

Reactivity in Infants: A Cross-National Comparison

Jerome Kagan, Doreen Arcus, Nancy Snidman, Wang Yu Feng, John Hendl, and Sheila Greene

Four-month-old infants from Boston, Dublin, and Beijing were administered the same battery of visual, auditory, and olfactory stimuli to evaluate differences in level of reactivity. The Chinese infants were significantly less active, irritable, and vocal than the Boston and Dublin samples, with American infants showing the highest level of reactivity. The data suggest the possibility of temperamental differences between Caucasian and Asian infants in reactivity to stimulation.

Most developmentalists agree that ease and intensity of behavioral arousal to external stimulation is a salient temperamental quality of infants. Rothbart (1989), who calls this characteristic *reactivity*, regards it as one of the two basic temperamental dimensions, along with *self-regulation*. However, because *ease of arousal* is a general term, several issues arise. The first of these is the need to specify response mode and class of incentive. Infants differ with respect to the behaviors that reflect their level of arousal (e.g., motor activity, vocalization, smiling, fretting, or crying) as well as the type of incentive that most often elicits arousal (e.g., visual, auditory, or olfactory stimulation). Most infants show more vigorous motor activity to moving objects than to sounds or speech, but they smile more often to speech than to visual stimuli during the first few months of life (Kagan & Snidman, 1991a). A second issue is whether the early variation in arousal, specified by incentive and behavioral modality, predicts variation in other psychological qualities displayed in later infancy or childhood. A third, related question asks whether infants of different nationalities, differing in genetic or cultural backgrounds or both, differ in ease of arousal.

This article reports a difference between Chinese and Caucasian 4-month-old infants in the level of reactivity to stimulation. Anticipation of the major result can be found in earlier reports that suggested that Asian infants are at a lower level of arousal than Caucasian infants (Caudill & Weinstein, 1969; Freedman, 1974; Freedman & Freedman, 1969; Kagan, Kearsley, & Zelazo, 1978; Lewis, 1989). Over 20 years ago, Freedman and Freedman (1969) reported that newborn Asian-American infants, compared with European-Americans, were calmer, less labile, less likely to remove a cloth placed on their face, and

more easily consoled when distressed. Kagan et al. (1978) found that Chinese-American infants living in Boston were less active, less vocal, and smiled less often to the presentation of visual and auditory events during the first year than did European-American infants.

Caudill and Weinstein (1969), who observed Japanese infants in their homes, reported them to be less easily aroused than European-American infants in the United States. More recently, Lewis (1989) and Lewis, Ramsay, and Kawakami (in press) reported that Japanese infants were less reactive than European-American infants during well baby examinations and were less likely to display intense distress to an inoculation.

These differences observed during the first year have some parallels in older children. The parents of school-age Thai children, compared with those of European-American children, were more concerned over low energy, low motivation, somatic problems, and forgetfulness; the parents of European-American children reported more concerns over disobedience, aggression, and hyperactivity (Weisz et al., 1987; Weisz et al., 1988). These differences have been viewed as reflecting only cultural variation in socialization values. However, assuming the validity of the parental descriptions, one might also regard the evidence as a result, in part, of differences in the temperaments of the children.

The aim of this article is to affirm, in a different experimental context, the possible differences in the level of reactivity between Asian and Caucasian infants. An additional incentive for the research was the recent discovery that 4-month-old European-American infants who showed very low levels of reactivity to the battery used in this study became outgoing and sociable in the second year, whereas infants with high reactivity became fearful to unfamiliar people and events (Kagan & Snidman, 1991a, 1991b). Green's (1969) observation that Chinese-American children in Chicago nursery schools were more subdued compared with European-American infants raises the possibility that the relation between reactivity in infancy and sociability in childhood is different in the two groups. Because the idea of ethnic differences in temperament is controversial, it is useful to examine its validity in varied experimental contexts.

