version of the Infant Temperament Questionnaire (Carey & McDevitt, 1978). Mothers indicated on a 6-point rating scale ranging from *almost never* to *almost always* how often their baby's behavior corresponded to the behavior described in each item. The 55 selected items provided scores for five of the original nine subscales: (a) activity (e.g., "My baby plays actively with parents—much movement of arms, legs, body"), (b) approach (e.g., "My baby's initial reaction to a new babysitter is rejection [crying, clinging to mother, etc.]"), (c) adaptability (e.g., "My baby is still wary or frightened of strangers after 15 minutes"), (d) mood (e.g., "My baby cries when left to play alone"), and (e) intensity (e.g., "My baby reacts strongly to strangers: laughing or crying"). An overall score of difficult temperament was created by averaging the 55 items (after reflecting appropriate items) with higher values reflecting higher negative emotionality and greater temperamental difficulty. Internal consistency of the 55 items was $\alpha = .81$ with the present sample. In contrast to two studies that investigated differential susceptibility through the transition to school using NICHD SECCYD data (Bradley & Corwyn, 2008; Pluess & Belsky, in press), temperament scores at 6 months were not averaged with a similar measure at 1 month because preliminary hierarchical linear model (HLM) analyses showed that the 6-month measure yielded stronger interaction effects than the average of the two measurement occasions.

Parenting quality at 6–54 months. Parenting quality was assessed by maternal sensitivity in semistructured play and observation at home.

Mother–child interactions were videotaped in semistructured 15-min observations at 6, 15, 24, 36, and 54 months. The tasks provided a context for assessing age-appropriate qualities of maternal behavior. The observation task at the 6-month visit had two free-play components, one with toys available in the home and the other with a standard set of toys provided by the experimenter. At 15, 24, and 36 months, the videotaped session involved a three-boxes procedure in which the mother was asked to show the child age-appropriate toys in three containers in a set order (see Vandell, 1979). Data were collected by research assistants who had attended centralized training sessions.

Videotapes of the mother–child interactions were shipped to a central location for coding by raters blind to other information about the families. Intercoder reliability was determined by assigning two coders to 19%–20% of the tapes randomly drawn at each assessment period. Intercoder reliability was calculated as the intraclass correlation coefficient. Reliability for the composite scores used in the current report exceeded .83 at every age. At 6, 15, and 24 months, composite maternal sensitivity scores were created from the sums of three 4-point ratings (maternal sensitivity to child nondistress, intrusiveness [reversed], and positive regard). At 36 and 54 months, the maternal sensitivity composite was the sum of the three 7-point ratings of supportive presence, hostility (reversed), and respect for autonomy. Cronbach alphas exceeded .70 at every age.

The Home Observation for Measurement of the Environment (HOME; Caldwell & Bradley, 1984) was administered during home visits at 6, 15, 36, and 54 months. The focus is on the child *in* the environment, as a recipient of inputs from objects, events, and transactions occurring in connection with the family surround-

ings. The Infant/Toddler version of the inventory is aimed for use during infancy (birth to age 3). It is composed of 45 items clustered into six subscales: (a) Parental Responsivity, (b) Acceptance of Child, (c) Organization of the Environment, (d) Learning Materials, (e) Parental Involvement, and (f) Variety in Experience. Each item is scored in binary fashion (yes or no). Information used to score the items is obtained during the course of the home visit by means of observation and semi-structured interview. The Early Childhood version of the inventory is aimed for use during early childhood (age 3–6 years). It is composed of 55 items clustered into eight subscales: (a) Learning Materials, (b) Language Stimulation, (c) Physical Environment, (d) Responsivity, (e) Academic Stimulation, (f) Modeling, (g) Variety, and (h) Acceptance. Each item is scored in binary fashion (yes or no). A centrally located system of training was used for data collectors at each age. Cronbach alphas for the total score at each age exceeded .77.

The HOME and maternal sensitivity ratings were standardized at each age and then averaged at each age to create a composite score. Together these combined scores reflect parenting in two contexts: in the home and during semistructured play. Two indices of parenting quality at 6–54 months—intercept and slope—were created from the mean of the standardized scores at each age with an unconditional HLM analysis. For the present investigation, interactions between temperament and parenting quality were tested with individual intercepts of the parenting quality composite scores.

Child Care Characteristics

Nonmaternal child care. Was defined as regular care by anyone other than the mother—including fathers, relatives, nannies (whether in home or out of home), and family day care providers and centers. Besides child care quality, two further aspects of child care were measured from birth through 54 months: quantity of care and type of care.

Child care quality. Observational assessments were conducted in the primary child care arrangement at ages 6, 15, 24, 36, and 54 months. Quality was assessed during two half-day visits scheduled within a 2-week interval at 6–36 months and one half-day visit at 54 months. Observers completed four 44-min cycles of the Observational Record of the Caregiving Environment (ORCE) per child age through 36 months and two 44-min ORCE cycles at 54 months. Detailed descriptions of the ORCE assessments can be found in NICHD ECCRN (2002), including coding definitions, training procedures, and interobserver agreement. As with quantity, individual measures of level and change in quality were estimated with an unconditional HLM analysis. Many children were not in child care for at least three time points, so their slopes could not be computed, and preliminary analyses based on imputed data did not suggest that the quality slope contributed to analysis models. Therefore, following Belsky, Vandell, et al. (2007), whose analyses we replicate and extend, we included only the quality intercept (centered at 27 months) in the analyses presented.

Child care quantity. Parents reported children's hours of routine nonmaternal care during phone and personal interviews conducted at 3-month intervals through 36 months and at 4-month intervals thereafter, as well as the type of child care being used (see below). The hours spent in all settings were summed for each of the 17 intervals or "epochs" and parameterized on an hours-per-week basis. Individual measures of level and rate of change in quantity of care were computed as the individual intercepts and slopes from an unconditional HLM analysis of these 17 repeated measures. Age was centered at the measurement midpoint, 27 months, so the estimated intercept reflected that child's hours per week at 27 months of age.

Child care type. For each epoch, each of the child's care arrangements was classified as center, child care home (any home-based care outside the child's own home except care by grandparents), in-home care (any caregiver in the child's own home except father or grandparent), grandparent care, or father care. The proportion of epochs in which the child received care in a center for at least 10 hr per week and the proportion of epochs in a child care home for at least 10 hr per week were determined and used as variables to represent type of care.

