Published in final edited form as:

J Pers Soc Psychol. 2004 November ; 87(5): 698–706.

Night and Day: Are Siblings as Different in Temperament as Parents Say They Are?

Kimberly J. Saudino

Department of Psychology, Boston University

Annie E. Wertz

Department of Psychology, Boston University

Department of Psychology, University of California, Santa Barbara

Jeffrey R. Gagne and Sonia Chawla

Department of Psychology, Boston University

Abstract

Twin studies suggest that parent ratings of temperament exaggerate differences between twins. The present study examined whether such contrast effects also operate for nontwin siblings. The activity level (AL) and shyness of 95 nontwin sibling pairs (ages 3 to 8 years) were assessed via parent ratings and objective measures (actigraph and observer ratings). Siblings showed no resemblance in either parent-rated AL or shyness; however, sibling resemblance for actigraph AL and observer-rated shyness was substantial. Thus, parents do contrast their nontwin siblings when rating these 2 temperament dimensions. Moreover, the importance of sibling differences in temperament to the sibling relationship and differential maternal treatment varied across the different measures of AL and shyness, suggesting that parent perceptions may play a role in these associations.

Parents are keenly aware of the behavioral differences between their children. This is particularly true for dimensions of temperament. When asked about their children's temperaments, parents frequently comment on how different their children are (e.g., "different as night and day" or "as if they came from different families"). Such statements imply that siblings show no resemblance in their temperaments. However, behavioral genetic research consistently reveals that temperament is genetically influenced (Saudino, 1997); therefore, genetically related siblings should show *some* similarity in temperament. Could parents be exaggerating the behavioral differences between their children? If so, do the observed relations between sibling differences in temperament and child outcome reflect parental expectations rather than actual child behavior?

The present study examines whether parents' perceptions of sibling differences are veridical. Evidence from twin studies suggests that they may not be. Twin studies that assess temperament with parent rating measures frequently produce an unusual pattern of results. Identical (MZ) co-twin resemblance for temperament dimensions is typically moderate, whereas fraternal (DZ) co-twin resemblance is very low, often near zero or even negative (e.g., Goldsmith, Buss, & Lemery, 1997; Neale & Stevenson, 1989; Plomin et al., 1993). This puzzling outcome is particularly evident with, but not exclusive to, rating measures that require parents to make global judgments of their child's behavior (e.g., Goldsmith et al., 1997). The low DZ twin

Correspondence concerning this article should be addressed to Kimberly J. Saudino, Department of Psychology, Boston University, 64 Cummington Street, Boston, MA 02215. E-mail: ksaudino@bu.edu.

This research was supported by Grant MH60816-01 from the National Institute of Mental Health. We gratefully acknowledge the help of Susan Fenstermacher, Denise Hines, and the families who participated in the study.

resemblance is puzzling because the simple genetic model predicts that DZ twin similarity should be at least *half* that of the MZ twins. The pattern of very low DZ correlations that emerges with parent ratings of temperament is significant because it implies that DZ twins are perceived as no more similar than two randomly paired children and, in some instances, are regarded as having *opposing* temperaments. The use of more objective measures of temperament, however, tells a different story. When temperament is assessed via behavioral observation or mechanical measures DZ similarity is not inappropriately low. That is, DZ correlations for temperament dimensions tend to be positive and significantly different from zero (e.g., Cherny, Fulker, Corley, Plomin, & DeFries, 1994; Plomin et al., 1993; Saudino & Eaton, 1991, 1995). The fact that DZ co-twins show substantial resemblance for objectively assessed temperament but not for parent-rated temperament suggests that parent ratings are prone to *contrast effects*, rater biases that magnify existing behavioral differences between co-twins. Twin and adoption studies have found that contrast effects influence a range of temperament dimensions (Goldsmith et al., 1997; Saudino, Cherny, & Plomin, 2000; Saudino, McGuire, Reiss, Hetherington, & Plomin, 1995); however, activity level (AL) is the temperament dimension that most frequently displays contrast effects.

Most temperament researchers have considered contrast effects to be a problem specific to behavioral genetic research. However, the finding that parent rating measures of temperament are prone to contrast effects has important implications more generally for temperament research. Parent ratings are the most widely used measures of temperament in the literature. Many conclusions regarding child behavior and developmental outcomes (e.g., behavior problems, sibling relationships) have been based on studies using such measures. Yet, twin research suggests that such measures may be flawed. Contrast effects may not be limited to parents' ratings of twins. That is, it is possible that whenever parents rate the temperaments of their children, they evaluate each child in the context of other children that they know well—most likely other children within the family. Thus, just as with twins, the rating of one sibling's temperament is likely to be influenced by the perceived temperament of another sibling. Because most temperament research has involved the study of only one child per family, contrast effects have not been evident even when they do exist.

Evidence for contrast effects in nontwin samples comes from the Colorado Adoption Project (CAP; DeFries, Plomin, & Fulker, 1994; Plomin & DeFries, 1985; Plomin, DeFries, & Fulker, 1988), in which parent ratings of temperament in infancy and middle childhood have yielded a pattern of near-zero or slightly negative correlations for both nonadoptive and adoptive sibling pairs (Plomin, Coon, Carey, DeFries, & Fulker, 1991; Schmitz, 1994). Although consistent with twin research, these results may underestimate parents' tendency to contrast siblings because CAP parents were required to rate the temperaments of each child at the same chronological age, and consequently, there was an average 3-year interval between the ratings of the two siblings. Thus, the CAP results do not reflect parents' current perceptions of differences between siblings.

Although not well studied, there are hints that a lack of sibling resemblance for parent ratings of temperament is not limited to behavioral genetic research designs. Schachter and colleagues (Schachter, Gilutz, Shore, & Adler, 1978; Schachter, Shore, Feldman-Rotman, Marquis, & Campbell, 1976; Schachter & Stone, 1987) have observed that family members tend to describe siblings as different or contrasting and refer to this as “sibling deidentification.” As with the findings of twin and adoption studies, in families with two or three children, siblings displayed significant *negative* correlations for maternal ratings on a single item assessing easy–difficult temperament (Schachter & Stone, 1985). The sibling deidentification perspective suggests that these sibling differences are real behavioral differences resulting from a desire to mitigate sibling rivalry (Schachter et al., 1976, 1978); however, we suggest that rater contrast biases may be operating.