Method

Subjects

There were two Caucasian samples of infants, one from the Boston metropolitan area and one from Dublin, Republic of Ireland. The Chi-

Jerome Kagan, Doreen Arcus, and Nancy Snidman, Department of Psychology, Harvard University; Wang Yu Feng, Beijing Medical College, Beijing, People's Republic of China; John Hendl, Department of Psychology, University of Massachusetts, Amherst; Sheila Greene, Department of Psychology, Trinity College, Dublin, Republic of Ireland.

This research was supported in part by the John D. and Catherine T. MacArthur Foundation. We thank Anna Costes, Hilary Sokolowski, Gail Agronick, Jennifer Davis, Sherry Gardner, Carol Anderson, Ruth Bell, Lynn Goldsmith, and Kevin Nugent.

Correspondence concerning this article should be addressed to Jerome Kagan, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138.

nese sample lived in Beijing, People's Republic of China. There were 247 infants from the Boston area (48% male and 52% female). The Irish sample consisted of 106 infants (56% male and 44% female; Kagan, Snidman, Hendl, Greene, & Nugent, 1991). The Chinese sample consisted of 80 infants (59% male and 41% female).

The Dublin and Beijing samples were less heterogeneous in ethnic composition than the Boston sample because both parents and grandparents were born in each of those respective countries. By contrast, the parents of the Boston infants were ethnically more diverse and consisted of adults of both northern and southern European ancestries.

All three samples consisted of 4-month-old infants born at term with no pre- or perinatal complications to intact families who volunteered to participate. All the mothers in the three samples were married and between the ages of 20 and 40 years. All mothers in the American sample had college degrees or the equivalent. The Irish infants also came from predominantly middle-class families. Over half of the Irish parents—57%—either graduated from or had attended college, and 92% had completed their high school certificate. Seventy-six percent of the Chinese parents had attended or graduated college, and 99% had graduated high school. None of the American or Chinese infants had been in full-time day care before they were 4 months old; only 15% of the Irish infants had been in alternate care full time after 14 weeks (the length of paid maternity leave). It should be noted that the American sample, which was part of a larger study, was selected so that no infant was in full-time day care. Thus, there is comparability across the three samples with respect to the educational attainment of the mothers and day-care experience during the first 4 months of life.

Procedure

The experimenter for the Dublin sample and the supervisor for the Beijing sample were both trained in our laboratory in Cambridge, Massachusetts, on the battery that was administered. Thus, the staff in both sites were very familiar with the procedures. As noted, the procedures administered to all infants were based on previous research that discovered a relation between motor activity and crying to stimulation at 4 months and inhibited or uninhibited behavior in the second year (Kagan & Snidman, 1991a). Hence, it seemed appropriate to use the same battery in the Dublin and Beijing settings. A description of the battery follows.

Baseline 1. When the infant was well fed and relaxed, he or she was placed in an infant seat, and the mother looked down on the infant with a smiling face but did not touch or vocalize to her infant for a 60-s period.

Speech. The infant heard a tape recording of eight different, short sentences, each 5 s in duration. The first two sentences were spoken by a female adult at an average conversational loudness; the next two were spoken simultaneously by one female and one male adult, the next two by two female and one male adult; and the final two spoken by two female and two male adults. Each time the composition of the voices changed, the intensity increased by approximately 5 dB. The same tape recording was used for the Boston and Dublin samples. The sentences were recorded in Chinese for the Beijing sample and were read by a Chinese female adult.¹ The speaker baffle, located 1 ft from the infant, was covered with a schematic human voice to attract the infant's attention to the source of the voice. The phrases were statements like, "Hello pretty baby, how are you today?"

Mobiles. Each infant was presented with a series of three different mobiles constructed from colorful plastic animals and dolls. A mobile had one, three, or seven elements. On the first trial, the mobile contained one element; on the second trial, three elements; and on the third trial, seven elements. Each mobile was moved back and forth slowly (approximately 1 cycle per second) in front of the infant's face for trials that were 20 s in duration. The three different mobiles were then presented for two additional series for a total of nine 20-s trials.

