

STABILITY OF ATTACHMENT ACROSS TIME AND CONTEXT IN A DUTCH SAMPLE¹

F. A. GOOSSENS, M. H. VAN IJZENDOORN, L. W. C. TAVECCHIO,
P. M. KROONENBERG

Rijksuniversiteit Leiden

Summary.—Children were observed twice in Ainsworth's Strange Situation to measure the stability of their attachment relationship with the caregiver across 1 mo. To get some insight into the ecological validity of the procedure, the children were randomly divided over four sets of conditions: the home-home, home-lab, lab-home, and lab-lab conditions. The test-retest reliability of the interactive scores was high as was the stability of the classifications. The results of the Strange Situation at home and in the laboratory remain non-comparable.

The success of the attachment theory is to a large extent accounted for by the ready availability of an instrument to measure its central construct, quality of attachment (Ainsworth & Wittig, 1969; Ainsworth, Blehar, Waters, & Wall, 1978). Given the measuring procedure and Ainsworth's elucidations, one usually speaks of the Bowlby-Ainsworth theory of attachment (Sroufe, 1983).

The Strange Situation is a standardized observational procedure, based on two assumptions: (a) staying in a strange environment, being confronted with a stranger, and being left by the attachment figure (mother, in our case) are stressful circumstances to the child and elicit optimal activation of attachment behavior; (b) the return of the mother is sufficient to relieve the stress for children with a secure attachment relationship but not for those with an insecure one (Grossmann, Grossmann, Huber, & Wartner, 1981). The Strange Situation enables a trained observer to distinguish three different types of securely and anxiously attached children.

In the past, we have seen a host of investigations into the reliability and validity of the Strange Situation; see Ainsworth, *et al.* (1978) for an overview. Despite the procedures' clinical and holistic character, the usual psychometric qualities have been assessed (Waters, 1978; Connell, 1977). The test-retest stability of the Strange Situation has become a controversial topic (Thompson, Lamb, & Estes, 1982; Waters, 1983; Thompson, Lamb, & Estes, 1983). Bowlby (1971, p. 411) expects the quality of attachment relationships, once they are formed, to remain relatively stable. Sroufe (1979, 1983) also favours stability and continuity of adaptation on the basis of three arguments: (a) there is continuity in the quality of care, (b) children interact with the environment in terms of previous adaptation, and (c) development is hierarchical. This, of

¹Requests for reprints should be sent to Frits A. Goossens, Department of Education, University of Leiden, PO Box 9507, 2300 RA Leiden, The Netherlands. Thanks is due Jan Swaan for his research assistance.

course, does not necessarily mean that the quality of attachment relationships will always remain stable. Both Connell (1977) and Waters (1978) found the stability of attachment relationships across a 6-mo. period to be very high. Vaughn, Egeland, Sroufe, and Waters (1979) and Thompson, *et al.* (1982) found a much lower stability across the same time span, however. They traced this finding to drastic changes in the caregiving arrangements (i.e., discontinuities in the quality of care). Ainsworth and associates (1978) were less successful. In their test-retest study, with a fortnight between T_1 and T_2 , only 13 out of 23 children received the same classification. How to explain this result was a puzzle. They hypothesized that the period between the two measurements was too short to erase all traces of the children's memories of what went on at T_1 . Yet, it seemed important to investigate the stability of the Strange Situation in The Netherlands, where it had not been used before. As we wanted to have this information quickly, looking for the short-term stability was indicated. It had the added advantage of enabling us to exclude rival hypotheses about possible contextual causes of instability. To diminish the chance of a memory carry-over from T_1 to T_2 , we decided to extend the interval between the two measurements to 1 mo. over which we expected the classifications to be stable.

