Cross-cultural consistency in coding the Strange Situation

IJzendoorn, M.H. van; Kroonenberg, P.M.

INFANT BEHAVIOR AND DEVELOPMENT **13,** 469 485 (1990)

Cross-Cultural Consistency of Coding the Strange Situation

MARINUS H. VAN IJZENDOORN AND
PIETER M. KROONENBERG
Leiden University

This study tested whether or not cross-cultural differences in attachment classification distributions result from systematic differences in coding practices First, we investigated whether or not the interactive scales have been scored consistently in several different cross-cultural samples Second, the Richters, Waters, and Vaughn (1988) functions were applied to address the question of whether or not attachment classifications were consistently based upon the same pattern of interactive behaviors Third, cross-cultural coding differences were described from a multivariate perspective Data sets from seven investigators in six countries were available for analysis Analyses on this "multinational data set" revealed that except for distance interaction, the interactive scales in the two reunion episodes were scored in accordance with the original coding rules Furthermore, a good to reasonable agreement appeared to exist between the original classifications and those computed by the functions, except for infants older than 20 months of age The multivariate principal component analysis showed that classification groups across cultures were more alike than cultures across classification groups Our data showed, therefore, that attachment classifications have been consistently coded across cultures

cross-cultural studies attachment Strange Situation coding system cross-cultural validity classification functions crying replicated principal component analysis

Because the Strange Situation procedure (Ainsworth, Blehar, Waters, & Wall, 1978) has produced attachment classification distributions that are highly divergent across and within countries (Van IJzendoorn & Kroonenberg, 1988), serious questions have been raised about the cross-cultural validity of the procedure as an instrument to measure attachment and about its underlying theoretical framework (Lamb, Thompson, Gardner, & Charnov, 1985; LeVine & Miller, 1990). The Strange Situation is

We would like to thank Kuno Beller, Klaus Grossmann, Michael Lamb, Kazuo Miyake, and Ross Thompson, who made the data available to Avi Sagi. Special thanks are due to Avi Sagi who, in addition to supplying two data sets himself, took care of the collation of the complete data and gave permission for our use of these data. We would also like to thank John Richters for computing the revised discriminant functions. The study was partially supported by a PIONEER grant from the Netherlands Organization of Scientific Research (NWO) and a Fulbright grant to Marinus Van IJzendoorn for his stay at the National Institute of Child Health and Human Development, Bethesda, MD.

Correspondence and requests for reprints should be sent to Marinus H Van IJzendoorn, Center for Child and Family Studies, Leiden University, P O. Box 9555, 2300 RB Leiden, The Netherlands

designed to observe parent–infant attachment during a series of increasingly stressful episodes. Infants' behavior during reunion with the parent after a short separation is classified into three main categories of attachment. The securely attached group (B) shows minimal resistant and avoidant behavior. These children are somewhat upset when their caregiver has left, but his or her return has a calming effect. Avoidant children (A) do not seek proximity or contact to their returning caregiver, instead they show avoidant behavior. Resistant or ambivalent children (C) seek contact but resist the caregiver at the same time. Some resistant children are unable to settle within the 3-min reunion episodes (Ainsworth et al., 1978).

Implicit in the debate about the cross-cultural validity of the Strange Situation procedure is the existence of a satisfactory reliability of the measurement of attachment. In particular, it is taken for granted that cross-cultural differences in attachment classification distributions do not result from systematic differences in coding practices. Although satisfactory intercoder agreements have been reported in all studies relevant to the cross-cultural debate, (informal) "interlaboratory" agreement has only been established for those studies in which the same researcher (Lamb et al., 1985) scored the Strange Situation procedure or trained coders to apply the classification guidelines of Ainsworth et al. (1978). Although Ainsworth et al. described their coding rules in great detail, a set of precoded Strange Situation videotapes to standardize coding practices across laboratories is not yet available (see Richters, Waters, & Vaughn, 1988).

In general, coders expect to find a distribution comparable to the Ainsworth distribution (see Ainsworth et al., 1978, p. 98). However, in the case of deviating distributions, the possibility of unintentionally changing the interpretation of the coding rules cannot be excluded, even in the case of the same coder participating in different studies. For example, in a sample with an overrepresentation of resistantly attached infants, a child showing only a modest degree of resistance behavior may have a greater chance of being classified as securely attached than in a sample with an overrepresentation of avoidantly attached infants. A drift in classification criteria across laboratories that find differently skewed distributions may, therefore, be present and may determine discrepancies with the true distributions to an unknown degree. Richters et al. (1988) developed a series of classification functions to objectively classify infants' attachments only on the basis of the interactive scales and crying behavior in the two Strange Situation reunion episodes. The functions were developed and validated on data from 255 Strange Situations conducted and coded by Ainsworth and her colleagues. The cross-validation showed high agreement between computed and original classifications. Therefore, these functions may effectively prevent the drift in classification criteria and constitute an important way to calibrate applications of Ainsworth's classification system to her original codings and to enhance interlaboratory agreement.