Olfactory stimulation. The infant was presented with eight trials in

which a cotton swab was applied under the child's nostrils without touching the skin. On Trial 1, the swab was dipped in water. On Trials 2, 3, and 4 the swab was dipped in a dilute solution of butyl alcohol of 8 parts per million (ppm); on Trials 5, 6, and 7 the swab was dipped in a stronger concentration of butyl alcohol at 16 ppm, and on Trial 8 the swab was dipped in water. Each trial was 5 s in duration.

Auditory stimulation: Syllables. Each infant heard a tape recording of a female voice speaking three different consonant vowel syllables (*ma, pa, ga*), each repeated for 10-s durations at three different loudness levels. As with the first auditory episode, the voice was in Chinese for the Beijing sample. Each stimulus increased by approximately 5 dB over the initial conversational level for a total of nine trials. The stimuli were presented through the same speaker used for Episode 2.

Startle. When the infant was relaxed, the examiner burst a balloon directly behind the infant's head; behavior was recorded for 15 s.

Baseline 2. Same as Baseline 1.

Conditions. The instructions to the examiners at each site were identical, especially with regard to the interruption of an episode. An episode was stopped when the infant cried for more than 2 s, and testing resumed when the infant was neither fretting nor crying and was quiet. The examiner also intervened occasionally to reposition an infant in the infant seat. Finally, the examiner told the mother she could interrupt whenever she felt the infant might be hungry or fatigued—this event was infrequent in all sites.

Coding. Videotapes of the infants from all three sites were coded in our laboratory by different observers, none of whom had been an examiner. Six coders scored 94% of the tapes for the Boston sample. The remaining 16 videotapes were coded by four coders, each of whom left the project shortly after the completion of their training; the cases they coded were also coded by the project supervisor. Because the Irish and Chinese samples were smaller, these infants were scored by a smaller number of individuals. There were two coders for the Irish sample and one coder for the Beijing sample. All coders were trained and monitored following the same scoring protocol. The primary coder for the Boston sample trained each coder until all paired coders produced results that had less than 5% disagreement. This phase of training always began with infants from the Boston sample, and it was repeated with Irish or Chinese infants for coders scoring those ethnic groups. After the training was completed, periodic reliability checks were conducted, and tapes were reviewed for discrepancies. Toward the end of the collection of data, after the primary coder had left the laboratory, the project supervisor continued to use her tapes as the training and reliability standard for the other coders. Either the main coder or the project supervisor was always available for consultation when judgments were difficult.

The variables coded from the videotapes were as follows:

1. *Motor activity:* A total motor score was a weighted sum of the frequency of movements of both arms, both legs, bursts of movement of either arms or legs (three or more movements in rapid succession), extensions of both arms or legs, and arches of the back. Each occurrence of a movement of both arms, both legs, or extensions was awarded 1 point; limb bursts and arches of the back were awarded 2 points. The rationale for this decision was based on the fact that bursts of activity and arches of the back reflected more intense motor activity.

2. *Vocalization:* Proportion of trials on which a vocalization that was neutral or happy in affective tone and not a fuss, fret, or protest was coded.

3. *Smiling:* Proportion of trials on which a smile occurred, where a smile was defined as a bilateral retraction and elevation of the lips.

4. *Fretting:* Proportion of trials on which the child fretted, where fretting was judged as a vocalization characterized by an increase in

¹ The Chinese tape used only one female voice for all sentences. However, because this episode contributed little to the motor and cry scores in all samples, we suggest that this subtle deviation in procedure had a minimal effect on the results.

pitch in relation to a vocalization, a constrained quality, and audible negative affect.

5. *Crying*: Number of seconds the child cried. If an infant cried for more than 2 s on a particular trial, the examiner or the mother or both terminated the trial and attempted to soothe the infant.