To answer the question of whether perceived behavioral differences reflect actual behavioral differences, researchers need to compare parent perceptions with objective measures of temperament. In the present study, we explored parent contrast effects for AL and shyness, two temperament dimensions in which the discrepancy between parent and observer ratings is particularly striking in both twin and adoption studies (e.g., Braungart, Plomin, DeFries, & Fulker, 1992; Plomin et al., 1991; Saudino & Cherny, 2001a, 2001b). Moreover, both dimensions can be objectively assessed with reasonable ecological validity. For example, activity can be assessed using motion recorders that can be worn in the home for extended periods. Similarly, the situation of the child entering the laboratory and meeting the examiner provides a natural opportunity to assess shyness, the child's initial reaction to unfamiliar people.

We also considered the possible implications of parental contrast effects by exploring the relation between sibling differences on parent ratings and our more objective measures of AL and shyness with two outcome measures, specifically sibling relationships and parental differential treatment. Sibling differences in temperament appear to be systematically related to experiential differences between siblings. For example, activity level is a temperament dimension that most reliably accounts for variance in the quality of sibling relationships (Stoneman & Brody, 1993). Highly active children tend to have more conflictual relationships with their siblings. In fact, one study found that young children with high ALs had four times as much sibling conflict as low-activity children (Mash & Johnson, 1983). Children who are higher in activity both give and receive more antagonistic behaviors within the sibling relationship (Brody, Stoneman, & Burke, 1987). Additionally, sibling differences in activity appear related to the relationship between siblings. Siblings who differ in activity experience high levels of conflict and negativity and less warmth in the sibling relationship (Stoneman & Brody, 1993). Shyness is also related to the quality of the sibling relationship. When older siblings are shy, sibling relationships are characterized as being less controlling and less competitive (Stocker, Dunn, & Plomin, 1989). Similarly, sibling differences in shyness have been related to differential sibling closeness and caretaking. That is, the sibling who is shyer experiences less closeness and less caretaking within the sibling relationship (Daniels, 1986). The question of whether sibling differences in temperament are associated with differential parental treatment is less clear. AL in childhood has been found to be negatively associated with maternal warmth (McGuire & Dunn, 1994) and favoritism (Stocker, 1995), suggesting that the sibling who is more active has a less favorable relationship with the mother. However, these results did not replicate in two studies of young adults' self-report ratings of temperament and differential parental treatment (Daniels, 1986; Mann, 1993).

Although the research suggests that sibling differences in temperament are related to child outcome, these studies have generally relied on parent ratings of temperament. If parent rating measures are prone to contrast effects, then it is possible that parent *expectations* about their children's temperament, rather than actual child behavior, are driving these observed relations. Indeed, Brody, Stoneman, and Burke (1988) found that parents' perceptions of differences between siblings' ALs were a more important predictor of adjustment than the absolute AL of the child. These authors suggest that parents may label one child as maladjusted on the basis of perceptions of behavioral differences between siblings that are not objectively valid. Therefore, in an attempt to explore a potential consequence of parent contrast effects, we examine whether the association between sibling differences in temperament and the quality of sibling relationships or differential parental treatment differ across parent and objective measures of AL and shyness.

Method

Sample

The sample included 95 sibling pairs (33 male–male; 20 female–female; 42 male–female). The mean ages of older and younger siblings were 7.6 years ($SD = 2.9$) and 5.2 years ($SD = 2.2$), respectively. The sample was predominantly Caucasian, highly educated (50% of parents had some postgraduate education) and middle-class according to the Hollingshead (1975) index of socioeconomic status (SES).

Procedure

The procedure involved two visits, 48 hr apart, to the laboratory. On arrival at the laboratory at the initial visit, the children engaged in freeplay while the motion recorders (actigraphs) were demonstrated to the parent. The children were then given a series of tasks to determine each child's dominant arm and leg (e.g., throw a ball, kick a ball, draw, step on a riser). Motion recorders were attached and parents were provided with instructions regarding the care of the actigraphs and record sheets for logging the time that the actigraphs were off limbs (e.g., for baths, swimming). Parents were also given questionnaires regarding the sibling relationship, differential treatment, parent personality, and family demographics to be completed by the next visit. At the second visit, the motion recorders were removed and questionnaires collected.

Measures

Activity measure—During the 2-day data collection period, AL was assessed with Computer Science and Applications' Activity Monitors Model 7164 V1.2. This device is a miniature single-channel accelerometer that has been designed to detect motion within the range of normal human movement (Activity Monitor Operator's Manual Model 7164, Version AM7164–2.2; Computer Science and Applications, 1999). Actigraphs have been shown to be valid and reliable instruments for recording the activity of children (Freedson et al., 1997; Janz, 1994). Each child was randomly assigned two activity monitors (one arm and one leg). To differentiate motion recorders within sibling pairs, we covered each monitor with white surgical tape on which the child's name was printed. The limb and direction of attachment was also printed on the tape. To prevent the children from easily removing the monitors, they were attached by means of Tyvek (Dupont Tyvek, Richmond, VA) adhesive wristbands. Arm attachment, at the wrist, was on the dorsal aspect of the forearm proximal to the radial carpal joint. Leg attachment, at the ankles, was superior to the lateral malleoli.

Parents were encouraged to engage in normal activities and to maintain daily routines with their children during the 2-day data collection period. Parents were asked to record, on the sheets provided, the times when an activity monitor was off a limb. The average length of time that the motion recorders were worn was 44.9 hr ($SD = 5.6$). Data for 2 sibling pairs were lost because of equipment failure and another was lost because of the refusal of 1 child to wear the motion recorders.

To adjust for variations in the total time that each instrument was worn, we converted the number of activity units to a rate per minute real time. Arm and leg activity counts were correlated ($r = .59, p < .001$); therefore, a composite actometer score (Actigraph Total) reflecting overall motor activity was calculated by averaging the two limb actometer scores. Two additional actigraph scores were calculated. Because the salient activity for parents might be the activity that they observe, a “daytime” measure of actigraph activity that included activity counts for the hours from 7 a.m. to 9 p.m. for both days was computed (Actigraph Day). Similarly, parents may be more attuned to how often a child's AL exceeds a high activity threshold rather than the overall rate of AL; therefore, a threshold measure (Actigraph Threshold) was computed as the total number of episodes in which each child's activity counts

exceeded the 75th percentile for daytime activity divided by the total number of episodes. All three actigraph scores were highly correlated (Actigraph Total—Actigraph Day: $r = .89, p < .001$; Actigraph Total—Actigraph Threshold: $r = .89, p < .001$; Actigraph Day—Actigraph Threshold: $r = .80, p < .001$).