In this study, we evaluated the cross-cultural consistency of coding Strange Situation data in three ways. First, we investigated whether the interactive scales, such as proximity seeking, contact maintaining, avoidance, resistance, and distance interaction, have been scored consistently. Coding these scales is the first step in classifying Strange Situations (Ainsworth et al., 1978). More specifically, we tested per classification category whether mean scores on the interactive scales in several cross-cultural samples systematically deviate from the mean scores on the same scales reported by Ainsworth et al. (1978). Second, we addressed the question of whether or not attachment classifications in several different cross-cultural samples are consistently based upon the same patterns of interactive behaviors. More specifically, we applied the Richters et al. (1988) classification functions to evaluate whether or not cross-cultural differences exist in reproducing attachment classifications on the basis of interactive scales. Finally, we performed a replicated principal component analysis on the two sample-by-interactive-scales matrices with means from the two reunion episodes to find a graphical representation for the samples and interactive scales jointly. This provided further insight into cross-cultural coding differences from a multivariate perspective.

METHOD

Subjects

Data sets from seven investigators in six countries were available for analysis (see Sagi, 1990), and the combined samples of Ainsworth et al. (1978) served as a standard for the first part of the study. In Table 1, attachment classification distributions of each sample as used in this study are presented; effective total sample size was 442. These distributions are not identical to previously published distributions for the data set due to the regrettable elimination of several subjects who had no valid data in one or more reunion episode or directly preceding stranger episodes. We have also included the "global" distribution of attachment classifications based on samples from eight different countries compiled by Van IJzendoorn and Kroonenberg (1988).

Because age effects can be especially large for some interactive scales (e.g. this distance interaction), the Dutch sample was divided into two subsamples—younger (under 20 months) and older (20 months and over) infants—in order to differentiate age effects from cultural differences.

Means

Our analyses concentrate on proximity seeking (PROX), contact maintaining (CM), avoidance (AVOI), resistance (RES), and distance interaction (DIS) in the two reunion episodes, referred to as Reunion 1 and Reunion

TABLE 1
Description of Data Sets

Researchers	Country	Age (in Months)	Frequencies				%		
			A	B	C	*N*	A	B	C
Beller & Pohl (1986)	FRG (Berlin)[a d]	12	7	30	2	39	18	77	5
Grossmann, Grossman, Huber, & Wartner (1981)	FRG (Bielefeld)[a d]	12	23	18	2	43	54	42	5
Lamb, Hwang, Frodi, & Frodi (1982)	SWEden[a]	11–13	10	36	2	48	21	75	4
Miyake, Chen, & Campos (1985)	JAPan (Sapporo)[a]	12	0	38	16	54	0	70	30
Sagi et al. (1985)	ISRael (Kibbutz)[a]	11–14	7	36	9	52	14	69	17
Sagi et al. (1985)	ISRael (Day Care)[a]	12	1	24	3	28	4	86	11
Thompson & Lamb (1983)	USA[a]	12	7	30	5	42	17	71	12
Van IJzendoorn, Goossens, Kroonenberg, & Tavecchio (1985)	NL (Young)[b d]	12–19	14	27	0	41	34	66	0
Van IJzendoorn et al. (1985)	NL (Old)[b d]	19–25	19	71	5	95	20	75	5
Ainsworth et al. (1978)	USA	12	22	70	13	105	21	67	12
Global distribution[c]		—	423	1294	273	1990	21	65	14

[a] Original distributions were derived from Sagi, Van IJzendoorn, and Koren-Karie (1989). However, only those Strange Situation sessions are included for which complete data were available in the two reunion and the preceding stranger episodes. For three children, a single missing value was substituted

[b] The Dutch sample, that was coded by Frits A. Goossens, has been divided into a young sample (infants younger than 20 months) and an old sample (20 months and older).

[c] Derived from Van IJzendoorn and Kroonenberg (1988).

[d] FRG = Federal Republic of Germany, NL = The Netherlands

2. These episodes and scales are the most important ones for classifying infants in the Strange Situation.

The means of the interactive scales for the cross-cultural samples were evaluated against the means reported in Ainsworth et al. (1978, p. 107, Table 13, p. 365, Table 32). Although the first studies by Ainsworth and her colleagues (combined sample size 105) were exploratory, we decided to use the Ainsworth et al. data as a standard because the system for coding the Strange Situation was based on this data set. The required standard deviations necessary for the evaluation were calculated from the sizes and standard deviations of the classification subgroups presented in Table 33 (Ainsworth et al., 1978, p. 367). Treating the Ainsworth means and the calculated standard deviations as population values, per scale confidence intervals were calculated for each cross-cultural sample-classification group as if it were a sample from the Ainsworth population. In this way, a single-sample z test or standard normal deviate test was carried out to investigate whether the cross-cultural samples could be considered samples from a standard (i.e., Ainsworth et al.'s) population. The critical value of the z test was adjusted for multiple testing by applying the Bonferroni correction for each scale and classification group separately, that is, by dividing the standard α=.05 by the number of samples (9), giving α=.006, leading to a critical value of z=2.78.