Equally controversial is the ecological validity of the Strange Situation (Van IJzendoorn, Tavecchio, Goossens, Vergeer, & Swaan, 1983). Since it takes place in an experimental room in the laboratory with strange people playing standardized roles, it has everything experimental psychologists are renowned for: a highly contrived situation, hardly representative of real life. In their standard work about the Strange Situation Ainsworth, *et al.* (1978) go to great lengths to demonstrate the superiority of the Strange Situation over natural observation in the living room. The Strange Situation is designed to highlight patterns of attachment, which are related to behavior at home, but would take a great many more hours to observe. It provides a shortcut to observing patterns of attachment behavior. In their view, the Strange Situation is comparable to a daily-occurring separation of the type 'mother leaves child with neighbour to do some quick shopping'. They also argue that the Strange Situation has the characteristics of a 'compound situation', about which Bowlby (1975, p. 146) has the following to say: "... again and again, it is found that a child or adult is especially frightened in a situation characterized by two or more potentially alarming features." According to aforementioned authors, the 'compound' characteristics of the Strange Situation are the strangeness of the environment, a confrontation with a stranger, and the separations from the attachment figure. One might, however, question whether the Strange Situation has to take place in the laboratory. Perhaps, assessment of the quality of the relationship could take place in the home. After all, the child would

still be confronted with more than one alarming feature, i.e., being separated from the attachment figure and being left with a strange visitor. This was tested by measuring attachment twice for all children concerned either by submitting them to the Strange Situation in the laboratory or in the home or by alternating the context between sessions.

METHOD

Children (23 boys and 21 girls) were randomly divided into four groups and observed twice between September and December 1980. All procedures were recorded on videotape. Mean duration between the sessions was 28.9 days ($SD = 6$). Mean age of the children at the first sessions was 17.9 mo. ($SD = 4.7$). Socioeconomic status was determined according to an occupational index currently used in The Netherlands (Van Westerlaak, Kropman, & Collaris, 1975), in which the occupational level is scored on a scale extending from unskilled labour (1) to high-level and academic occupations. It averaged 3.9 ($SD = 1.4$). The laboratory sessions took place in a playroom of 515 by 317 cm, amply provided with a wide range of toys (ranging from a soft, cuddly rabbit to puzzles, building blocks to a rack containing six puppets). The home sessions always took place in the living room. To prevent the stranger being associated with specific toys, none were supplied. But the children had their own at their disposal.

The home sessions also differ from the laboratory sessions in two other ways. There is more furniture in the living room than in the laboratory, and the living room lacks a one-way screen through which the mother can follow the behavior of her child during separation episodes. Instead we supplied her with a small monitor placed in the hall. Otherwise, we adhered rigorously to the instructions for use of the Strange Situation. Due to illness and technical failures, not all children were observed twice or could be classified exactly. In the final sample were 39 mother-child pairs, as Table 1 shows.

Ainsworth, *et al.* (1978) warn that scoring the behavior in the Strange Situation is no mean feat. Especially the scoring of avoidance is problematical for untrained observers. Much training is desired to carry out Strange Situa-

TABLE 1
DIFFERENT ORDERS IN WHICH SESSIONS TOOK PLACE

Group*	Session 1	Session 2	n	Boys	Girls	Mean Age (mo.)	
						T ₁	T ₂
1	home	home	10	4	6	17.5	18.6
2	home	lab	9	2	7	18.4	19.3
3	lab	home	11	7	4	17.9	18.9
4	lab	lab	9	5	4	17.6	18.7

*Note the mother-child pairs are randomly divided over the four conditions.

tion assessments. The fact that coders agree with each other or that the data are stable is not a sufficient guarantee for assessing quality of attachment. Klaus Grossmann kindly invited us to check our expertise as observers by scoring tapes which had already been rated by both trained German observers and checked by American observers (i.e., Main and Weston). A training session with Dr. Escher-Gräub was started, after which we computed the interinvestigator reliability between Leiden and Regensburg on the Regensburg material. Pearson interinvestigator coefficients for the four interactive scales ranged from .77 to .96 (lowest $n = 15$). The percentage of agreement at the nominal level of classification was 94% for each coder (Tavecchio, 1977). In Table 2 the intercoder reliabilities are given for the Leiden material.