Shyness measure—Following procedures from the MacArthur Longitudinal Twin Study (Plomin et al., 1990), we obtained an observational measure of shyness based on the children's initial reactions to the examiner on arrival at the laboratory. The initial visit was videotaped through a one-way mirror, and the first 5 min after entry served as the shyness episode. Observers rated each child (different rater for each member of a sibling pair) for the occurrence of discrete behaviors during each minute of the 5-min shyness episode. Observed behaviors included approach to the examiner, approach to a proffered toy, proximity to mother, clinging to mother, self-soothing, vocalization, and crying. Global ratings of shyness and hesitation were also obtained. All items were entered in a principal-components analysis and the first unrotated principal component was used as an index of shyness. To assess reliability of this measure, two observers rated the same child for 20% of the sample. The correlation between raters was high ($r = .91, p < .05$), indicating excellent interrater agreement. Although this measure of shyness is brief, previous research has found it to show significant stability across age and contexts (Cherny et al., 1994; Saudino & Cherny, 2001b).

Parent-report measures of temperament—Because twin research has suggested that low DZ correlations are particularly evident with rating scales that require parents to make global judgments about their child's behavior (Plomin, 1982), we included two parent rating measures of temperament: The Colorado Childhood Temperament Inventory (CCTI; Rowe & Plomin, 1977), which had been modified to include separate and distinct scales of Shyness and Sociability (Buss & Plomin, 1984). This measure contains 30 general statements, for example, “child cries easily” or “child is very energetic,” describing the temperament dimensions of Emotionality, Activity, Sociability, Shyness, and Persistence. Parents were asked to rate each statement on a 5-point Likert scale ranging from 1 (*strongly disagree; not at all like the child*) to 5 (*strongly agree; a lot like the child*). Internal consistencies, as measured by Cronbach's alpha, ranged from .70 for Sociability to .87 for Shyness. The second parent-report measure of temperament, the Child Behavior Questionnaire (CBQ; Rothbart, Ahadi, & Hershey, 1994) asks parents to rate specific behaviors in specific situations (e.g., “tends to run rather than walk from room to room”). The CBQ consists of 195 items assessing 16 behavioral dimensions (e.g., Activity Level, Anger/Frustration, Approach, Attentional Focusing, Attentional Shifting, Discomfort, Reactivity/Soothability, Fear, High Intensity Pleasure, Impulsivity, Inhibitory Control, Low Intensity Pleasure, Perceptual Sensitivity, Sadness, Shyness, and Smiling and Laughter). For each item, parents are asked to indicate on a 7-point scale whether the statement is a “true” or “untrue” description of their child. The CBQ demonstrated excellent reliability with internal consistencies ranging from .71 for Low Pleasure to .95 for Shyness. In addition to the CCTI and the CBQ, a global rating of easy or difficult temperament was obtained on a single 5-point Likert item (1 = *easy* to 5 = *difficult*).

Sibling relationship—Positivity and negativity within the sibling relationship were assessed through a maternal report measure of the sibling relationship adapted from Stocker et al. (1989). For each child, parents were asked to indicate the frequency of 11 discrete behaviors displayed within the child's relationship with his or her sibling. Items assessed display of affection, helping, sharing, quarrels, physical fighting, jealousy, competition, and conflict resolution. With the exception of the display of affection item, which was scored on an 8-point scale (0 = *almost never*; 7 = *many times a day*), items were rated on a 6-point scale (0 = *almost never*; 5 = *regularly*). To adjust for the differential scoring, we standardized all items prior to analyses. Factor analyses yielded factors representing sibling positivity (share,

affection, help, make up, and acquiesce; $\alpha = .72$) and negativity (provoke, fight, compete, jealous, quarrel, and run away; $\alpha = .69$).

Differential treatment—A measure of maternal differential treatment was obtained using a modified version of the Sibling Inventory of Differential Experience (SIDE; Daniels & Plomin, 1984). The parent version of the SIDE consists of 12 questions regarding differential parental affection and control (e.g., “Who are you more affectionate toward?” “Who do you need to discipline more?”). Four additional items assessing differential privileges, discipline, affection, and arguing were included. Items were scored using a 5-point relative scoring system such that higher scores indicate that the item applies more to the younger child (i.e., 1 = *older child much more*; 3 = *both children the same*; 5 = *younger child much more*). Consistent with Daniels and Plomin, factor analyses yielded a maternal Control factor ($\alpha = .90$), however, maternal affection items loaded on two separate factors: Interest (“enjoy doing things with,” “proud of,” “interested in”; $\alpha = .63$) and Sensitivity (“particularly sensitive to,” “needs more support/leniency,” “needs more attention”; $\alpha = .74$).

Parent personality—Because previous research has demonstrated that parental characteristics such as personality can bias perceptions of child temperament (e.g., Bates & Bayles, 1984; Vaughan, Bradley, Joffe, Seifer, & Barglow, 1987), parent personality was also assessed. Neuroticism and extraversion were assessed on the Eysenck Personality Inventory (EPI; Eysenck & Eysenck, 1975). Each scale consists of 24 items, scored either 1 = yes or 0 = no. Internal consistency was .81 for Neuroticism and .74 for Extraversion. The EAS Temperament Survey (Buss & Plomin, 1984) provided measures of parent Emotionality, Anger, Fear, Activity, and Sociability. Parents were presented with 20 items and asked to indicate on a 5-point scale (1 = *not at all typical*; 5 = *very typical*) how well each item described their own behavior. Internal consistencies for the EAS ranged from .59 for Fear to .71 for Sociability.