Covariates

Early childhood. Measures of maternal, child, and family characteristics during infancy and early childhood were collected and used as controls for possible selection bias: maternal education (in years), the study child's gender and ethnicity, the proportion of (five measurement) epochs through 54 months in which the mother reported that a husband or partner was present, family income through 54 months calculated as the mean income-to-needs ratio, and the intercept and slope of maternal depressive symptoms assessed by the Center for Epidemiological Studies Depression Scale (Radloff, 1977) reported by the mother at 6, 15, 24, 36, and 54 months, determined by unconditional HLM analyses. These early childhood covariates were included in the reported HLM analyses as time-invariant controls.

Primary grades. Measures of family demographic and psychological characteristics also were obtained when children were in kindergarten and in first, third, and fifth grades. These were included as time-varying covariates in the HLM analyses of child outcomes. These factors were presence of a husband or partner in the household, income-to-needs ratio, maternal depressive symptoms, and parenting quality. Composite parenting quality scores for the primary grades were created similarly to parenting quality for 6–54 months (see above) by first averaging standardized ratings of observed maternal sensitivity and standardized ratings of observed home environmental quality (HOME; Caldwell & Bradley, 1984), which were assessed at 54 months and first (only maternal sensitivity), third, and fifth grades. As an index for parenting quality, unconditional individual HLM intercepts were then created from the mean of the standardized composite scores for each age and entered as time-varying concurrent control in the analysis. In addition to family-related covariates, classroom quality and after-school experience during the primary grades were included as controls. Children's classroom experiences were measured with the Classroom Observation System for First Grade (NICHD ECCRN, 2004b), the Classroom Observation System for Third Grade (NICHD ECCRN, 2005b), and the Classroom Observation System for Fifth Grade (NICHD ECCRN, 2004a). These observations focused on the classroom as well as the specific study child and his or her classroom experiences. In first grade, two 44-min observations were conducted during the morning. In third and fifth grades, classrooms were observed for eight 44-min cycles

distributed across the school day. Three 7-point global ratings of the classroom environment were made at the end of each observation cycle: overcontrol by teacher, teacher's emotional detachment, and teacher's sensitivity to student needs.

For more details on the derivation of these composites and the individual scales of which they are composed, see NICHD ECCRN (2004b, 2005b). Regarding after-school experience, mothers were interviewed by telephone in the fall and spring of kindergarten and first, third, and fifth grades about the study children's out-of-school care. They were asked a series of questions about a number of possible out-of-school care arrangements. In the present study analyses, hours of nonparental out-of-school care arrangements (here named after-school hours) were obtained for each school year from the average across the spring and fall reports of the total hours mothers reported across all nonparental out-of-school care arrangements.

Child Outcome

Cognitive-academic and social adjustment outcomes were assessed repeatedly over time. Standardized cognitive-academic achievement tests were administered at 54 months and at the end of first, third, and fifth grades. Caregiver–teacher reports of children's behavior problems were collected at 54 months, in kindergarten, and annually in first, second, third, fourth, fifth, and sixth grades; social skills from kindergarten through sixth grade; and emotional functioning and academic work habits annually in first through sixth grade.

Academic achievement. With respect to cognitive-academic achievement, children were administered several subtests from the Woodcock–Johnson Psycho-Educational Battery–Revised (Woodcock & Johnson, 1989): Letter–Word Identification (54 months and first grade), which assesses prereading skills in identifying isolated letters and words, and Broad Reading (third and fifth grades), which adds assessment of passage comprehension to the assessment of identification of words; Applied Problems (54 months and first grade) and Broad Math (third and fifth grades), which measure skill in analyzing and solving practical problems in mathematics; and Picture Vocabulary (54 months and first, third, and fifth grades), which measures children's ability to name objects depicted in a series of pictures. Items are presented in order of increasing difficulty and are scored 0 = *incorrect or no response* or 1 = *correct response*, with basal and ceiling levels established. Typically, raw scores are converted to standard scores with a mean of 100 and a standard deviation of 15, but for this study we relied upon *W* ability scores so that change over time could be more easily documented. The *W* ability scores are transformations of the Rasch raw ability scores designed to eliminate the need for decimal fractions and negative values. The overall Woodcock–Johnson battery of tests has been standardized on a nationally representative sample from 24 months to 95 years of age. For the current analysis, the Letter–Word Identification and Broad Reading subtests were summarized into the variable reading, and the Applied Problems and Broad Math were summarized into outcome variable math, resulting in three academic achievement outcome variables: reading, math, and picture vocabulary.

The Social Skills Questionnaire from the Social Skills Rating System (Gresham & Elliott, 1990) was used to assess social competence and social skills. This instrument is composed of 38 items describing child behavior, each rated on a 3-point scale reflecting how often the child exhibited each behavior. Items are grouped into four areas: (a) cooperation (e.g., "keeps room neat and clean without being reminded"), (b) assertion (e.g., "makes friends easily"), (c) responsibility (e.g., "asks permission before using someone else's property"), and (d) self-control (e.g., "controls temper when arguing with other children"). The total score used in this report represents the sum of all 38 items, with higher scores reflecting higher levels of perceived social skills (alphas range from .86 to .94).

Behavior problems. The Child Behavior Checklist Teacher Report Form (Achenbach, 1991) was used to evaluate problem behavior. The form lists 100 problem behaviors that generate two subscales: Internalizing Problems (e.g., "too fearful and anxious") and Externalizing Problems (e.g., "hits others," "disobedient at school," "argues a lot"). Achenbach reports test–retest reliability of .89, interparent agreement of .70, and stability of .71 over 2 years. Raw scores were converted into standard T scores based on normative data for children of the same age. For the present analysis, only the Externalizing Problems subscale was used.

Teacher–child conflict. The Student–Teacher Relationship Scale: Short Form (Pianta, 2001) was completed by caregivers or teachers. The form is a widely used indicator of a teacher's perceptions of the quality of his or her relationship with a specific child. In the current report, we focused on teacher–child conflict (e.g., "dealing with this child drains my energy"), which was assessed by seven items rated with 5-point Likert scales. Coefficient alphas for the Conflict subscale ranged from .88 to .91 across grades.

Academic work habits. Teachers completed a 19-item mock report card. This questionnaire includes six items addressing the child's work habits. The items, each rated on a 5-point scale (1 = *very poor* to 5 = *very good*), were taken from Madison (Wisconsin) Metropolitan School District report cards. The six items include "follows classroom procedures," "works well independently," "works neatly and carefully," "uses time wisely," "completes work promptly," and "keeps material organized." Scores were based on the mean of the items at each age. Coefficient alphas were high, ranging from .94 to .95.