TABLE 2
INTEROBSERVER RELIABILITIES

Scale	Interobserver Reliability		
	Ep. 5	Ep. 8	n
Proximity seeking	.77	.91	22
Contact maintaining	.95	.97	22
Resistance	.88	.92	22
Avoidance	.86	.91	22
% agreement at level of classification	100		
subclassification	96.2 ($n = 27$)		

Following our visit to Germany, descriptive notes were made on what happened in each episode, while scores were only assigned in Episodes 5 and 8 (reunion episodes) on the four interactive scales. The children were then (sub)classified. In this study, all first sessions were scored by one observer and all last session by another observer, to ensure independence of assessment.

RESULTS

Data are displayed in Table 3. The stability of the classifications across a 1-mo. period appears to be high, provided the setting in which the Strange Situation takes place remains the same. Groups 1 (home-home) and 4 (lab-lab) display a very high stability, while the classifications in Groups 2 (home-lab) and 3 (lab-home) do not correspond beyond chance-level. Ninety percent of the classifications remained the same in Group 1 (home-home), 33.3% in Group 2 (home-lab), 54.5% in Group 3 (lab-home), and 100% in Group 4 (lab-lab). Corresponding kappas (Cohen, 1960) are .82 ($z = 3.00$, $p < .05$), $-.26$ ($z = -.82$, n.s.), $-.27$ ($z = 1.12$, n.s.), and 1.0 ($z = 3.45$, $p < .05$). Note that subgroup B4 has been set apart. This is because we found B4-children differ from the other B-groups (Van IJzendoorn, *et al.*,

TABLE 3

STABILITY OF FINAL CLASSIFICATIONS FOR ONE MONTH ACROSS SETTINGS

T ₂	A	B1-3	B4	C	Σ	T ₂	A	B1-3	B4	C	Σ
Group 1 (home-home)						Group 2 (home-lab)					
A	3	1			4	A		2			2
B1-3		5			5	B1-3	1	3	1		5
B4					0	B4		1			1
C				1	1	C		1			1
Σ	3	6	0	1	10		1	7	1	0	9
Group 3 (lab-home)						Group 4 (lab-lab)					
A	2	2		1	5	A	3				1
B1-3	2	3			5	B1-3		5			5
B4			1		1	B4			1		1
C					0	C					0
Σ	4	5	1	1	11		3	5	1	0	9

1983). It remains to be seen whether the lack of stability in Groups 2 and 3 must be ascribed to differences in setting and to (systematic) differences in scores on the interactive scales upon which the classifications are based.

In Table 4, the means and standard deviations of the interactive scores of the two reunion episodes are given. Apparently, no large differences between the laboratory and the home setting seem to exist. This obtains for all relevant interactive scale scores. An order effect between the laboratory and home setting could not be discerned either. Significant differences between Groups 2 (home-lab) and 3 (lab-home) were not found. An effect of time of measurement seems not to be present either, since significant differences between the two sessions were not obtained. A multivariate analysis of variance for a repeated-measures design confirmed this conclusion. (Scores in Episodes 5 and 8 were combined for this analysis.) In interpreting these results, one must remember that the sample sizes are small which reduces the chances of rejecting the various null hypotheses. The power of the analysis is bound to be low. On the whole, it does not seem possible to attribute differences in stability between conditions to systematic differences on the interactive scores. Perhaps they originate from differences in correlations between the interactive scale scores at times T1 and T2. These are presented for both reunion episodes and for a summation of the reunion episodes; see Table 5.

Once again, we find that in the same setting (home-home, or lab-lab) all correlations are very high, with a few exceptions in the first group, while the correlations in Groups 2 (home-lab) and 3 (lab-home) with a few exceptions, are low or even negative. This is also illustrated by the mean correlations for the four groups. Obviously the stability of the classifications in Groups 1 and 4 corresponds to a high stability of the interactive scores across time in the same setting.