Results

Agreement Between Parent Ratings and Objective Measures

Table 1 presents the intercorrelations between parent ratings and objective measures. For both AL and shyness, parent ratings on the two different rating scales were highly correlated (Activity: $r = .68, p < .05$; Shyness: $r = .86, p < .05$), but the correlations between parent ratings and the more objective measures were modest. Although parent ratings of AL on the CCTI were significantly correlated with actigraph scores, none of the correlations indicating the agreement between CBQ AL and actigraph scores were significant. Moreover, the correlation between the two parent ratings of AL was significantly stronger than the correlations between the actigraph scores and either the CCTI (Actigraph Total: $z = 5.89, p < .001$; Actigraph Day: $z = 5.52, p < .001$; Actigraph Threshold: $z = 5.92, p < .001$) or the CBQ (Actigraph Total: $z = 4.67, p < .001$; Actigraph Day: $z = 4.63, p < .001$; Actigraph Threshold: $z = 4.84, p < .001$). Similarly, for shyness the association between parent rating measures was significantly stronger than that between parent ratings and observations, CCTI ($z = 8.83, p < .001$) or the CBQ ($z = 8.47, p < .001$).

Sibling Resemblances

Sibling intraclass correlations indexing sibling resemblances for temperament are presented in Table 2. Although the present research focuses specifically on AL and shyness, Table 2 includes the correlations for all dimensions as assessed by parent ratings on the CCTI and the CBQ in addition to those for actigraph AL and observed shyness. Overall, the significant negative correlations for the CCTI subscales (first column Table 2) indicate that when parents made global judgments about their children's temperament behaviors they rated their children as

differing in temperaments. Similar results emerged for our global rating of easy–difficult temperament ($r = -.16, p < .05$). The pattern of negative correlations between siblings was less prominent when parents made more specific behavioral ratings on the CBQ (second column Table 2); however, there was no significant sibling resemblance for 8 of the 16 scales. In general, parents reported more sibling similarities on emotion-based CBQ scales.

The sibling results for AL and shyness are remarkably consistent with the findings of twin studies. Siblings showed no resemblance in parent-reported AL or shyness. The significant negative correlations for parent-rated AL and shyness are particularly noteworthy because they indicate that parents see their children as having *opposing* temperaments. The significant positive correlations for all three actigraph AL scores and observer-rated shyness indicate that siblings do, however, show substantial resemblance for the same temperament dimensions when assessed with more objective methods.

Relative magnitude of sibling differences—Are those siblings whose parents reported them most different in AL or shyness those that were most different on the observational measures? The intraclass correlations clearly show that parents do exaggerate differences in AL and shyness between their children, but perhaps they are correct about the magnitude of sibling differences relative to other sibling pairs. To investigate this, we created sibling difference scores for each measure (i.e., parent rated and observed) by taking the absolute value of the older sibling's score minus that of the younger sibling. Perceived sibling differences on parent-reported temperament were then correlated with observed differences in objectively assessed temperament (see Table 3). For both AL and shyness, sibling differences on the CCTI were highly correlated with differences on the CBQ (Activity: $r = .55, p < .05$; Shyness: $r = .70, p < .05$), but sibling differences on these parent-report measures were not correlated with differences on our more objective measures. In other words, parent-reported sibling differences are unrelated to observed sibling differences in AL or shyness.

Rater characteristics and contrast effects—Are rater characteristics related to the tendency to contrast siblings? To investigate this, we correlated parent personality, SES, and education with sibling differences on the parent rating measures. The only significant correlations to emerge were between sibling differences on the CBQ Activity and parent Extraversion ($r = -.21, p < .05$) scales and sibling differences on CBQ Shyness and SES ($r = .23, p < .05$). Thus, there is little evidence that parent personality or demographics systematically influence the contrast biases.

Sibling age differences and contrast effects—Is there a greater tendency to contrast siblings who are more different in age? No. For both AL and shyness, reported sibling differences on the CCTI and CBQ were unrelated to age differences between siblings (Activity: CCTI_{diff}: $r = -.13$; CBQ_{diff}: $r = .00$; Shyness: CCTI_{diff}: $r = -.03$; CBQ_{diff}: $r = -.04$).

Sex differences and contrast effects—Are same-sex (SS) sibling pairs contrasted more than opposite-sex (OS) pairs? No. There were no significant mean differences between sibling differences for SS and OS sibling pairs for either parent rating measure of activity or shyness (Activity CCTI: SS $M_{\text{diff}} = 4.7$, OS $M_{\text{diff}} = 4.7$, $t(88) = -0.09$, *ns*; Activity CBQ: SS $M_{\text{diff}} = 12.6$, OS $M_{\text{diff}} = 14.0$, $t(92) = -0.71$, *ns*; Shyness CCTI: SS $M_{\text{diff}} = 6.1$, OS $M_{\text{diff}} = 6.1$, $t(90) = -0.04$, *ns*; Shyness CBQ: SS $M_{\text{diff}} = 21.2$, OS $M_{\text{diff}} = 24.4$, $t(92) = -0.94$, *ns*). Similarly, there were no significant differences between intraclass correlations for SS and OS sibling pairs (Activity CCTI: $r_{\text{OS}} = -.13$, $r_{\text{SS}} = -.28$; Activity CBQ: $r_{\text{OS}} = -.14$, $r_{\text{SS}} = .08$; Shyness CCTI: $r_{\text{OS}} = -.16$, $r_{\text{SS}} = -.21$; Shyness CBQ: $r_{\text{OS}} = -.14$, $r_{\text{SS}} = -.25$).

Sibling Differences in AL and Shyness and Outcomes

Quality of sibling relationship—To explore the relation between sibling differences in temperament and the sibling relationship, we created overall negativity and positivity scores by averaging the negativity or positivity scores for both siblings. Overall negativity and positivity scores were then correlated with sibling differences (unsigned) in temperament (see Table 4). Neither actigraph AL nor observed shyness was associated with overall positivity of the sibling relationship. However, the significant correlations between CBQ activity and overall positivity in the sibling relationship indicate that siblings who differed more in parent-rated AL had less positive relationships. Sibling differences in temperament were generally unrelated to overall negativity in the sibling relationship; however, two associations approached significance. Sibling difference scores for negativity and positivity were also calculated (e.g., older positivity minus younger positivity) and the correlations between signed sibling difference scores for temperament and for negativity or positivity examined (see Table 4). Signed difference scores express both the amount and direction of difference (i.e., a negative score means the younger child scores higher for the measured temperament dimension). The difference score correlations in Table 4 give an indication of the relationship between a child's temperament relative to his or her sibling and that child's positivity or negativity in the relationship relative to his or her sibling. Once again, parent ratings and our more objective measures yield different results. Sibling differences on the actigraph measures, but not parent ratings, were positively associated with sibling differences in negativity. The sibling who was more active on the 2-day mechanical measure of total AL or of daytime AL, or who had more episodes above the AL threshold, behaved more negatively within the sibling relationship.