Classifications

Because precise coding of crying is not considered necessary for classifying the Strange Situation "clinically," data on crying in the reunion episodes were not available in our multinational data set. Therefore, the Richters et al. (1988) classification functions could not be applied in their original form. The original functions are based on weighted scores for proximity seeking, contact maintaining, avoidance, resistance, and crying in the reunion episodes. Richters et al. first used functions to classify children either as B or non-B and subsequently classified the originally A or C children as A or C with new classification functions. High agreements between computed classifications and original classifications were found by Richters et al. (p. 517; approximately 91% agreement on B vs. non-B decisions, and 94% on A vs. C decisions; Cohen's, 1960, kappa values were .82 and .88, respectively). Because Richters et al. carried out the A versus C discriminant analysis only with those cases that originally were coded as anxiously attached and deleted those cases that were wrongly classified as non-B in the first round, their agreement figures are somewhat inflated. Percentage agreement and kappa for the complete A, B, C classifications (below referred to as A-B-C agreement and A-B-C kappa) were 88% and .79, respectively (Richters, personal communication, June 1989).

At our request, Richters recomputed the functions without including crying. The A-B-C percentage agreement for these functions is 85% and the A-B-C kappa is .73. These values were remarkably similar to those for the functions including crying. The new functions tended to allocate more A children to B and more B children to C (see Appendix A for more details).

Replicated Principal Component Analysis

In replicated principal component analysis, the data from two or more occasions are supposed to have the same configuration of variables and subjects for each occasion, but the relative size or importance of the configuration is allowed to vary from one occasion to the next. With this technique it is possible to combine data from two or more occasions without first averaging over occasions. The model is a special case of the more general three-mode principal component model (Kroonenberg, 1983; Kroonenberg & De Leeuw, 1980; Tucker, 1966).

More specifically, the model used has one single set of principal components for the samples and one single set of component loadings for the scales. In addition, a weight for each of the reunion episodes is estimated which indicates the relative contribution of the data of that episode to the total solution. By plotting the component scores and loadings in a single graph in the same manner as in Gabriel's (1971) biplot, it is possible to study the relationships between the samples, the classification groups, and the scales jointly.

In our study per reunion episode, five scale means were calculated for each of the three classification groups (A, B, and C) crossed with each of 10 samples. Thus, in principle for each episode, a (3*10)-by-5 matrix was available for analysis. However, as no A infants were present in the Japanese sample and no C infants in the younger Dutch sample, the actual size of each matrix was 28-by-5. These two matrices with means of the reunion episodes were simultaneously analyzed using a program for three-mode principal component analysis developed by Kroonenberg and De Leeuw (1980; Kroonenberg, 1983).

RESULTS

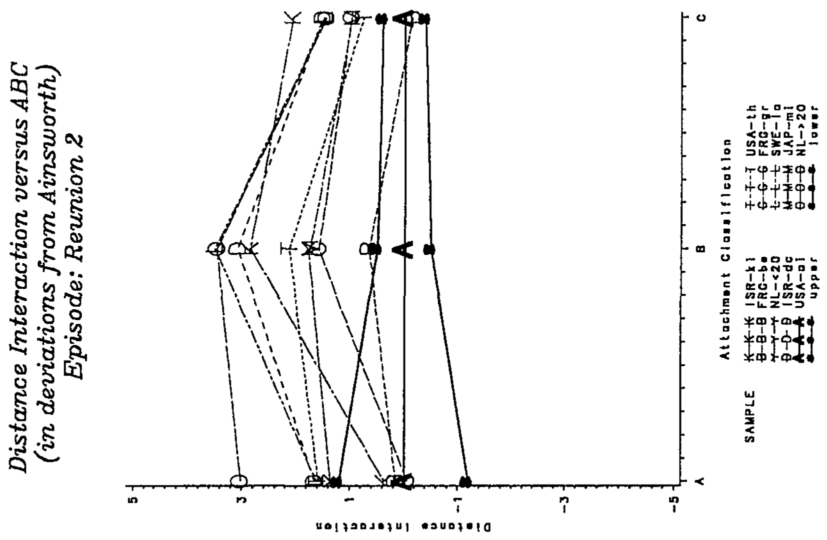
Cross-Cultural Consistency in Coding the Interactive Scales

First, we tested whether mean scores on the interactive scales in the cross-cultural samples were within the confidence interval of the (population) means of the Ainsworth scales. Tests were carried out for each classification separately to avoid differences in classification distributions influencing the number of interactive scale means falling outside the confidence bounds. In Table 2, the number of means falling outside confidence bounds has been given, for each sample-classification combination. The number of deviations from the scale means was summed over the first and second reunion episode.