For the infants who became irritable such that the session had to be terminated before the end of the battery, a prorated score was computed across the entire battery. For vocalization, smiling, and fretting, the variable was the proportion of trials administered during which the response occurred. The prorated motor and cry score was based on the formula: raw score + (trials missed × raw score)/trials completed. However, any infant who did not complete at least one half of the battery was eliminated from the analysis. All of the Chinese infants and 96% of the Irish infants completed the battery; 68% of the Boston sample completed the battery. The larger number of terminations in the Boston sample reflects the greater irritability in these infants.

We assessed the reliability of the scoring of behaviors from the videotapes by comparing each coder's judgments with those of the primary coder. This comparison was done both within and across the three ethnic groups. The within-group reliabilities (product-moment correlations) are based on 6% of the American sample, 8% of the Irish sample, and 9% of the Chinese sample. For motor activity, the average reliability was .75 for the American sample, .96 for the Irish sample, and .85 for the Chinese sample. For crying, reliability was .74 for the American sample, .70 for the Irish sample, and .99 for the Chinese sample. For fretting, the reliability was .93 for the American sample, .90 for the Irish sample, and .95 for the Chinese sample. For vocalizing, the reliability was .92 for the American sample, .63 for the Irish sample, and .79 for the Chinese sample. Finally, for smiling, the reliability was .79 for the American sample, .27 for the Irish sample, and .82 for the Chinese sample. Note the low reliability for smiling for the Irish sample.² The coders of the Irish and Chinese samples also checked their reliability with tapes from the Boston sample to ensure that agreement was maintained across ethnic groups. The reliabilities of the coding for those observers who scored the Irish and Chinese infants with the primary coder using the videotapes of American infants were .79 and .95 for motor activity, .97 and .99 for crying, .87 and .92 for fretting, .81 and .89 for vocalizing, and .57 and .92 for smiling, respectively. These generally high reliabilities for coding across different samples indicate that there is minimal bias in the coding of each of the three ethnic groups.

Results

Table 1 contains the average values for the total motor score, percentage of trials on which vocalization, fretting, or smiling occurred, and number of seconds crying. The three ethnic groups differed in the distributions of most of the variables. With one exception, scores for the American infants had significantly greater variance than those for the Irish sample, which, in turn, had greater variance than those for the Chinese sample. The *F* tests for the equality of variance yielded results significant at $p < .001$ for each group on each index of arousal, with the exception of American and Irish infants for motor activity ($p < .05$) and number of vocalization trials (ns). In addition to unequal variance, the distributions for crying and fretting were notably skewed. Therefore, the square-root transformation of each score was used in an analysis of variance (ANOVA) by ethnic group and sex. The ANOVAs were computed with partial sums of squares to account for the unequal number of subjects per cell. Duncan's multiple range tests were used to test contrasts between ethnic groups.

American infants showed the highest levels of behavioral arousal, followed by the Irish infants and then by the Chinese

Table 1
Means and Standard Deviations for Motor Activity, Crying, Fretting, Vocalizing, and Smiling for American, Irish, and Chinese 4-Month-Old Infants

Variable	American (n = 247)	Irish (n = 106)	Chinese (n = 80)
Motor activity			
<i>M</i>	48.6	36.7	11.2
<i>SD</i>	31.9	26.2	1.3
Range	0-194	0-116	0-57
Cry (seconds)			
<i>M</i>	7.0	2.9	1.1
<i>SD</i>	14.7	11.0	0.5
Range	0-89	0-98	0-32
Fret (% trials)			
<i>M</i>	10.0	6.0	1.9
<i>SD</i>	12.3	8.9	5.2
Range	0-70	0-46	0-22
Vocalize (% trials)			
<i>M</i>	31.4	31.1	8.1
<i>SD</i>	15.9	16.8	1.0
Range	0-78	0-67	0-54
Smile (% trials)			
<i>M</i>	4.1	2.6	3.6
<i>SD</i>	6.0	4.2	0.6
Range	0-38	0-30	0-24