TABLE 4
MEANS AND STANDARD DEVIATIONS OF INTERACTIVE SCORES FOR EPISODES 5 AND 8

Group	Proximity				Contact				Resistance				Avoidance			
	5		8		5		8		5		8		5		8	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Group 1 (home-home)																
T ₁	2.4	1.4	3.1	1.2	1.9	1.4	2.7	2.1	2.3	0.9	2.4	1.3	2.7	1.2	2.0	0.8
T ₂	2.4	1.2	2.7	1.2	1.6	1.3	2.5	1.6	2.3	1.2	2.8	1.4	3.3	1.6	2.2	1.0
Group 2 (home-lab)																
T ₁	2.9	1.6	3.4	1.4	2.2	1.6	3.0	2.2	2.2	1.4	2.3	1.3	2.9	1.8	2.6	1.2
T ₂	3.2	1.8	2.9	1.6	2.7	2.1	2.9	2.5	2.0	1.7	2.3	1.7	2.8	1.9	2.4	1.1
Group 3 (lab-home)																
T ₁	2.3	1.7	3.5	1.9	1.4	0.9	2.4	2.1	1.4	0.8	1.8	1.0	3.5	1.5	3.1	1.3
T ₂	2.5	2.2	3.2	1.5	1.8	1.8	3.1	2.5	2.6	1.9	2.9	1.8	2.9	1.8	3.3	1.9
Group 4 (lab-lab)																
T ₁	2.0	1.2	2.7	2.4	1.0	0.0	1.8	1.4	1.3	0.7	2.5	1.7	3.7	2.1	2.7	1.7
T ₂	2.3	2.2	2.8	2.2	1.3	0.7	1.6	0.8	1.7	0.9	1.8	1.1	3.3	1.8	3.1	1.5

TABLE 5
PEARSON'S CORRELATIONS OF INTERACTIVE SCORES ACROSS TIME

Scale	Group 1 (h-h)	Group 2 (h-l)	Group 3 (l-h)	Group 4 (l-l)
Proximity Seeking				
Episode 5	.59*	.48	.08	.72*
Episode 8	.44	.21	-.11	.86
Episode 5 + 8	.61*	.41	.16	.95*
Contact Maintaining				
Episode 5	.76*	.67*	.39	.00
Episode 8	.70*	.46	.05	.68*
Episode 5 + 8	.53	.55	.47	.82*
Resistance				
Episode 5	-.04	-.27	.12	.59*
Episode 8	.62*	-.41	-.09	.59*
Episode 5 + 8	.61*	-.46	-.07	.72*
Avoidance				
Episode 5	.50	.68*	.57*	.80*
Episode 8	.61*	.54	.43	.72*
Episode 5 + 8	.75*	.55	.65*	.81*
r_m	.52	.30	.18	.71
n	10	9	11	9

* $p < .05$.

DISCUSSION

The results point to a large test-retest reliability of the Strange Situation, both at the level of the classifications and at the level of the interactive scores. An interval of 1 mo. seems to be long enough to prevent possible contamination of the second measurement by the first, while it may be assumed to be too short for drastic changes in the care-giving arrangements, which may produce instability.

Ainsworth, *et al.* (1978) suggested that the first Strange Situation session seemed to sensitize the children to separation from their caregiver, so that they were more distressed and showed more proximity-seeking and contact-maintaining behavior during the second session. Antithetical behavior (resistance and avoidance), on the other hand, should remain the same or become even less intense. In this situation, the coders will be inclined to give a B-classification. In this study, that cannot be the case. The second sessions (in the home-home as well as in the lab-lab group) do not activate more intense proximity-seeking or contact-maintaining behavior, and they do not have any noticeable influence upon the scores on the antithetical scales. If the effect of anticipation is present at all, it appears already to have faded after 1 mo. The scores, however, are only stable when the setting remains the same. In discussing Brookhart and Hock's study (1976), which resembles this one in that it also used the