From Table 2 it can be derived that in the avoidantly attached category (A) 22 means, in the securely attached category (B) 33 means, and in the resistantly attached category (C) 17 means fell outside the confidence intervals. Ideally, one would like to test whether there are any regularities in the deviations from the Ainsworth sample, for instance, whether certain samples show more deviations in certain scales than other samples do. Loglinear analysis on the 9 (Samples—*S*) by 5 (Interactive Scales—*I*) by 3 (Classifications—*C*) three-way contingency tables is not appropriate because of small expected values in virtually all cells (Bishop, Fienberg, & Holland, 1975). Therefore, it was decided to sum over each way of the three-way contingency table in turn and test each of the resulting two-way tables (i.e., *S* by *I*, *S* by *C*, and *I* by *C*) separately for independence. Again, the usual assumptions are not completely fulfilled, but more so than in the complete three-way table. Thus, the $S \times C$, the $S \times I$, and the $C \times I$ tables were tested separately with the Pearson χ^2 statistic, with the following results: $S \times C$: $\chi^2(14, N=71)=9.20$, $p=.82$; $S \times I$: $\chi^2(32, N=71)=17.10$, $p=.99$; $C \times I$: $\chi^2(8, N=71)=23.60$, $p=.003$. The significant interaction of Classifications by Interactive Scales is primarily caused by the fact that, in contrast with other attachment categories, avoidantly attached groups have means outside the confidence intervals on the resistance scale. Because there are no significant two-way interactions involving the samples, differences between samples can be tested on the one-way margin by comparing the observed deviations from the Ainsworth sample with the uniform distribution. Such a test will establish whether some cross-cultural samples deviate significantly more from the Ainsworth sample than other samples. From the Pearson's chi-square goodness-of-fit test, $\chi^2(8, N=71)=9.00$, $p=.34$, it follows that the distribution of deviations across samples does not significantly deviate from the expected values of a uniform distribution. That is, we did not find cross-cultural differences in the number of deviations from Ainsworth's scale means.

The most conspicuous aspect of Table 2 is that distance interaction accounts for half of the deviations (i.e., 36 out of 72) from the Ainsworth sample. Testing the deviations of scale means from the Ainsworth's scale means against the uniform distribution gives a highly significant chi square: χ^2 $(4, N=71)=44.90$, $p<.001$. When distance interaction is deleted, no significant effects remain. To show the nature of the deviations, a graph has been made depicting the deviations from the Ainsworth means. The latter are shown in Figure 1 as a horizontal line, and the upper and lower confidence bounds based on the average sample size of the cross-cultural samples are shown as well.

For both reunion episodes, almost all mean scores for distance interaction, including all significant ones, *exceeded* the Ainsworth means (Figure 1), irrespective of culture or classification group.



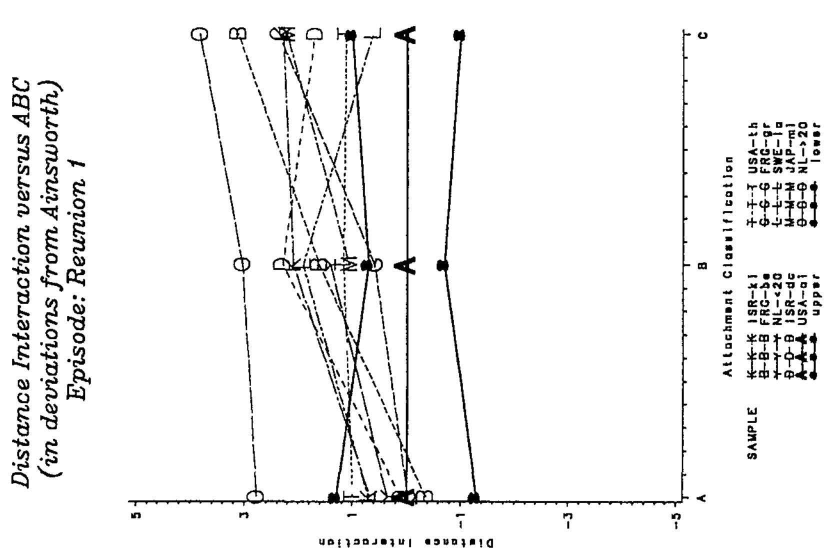


Figure 1. Deviations of sample means from the Ainsworth means with Bonferroni-adjusted 95% confidence bands based on average sample sizes: distance interaction.