For shyness, sibling differences on both the CCTI and CBQ were positively associated with differences on parent-rated negativity. The sibling who was rated most shy on these measures was also rated as behaving more negatively within the sibling relationship. When shyness was assessed by observation, the opposite pattern emerged—the sibling who was observed to be most shy was rated as behaving *less* negatively within the sibling relationship; however, this association only approached significance ($p < .08$).

Differential treatment—Maternal reports of control, interest, and sensitivity on the SIDE were used to predict sibling differences in temperament. The SIDE provides a relative measure of differential treatment such that high scores indicate differential treatment in favor of the younger child and low scores differential treatment in favor of the older child. Thus, correlations between signed sibling differences in temperament and SIDE maternal control, interest, and sensitivity provide information not only on the relation of temperament and treatment differences but also on the direction of the effects. As can be seen in Table 5, sibling differences in AL on all activity measures were related to differential experiences in maternal control. The negative correlations between sibling differences in activity and control indicate that mothers report that the sibling who is more active experiences more parental control relative to his or her sibling (i.e., high scores on the activity difference measure indicate that the older child is more active; low scores on the SIDE indicate parental behavior favoring the older child). Sibling differences in CBQ activity were also related to differential maternal interest. The sibling who is perceived as more active on the CBQ is the sibling in whom the mother is relatively less interested. Sibling differences in shyness were also related to differential treatment, but only when shyness was assessed with parent ratings. The negative correlations for CCTI and CBQ shyness indicate that the sibling that is perceived as more shy is treated with relatively more sensitivity.

Discussion

The data from the present study clearly indicate that parents' perceptions of sibling differences in the temperament dimensions of AL and shyness do not reflect observed behavioral

differences. Parent ratings of both temperament dimensions on two different measures indicated that parents see no resemblances or, indeed, significant dissimilarities in their children's ALs and shyness. However, consistent with temperament theory that assumes a genetic basis (Goldsmith et al., 1987), and therefore familial resemblance, when these two temperament dimensions are assessed with more objective methods siblings showed substantial similarity. Thus, as has been found with twin studies, parents tended to exaggerate the differences between their children's temperaments. In other words, parents' ratings of temperament in nontwin siblings are prone to contrast effects.

Little is known about the factors that influence parents' tendencies to contrast their children. Twin research suggests that some measures may be more prone to contrast effects than others. For example, the very low DZ correlations that emerge from twin studies tend to be more evident when parents are required to make global judgments (e.g., "always on the go") as opposed to specific judgments (e.g., "splashes when in the tub") about their children's behavior (Goldsmith & Hewitt, 2002; Hwang & Rothbart, 2002; Plomin, 1982); however, even specific measures can yield too low DZ correlations (e.g., Goldsmith et al., 1997). To our knowledge, there has been no prior research that directly compares contrast effects for global versus specific measures—certainly not in non-twin siblings. In the present study, evidence of contrast effects emerged for both the global and specific parent rating measures of activity and shyness, but there was a tendency for contrast effects to be more pervasive when global ratings were required. That is, there was substantial sibling *dissimilarity* for *all* temperament dimensions on the CCTI (i.e., all sibling correlations were negative and most were significant), whereas on the CBQ, there was significant sibling resemblance for half of the dimensions assessed. Moreover, although there was a lack of resemblance for half of the dimensions assessed on the CBQ, fewer significant negative correlations emerged. Interestingly, a recent twin study using the CBQ found a very similar pattern of results (Goldsmith et al., 1997). MZ twins showed significant and substantial similarity for *all* CBQ dimensions, whereas DZ twins displayed no resemblance for 10 of the 16 scales, including Activity Level and Shyness. Seven of these scales yielded negative correlations for DZ twins, but only one, Activity Level, was significant. Moreover, the pattern of CBQ dimensions showing a lack of sibling resemblance was largely overlapping across our sibling and Goldsmith et al.'s (1997) twin samples (e.g., near zero or negative correlations for Activity Level, Shyness, Attentional Focusing, Attentional Shifting, Impulsivity, Soothability, and Discomfort). Thus, contrast effects are problematic for both types of ratings, but they are more prodigious when global ratings are required.

There are also hints that the tendency to contrast siblings differs across temperament dimensions. As indicated above, certain scales on the CBQ consistently showed no resemblances for siblings or DZ twins. However, the siblings in our study showed significant similarities on the more affective behavioral dimensions on the CBQ. Again, these results mirror those found with DZ twins assessed on the same instrument (Goldsmith et al., 1997). In both studies, there were significant positive correlations for Approach, Fear, Low Pleasure, Smiling & Laughter, and Perceptual Sensitivity. The substantial sibling similarities for these dimensions suggest that they are less likely to be affected by contrast effects. Of course, without comparing these results to more objective measures, we cannot be certain that contrast effects are not operating (i.e., it is possible that objective measures yield even stronger resemblances). Moreover, it is also possible that parents are *overestimating* sibling similarities for these affective dimensions. Again, comparison with objective measures is needed. Nonetheless, the possibility that some dimensions are more prone to contrast effects than others is intriguing and warrants further attention.

What factors influence the tendency for parents to contrast their children's temperaments? Sibling constellation variables such as sex, age, or birth order have been suggested as factors that might affect the contrast process (Carey, 1986). Neither sibling differences in sex nor age

were associated with sibling differences in AL or shyness in the present sample. Because over two thirds of our sample were first-second sibling pairs, it was not possible to evaluate differences in birth order, but the high correlation between birth order and age ($r = -.66, p < .01$) suggests that it would contribute little unique variance above and beyond age. Parent personality and demographics were also not related to their reports of sibling differences. In retrospect, this makes sense given the observation of the general tendency of parents to describe their children as different (“night and day”) in casual conversations. That is, if contrast biases are common to most parents, then there is little variance to be explained by personality or demographics.

