Chapter 2

Emotional instability as an indicator of strictly timed infantile developmental transitions

C. de Weerth and P. van Geert

Developmental Psychology, University of Groningen
The Netherlands

ABSTRACT

This chapter presents the results of the first validation attempt of the pattern of 10 strictly timed periods of infant emotional instability, proposed by Van de Rijt-Plooij & Plooij (1992a). These authors base their hypothesis mainly on the analysis of maternal reports. The goal of the present study was to validate the hypothesized pattern by means of both maternal reports *and* ethological observations. A longitudinal study was carried out on 4 mother-infant pairs. The infants were followed weekly from birth till the age of 15 months. The following behaviors were scored during the 3-hour long home observations: infant crying, fretting/fussing and smiling, and body contact between mother and infant. The results failed to support the 10-periods pattern. Possible explanations for these findings are presented, together with a discussion about methodological aspects of the data analysis.

INTRODUCTION

In this study the criterion of *instability* is used to identify the occurrence of a number of transitions in infantile development. Instability can be roughly defined as a temporary loss of the actual behavioral competence, resulting in the upsurge of behavioral patterns characteristic of an earlier developmental level, i.e. regression. The starting point of our study lies in a claim by Van de Rijt-Plooij & Plooij (1992a) that there exist 10 periods of instability, which they called 'regressions'[1], in the first twenty months of life. According to the authors, these regressions precede the onset of developmental transitions and are found in the emotional domain where they are characterized by temporary decreases, or even disappearances, of the growing independence of the baby, giving rise to instability in the above defined sense. In their longitudinal study, in which 15 mothers independently filled in a weekly questionnaire, an impressive consensus was found on the ages when their infants were reported to cry more and spend more time in body contact. Strikingly, there were even weeks in which none of the mothers reported these behaviors while during others all of the mothers reported them. These regression periods took place around weeks 5, 8, 12, 17, 26, 36, 44, 52, 61-62 and 72-73, and were also often characterized by a decrease in amount of sleep, fear for other people, childish behaviors, decrease in amount of food intake, problems with changing/dressing, decrease in activity, peak in cuddling mother, and peak in cuddling objects. Thus, the infants were unstable and easily put off balance, and most mothers reported experiencing these periods as very tiring and difficult to cope with. In two mother-infant pairs the questionnaire data were compared with objective recordings of mother-infant interactions obtained through monthly home observations. The authors found that the observational data corroborated the questionnaire data. They concluded that the mothers' reports describe real variations in mother-infant behavior, instead of mere variations in maternal perceptions (van de Rijt-Plooij & Plooij, 1993).

The present chapter reports the findings of a project which was designed as a partial replication of the 1992 study by Van de Rijt-Plooij and Plooij. It differed from the latter study in that the accent lay on an intensive regime of weekly home observations instead of on maternal reports. Because of the replicative nature of the study, we have adopted the definitions of regression as stated by the Plooijs without speculating on their adequacy as indicators of instability nor passing judgement on their suitability as criteria indicative of transitions. The data presented in this chapter were therefore not used in a direct search for transitions, but in a search for a pattern of well-defined periods of increased crying/fretting/fussing and increased body contact, which according to the Plooijs are related to developmental transitions.

The Plooijs' evidence for the existence of a number of transition periods in infancy does not stand alone. We shall first give a short overview of the literature on transitions during infancy.

---

[1] Note that although Van de Rijt-Plooij & Plooij have labeled these periods of instability with the term 'regression', they have not used it with its classical psychoanalytical meaning. Their regression periods may, but do not necessarily, include classical regressive phenomena, and should therefore be interpreted solely as the above mentioned authors have defined them. Thus, the label of 'regression' for these periods of emotional instability may be confusing to some. Nonetheless, because this study is a partial replication and a validation of the phenomena described by the Plooijs, we have decided to adopt the term throughout this paper thus respecting their choice of terminology.