FRG= Federal Republic of Germany (G=Grossmann; B=Beller)
ISR= Israel (K=Kibbutz; D=Day Care)
USA= United States of America (A=Ainsworth; T=Thompson)
NL= The Netherlands (Y=younger children; O=older children)
JAP= Japan (M=Miyake)
SWE= Sweden (L=Lamb)

TABLE 2

Number of Interactive Means Falling Outside Confidence Boundaries of the Ainsworth Sample

	Interactive Scales[b]																		
	P			C			R			A			D			Total			
Sample[a]	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	Total/Sample
NL—Old	0	1	1	0	2	1	1	0	0	1	1	0	2	2	2	4	6	4	14
SWE—Lamb	1	0	0	1	1	0	1	0	0	1	1	0	1	2	1	5	4	1	10
NL—Young	0	2	—	0	1	—	1	0	—	1	1	—	1	2	—	3	6	—	9
ISR—Kibbutz	0	0	0	1	1	0	1	0	0	1	0	0	0	2	2	3	3	2	8
USA—Thompson	0	1	0	0	1	0	0	0	0	1	0	0	1	2	2	2	4	2	8
FRG—Grossmann	1	0	0	2	1	0	1	0	0	0	0	0	0	1	2	4	2	2	8
ISR—Day Care	0	0	0	0	1	0	0	0	0	0	1	0	0	2	2	0	4	2	6
JAP—Miyake	—	0	1	—	0	0	—	0	0	—	0	0	—	2	2	—	2	3	5
FRG—Beller	0	0	0	1	0	0	0	0	0	0	0	0	0	2	1	1	2	1	4
Totals across samples	2	4	2	5	8	1	5	0	0	5	4	0	5	17	14	22	33	17	72
Total per scale		8			14			5			9			36			72		

[a] FRG = Federal Republic of Germany; ISR = Israel; JAP = Japan; NL = The Netherlands; SWE = Sweden.

[b] P = proximity seeking; C = contact maintaining; R = resistance; A = avoidance; D = distance interaction.

Cross-Cultural Consistency in Coding the Strange Situation Classification

Scoring the interactive scales consistently across cultures is only the first step for consistently coding classifications in the Strange Situation. Until the article by Richters et al. (1988), such coding was only performed using the extensive coding rules for deriving the classifications from the interactive scales by Ainsworth et al. (1978). Using Richter's revised functions (without crying parameters), we tried to reproduce the original classifications for all 442 infants from the multinational data set (see Table 3). Differences and correspondences between the original and computed classifications are given, as well as A-B-C percentage agreement and A-B-C kappas to measure chance-corrected agreement. Confidence intervals for the kappas are presented as well, based upon the formula (13) of Fleiss, Cohen, and Everitt (1969). The confidence intervals are very wide for these (small) samples, and their stability is not entirely clear (see Fleiss & Cicchetti, 1978).

In most samples, the functions reproduced the original classifications rather well, compared to the overall recovery rate of 85% and a kappa of .73 in Richters et al.'s (1988) cross-validation sample. A notable exception was the Dutch sample with older subjects, which did not show any overlap with the Richters et al. confidence interval. Of the 17 B children in the Dutch sample with older infants classified as A by the functions, 12 were originally coded as B1. The confidence intervals for kappa of three other samples showed overlap with the Richters et al. confidence interval, but their actual kappas were not contained in the latter. In the Japanese sample, the functions produced more A and fewer C classifications than the original coders did. In the Swedish and Dutch (young) samples, the functions produced more A and more C classifications than the original coders did. Because of the very skewed distribution of the Dutch sample with younger subjects, the rather low kappa is equal to the maximum value of the kappa for this distribution. The mean percentage of (A-B-C) agreement for the whole multinational data set is 83%, and the mean kappa is .66. The overall A-B-C percentage agreement is 82%, and the corresponding kappa is .65.

Cross-Cultural Consistency in the Structural Relations Within the Strange Situation

In this section, we will simultaneously analyze the coding differences in means on the interactive scales and in classification groups through a replicated principal component analysis. Because of the aberrant behavior of distance interaction, this scale has not been included. This exclusion is also in accordance with its exclusion from the Richters et al. (1988) classification functions.

TABLE 3
Shifts Between Original and Algorithmic Classifications

				Concordances[c]			Discordances[c]					
Sample[a]	%[b]	K	Interval[d]	AA	BB	CC	AB	BA	BC	CB	AC	CA
FRG—Grossmann	93	87	.72–1.00	23	15	2	0	3	0	0	0	0
USA—Thompson	89	.72	.49– .94	5	29	1	2	0	1	1	0	1
ISR—Kibbutz	85	.72	.54– .90	6	29	6	0	1	6	0	1	0
ISR—Day Care	93	72	34–1.00	1	23	2	0	0	1	1	0	0
FRG—Beller	87	.70	.45– .95	6	26	2	1	3	1	0	0	0
SWE—Lamb	83	.64	.43– .84	9	30	1	0	3	3	1	1	0
NL—Young	76	.57	.36– .77	14	17	—	0	6	4	—	0	—
JAP—Miyake	80	54	.31– .76	—	33	10	—	3	2	5	—	1
NL—Old	72	.46	.31– 62	17	48	3	0	17	6	2	2	0
Overall	82	65	.59– .72	81	250	32	3	36	24	10	4	2
Average	83	.66	—									
Richters et al.	85	.73	.65– .80	54	137	25	1	10	2	19	3	4