Socioemotional functioning. Teacher-reported socioemotional functioning was obtained from the mock report cards. The seven included items that addressed socioemotional functioning came from the Teacher Checklist of Peer Relations (Coie & Dodge, 1988) and were rated on a 5-point scale (1 = *very poor* to 5 = *very good*). Items addressed the children's social skillfulness with peers, such as "generates good quality solutions to interpersonal problems" and "is aware of the effects of his/her behavior on others." Cronbach alphas ranged from .94 to .95.

Data Analysis Plan

Data analysis focused on testing the moderating effect of infant temperamental difficulty on the long-term associations between parenting quality and child care quality during the first 4.5 years and children's cognitive-academic achievement and social adjustment from that age through the spring of fifth or sixth grade. The same analytic strategy adopted in Belsky, Vandell, et al. (2007) to evaluate long-term effects of child care was implemented with minor modification to the same statistical program that generated

the Belsky, Vandell, et al. results. These modifications involved the inclusion of a main effect for infant temperament and two two-way interaction terms—Temperament × Parenting Quality and Temperament × Child Care Quality—in the analyses to be presented, after checking to see whether these two-way interactions were themselves moderated by age, which proved not to be the case, or whether a three-way interaction involving temperament, parenting, and child care quality proved significant, which also proved not to be the case.

HLMs (Bryk & Raudenbush, 2002; Singer & Willett, 2003) were fitted to estimate individual and group linear and quadratic growth curves. The models included both preschool and concurrent family and child care–school experiences, as well as 6-month temperament and the two two-way interactions involving temperament. Individual intercepts and linear slopes with respect to age were estimated as correlated random effects for each child for each outcome. These individual-level parameters were related to 22 predictors: In addition to the primary predictors of interest in this study, the two two-way interactions involving temperament, there was 6-month temperament, parenting intercept (across the period 6–54 months), child care quality intercept (estimated quality of care at 27 months, reflecting the midpoint between 6 and 54 months), four additional indices of child care experiences across the period 3–54 months (proportion of 3–4-month epochs in center-based child care for at least 10 hr per week, proportion of 3–4-month epochs in a child care home for at least 10 hr per week, the hours-per-week intercept [estimated from HLM analyses in which the intercept was set at 27 months], and the hours-per-week slope [estimated linear change over time in hours per week]), eight preschool time-invariant covariates (site, child ethnicity, child gender, maternal education, mean income-to-needs ratio between 1 and 54 months, parenting quality slope from 6 to 54 months, maternal depressive symptoms intercept, and slope from 6 to 54 months), and five concurrent time-varying covariates from 54 months through sixth grade (income-to-needs ratio, parenting quality, maternal depression, observed school classroom quality, and hours per week of after-school care [set to zero for 54 months]). In the interest of space and owing to the fact that main effects of child care variables and parenting detected in the original Belsky, Vandell, et al. (2007) analysis were replicated, reporting of results focuses exclusively upon the Temperament × Parenting Quality and the Temperament × Child Care Quality interactions, that is, those effects that extend findings previously reported by Belsky, Vandell, et al. and are of primary interest presently.

Several modeling decisions were made. Most importantly, age was centered at the mean age for spring of the final assessment period: fifth grade for academic achievement and sixth grade for teacher-rated socioemotional outcomes. Thus, the two-way interaction effect coefficients to be presented indicate whether the interaction was related to the outcome in the spring of either fifth or sixth grade, that is, the final middle-childhood time of measurement for that outcome.

Missing data occurred in this longitudinal project because of attrition and failure to complete all assessments. For the most critical variables used in the analyses to be reported, missing data were as follows: difficult temperament (6.2%), parenting quality (6.2%–23.8%, representing lowest and highest at any one time across repeated measurement occasions), child care quality (8.4%–19.4%), reading achievement (22.6%–27.2%), math achievement (22.8%–27.2%), picture vocabulary (22.3%–27.3%), behavior problems (26.2%–47.7%), social skills (26.7%–38.3%), teacher–child conflict (26.3%–47.5%), academic work habits (26.3%–36.1%), and socioemotional functioning (26.3%–36.5%).

Missing data were imputed through multiple imputation (Rubin, 1987; Schafer, 1997; Schafer & Graham, 2002) under the assumption that missing data were ignorably missing. Using Schafer's (1997) and Schafer & Graham's (2002) recommended procedure—an iterative expectation-maximization algorithm—we estimated missing values for each variable iteratively by means of a logistic or multiple regression based on all other variables using data from all individuals with observed values on that variable and adding random variability to the predicted data. The process is repeated for each variable until the differences in predicted values across iterations are minuscule. Five data sets were created in which all observed data are represented and missing data are estimated. Consequently, analyses were conducted five times with each of the five imputation data sets. Results of these analyses were combined with the recommended procedures of Schafer (1997) to ensure that variability in imputed values across the data sets was considered as well as variability in the variables within the data set. The test statistics and regression coefficients were averaged across the five analyses, and the standard errors for the coefficients were summarized by combining within- and between-model variability. For the follow-up analyses of significant interactions, nonadjusted intercepts for each single outcome variable were estimated by means of separate unconditional HLM analyses for each imputed data set. These intercepts were then averaged across the five imputed data sets. Imputed parenting quality intercepts, child care quality intercepts (6–54 months), and temperament scores were averaged in the same way. Of importance is the fact that when the analyses to be reported were run only with cases with complete data, results were not appreciably different. More specifically, one of the significant interactions to be reported proved no longer significant; one became significant having not achieved significance in the imputed-case analyses; and three became more significant (and none less) than was the case with the imputed data. These results are available on request.

Effect sizes were computed to show the anticipated difference in standard deviation units of the outcome measure between children who had child care experiences that differed by one standard deviation (for details, see NICHD ECCRN & Duncan, 2003). The effect sizes were computed as the product of the estimated child care coefficient and the standard deviation for the child care index divided by the standard deviation for the outcome measure, and can be interpreted somewhat like a correlation. For example, the standard deviation for the child care quality measure was .22, so the effect size compared predicted outcome scores for children whose child care differed in quality by .22 points on the ORCE measure of child care quality.

Results

Four sets of results are presented: the first preliminary and addressing simple bivariate relations among variables; the second primary, involving HLM analyses with both interaction terms included simultaneously; and the third secondary, involving separate HLM models for the two moderating variables. In the fourth set of results, significant interactions are probed by testing differ-

ences between the simple slopes (linking parenting or child care quality with the outcome of interest) of children scoring one standard deviation above and below the mean. The level of significance for all analyses was set at $\alpha = .05$, except in the case of the tested interactions in the HLM analyses where $\alpha = .10$ was used, given widely acknowledged challenges to detecting interaction effects in field studies (McClelland & Judd, 1993).