Evidence for transitions

Developmental transitions have been observed in many different areas of infant development, such as cognition and emotion, and motor and language development (Spitz, 1959; Corrigan, 1983; Fischer, 1982; von Hofsten, 1984; Trevarthen, 1974; see also Ruhland & van Geert, 1998, and Wimmers, Beek, Savelsbergh & Hopkins, 1998). During the first two years of life, there are four major transitions which are agreed upon in the literature: around 2, 7, 12 and 18-21 months (Bates, 1976; Bloom, 1973, Emde and Gaensbauer, 1981; Fischer, 1980, 1987; Kagan, 1984; Konner, 1976; Lamborn and Fischer, 1988; McCall et al., 1977; Siegler, 1989; Trevarthen, 1988; Volterra, 1987; Zelazo and Leonard, 1983). Also, infant physical growth occurs in spurts, as measured by the length and head circumference (Lampl & Emde, 1983; Lampl, Veldhuis & Johnson, 1992). Spurts in brain growth occur at approximately 3-4, 7-8, 10-11, and 15-18 weeks, and at approximately 8, 12, and 20 months of age (Fischer & Rose, 1994). Thus, the transitions mentioned above tend to go together with these cerebral growth spurts. However, as there are more than four growth spurts of the brain it is possible that these also accompany transitions which are yet to be determined. Indeed, there is evidence from different sources which indicates that more than four transitions take place in the first years of life (Mahler et al., 1975; Trevarthen, 1977; Proctor, 1989; Fischer and Hogan, 1989).

Given that Van de Rijt-Plooij and Plooij (1992a) provided evidence of no less than ten transition periods, we may conclude that the questions of how many transitions occur, in which developmental domains and when, are still open issues.

Evidence for regressions

Various researchers have found developmental transitions to be characterized by regression during which certain infantile capacities disappear or decrease temporarily, only to reappear (often in a modified, more sophisticated form) at a later age (Spitz, 1959; von Hofsten, 1984; Trevarthen, 1974; van Geert, 1991). For example, Von Hofsten (1984) observed how the amount of infantile prereaching declined at the age of 7 weeks, to later increase again as a behavior characterized by more precision. The theoretical interpretation of these regressive phenomena, which immediately strike as being the counterpart of development, is still a matter of dispute. According to Kozulin (1990) a central notion in Vygotsky's ideas on development is that development constitutes the formation of higher functional systems out of lower separate functions, whereas regression is the disintegration of functional systems and a return to isolated functions at a lower level. Other authors, partly inspired by the Piagetian tradition, have claimed that development consists of a reconstruction of the existing cognitive means and tools. It is obvious that the quality of performance suffers from the reconstruction process and thus shows a temporary regression (Bever, 1982; Strauss, 1982; Strauss and Stavy, 1982). Another possible explanation is that of Van Geert (1991; 1994) who interprets periods of instability within the frame of the dynamic systems model of competitive (cognitive) growth. According to this theory, regressions in cognitive and language development are part of take-over phenomena between two alternative strategies or 'growers'. Temporary regressions to a former state of development would occur as a result of the competitive relationships between growers.

In summary, various authors claim that several periods of transition take place during infancy, although there is at present no general agreement about the number and the nature of those transitions. Second, regressions and comparable forms of instability can be viewed as indicators of developmental transitions. The strongest claims are made by Van de Rijt-Plooij and Plooij (1992a) who distinguish ten transitions, each characterized by observable ‘regressive’ behaviors.

Goal of the present study

As already stated, the present study was designed with the primary goal of collecting detailed data which should confirm the regression periods found in the 1992 study by Van de Rijt-Plooij & Plooij. Particularly interesting was the question whether weekly ethological observations would confirm the results of the data of the former study, which consisted mainly of interpretations of maternal reports. We argue that because of their objectivity, ethological observations may constitute important corroborations of questionnaires. The Plooijs themselves used this approach, albeit with monthly observations, and found that their observational data corroborated their questionnaire data.

A design was chosen in which 4 mother-infant pairs were intensively followed during a period of 15 months. Apart from weekly questionnaires and interviews with the mothers, weekly home observations in naturalistic settings were carried out throughout the whole study period. Also, the infants' sleeping behavior was automatically registered 24 hours/day by equipment placed next to the crib.