infants. American infants displayed more motor activity and more fretting than Irish infants who, in turn, were more active and more fretful than Chinese infants: for motor activity, $F(2, 427) = 80.19, p < .0001$, all contrasts significant at $p < .01$; for fretting, $F(2, 427) = 26.79, p < .0001$, all contrasts significant at $p < .01$. American infants cried more often than either Chinese or Irish infants, $F(2, 427) = 13.71, p < .0001$; contrasts between American infants and other groups were significant at $p < .01$. Both American and Irish infants vocalized more often than Chinese infants, $F(2, 427) = 126.96, p < .0001$; contrasts between Chinese infants and other groups were significant at $p < .01$. Nineteen percent of the Chinese infants never cried or vocalized once compared with less than 1% of the American infants, $\chi^2(1, N = 327) = 28.3, p < .0001$. Smiling was the only behavior for which the Chinese infants did not differ from the other two groups, although American infants smiled slightly more often than Irish infants, $F(2, 427) = 3.71, p < .05$; contrast between American and Irish infants were significant at $p < .05$.

There were no sex differences for motor activity, fretting, or crying, but male infants vocalized and smiled more than female infants: for vocalization, $F(1, 427) = 9.85, p < .01$; for smiling, $F(1, 427) = 61.9, p < .01$. Although the mean vocalization and smiling scores were higher for all male infants compared with all female infants (pool of all three ethnic groups), the sex difference was only significant for the Chinese sample when each group was analyzed separately: for vocalization, $t(78) = 2.20, p < .05$; for smiling, $t(78) = 2.06, p < .05$. There were no Sex \times Sample interactions.

² Although the reliability for smiling in the Irish infants was low, we believe that the coding is more accurate than this index implies (percentage agreement on a trial-by-trial analysis was .97). It is possible that a range restriction in smiling among Irish infants contributed to the low reliability.

Intercorrelations among the response variables were examined separately for each ethnic group. With the exception of a positive relation between fretting and crying, which is to be expected, the intercorrelations within each group were very similar and, although significant, were generally low ($r < .3$).

Discussion

The data indicate a significant difference in both level and variability of arousal to stimulation between Caucasian and Chinese infants, with the Chinese infants showing both lower levels of arousal as well as less variability. Although variation in experience during the first 4 months probably makes a contribution to this result, it is possible that temperamental factors are also of some influence. This claim rests on several arguments.

First, Freedman and Freedman (1969) reported very similar differences with infants who were less than 1-week-old; our results suggest that these differences are maintained beyond the neonatal period. A second reason for not attributing this result exclusively to prior experience rests on the presumption that infant smiling is unusually sensitive to differential reinforcement through social interactions with adults (Watson, 1972). However, the Chinese and Caucasian infants were minimally different on frequency of smiling and maximally different on motor activity, vocalization, and irritability. The latter three behaviors are more clearly indexes of degree of arousal to stimulation compared with smiling. In addition, the ethnic differences in vocalization and activity were also present in an independent sample of Chinese-American and European-American infants who attended the same experimental day-care center, 5 days a week, from early infancy to 2 years of age (Kagan et al., 1978).

Other investigators have reported biological differences between Asians and Caucasians. For example, Asian-American psychiatric patients require a lower dose of psychotropic drugs than do European-American patients (Lin, Poland, & Lesser, 1986), suggesting that Asian populations may be at a lower level of limbic arousal. Furthermore, there are known genetic differences between Asians and Europeans. For example, the proportion of Rh negative individuals is less than 1% in China but greater than 15% in Europe (Cavalli-Sforza, 1991). In addition, there is greater genetic diversity in many loci related to blood groups and proteins among Caucasians than Asians, a fact in accord with the greater variation in behavior discovered in this investigation (Latter, 1980). Even though all scientists agree that the genetic differences within an ethnic-geographic group is greater than the variation between groups, nonetheless, it is reasonable to at least entertain the hypothesis of genetically influenced behavioral differences in infants belonging to populations that have been reproductively isolated for a long time. Europeans and Asians have been reproductively isolated for about 30,000 years, or about 1,500 generations; it requires only 15 to 20 generations of selective breeding to produce obviously different behavioral profiles in many animal species (Mills & Faure, 1991; Plotkin, 1988).