Preliminary Analysis: Unadjusted Associations

As the means and standard deviations of almost all variables included in this investigation are presented in Belsky, Vandell, et al. (2007), as are their intercorrelation, we limit the reporting of such findings here, emphasizing results pertaining to difficult temperament, as this variable was not included in that report. Both environmental measures of interest—parenting quality and child care quality—were significantly ($p < .01$) related to all outcome variables, with parenting quality generally showing stronger associations with outcome variables ($r_{(1362)} = .41–.56$) than child care quality ($r_{(1362)} = .15–.23$). Parenting quality and child care quality were positively correlated ($r_{(1362)} = .30$, $p < .01$). Child temperament at 6 months predicted parenting quality ($r_{(1362)} = -.23$, $p < .01$) but not child care quality ($r_{(1362)} = -.02$, $p > .05$). Greater difficulty was related to poorer reading ($r_{(1362)} = -.14$, $p > .01$), math ($r_{(1362)} = -.17$, $p > .01$), picture vocabulary ($r_{(1362)} = -.20$, $p > .01$), social skills ($r_{(1362)} = -.11$, $p > .01$), work habits ($r_{(1362)} = -.09$, $p > .01$), and socioemotional functioning ($r_{(1362)} = -.11$, $p > .01$) and to more behavior problems ($r_{(1362)} = .09$, $p > .01$) and teacher–child conflicts ($r_{(1362)} = .08$, $p > .01$). Multiplicative Parenting × Temperament and Child Care Quality × Temperament interaction terms were positively correlated ($r_{(1362)} = .26$, $p < .01$).

Primary Analysis: HLM Analysis

For each outcome measure a hierarchical linear regression model (HLM) analysis was computed that included, as stipulated above, the two interaction terms, Temperament × Parenting Quality and Temperament × Child Care Quality, as well as the other predictors previously mentioned. Consequently, there were altogether 16 tests of two-way interactions involving temperament across eight HLM analyses; 7 of these 16 interactions met statistical criteria for follow-up probing. More specifically, temperament interacted with parenting quality on all three academic outcomes (reading, math, and picture vocabulary) and on two of the five social adjustment outcomes (social skills and work habits). Temperament interacted as well with child care quality on two social adjustment outcomes (behavior problems and teacher–child conflicts). HLM coefficients, 95% confidence intervals (CIs), and effect sizes (ESs) of the main effects and both interaction terms for each outcome variable are displayed in Table 1.

Secondary Analysis: Separate Follow-Up HLM Analyses

To investigate whether including both interaction terms—which were significantly and positively correlated (see above)—in a single model affected results just reported, we conducted a secondary analysis testing each of the two interactions in separate equations for each outcome. All seven interactions that met criteria previously continued to do so, but so did two new interactions: Temperament × Child Care Quality predicting reading ($B = 8.45$, $p < .10$, 95% CI of $B = -1.14$ to 18.04, ES = .05) and Temperament × Parenting Quality predicting socioemotional functioning ($B = 0.12$, $p < .10$, 95% CI of $B = -0.01$ to 0.25, ES = .04). Moreover, the previously reported marginally significant interaction between temperament and child care quality predicting behavior problems achieved conventional levels of significance ($B = -4.37$, $p < .05$, 95% CI of $B = -8.24$ to -0.50, ES = $-.04$).

Tertiary Analysis: Probing Significant Interactions

The significant interactions of the primary HLM analysis were followed up by estimating the relation between each predictor (i.e., parenting or child care quality) and each outcome separately for children who scored high and low on difficult temperament at 6 months, defined in terms of plus or minus one standard deviation from the sample mean (Aiken & West, 1991). Differences between the slopes of the two groups were tested for significance after z transformation of the standardized regression coefficients (Fisher, 1924).

Higher quality parenting predicted greater reading, vocabulary, work habits, math, and social skills for children who scored high as well as for children who were low on difficult temperament as infants, but all these parenting effects were significantly stronger (i.e., steeper slopes) for the children with histories of temperamental difficulty ($r_{(191)} = .47–.62$, $p < .001$ vs. $r_{(210)} = .22–.43$, $p < .01$), with the exception of broad math ($p < .06$; see Figure 1).

Evidence of differential susceptibility was even more pronounced in the case of child care quality effects. Whereas virtually no relation emerged between quality of care and behavior problems ($r_{(210)} = -.04$, $p > .05$) and teacher–child conflicts ($r_{(210)} = -.004$, $p > .05$) in the case of children who did not manifest temperamental difficulty as infants, this was decidedly not so for their counterparts with contrasting temperamental histories. For these latter children with histories of difficult temperament, greater quality of child care predicted fewer problems ($r_{(191)} = -.33$, $p < .001$) and less conflict ($r_{(191)} = -.25$, $p < .001$), and as result, in both cases the regression slopes were significantly different across the two temperament groups (see Figure 2).

Follow-up of the two marginally significant interactions that emerged in the secondary HLM analyses in which only one interaction at a time was tested generated results like those just described. That is, positive effects of parenting on socioemotional functioning differed in degree between high- and low-difficulty temperament groups, whereas the positive effect of child care quality on reading differed, with the low-difficulty group manifesting no association and the high-difficulty group a substantial one (details available on request).

Discussion

This study was designed to extend research on differential susceptibility to rearing by investigating in a single inquiry the effects of both parenting and child care quality, testing the hypothesis that children with histories of difficult temperament as infants would prove more susceptible to rearing effects when their

Table 1
Hierarchical Linear Model Analysis of Child Outcomes at Fifth or Sixth Grade: Coefficients Predicting Outcomes From Temperament, Parenting Quality, and Child Care Quality and Their Interactions With Temperament (N = 1,364)