[a] FRG = Federal Republic of Germany, ISR = Israel; SWE = Sweden; NL = The Netherlands; JAP = Japan.
[b] % = Percentage agreement.
[c] The first mentioned classification is the original classification, the second one is the algorithmic classification
[d] 5% confidence intervals of kappa, based on Fleiss, Cohen, and Everitt (1969).

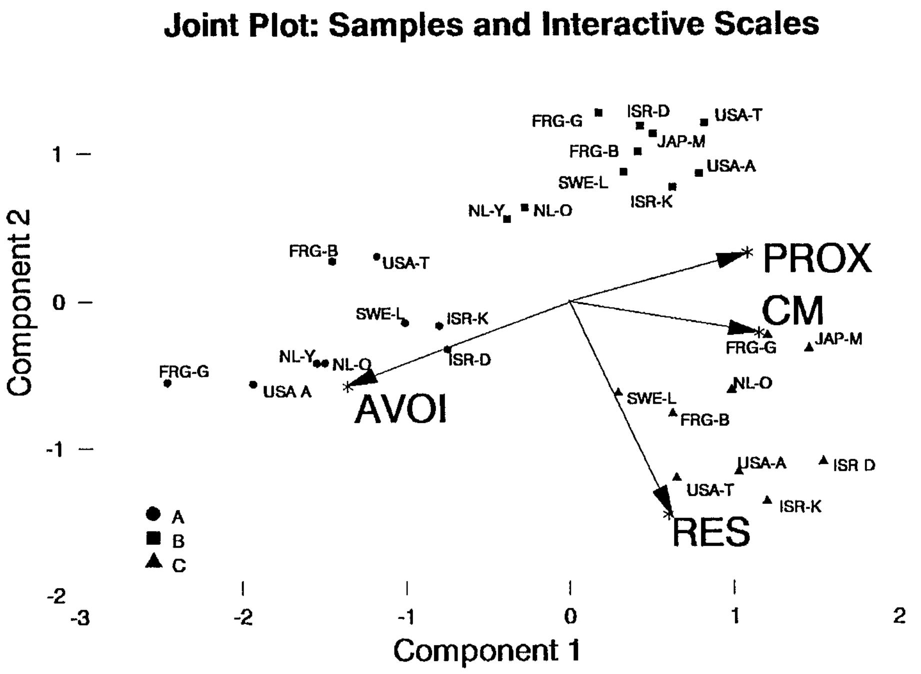


Figure 2. Joint representation of samples and interactive scales derived by a replicated principal component analysis, representing 72% of the variability of the two reunion episodes (68% of the first, 75% of the second)

FRG= Federal Republic of Germany (G=Grossmann, B=Beller)
ISR= Israel (K=Kibbutz, D=Day Care)
USA= United States of America (A=Ainsworth, T=Thompson)
NL= The Netherlands (Y=younger children, O=older children)
JAP= Japan (M=Miyake)
SWE= Sweden (L=Lamb)

In Figure 2, the first two components are portrayed showing both the scores for each sample-classification group combination and the loadings for the interactive scales. The relative weight of the Reunion 1 and Reunion 2 episodes is .42 and .58, respectively. The loadings have been drawn as vectors to clearly mark the difference between scores and loadings.

The most striking feature of Figure 2 is the strong clustering of the classification groups. Clearly, classification groups across cultures are more alike than cultures across classification groups. The Dutch A and C infants are clustered with their counterparts from other countries, but the Dutch B infants show on average somewhat less proximity and contact maintaining and more avoidance than B infants from other samples. Their profile is, however, clearly distinct from the profile of A infants. The profile of the Japanese C infants shows somewhat less resistance and more proximity than that of most other C infants.

DISCUSSION

With a few exceptions, we were able to show that the Strange Situation procedures have been scored consistently across samples from different countries. First, we showed that the interactive scales in the two reunion episodes were scored in accordance with Ainsworth et al.'s (1978) coding rules. That is, mean scores on the interactive scales in our multinational data set did not deviate from the mean scores given in Ainsworth et al., except for distance interaction. All samples show more distance interaction than the original sample, indicating that a cross-cultural explanation is not required here. We suggest that the coding rules as described in Ainsworth et al. do not reflect the intentions of the authors in such a way that systematically different interpretations have been excluded. Because of the less central position of distance interactions in classifying attachment relationships, coding rules may have been described somewhat more loosely than those for the other scales.