Strange Situation procedure at home, Ainsworth, *et al.* (1978) claim that the familiar environment is best comparable to Episode 2 in the laboratory, i.e., a situation that is not stressful enough because it lacks the characteristics of a 'compound situation'. We do not agree with this point of view, as both the presence of a stranger and the absence of the mother in the separation episodes (4, 6, and 7) must be sufficient conditions to speak of a 'compound situation'. The empirical material of our study supports this point of view, since differences on the interactive scores for the home and the laboratory procedures do not seem present. Ainsworth and colleagues (1978) suggest that precisely the securely attached B-children will show avoidant behavior at home. In other words, they predict that those children who in the laboratory are quick to express their preference for the mother by a ready display of proximity-seeking and contact-maintaining while remaining low on avoidance (ignoring, turning away, averting the eyes) following a short separation, will be less affected in the home and show more behavior that could be interpreted as avoidant. This was not confirmed by our data. The change from an A-classification to a B-classification was made as easily as the change from B to A. Such a trend does not show up in the means of the interactive scores. Because the B-children are in the majority, one would expect to find a higher mean on the avoidance scale in the home procedure, but such is not the case. The 'home procedure' seems to be stressful enough to activate attachment behavior. Yet, the results of both settings, home and lab, seem noncomparable. In that respect we are in agreement with Ainsworth, *et al.* (1978). The question as to which construct is operationalized by the home procedure still remains to be answered. The results show that the construct measured can show some reliability, assuming that quality of attachment is validly measured in the laboratory. It might be that at home, another stable aspect of the caregiver-child relationship is being operationalized. This is, however, not supported by the interactive scores, which present normal patterns. We think it is most likely that, although both laboratory procedure and the home procedure correspond in creating 'compound situations', they also differ in their extent of 'compoundness' and are appraised differently by the child. Bowlby (1971) attributed three roles to appraising processes, i.e., appraising changing environments and/or organismic states, providing a monitoring service to the individual as a sentient being and providing a communicative service to others. It is, of course, to the first of these roles that we refer. Ross, Kagan, Zelazo, and Kotelchuck (1975) found in the home and the laboratory, both similarity of responses (i.e., being left alone with a stranger is a critical incentive for distress in both situations) and differences (the unfamiliar laboratory elicits greater fear). Rinkoff and Corter (1980) experimentally manipulated mothers' accessibility by instructing the mother to either leave the door open or close it on departure in both the home and laboratory situation. They found the closed-door condition led to more

distress as measured by crying, irrespective of the familiarity of the environment. In both settings the door was closed all the time except when someone entered or left, as is characteristic of the Strange Situation procedure. Bowlby (1975) contends that proximity-seeking must be seen as situation dependent. But what a situation is depends on the child's assessment. Familiarity or unfamiliarity is in itself not sufficient to define a situation. Mussen, Conger, and Kagan (1974) see separations as events, which may be more tolerable to the extent that a child has more control over the separation (which is the case in the open-door condition) or perhaps to the extent a child knows he can rely upon the return of the attachment figure on the basis of his previous experiences (as is the case with regular alternative care).

Obviously compound situations are not all alike, as witnessed by the lack of stability in Groups 2 (home-lab) and 3 (lab-home) at both level of interactive scores and the classifications. Meanwhile, we would do well to take account of the child's 'interpretations' of the situation. It may well be possible to adapt the classification system to the particular qualities of the environment in which the Strange Situation procedure is taking place. Ainsworth, *et al.* (1978) suggested a different weighting of especially avoidant behavior at home, which they expected to be higher. This we did not find, however. Similar adaptations might be worked out to classify older children. Attachment behavior exhibits changes during the course of development, and the classification system has to be adapted to these age-bound transformations. Proximal attachment behavior gradually gives way to more distal behavior (talking, smiling, showing a toy), and such a replacement also has to be taken into account. How this should be done remains a question for further research. The main conclusion to be drawn from this study is that quality of attachment can be reliably measured in the laboratory with the help of the Strange Situation.