	Cognitive outcomes (direct assessment, fifth grade)			Socioemotional outcomes (teacher report, sixth grade)				
Predictor variable	Reading	Math	Picture vocabulary	Behavior problems	Social skills	Teacher conflict	Work habits	Socioemotional functioning
Main effects								
Temperament								
B (*SE*)	−1.52 (1.01)	−0.91 (0.77)	−0.99 (0.75)	−0.28 (0.51)	0.80 (0.72)	−0.16 (0.30)	0.05 (0.07)	0.01 (0.05)
95% CI of *B*	−3.51 to 0.47	−2.41 to 0.60	−2.47 to 0.49	−1.27 to 0.71	−0.60 to 2.21	−0.75 to 0.43	−0.08 to 0.18	−0.09 to 0.12
Effect size	−.05	−.03	−.04	−.01	.03	−.01	.02	.01
Parenting quality								
B (*SE*)	**4.04** (1.04)**	**2.30* (0.90)**	**4.46** (0.71)**	**−2.60** (0.59)**	**3.42** (0.97)**	**−1.07** (0.39)**	**0.30** (0.06)**	**0.32** (0.05)**
95% CI of *B*	2.00 to 6.07	0.54 to 4.07	3.07 to 5.85	−3.76 to −1.43	1.53 to 5.32	−1.83 to −0.31	0.18 to 0.42	0.23 to 0.42
Effect size	.19	.12	.24	−.19	.16	−.13	.18	.24
Child care quality								
B (*SE*)	0.89 (2.24)	1.97 (1.76)	**3.76* (1.53)**	0.09 (1.27)	1.47 (1.74)	−0.04 (0.82)	0.16 (0.13)	0.07 (0.10)
95% CI of *B*	−3.50 to 5.29	−1.48 to 5.42	0.75 to 6.77	−2.40 to 2.57	−1.93 to 4.88	−1.65 to 1.57	−0.09 to 0.41	−0.11 to 0.27
Effect size	.01	.03	.07	.00	.02	.00	.03	.02
Interaction terms								
Temperament × Parenting Quality								
B (*SE*)	**4.23* (1.94)**	**2.62* (1.23)**	**2.35* (1.04)**	−0.67 (0.77)	**1.88† (0.99)**	0.03 (0.40)	**0.15† (0.08)**	0.10 (0.07)
95% CI of *B*	0.42 to 8.04	0.21 to 5.03	0.32 to 4.39	−2.18 to 0.84	−0.05 to 3.81	−0.75 to 0.80	−0.01 to 0.30	−0.03 to 0.24
Effect size	.09	.06	.06	−.02	.04	.00	.04	.04
Temperament × Child Care Quality								
B (*SE*)	5.23 (5.34)	2.06 (4.08)	1.17 (3.23)	**−3.87† (2.02)**	2.02 (2.92)	**−2.61* (1.27)**	−0.01 (0.28)	0.21 (0.19)
95% CI of *B*	−5.25 to 15.70	−5.94 to 10.06	−5.17 to 7.50	−7.83 to 0.10	−3.70 to 7.74	−5.09 to −0.12	−0.83 to 0.82	−0.35 to 0.77
Effect size	.03	.01	.01	−.04	.01	−.04	.00	.02

Note. Model includes as time-invariant covariates site, gender, ethnicity, maternal education, 6–54 months income to needs, proportion time mother had partner in household, 6–54 months maternal depression intercept and slope, 3–54 months child care hours intercept and slope, 6–54 months parenting slope, 6–54 months proportion center child care, 6–54 months proportion home child care. Model includes as time-varying covariates concurrent measures of income to needs, partner in household, parenting quality, maternal depression, classroom quality, and after-school hours. The model further includes interaction terms of all covariates, 6–54 months parenting intercept, and 6–54 months child care quality intercept with age (first–sixth grade). Significant effects are displayed in bold. CI = confidence interval.
† $p < .10$. * $p < .05$. ** $p < .01$.

cognitive-academic achievement and social adjustment were evaluated late in middle childhood. At this point, it should be made explicit that as this was a field study and not an experimental one, effects detected were statistical and cannot be presumed to be causal in nature, even with all the covariates included. Nevertheless, for heuristic purposes, causal language is often used in this discussion, much as it has been throughout this report.

Also important to note is that the NICHD SECCYD does not include a nationally representative sample. Therefore, no inference can or should be made that the results to be discussed would be replicated were such a sample studied. What seems important to keep in mind in light of this comment, however, is that the evidence of differential susceptibility that emerged in this investigation is consistent with findings from rather diverse studies, including investigations of mostly white families in Iowa (Kochanska, 1993), poor African American low-birth-weight infants (Blair, 2002), and even rhesus macaques (Suomi, 1995, 1997). This clearly suggests that whatever the limits of the sample, the results reported seem unlikely to be an artifact of this study limitation. Indeed, it would seem just as conceivable that had a better sample been available for investigation, incorporating thereby a fuller range of family and child care rearing experiences, as well as of infant temperaments and developmental outcomes, evidence of differential susceptibility might have been greater than proved to be the case.

Findings resulting from testing two Rearing × Temperament interaction terms were generally in accord with results of studies cited in the introduction in that both parenting quality and child care quality predicted child functioning more strongly in the case of children who manifested difficult temperaments as infants than those who did not. The present work extends past research, though, by providing evidence of differential susceptibility beyond the early childhood and even early elementary school years. The fact that heightened susceptibility of infants with difficult temperament to the developmental benefits of sensitive parenting and quality child care extends to fifth and sixth grade raises the prospect that it could endure into adolescence and even beyond. Certainly consistent with this possibility are Gene × Environment interactions linking maltreatment in childhood with antisocial behavior in young adulthood (Caspi et al., 2002).