Secondly, we found a remarkable cross-cultural consistency in classifying the Strange Situation. Using the revised Richters et al. (1988) functions (i.e., without crying), good agreement appeared to exist between the original classifications and those computed by the functions. The average kappa for the reclassified Strange Situations from our data set is not much lower than the kappa found by Richters et al. in their cross-validation study. Because some shrinkage is likely to occur in applying the functions to other samples, the average agreement figure is quite satisfactory. Our average kappa is in the range of values indicating substantial agreement (.61–.80 according to Landis & Koch, 1977), even if the only sample from the U.S.A. (the Thompson & Lamb, 1983, sample) is not taken into account. There is, therefore, evidence that the coding system has been consistently applied in these samples from several non-American countries. The multivariate analysis confirms this outcome in showing that classification groups are clustering, irrespective of country. Classification groups across countries are, therefore, more alike than are countries across classification groups.

Only the Dutch sample with older infants can be considered an exception to the general rule of substantial agreement between computed and original classifications. Note that it is also the sample with the largest number of deviations from the Ainsworth et al. scale means (Table 2). The sample shows that the functions should be used with caution for infants that are 20 months or older. The behavioral and cognitive abilities of older infants allow them to depend more on distal modes of interaction and less on proximal modes, but the functions cannot take these developments into account as can experienced coders. The Dutch (young) and Japanese samples show kappas in the range of values indicating moderate agreement (Landis & Koch, 1977). We suspect that the modest kappa for the Dutch sample with younger infants is an artifact of the skewed distribution of the original classifications leading to a quite low maximum kappa. In the Japanese sample, the functions produce some A classifications that were originally coded as B, and some B classifications that were originally coded as C, thereby resulting in a more "standard" distribution of attachment categories. Takahashi (1990) noted that different Japanese coders do not always agree on the classifications, implying that at least some of the Japanese classifications are disputable. Grossmann and Grossmann (1988) showed that their recoding of part of the Japanese sample led to more B and fewer C classifications. Furthermore, Japanese C infants are reported to cry longer and show less intensive resistant behavior (Takahashi, 1990). The revised functions do not take crying into account, which could result in C infants being wrongly reclassified as B. The principal component analysis confirms the particular status of the Japanese and Dutch samples, but also shows that they are very similar to the other samples in several other respects, that is, the normal status of Dutch A and C infants and Japanese B infants.

The cross-cultural applicability of the functions is quite good and might have been better if for all samples crying had been available. The functions can, therefore, not only be used for the purpose of establishing interlaboratory agreement within the U.S. (Richters et al., 1988), but also between cultures. For older subjects, however, the functions appear to be of less value, and we suggest that for infants 20 months and older, new functions must be constructed to take the development of proximal to distal interaction patterns into account.

We do not consider the functions as a valid replacement of clinically coding the Strange Situation. The functions may be used effectively to calibrate (cross-cultural) applications of the original coding system, but it has to be kept in mind that the correspondence between original and algorithmic classifications is by no means perfect.

In sum, our data show that the debate about the cross-cultural validity of the Strange Situation can be based on the assumption that attachment relationships have by and large been coded consistently across cultures. We would recommend the use of the functions in cross-cultural attachment research to detect possible systematic deviations from the coding rules. However, it would not be wise to rely blindly on the algorithmic classifications. If correspondence between original and algorithmic classifications is low or absent, researchers should be encouraged to have their Strange Situation classifications checked by experienced coders.

REFERENCES

Ainsworth, M.D.S., Blehar, M.C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation.* Hillsdale, NJ: Erlbaum.

Beller, E.K., & Pohl, A. (1986, April). *The Strange Situation revisited* Paper presented at the meeting of the International Conference on Infant Studies, Beverly Hills, CA.

Bishop, Y.Y., Fienberg, S.E., & Holland, P.W (1975). *Discrete multivariate analysis Theory and practice* Cambridge, MA: MIT Press.

Cohen, J.A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 60,* 37–46.

Fleiss, J.L., & Cicchetti, D.V. (1978). Inference about weighted kappa *Applied Psychological Measurement, 2,* 113–117

Fleiss, J.L., Cohen, J., & Everitt, B.S. (1969). Large sample standard errors of kappa and weighted kappa *Psychological Bulletin, 72,* 323–327.

Gabriel, K.R. (1971). The biplot graphical display of matrices with applications to principal components *Biometrika, 58,* 453–467.

Grossmann, K.E., & Grossmann, K. (1988). Preliminary observations on Japanese infants' behavior in Ainsworth's Strange Situation. *Annual Report of the Research and Clinical Center for Child Development, 11,* 1–12.

Grossmann, K.E., Grossmann, K., Huber, F., & Wartner, U. (1981). German children's behavior towards their mothers at 12 months and their fathers at 18 months in Ainsworth's Strange Situation. *International Journal of Behavioral Development, 4,* 157–181.