The design of the motion recorders, which collect real-time data, allowed us to evaluate other possible explanations for differences between findings for actigraph versus parent-rated AL. For example, we wondered if sibling resemblance for overall actigraph AL was inflated because of the fact that the children wore the motion recorders to bed and almost all children in our sample showed very little activity during nighttime hours. This could also explain modest parent-actigraph agreement because parents were asked to report on waking activities, whereas the actigraph assessed AL continuously (i.e., awake and asleep). Alternatively, we considered whether the low parent-actigraph agreement and the discrepancy between sibling resemblances across the parent report and actigraph measures might be due to parents being more sensitive to bouts of high activity than to the overall rate of activity. Neither of these possibilities appears to be the case. Restricting the actigraph scores to daytime activity or to the proportion of episodes above a high AL threshold resulted in the same or slightly *higher* estimates of sibling similarities and in no significant differences in parent-actigraph agreement. Clearly, other explanations are needed. It is possible that contextual differences between parent ratings and actigraph measures of AL are responsible for the different outcomes when these measures are used (i.e., parents are generalizing across multiple situations that go beyond those assessed during the 48-hr actigraph session). However, this seems unlikely given that similar results have emerged from twin studies of AL in which parent ratings and mechanical measures of activity are based on the *same* 48-hr period (Saudino & Eaton, 1991; Saudino, Gagne, & Thompson, 1999).

Perhaps labeling siblings' temperaments as “active/not active” or “shy/not shy” on the basis of subtle behavioral differences provides parents with a heuristic for understanding and interacting with their different children. One sibling's temperament may serve as an anchor for evaluating the temperament of other siblings. Alternatively, contrast effects might be due to the fact that parents value and seek to promote the development of the individuality of each child within a family. It would be interesting, therefore, to see whether contrast effects are apparent in parent ratings of sibling temperament in more collectivist cultures.

As is usually the case with developmental research, our sample was predominantly Caucasian, middle-class, and well-educated. This limitation raises the question of the generalizability of results. Replication with a more diverse sample is needed. However, given that much of what we know about child temperament is based on research with similar sample characteristics, our findings of sibling contrast effects have considerable implications. Parents were asked to rate each child's temperament individually, yet the ratings of one child's AL and shyness were systematically related to the other child's such that the higher the score for one sibling, the lower for the other. In other words, the ratings of siblings' temperaments are not independent of each other. This finding raises important questions about the validity of parent ratings of temperament, especially when such measures are used to evaluate sibling differences. As indicated in the present study, parent ratings and our more objective measures produced different patterns of associations between sibling differences in activity and shyness and the outcome measures of sibling relationship and differential maternal treatment. Thus, as suggested by Brody et al. (1988), links between sibling differences on parent-rated

temperament and outcome may reflect parent perceptions rather than actual child behavior. The fact that parent perceptions of sibling differences in temperament do not correspond with our more objective measures calls into question the interpretations of the abundance of sibling research that has relied solely on parent-report measures of temperament.

Contrast effects reveal themselves when parents are asked to rate the temperaments of more than one child; however, this is not simply an epiphenomenon of a concurrent rating process. In the CAP, zero and negative sibling correlations for parent ratings of temperaments emerged even though there was an average 3-year interval between ratings (Plomin et al., 1991; Schmitz, 1994). It is likely that parents contrast and compare their children and form perceptions about their temperaments outside the rating process. Their ratings on temperament scales reflect these perceptions. Thus, contrast effects would be an important threat to validity even when obtaining parent ratings of temperament based on a single child and may explain the low associations between parent-rated temperament and observer ratings of the same behaviors that are frequently found in the literature (e.g., Hubert, Wachs, Peters-Martin, & Gandour, 1982; Rothbart, 1986; Saudino & Cherny, 2001a; Seifer, Sameroff, Barrett, & Krafchuk, 1994). Comparing temperament data for families with only one child (i.e., no opportunity for contrast to siblings) to those for families with more than one child could provide valuable information about the potential threats to validity that contrast effects pose. For example, if the associations between parent and observer ratings are higher for single-child families, it would suggest that the low parent–observer agreement observed in the literature is not simply a result of contextual differences, but that parental contrast biases may well be operating.

The issue of contrast effects in parent-rated temperament has long been recognized in behavioral genetic research. Statistical models incorporating such rater biases have been developed and are routinely used whenever correlational or variance analyses suggest that contrast effects may be present. However, the significance of contrast effects for temperament research more generally has not been considered. The results from the present study suggest that contrast effects are a potential problem whenever parent ratings of temperament are used. On the basis of the present findings, we recommend that researchers consider sibling analyses as an important part of the validation process for temperament measures. In addition, researchers need to investigate the factors that influence the tendency for parents to contrast their children. The more we know about causes of contrast effects, the better we will be able to control for them.

References

- Bates JE, Bayles K. Objective and subjective components in mothers' perceptions of their children from age 6 months to 3 years. *Merrill-Palmer Quarterly* 1984;30:111–130.
- Braungart JM, Plomin R, DeFries JC, Fulker DW. Genetic influence on tester-rated infant temperament as assessed by Bayley's Infant Behavior Record: Nonadoptive and adoptive siblings and twins. *Developmental Psychology* 1992;28:40–47.
- Brody GH, Stoneman Z, Burke M. Child temperaments, maternal differential behavior, and sibling relationships. *Developmental Psychology* 1987;23:354–362.
- Brody GH, Stoneman Z, Burke M. Child temperament and parental perceptions of individual child adjustment: An intrafamilial analysis. *American Journal of Orthopsychiatry* 1988;58:532–542. [PubMed: 3228160]
- Buss, AH.; Plomin, R. *Temperament: Early developing personality traits*. Erlbaum; Hillsdale, NJ: 1984.
- Campbell SB, Breaux AM. Maternal ratings of activity level and symptomatic behaviors in a nonclinical sample of young children. *Journal of Pediatric Psychology* 1983;8:73–82. [PubMed: 6842352]
- Carey G. Sibling imitation and contrast effects. *Behavior Genetics* 1986;16:319–341. [PubMed: 3753368]