In this chapter, the results of the analyses of the questionnaire and the observational data will be presented. The goal of the analyses was to test the following hypothesis: (a) The data obtained through direct home observations will show increases or peaks in crying/fretting/fussing and in mother-infant body contact around weeks 5, 8, 12, 17, 26, 30, 36, 44, 52, 61-62; (b) The mothers will report periods in which their infants display behavior characteristic of emotional instability which will be consistent with those found by Van de Rijt-Plooij & Plooij (1992a) and therefore take place in the same weeks .

METHOD

Subjects

The subjects were four Caucasian mother-infant dyads from two-parent middle-class families. The mothers were recruited through midwives, a local newspaper, letters sent through the Green Cross (state agency that provides help at home for mothers with new-borns) and by visiting pre-natal gym classes. All the mothers had completed secondary school at least and one had a college degree. They were selected on the basis of the following criteria: a) they were married, financially secure, and had an extended support system (family and friends nearby); b) during the research period they would not work outside the home and would be the primary caretakers; c) the pregnancies had been uneventful and had resulted in deliveries without major complications (three vaginal and one Caesarean); d) the infants, two males (E and F) and two females (J and S), were healthy, full-term firstborns; e) the infants had adequate Apgar scores (8 and 9 at 1 min. and 9 and 10 at 5 min.), and they showed no abnormalities during the neurological examination that was performed by a physician 7 days later.

In order to determine whether the infants were developing normally, they were tested with the Bayley Scales of Infant Development (BOS 2-30) around the age of 15 months. The resulting Mental/Motor scores for the 4 infants were: 117/45, 114/47, 115/57, 115/49 which indicate developmental ages between 14 and 16 months, with the exception of the motor score of 57 which indicates a developmental age of 22 months.

Materials

The ethological observations were carried out by typing in behavior codes on a laptop computer while observing the mother and infant. The one-minute periods of the interval sampling were identified by a clip-on timer and were communicated to the observer through an inconspicuous ear-phone. These materials gave the observer a relatively high degree of freedom of movement in order to follow the mother and her infant around the house. The software package used for the observations was The Observer 3.0 (Noldus, 1991). The observations were not videotaped.

The questionnaire that the mothers filled in weekly was based on the one used by Van de Rijt-Plooij & Plooij (1992a, 1993) and on the results of their data analysis. It was made up partly by open questions and partly by multiple choice and yes/no questions (see Appendix A for the complete text of the questionnaire). The interview consisted of the observer reading the questionnaire and asking the mother for additional information, clarification and examples.

Procedure

Weekly data collection

The four mother-infant pairs were visited weekly by the first author, who carried out all of the observations. The first observation took place in the second week of life for two of the infants and in the third week for the other two infants. The last observation took place in the sixty-fifth week for three of the infants and in the sixty-fourth week for the fourth (corrected ages, see Analysis). Observations were missed only during the weeks the families went on vacation and one week when the observer was ill. Each mother had her "own" observation day and it was very rarely necessary to reschedule the weekly visit for another day of the week. The observations always began around 9.00 in the morning and continued until 3 hours of data collection had been completed. If the child fell asleep, the observation was stopped until the child was awake again. The interview either took place while the infant was asleep or after the observation was completed.

The mothers were instructed to ignore the observer as much as possible and to go about their everyday household routine. The goal was to observe mother and infant under naturalistic conditions. The only constraints on their activities was to ask the mothers not to go for walks or to receive visitors during the observation period.

Due to the lengthy observations and the extended study period, it was impossible to keep the mother from engaging in conversation with the observer during the observations. However, because this gave the situation a fairly normal character and it remained unchanged during the 15 months, we do not feel it could have greatly influenced the subjects' behavior.

During the observations, the observer's interactions with the infants were limited to smiling back when they smiled at her and fending them gently off when, at an older age, they tried to touch the keyboard of the computer. In practice, the infants largely ignored the observer during the observations.

Although ideally the mothers should have been naive to the object of investigation, this was in practice impossible due to the publication of a popular book by Van de Rijt-Plooij & Plooij (1992b). This book explains the 10-regression period pattern to parents and was at the time of initiation of the project extremely successful in The Netherlands, with plenty of newspaper and television coverage. Therefore, because it was impossible to avoid the mothers hearing about the book, or even receiving it as a present, a copy was given to each of the four mothers in order to keep the study group as homogeneous as possible.