Kroonenberg, P.M. (1983). *Three-mode principal component analysis Theory and applications* Leiden, The Netherlands: DSWO Press.

Kroonenberg, P.M., & De Leeuw, J (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika, 45,* 69–97.

Lamb, M.E., Hwang, C.P., Frodi, A., & Frodi, M. (1982). Security of mother–and father–infant attachment and its relation to sociability with strangers in traditional and non-traditional Swedish families. *Infant Behavior and Development, 5,* 355–367.

Lamb, M.E., Thompson, R.A., Gardner, W., & Charnov, E.L. (1985). *Infant–mother attachment The origins and developmental significance of individual differences in Strange Situation behavior* Hillsdale, NJ: Erlbaum.

Landis, J.R., & Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

LeVine, R.A., & Miller, P.M. (1990). Commentary on "Cross-cultural validity of attachment theory." *Human Development, 33,* 73–80.

Miyake, K., Chen, S.-J., & Campos, J.J. (1985) Infant temperament, mother's mode of interaction, and attachment in Japan: An interim report. In I. Bretherton & E. Waters (Eds.), Growing points of attachment theory and research. *Monographs of the Society for Research in Child Development, 50*(1-2, Serial No. 209).

Richters, J.E., Waters, E., & Vaughn, B.E. (1988). Empirical classification of infant–mother relationships from interactive behavior and crying during reunion. *Child Development, 59,* 512–522

Sagi, A. (1990). Attachment theory and research from a cross-cultural perspective. *Human Development, 33,* 10–22.

Sagi, A., Lamb, M.E., Lewkowicz, K.S., Shoham, R., Dvir, R., & Estes, D. (1985). Security of infant–mother,–father,–metapelet attachments among kibbutz-reared Israeli children. In I. Bretherton & E. Waters (Eds.), Growing points of attachment theory and research. *Monographs of the Society for Research in Child Development, 50*(1–2, Serial No. 209).

Sagi, A., Van IJzendoorn, M.H., & Koren-Karie, N (1989). *Primary appraisal of the Strange Situation A cross-cultural analysis of preseparation episodes* (Internal report). Haifa, Israel: University of Haifa.

Takahashi, K. (1990). Are the key assumptions of the Strange Situation procedure universal? A view from the Japanese research. *Human Development, 33,* 23–30.

Thompson, R.A., & Lamb, M.E. (1983). Security of attachment and stranger sociability in infancy. *Developmental Psychology, 19,* 184–191.
Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31,* 279–311.
Van IJzendoorn, M.H., Goossens, F.A., Kroonenberg, P.M., & Tavecchio, L.W.C. (1985). Dependent attachment: B4-children in the Strange Situation. *Psychological Reports, 57,* 439–451.
Van IJzendoorn, M.H., & Kroonenberg, P.M. (1988). Cross-cultural patterns of attachment: A meta-analysis of the Strange Situation. *Child Development, 59,* 147–156.

1 November 1989; Revised 27 March 1990 ■

Appendix

The Richters Functions Without Crying

As crying was not included in the multinational data set, at our request, John Richters was kind enough to recompute the discriminant functions without including crying. Furthermore, he allowed us to publish the revised functions as part of this article. The way they are presented in Table A-1 is, however, slightly different from the way they were presented in Richters et al. (1988, Tables 6 and 7). First, we rounded the coefficients to two decimals and subsequently subtracted the Non-B function from the B function to obtain a single function such that positive and zero values lead to a B classification and negative values lead to a Non-B classification. Similarly, the A function was subtracted from the C function

TABLE A-1
Simplified Richters Functions for Automatic Classification of Strange Situations (Without Crying)

	Coefficient[a]	
Interactive Scale	B v. Non-B	C v. A
Reunion 1		
Proximity	0.13	0.05
Contact maintaining	−0.41	0.38
Resistance	−0.17	0 41
Avoidance	−0.25	−0.43
Reunion 2		
Proximity	0.19	−0 39
Contact maintaining	0.54	1 00
Resistance	−1.44	0 81
Avoidance	−0.90	−1.59
Constant	4.57	0.42

[a] New cases can be classified with these functions by first multiplying the observed scores with the appropriate coefficient of the B/Non-B function and then adding the resulting values plus the constant If a Non-B classification results (i e., the sum is less than zero, B otherwise), the second function can be used in the same manner to decide between an A or C classification (C if sum is greater than or equal to zero, A otherwise).

to obtain a single function such that positive and zero values lead to a C classification and negative values lead to an A classification. Of our 442 cases, 441 received the same classification as with the six-decimal original functions. The one deviation resulted from a difference in the third decimal for the B/Non-B function (−0.00126 vs. 0.00000). In other words, this case was exactly on the borderline between B and Non-B.