REFERENCES

- AINSWORTH, M. D. S., BLEHAR, M. C., WATERS, E., & WALL, S. *Patterns of attachment, a psychological study of the Strange Situation*. Hillsdale, NJ: Erlbaum, 1978.
- AINSWORTH, M. D. S., & WITTING, B. A. Attachment and exploratory behavior of one-year olds in a strange situation. In B. M. Foss (Ed.), *Determinants of infant behavior IV*. London: Methuen, 1969. Pp. 111-136.
- BOWLBY, J. *Attachment and loss*. Vol. I. *Attachment*. Harmondsworth, Eng.: Penguin, 1971.
- BOWLBY, J. *Separation: anxiety and anger, attachment and loss*. Vol. II. Harmondsworth, Eng.: Penguin, 1975.
- BRONFENBRENNER, U. *The ecology of human development*. Cambridge, MA: Harvard Univer. Press, 1979.
- BROOKHART, J., & HOCK, E. The effects of experimental context and experiential background on infant's behavior toward their mothers and a stranger. *Child Development*, 1976, 47, 333-340.
- COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, 2, 119-122.

- CONNELL, D. B. Individual differences in attachment behavior: long-term stability and relationships to language development. Unpublished doctoral dissertation, Syracuse Univer., Syracuse, NY, 1977.
- GROSSMAN, K. E., GROSSMAN, K., HUBER, F., & WARTNER, U. 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*, 1981, 4, 157-181.
- MUSSEN, P. H., CONGER, J. J., & KAGAN, J. *Child development and personality*. (4th ed.) New York: Harper & Row, 1974.
- RINKOFF, R. F., & CORTER, C. M. Effects of setting and maternal accessibility on the infant's response to brief separation. *Child Development*, 1980, 51, 603-606.
- ROSS, G., KAGAN, J., ZELAZO, P. M., & KOTELCHUCK, M. Separation protest in infants in home and laboratory. *Developmental Psychology*, 1975, 11, 256-257.
- SROUFE, L. A. The coherence of individual development: early care, attachment and subsequent developmental issues. *American Psychologist*, 1979, 34, 834-841.
- SROUFE, L. A. Infant-caregiver attachment and patterns of adaptation in preschool: the roots of maladaptation and competence. In M. Perlmuter (Ed.), *Development and policy concerning children with special needs: The Minnesota Symposia on Child Psychology*. Vol. 16. Hillsdale, NJ: Erlbaum, 1983. Pp. 11-91.
- TAVECCHIO, L. W. C. Quantification of teaching behavior in physical education: a methodological study. Unpublished doctoral dissertation, Univer. of Amsterdam, 1977. (University Microfilms International, No. 77-70,039)
- THOMPSON, R. A., LAMB, M. E., & ESTES, D. Stability of infant-mother attachment and its relationship to changing life circumstances in an unselected middle-class sample. *Child Development*, 1982, 53, 144-148.
- THOMPSON, R. A., LAMB, M. E., & ESTES, D. Harmonizing discordant notes: a reply to Waters. *Child Development*, 1983, 54, 521-524.
- VAN IJZENDOORN, M. H. Operationaliseringsproblemen bij onderzoek naar de affectieve relatie tussen ouder en kind. *Pedagogische Studiën*, 1979, 56, 358-368.
- VAN IJZENDOORN, M. H., TAVECCHIO, L. W. C., GOOSSENS, F. A., & VERGEER, M. M. *Opvoeden in geborgenheid*. Deventer: Van Loghum Slaterus, 1982.
- VAN IJZENDOORN, M. H., TAVECCHIO, L. W. C., GOOSSENS, F. A., VERGEER, M. M., & SWAAN, J. How B is B4? Attachment and security of Dutch children in Ainsworth's Strange Situation and at home. *Psychological Reports*, 1983, 52, 683-691.
- VAUGHN, B., EGELAND, B., SROUFE, L. A., & WATERS, E. Individual differences in infant-mother attachment at twelve and eighteen months: stability and changes in families under stress. *Child Development*, 1979, 50, 971-975.
- WATERS, E. The reliability and stability of individual differences in infant-mother attachment. *Child Development*, 1978, 49, 483-494.
- WATERS, E. The stability of individual differences in infant attachment: comments on the Thompson, Lamb, and Estes contribution. *Child Development*, 1983, 54, 516-520.

Accepted November 8, 1985.