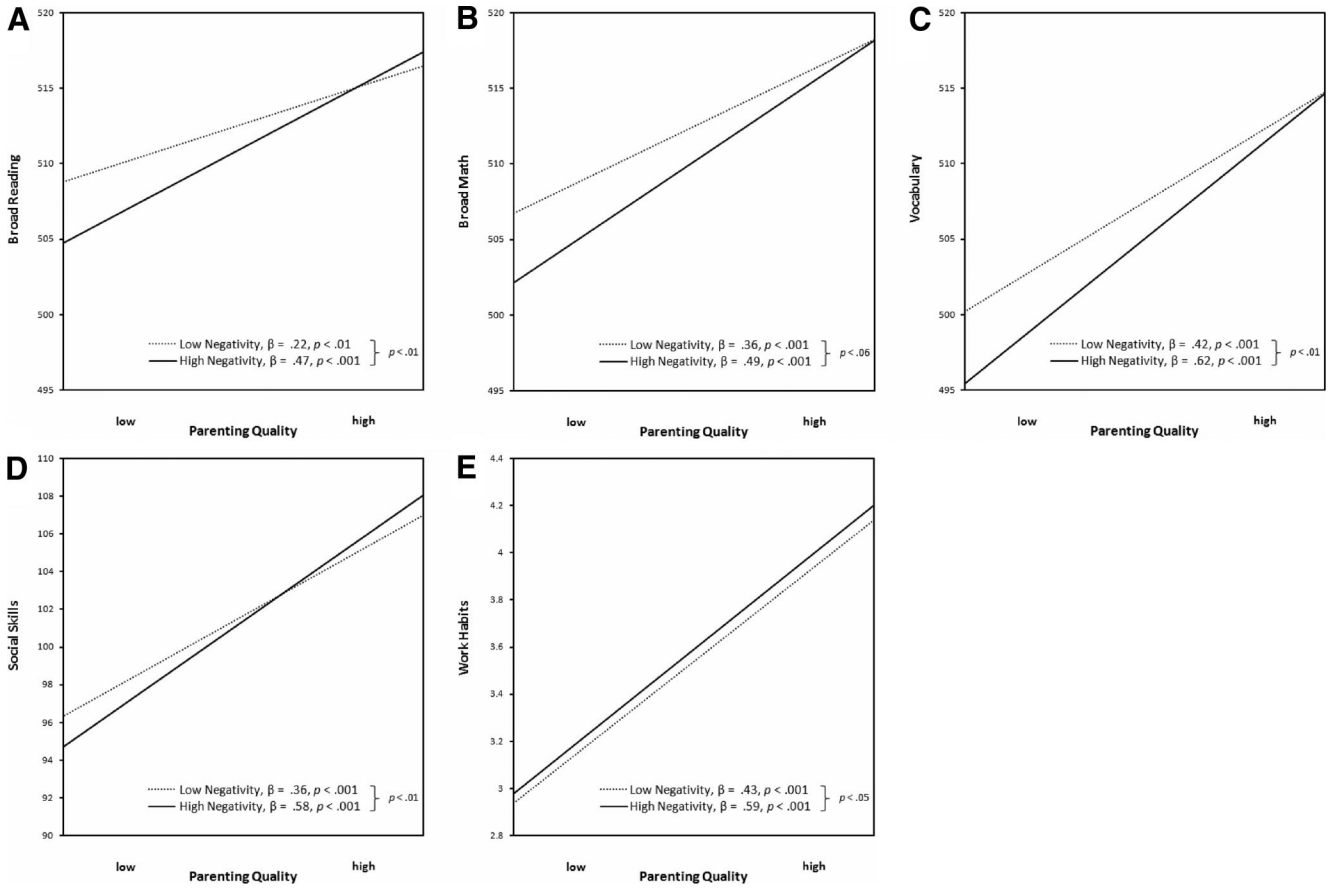

Figure 1. Linear relation between parenting quality and cognitive outcomes at fifth grade—(A) reading, (B) math, and (C) picture vocabulary—and between parenting quality and teacher-rated outcomes at sixth grade—(D) social skills and (E) work habits—for children one standard deviation below (n = 212) and one standard deviation above (n = 193) the mean of temperament scores at 6 months.

Although interactions between temperament and parenting met statistical criteria more frequently than interactions between temperament and child care quality, the latter were perhaps more notable. Whereas the former simply showed that anticipated effects of parenting were somewhat stronger in the case of children with histories of difficult temperament (see Figure 1), the more striking crossover interactions involved child care quality (see Figure 2), indicating that quality child care substantially predicted (less) problem behavior, (less) teacher–child conflict, and (better) reading in the case of children with difficult temperaments but did not predict these outcomes in the case of children scoring low on difficulty as infants. Such crossover interactions, including several Temperament × Parenting interactions that did not involve zero-order associations for the low-difficulty subgroup, that result in significant differences between slopes are in line with evidentiary requirements advanced by Belsky, Bakermans-Kranenburg, and van IJzendoorn (2007) for establishing differential susceptibility.

Another statistical criteria stipulated by Belsky, Bakermans-Kranenburg, and van IJzendoorn (2007) for documenting differential susceptibility is that predictor and moderator variables prove independent, as was the case in this inquiry with respect to child care quality (and temperament) but not parenting. This latter circumstance raised the possibility that evocative effects of child temperament on parenting simply masqueraded as moderating effects of temperament on parenting in the current inquiry. To test whether this might have been the case, we residualized parenting quality for the influence of temperament and reran the primary HLM analysis using this modified predictor (which no longer correlated with the temperament moderator). In all cases but one (i.e., academic work habits), previously detected Temperament × Parenting interactions continued to meet statistical criteria established for drawing differential susceptibility inferences: The interaction was significant (or marginally so), and simple slopes of contrasting temperament groups proved significantly different. Because the further requirements of independence between moderator and outcome variables were also met and because the moderator, temperament, did not predict any of the outcomes in the HLM analysis as a main effect, the overwhelming majority of interaction effects detected and probed met criteria promulgated for establishing differential susceptibility (Belsky, Bakermans-Kranenburg, & van IJzendoorn, 2007).

The evidence for differential susceptibility in the case of child care quality merits special consideration for several reasons. First, as already noted, it was here that the contrasting (putative) influ-