- Cherny SS, Fulker DW, Corley R, Plomin R, DeFries JC. Continuity and change in infant shyness from 14 to 20 months. *Behavior Genetics* 1994;24:365–380. [PubMed: 7993315]
- Computer Science and Applications. Activity monitor operator's manual: Model 7164, Version AM7164–2. 2. [Brochure]. Author; Shalimar, FL: 1999. Ambulatory Monitoring Applications Division.
- Daniels D. Differential experiences of siblings in the same family as predictors of adolescent sibling personality differences. *Journal of Personality and Social Psychology* 1986;51:339–346. [PubMed: 3746616]
- Daniels, D.; Plomin, R. *The Sibling Inventory of Differential Experience*. University of Colorado, Institute for Behavioral Genetics; Boulder: 1984.
- DeFries, JC.; Plomin, R.; Fulker, DW. *Nature and nurture during middle childhood*. Blackwell; Cambridge, MA: 1994.
- Eysenck, HJ.; Eysenck, SB. *Manual of the Eysenck Personality Inventory*. Educational & Industrial Testing Service; San Diego, CA: 1975.
- Freedson, PS.; Sirard, J.; Debold, N.; Pate, R.; Dowda, M.; Sallis, J. Validity of two physical activity monitors in children and adolescents; Paper presented at the 19th Symposium of European Group of Pediatric Work Physiology; Moretonhampstead, UK. Sep. 1997
- Goldsmith HH, Buss KA, Lemery KS. Toddler and childhood temperament: Expanded content, stronger genetic evidence, new evidence for the importance of environment. *Developmental Psychology* 1997;33:891–905. [PubMed: 9383612]
- Goldsmith HH, Buss AH, Plomin R, Rothbart MK, Thomas A, Chess S, et al. Roundtable: What is temperament? Four approaches. *Child Development* 1987;58:505–529. [PubMed: 3829791]
- Goldsmith HH, Hewitt EC. Validity of parental report of temperament: Distinctions and needed research. *Infant Behavior and Development* 2002;26:108–111.
- Hollingshead, A. *Four-factor Index of Social Status*. Yale University; New Haven, CT: 1975. Unpublished manuscript
- Hubert NC, Wachs TD, Peters-Martin P, Gandour MJ. The study of early temperament: Measurement and conceptual issues. *Child Development* 1982;53:571–600.
- Hwang J, Rothbart MK. Behavior genetics studies of infant temperament: Findings vary across parent-report instruments. *Infant Behavior and Development* 2002;26:112–114.
- Janz KF. Validation of the CSA accelerometer for assessing children's physical activity. *Medicine and Science in Sports and Exercise* 1994;26:369–375. [PubMed: 8183103]
- Mann TL. A failure of nonshared environmental factors in predicting sibling personality differences. *Journal of Psychology* 1993;127:79–86.
- Mash EJ, Johnson C. Sibling interactions of hyperactive and normal children and their relationship to reports of maternal stress and self-esteem. *Journal of Clinical Child Psychology* 1983;12:91–99.
- McGuire, S.; Dunn, J. Nonshared environment in middle childhood. In: DeFries, JC.; Plomin, R.; Fulker, DW., editors. *Nature and nurture during middle childhood*. Blackwell; Cambridge, MA: 1994. p. 201–213.
- Neale MC, Stevenson J. Rater bias in the EASI temperament scales: A twin study. *Journal of Personality and Social Psychology* 1989;56:446–455. [PubMed: 2926639]
- Plomin, R. Childhood temperament. In: Lahey, B.; Kazdin, A., editors. *Advances in clinical child psychology*. 6. Plenum; New York: 1982. p. 1–80.
- Plomin, R.; Campos, JJ.; Corley, R.; Emde, RN.; Fulker, DW.; Kagan, J., et al. Individual differences during the second year of life: The MacArthur Longitudinal Twin Study. In: Colombo, J.; Fagen, J., editors. *Individual differences in infancy: Reliability, stability, and predictability*. Erlbaum; Hillsdale, NJ: 1990. p. 431–455.
- Plomin R, Coon H, Carey G, DeFries JC, Fulker D. Parent–offspring and sibling adoption analyses of parental ratings of temperament in infancy and early childhood. *Journal of Personality* 1991;59:705–732. [PubMed: 1774616]
- Plomin, R.; DeFries, JC. *Origins of individual differences in infancy: The Colorado Adoption Project*. Academic Press; Toronto, Ontario, Canada: 1985.
- Plomin, R.; DeFries, JC.; Fulker, DW. *Nature and nurture during infancy and early childhood*. Cambridge University Press; New York: 1988.