Scientists should consider the ethnic composition of their samples when the psychological variables they quantify bear some relation to reactivity and ease of arousal. Investigators

who discover cross-cultural differences in behavioral or emotional qualities that are related to reactivity should consider the possibility of a modest biological contribution to their comparative observations.

References

- Caudill, W., & Weinstein, H. (1969). Maternal care and infant behavior in Japan and America. *Psychiatry*, 32, 12-43.
- Cavalli-Sforza, L. L. (1991). Genes, people and languages. *Scientific American*, 265, 104-111.
- Freedman, D. G. (1974). *Human infancy: An evolutionary perspective*. Hillsdale, NJ: Erlbaum.
- Freedman, D. G., & Freedman, M. (1969). Behavioral differences between Chinese-American and American newborns. *Nature*, 224, 1227.
- Green, N. (1969). *An exploratory study of aggression and spacing behavior in two preschool nurseries: Chinese-American and European-American*. Unpublished master's thesis, University of Chicago.
- Kagan, J., Kearsley, R. B., & Zelazo, P. R. (1978). *Infancy: Its place in human development*. Cambridge, MA: Harvard University Press.
- Kagan, J., & Snidman, N. (1991a). Infant predictors of inhibited and uninhibited behavioral profiles. *Psychological Science*, 2, 40-44.
- Kagan, J., & Snidman, N. (1991b). Temperamental factors in human development. *American Psychologist*, 46, 856-862.
- Kagan, J., Snidman, N., Handler, H., Greene, S. M., & Nugent, J. K. (1991). Predicting inhibited and uninhibited behavior, *Irish Journal of Psychology*, 12, 248-262.
- Latter, B. D. H. (1980). Genetic differences within and between populations of the major human subgroups, *American Naturalist*, 116, 220-237.
- Lewis, M. (1989). Culture and biology: The role of temperament. In P. R. Zelazo & R. G. Barr (Eds.), *Challenges to developmental paradigms: Implications for theory assessment and treatment* (pp. 203-223). Hillsdale, NJ: Erlbaum.
- Lewis, M., Ramsay, D. S., & Kawakami, K. (in press). Affectivity and cortisol response differences between Japanese and American infants. *Child Development*.
- Lin, K. M., Poland, R. E., & Lesser, I. M. (1986). Ethnicity and psychopharmacology. *Culture, Medicine and Psychiatry*, 10, 151-165.
- Mills, A. D., & Faure, J. M. (1991). Diverse selection for duration of chronic immobility and social reinstatement behavior in Japanese quail. *Journal of Comparative Psychology*, 105, 25-38.
- Plotkin, H. C. (1988). *The role of behavior in evolution*. Cambridge, MA: MIT Press.
- Rothbart, M. K. (1989). Temperament in childhood: A framework. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 59-76). New York: Wiley.
- Watson, J. S. (1972). Smiling, cooing, and "the game." *Merrill-Palmer Quarterly*, 18, 323-329.
- Weisz, J. R., Suwanlert, S., Chaiyosit, W., Weiss, B., Achenbach, T. M., & Walter, B. A. (1987). Epidemiology of behavioral and emotional problems among Thai and American children: Parent reports for ages 6-11. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 890-898.
- Weisz, J. R., Suwanlert, S., Chaiyosit, W., Weiss, B., Walter, B. R., & Anderson, W. W. (1988). Thai and American perspectives on over- and undercontrolled child behavior problems. *Journal of Consulting and Clinical Psychology*, 56, 601-609.

Received August 5, 1992
 Revision received October 13, 1993
 Accepted October 13, 1993 ■