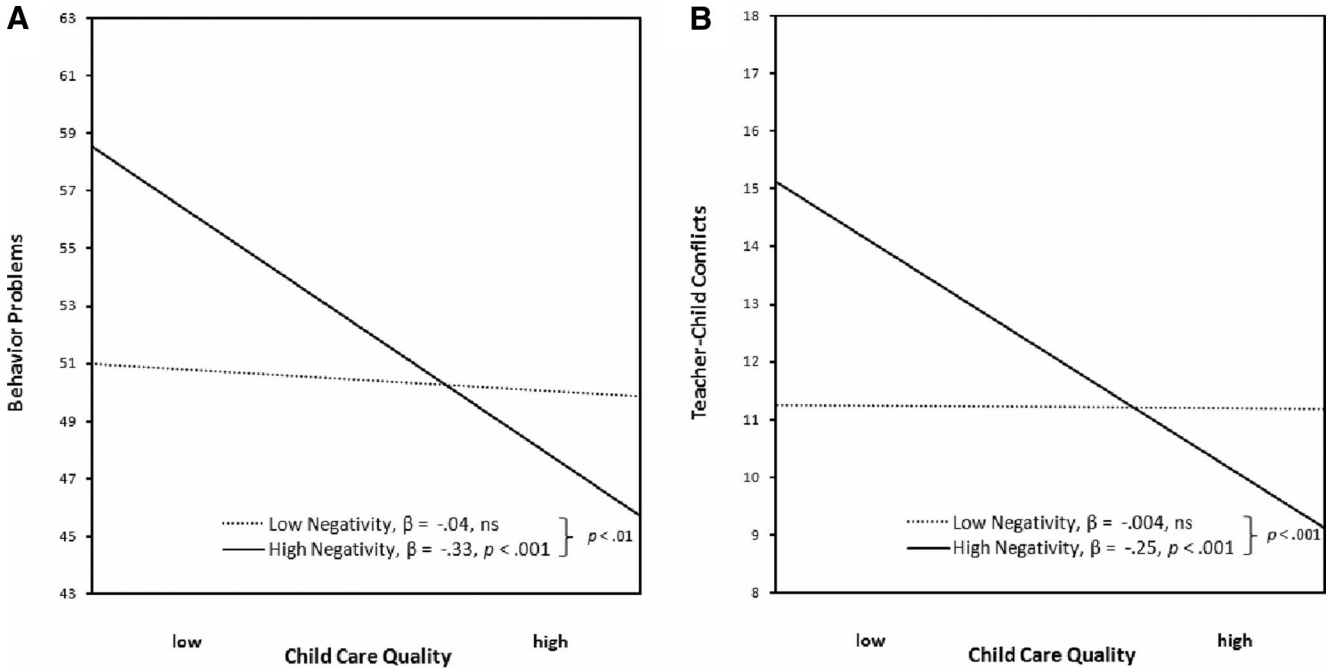

Figure 2. Linear relation between child care quality and teacher-rated outcomes at sixth grade—(A) behavior problems and (B) teacher–child conflict—for children one standard deviation below ($n = 212$) and one standard deviation above ($n = 193$) the mean of temperament scores at 6 months.

ence of child care quality proved most pronounced, with good quality care appearing to foster fewer behavior problems, less teacher–child conflict, and better reading and poorer quality appearing to exert the opposite effect, but only in the case of children with histories of difficult temperament in infancy. Second, as no main effects of child care quality on any social outcomes emerged, just as in past analyses of child care effects at 4.5 years and in kindergarten (NICHD ECCRN 2003a, 2006), in third grade (NICHD ECCRN, 2005c) and in sixth grade (Belsky, Vandell, et al., 2007), it should be clear that such null results may lead to erroneous conclusions about the breadth of quality-of-care effects. After all, these NICHD SECCYD reports have all noted that although quality predicts cognitive-academic functioning, it is quantity and/or type of care that predicts social functioning. Perhaps what is most important about this observation is not just that quality of care predicts social adjustment in the NICHD SECCYD, at least through the middle-childhood years (Pluess & Belsky, in press; this study), but that other investigations that fail to detect rearing effects may do so because of the implicit presumption that all children will be equally susceptible to the rearing effects under investigation. As a result, tests of the differential susceptibility hypothesis are simply not undertaken, thereby risking premature and inappropriate embracement of the null.

As for explanation, the question still meriting consideration arises: Why might difficult infants be more susceptible to positive and negative effects of parenting and child care quality? As suggested in the introduction, the specific characteristics of difficult temperament may be indicators of a general heightened sensitivity of the nervous system to environmental stimuli, such that experiences, be they supportive and nurturing or otherwise, register more strongly than in the case of infants with less sensitive nervous systems (Pluess & Belsky, in press). Thus, putatively difficult infants become easily overwhelmed in adverse, low-quality environments but benefit disproportionally from high-quality supportive environments that enable them, first, to coregulate their emotions with the help of a sensitive caregiver (at home and/or in child care) and later to self-regulate them, thereby facilitating the gathering of developmentally useful information and insights from the world around them. If differential susceptibility were truly a function of such heightened sensitivity of the nervous system, its effects should perhaps not be restricted to experiences in the infant, toddler, and early childhood years but influence reactions to later experiences as well. Of note in this regard is Aron and Aron's (1997) claim that a highly sensitive (and shy) personality is relatively common, representing some 20% of the American population, stable across the life span, and characteristic of individuals who are disproportionately affected in an adverse way by stressful experiences but in a positive manner by nurturant, supportive ones.

Quite conceivably, biological inheritance exerts an important influence on heightened sensitivity of the nervous system. Of interest in this regard is evidence that the same genotype found to make individuals more susceptible to the adverse effects of child maltreatment and negative life events on depression (Caspi et al., 2003) has also been linked in some molecular-genetic research to negative emotionality in the newborn (Auerbach et al., 1999), an important component of difficult temperament. This raises the prospect that research that operationalizes the organismic moderator of experiential influences in genotypic terms (e.g., Caspi et al., 2002, 2003), in endophenotypic terms (e.g., physiological stress reactivity; Ellis, Essex, & Boyce, 2005), or in phenotypic terms, as

in the current investigation (i.e., negative emotionality and difficult temperament), may be identifying the very same highly susceptible or highly sensitive individuals. Only research that measures each of these levels of analysis and tests for differential susceptibility will be in position to address this possibility.

Consideration of genetic determinants of negative emotionality and/or heightened susceptibility to rearing does not preclude the possibility that heightened susceptibility itself may be experientially induced. Indeed, this is a prospect raised by Boyce and Ellis (2005), which receives some support from research showing that cortisol-indexed stress reactivity in some young mammals is shaped, in part, by fetal experience (Gluckman & Hanson, 2005). In fact, it may be the case that for genetic reasons some individuals are more likely to be affected by both prenatal and postnatal experiences than others. More specifically, some individuals may be more likely to be affected by experience in general (most notably perhaps, fetal experience) in ways that subsequently affect whether or to what degree they will be influenced by the postnatal world they encounter. As Belsky and Pluess (in press) have recently argued, it may be appropriate to view differential susceptibility in terms of individual differences regarding the capacity for developmental plasticity, a capacity that can be born as well as made.

References

Achenbach, T. (1991). *Manual for the Child Behavior Checklist/4–18 and 1991 profile.* Burlington: University of Vermont, Department of Psychiatry.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Aron, E. N., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology, 73*(2), 345–368.

Auerbach, J., Geller, V., Lezer, S., Shinwell, E., Belmaker, R. H., Levine, J., et al. (1999). Dopamine D4 receptor (D4DR) and serotonin transporter promoter (5-HTTLPR) polymorphisms in the determination of temperament in 2-month-old infants. *Molecular Psychiatry, 4*(4), 369–373.

Belsky, J. (1997a). Theory testing, effect-size evaluation, and differential susceptibility to rearing influence: The case of mothering and attachment. *Child Development, 68*(4), 598–600.

Belsky, J. (1997b). Variation in susceptibility to rearing influences: An evolutionary argument. *Psychological Inquiry, 8,* 182–186.

Belsky, J. (2005). Differential susceptibility to rearing influences: An evolutionary hypothesis and some evidence. In B. Ellis & D. Bjorklund (Eds.), *Origins of the social mind: Evolutionary psychology and child development* (pp. 139–163). New York: Guilford Press.

Belsky, J., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). For better and for worse: Differential susceptibility to environmental influences. *Current Directions in Psychological Science, 16*(6), 300–304.