- Plomin R, Emde R, Braungart JM, Campos J, Corley R, Fulker DW, et al. Genetic change and continuity from 14 to 20 months: The MacArthur Longitudinal Twin Study. *Child Development* 1993;64:1354–1376. [PubMed: 8222877]
- Rothbart MK. Longitudinal observation of infant temperament. *Developmental Psychology* 1986;22:356–365.
- Rothbart MK, Ahadi SA, Hershey KL. Temperament and social behavior in childhood. *Merrill-Palmer Quarterly* 1994;40:21–39.
- Rowe DC, Plomin R. Temperament in early childhood. *Journal of Personality Assessment* 1977;41:150–156. [PubMed: 856967]
- Saudino KJ. Moving beyond the heritability question: New directions in behavioral genetic studies of personality. *Current Directions in Psychological Science* 1997;6:86–90.
- Saudino, KJ.; Cherny, SS. Parent ratings of temperament in twins. In: Emde, RN.; Hewitt, JK., editors. *The transition from infancy to early childhood: Genetic and environmental influences in the MacArthur Longitudinal Twin Study*. Oxford University Press; New York: 2001a. p. 73-88.
- Saudino, KJ.; Cherny, SS. Sources of continuity and change in observed temperament. In: Emde, RN.; Hewitt, JK., editors. *The transition from infancy to early childhood: Genetic and environmental influences in the MacArthur Longitudinal Twin Study*. Oxford University Press; New York: 2001b. p. 89-110.
- Saudino KJ, Cherny SS, Plomin R. Parent ratings of temperament in twins: Explaining the “too low” DZ correlations. *Twin Research* 2000;3:224–233. [PubMed: 11463143]
- Saudino KJ, Eaton WO. Infant temperament and genetics: An objective twin study of motor activity level. *Child Development* 1991;62:1167–1174. [PubMed: 1756660]
- Saudino KJ, Eaton WO. Continuity and change in objectively assessed temperament: A longitudinal twin study of activity level. *British Journal of Developmental Psychology* 1995;13:81–95.
- Saudino, KJ.; Gagne, JR.; Thompson, LA. A multi-method assessment of activity level in adolescent twins; 63rd meeting of the Society for Research in Child Development; Albuquerque, NM. Apr. 1999. Paper presented at the
- Saudino KJ, McGuire S, Reiss D, Hetherington EM, Plomin R. Parent ratings of EAS temperaments in twins, full siblings, half siblings, and step siblings. *Journal of Personality and Social Psychology* 1995;68:723–733. [PubMed: 7738774]
- Schachter FF, Gilutz G, Shore E, Adler M. Sibling deidentification judged by mothers: Cross-validation and developmental studies. *Child Development* 1978;49:543–546.
- Schachter FF, Shore E, Feldman-Rotman S, Marquis RE, Campbell S. Sibling deidentification. *Developmental Psychology* 1976;12:418–427.
- Schachter FF, Stone RK. Difficult sibling, easy sibling: Temperament and the within-family environment. *Child Development* 1985;56:1335–1344.
- Schachter FF, Stone RK. Comparing and contrasting siblings: Defining the self. *Journal of Children in Contemporary Society* 1987;19:55–75.
- Schmitz, S. Personality and temperament. In: DeFries, JC.; Plomin, R.; Fulker, DW., editors. *Nature and nurture during middle childhood*. Blackwell; Cambridge, MA: 1994. p. 120-140.
- Seifer R, Sameroff AJ, Barrett LC, Krafchuk E. Infant temperament measured by multiple observations and mother reports. *Child Development* 1994;65:1478–1490. [PubMed: 7982363]
- Stocker C. Differences in mothers' and fathers' relationships with siblings: Links with children's behavior problems. *Development and Psychopathology* 1995;7:499–513.
- Stocker C, Dunn J, Plomin R. Sibling relationships: Links with child temperament, maternal behavior, and family structure. *Child Development* 1989;60:715–727.
- Stoneman Z, Brody G. Sibling temperaments, conflict, warmth, and role asymmetry. *Child Development* 1993;64:1786–1800. [PubMed: 8112119]
- Vaughan BE, Bradley CF, Joffe LS, Seifer F, Barglow P. Maternal characteristics measured prenatally predict ratings of temperamental difficulty on the Caryl Infant Temperament Questionnaire. *Developmental Psychology* 1987;23:152–161.

Table 1
Parent Rating–Observed Measure Agreement

Observed measure	Parent rating measure	
	CCTI	CBQ
Activity Level		
Actigraph Total	.25 [*]	.11
Actigraph Day	.26 [*]	.16 [†]
Actigraph Threshold	.23 [*]	.11
Shyness	.23 [*]	.28 [*]

Note. CCTI = Colorado Childhood Temperament Inventory; CBQ = Child Behavior Questionnaire.

[†]
 $p < .10$.

^{*}
 $p < .05$.

Table 2
Sibling Intraclass Correlations for Temperament Behaviors

Dimension	Temperament measure					
	Parent rated		Observed measure			
	CCTI	CBQ	Shyness	Actigraph Total	Actigraph Day	Actigraph Threshold
Activity Level	-.21 [*]	.00		.58 [*]	.58 [*]	.66 [*]
Shyness	-.19 [*]	-.19 [*]	.37 [*]			
Emotionality	-.19 [*]					
Sociability	-.24 [*]					
Attention	-.13 [†]					
Attentional Focusing		-.10				
Attentional Shifting		-.13 [†]				
Inhibitory Control		.11				
Impulsivity		-.06 [*]				
Anger/Frustration		.22 [*]				
Approach/Anticipation		.45 [*]				
Discomfort		.14 [†]				
Fear		.38 [*]				
High Pleasure		.21 [*]				
Low Pleasure		.55 [*]				
Sadness		.27 [*]				
Smiling & Laughter		.41 [*]				
Perceptual Sensitivity		.44 [*]				
Soothability		.05				

Note. CCTI = Colorado Childhood Temperament Inventory; CBQ = Childhood Behavior Questionnaire.

[†]
 $p < .10$.

^{*}
 $p < .05$.

Table 3
Correlations Between Sibling Difference Scores Across Measures

Observed measure	Parent rating measure	
	CCTI	CBQ
Activity Level		
Actigraph Total	-.14	-.06
Actigraph Day	-.10	-.04
Actigraph Threshold	-.13	-.10
Shyness	-.07	-.04

Note. CCTI = Colorado Childhood Temperament Inventory; CBQ = Childhood Behavior Questionnaire.

Table 4
Correlations Between Sibling Differences in Temperament and Quality of the Sibling Relationship

Measure	Overall level ^a		Differences ^b	
	Positivity	Negativity	Positivity	Negativity
Activity				
CCTI	-.16	-.15	.17	-.03
CBQ	-.23*	-.01	.12	.10
Actigraph Total	.09	-.09	.03	.24*
Actigraph Day	.15	-.19 [†]	.09	.25*
Actigraph Threshold	.15	-.09	.03	.23*
Shyness				
CCTI	-.01	.08	-.11	.32*
CBQ	.01	.20 [†]	-.05	.30*
Observed	-.08	.08	.14	-.19 [†]

Note. CCTI = Colorado Childhood Temperament Inventory; CBQ = Childhood Behavior Questionnaire.

^a Difference scores are unsigned.

^b Difference scores are signed.

[†]
 $p < .10$.

*
 $p < .05$.

Table 5
Correlations Between Sibling Differences and Differential Treatment

Measure	SIDE Scale		
	Control	Interest	Sensitivity
Activity			
CCTI	-.29 [*]	.10	.03
CBQ	-.41 [*]	.28 [*]	-.09
Actigraph Total	-.34 [*]	.08	-.12
Actigraph Day	-.24 [*]	.03	-.06
Actigraph Threshold	-.30 [*]	-.01	-.19 [†]
Shyness			
CCTI	.03	-.02	-.30 [*]
CBQ	.03	-.07	-.28 [*]
Observed	.06	.04	-.08

Note. Difference scores are signed. SIDE = Sibling Inventory of Differential Experience; CCTI = Colorado Childhood Temperament Inventory; CBQ = Childhood Behavior Questionnaire.

[†]
 $p < .10$.

^{*}
 $p < .05$.