Belsky, J., Hsieh, K. H., & Crnic, K. (1998). Mothering, fathering, and infant negativity as antecedents of boys' externalizing problems and inhibition at age 3 years: Differential susceptibility to rearing experience? *Development and Psychopathology, 10*(2), 301–319.

Belsky, J., & Pluess, M. (in press). The nature (and nurture?) of plasticity in early human development. *Perspectives on Psychological Science.*

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., & Owen, M. T. (2007). Are there long-term effects of early child care? *Child Development, 78*(2), 681–701.

Blair, C. (2002). Early intervention for low birth weight, preterm infants: The role of negative emotionality in the specification of effects. *Development and Psychopathology, 14*(2), 311–332.

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology, 17*(2), 271–301.

Bradley, R. H., & Corwyn, R. F. (2008). Infant temperament, parenting, and externalizing behavior in first grade: A test of the differential susceptibility hypothesis. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 49*(2), 124–131.

Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Caldwell, B., & Bradley, R. H. (1984). *Home observation for measurement of the environment.* Little Rock: University of Arkansas Press.

Carey, W. B., & McDevitt, S. C. (1978). Revision of the Infant Temperament Questionnaire. *Pediatrics, 61*(5), 735–739.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., et al. (2002, August 2). Role of genotype in the cycle of violence in maltreated children. *Science, 297*(5582), 851–854.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003, July 18). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science, 301*(5631), 386–389.

Chess, S., & Thomas, A. (1989). Temperament and its functional significance. In S. I. Greenspan & G. H. Pollock (Eds.), *The course of life: Vol. 2. Early childhood* (pp. 163–227). Madison, CT: International Universities Press.

Coie, J. D., & Dodge, K. A. (1988). Multiple sources of data on social behavior and social status in the school: A cross-age comparison. *Child Development, 59*(3), 815–829.

Dienstbier, R. A. (1985). The role of emotion in moral socialization. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 484–514). New York: Cambridge University Press.

Early, D., Rimm-Kaufman, S., Cox, M., Saluja, G., Pianta, R., Bradley, R., et al. (2002). Maternal sensitivity and child wariness in the transition to kindergarten. *Parenting: Science and Practice, 2,* 355–377.

Ellis, B. J., Essex, M. J., & Boyce, W. T. (2005). Biological sensitivity to context: II. Empirical explorations of an evolutionary-developmental theory. *Development and Psychopathology, 17*(2), 303–328.

Feldman, R., Greenbaum, C. W., & Yirmiya, N. (1999). Mother–infant affect synchrony as an antecedent of the emergence of self-control. *Developmental Psychology, 35*(1), 223–231.

Fisher, R. A. (1924). *On a distribution yielding the error functions of several well known statistics.* Paper presented at the proceedings of the International Congress of Mathematicians, Toronto, Canada.

Gluckman, P., & Hanson, M. (2005). *The fetal matrix: Evolution, development and disease.* Cambridge, England: Cambridge University Press.

Gresham, F., & Elliott, S. (1990). *Social skills rating system.* Circle Pines, MN: American Guidance Service.

Kochanska, G. (1993). Toward a synthesis of parental socialization and child temperament in early development of conscience. *Child Development, 64*(2), 325–347.

Kochanska, G., Clark, L. A., & Goldman, M. S. (1997). Implications of mothers' personality for their parenting and their young children's developmental outcomes. *Journal of Personality, 65*(2), 387–420.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*(2), 376–390.

NICHD Early Child Care Research Network. (1998). Early child care and self-control, compliance, and problem behavior at twenty-four and thirty-six months. *Child Development, 69*(4), 1145–1170.

NICHD Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD

Study of Early Child Care. *American Educational Research Journal, 39*(1), 133–164.

NICHD Early Child Care Research Network. (2003a). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development, 74*(4), 976–1005.

NICHD Early Child Care Research Network. (2003b). Does quality of child care affect child outcomes at age 4½? *Developmental Psychology, 39*(3), 451–469.

NICHD Early Child Care Research Network. (2004a). *COS qualitative scales and reliabilities observational ratings of the fifth grade classroom* (Child Care Data Report 571). Research Triangle Park, NC: RTI International.

NICHD Early Child Care Research Network. (2004b). Does class size in first grade relate to children's academic and social performance or observed classroom processes? *Developmental Psychology, 40,* 651–664.

NICHD Early Child Care Research Network. (2005a). *Child care and child development: Results of the NICHD Study of Early Child Care and Youth Development.* New York: Guilford Press.

NICHD Early Child Care Research Network. (2005b). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *Elementary School Journal, 105*(3), 305–323.

NICHD Early Child Care Research Network. (2005c). Early child care and children's development in the primary grades: Follow-up results from the NICHD Study of Early Child Care. *American Educational Research Journal, 42*(3), 537–570.

NICHD Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD Study of Early Child Care and Youth Development. *American Psychologist, 61*(2), 99–116.

NICHD Early Child Care Research Network & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development, 74*(5), 1454–1475.

Pianta, R. C. (2001). *Student–Teacher Relationship Scale.* Odessa, FL: Psychological Assessment Resources.

Pluess, M., & Belsky, J. (in press). Differential susceptibility to rearing experience: The case of childcare. *Journal of Child Psychology and Psychiatry and Allied Disciplines.*

Radloff, L. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Journal of Applied Psychological Measurement, 1,* 385–401.

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 99–166). Hoboken, NJ: Wiley.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177.

Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Stright, A. D., Gallagher, K. C., & Kelley, K. (2008). Infant temperament moderates relations between maternal parenting in early childhood and children's adjustment in first grade. *Child Development, 79*(1), 186–200.

Suomi, S. J. (1995). Influence of attachment theory on ethological studies of biobehavioral development in nonhuman primates. In S. Goldberg, R. Muir, & J. Kerr (Eds.), *Attachment theory: Social, developmental, and clinical perspectives* (pp. 185–201). Hillsdale, NJ: Analytic Press.

Suomi, S. J. (1997). Early determinants of behaviour: Evidence from primate studies. *British Medical Bulletin, 53*(1), 170–184.

Vandell, D. L. (1979). Effects of a playgroup experience on mother–son and father–son interaction. *Developmental Psychology, 15*(4), 379–385.

Wachs, T. D., & Gandour, M. J. (1983). Temperament, environment, and six-month cognitive-intellectual development: A test of the organismic specificity hypothesis. *International Journal of Behavioral Development, 6*(2), 135–152.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery–Revised.* Itasca, IL: Riverside.

Received September 3, 2008
Revision received November 18, 2008
Accepted December 2, 2008 ■