Ethogram

The ethogram is based on the one used by Van de Rijt-Plooij & Plooij (1992a). The following list includes descriptions of only those behavioral elements which will be treated in this chapter:

Vocalizations. *Crying:* weeping. Often accompanied by intense facial coloring and production of tears. *Fretting/fussing:* protesting, whimpering, whining. Vocalizations of an audible negative affect which suggest complaint, irritation, frustration, worry, distress, annoyance, tiredness, etc. Vocalizations which were not clearly interpreted as having a negative affect were not scored.

Contact/distance. Body contact: the infant's body was touching the mother's. It included physical contact even if there was something between the mother and the baby's skin. For example: pull or tug at clothes, arrange clothes, wipe baby's mouth, tickle the baby with something (like a teddy bear), etc.

Contact/distance and *crying* were scored by continuous coding of events. *Contact/distance* was actually scored in 7 mutually exclusive categories and by registering the initiator of each change in category. However, because in this analysis of the data only the total time spent in body contact was used, the categories and initiator are irrelevant and will not be further referred to.

Fretting/fussing was recorded by means of interval coding. The observer categorized each 60 sec. interval of the observation according to the occurrence or non-occurrence of these behaviors. The decision to use interval sampling was based on the low reliability scores which were obtained when sampling these behaviors continuously.

Reliability

The observer received an intensive training by H.H.C. van de Rijt-Plooij during live home observations. Training was continued until the inter-observer reliability scores were satisfactory. The mean percentage agreements for the contact-distance categories was 91.1% and for crying 90.8%, and Cohen's kappa for fretting/fussing was .88.

To rule out the possibility of observer drift or decay, videos were made of two mother-infant pairs after 7 months of the observation period had elapsed. Fragments containing active mother-infant interaction were used to calculate intra-observer reliability by scoring each fragment twice, with an interval of a month between both occasions. The mean percentage agreements for the body contact categories was 89.3% and for crying 86.6% and the kappa for smiling-fretting was .71, levels which indicate little observer drift.

The maternal questionnaires were coded for analysis with the coder blind to the age of the infant. A naive coder was used to investigate the reliability of the coding system. The percentage agreement that was obtained over a sample of 12 questionnaires, was 90.69% which indicates the coding system to be reliable.

Analysis

Before the data could be used in our analysis, they had to be corrected for age and checked for the occurrence of illness (see sections *Age Correction* and *Illness*). Next we had to develop criteria for regression periods, both in the original Plooij data set and in our present study (sections *Regression periods and criteria* and *Determination of regression weeks in the present study*). In addition, procedures were sought for testing our major hypothesis in the observational data. We finally chose two approaches. One was aimed at determining the occurrence of peaks, as opposed to normal background variation, and was based on the difference between the value of a particular data point and the (average) value of its neighbors (section *Procedures: the Peak Method*). Whereas this first approach to the question of regression periods in infancy focused on the correspondence of peaks in the data and on the original Plooij data, the second approach was to check whether weeks that fell within predicted regression periods were characterized by general increases in regressive behavior in comparison to weeks that fell outside those periods. Thus, instead of displaying spikes of regressive behavior that occur only every *n*th day, a regressive period could as well show a general increase in regressive behaviors. Since the raw data displayed significantly varying non-linear trends, all data had to be detrended and residuals from those non-linear trends were used in order to test the hypotheses in the following three analyses (see *Procedures: the Residuals Method*). The first method checked for the general fit between the Plooij data and that of the total series of observed behaviors of these four infants by using Pearson's correlations (*Total series of observed behaviors*). The second method compared the average regressive behavior scores in the predicted regression weeks with that of the remaining weeks (*Means of residuals*). The third and final method compared the number of positive residuals of the polynomial curve falling in predicted regression weeks with the number falling in the remaining weeks (*Positive residuals*). These methods were also used to investigate relationships between the observational variables within and between infants, and also between each infant's observational variables and maternal reports.

In both the *Peak Method* and the *Positive Residuals Method*, the Chi-Square test (or the Fisher Exact Test when the minimum expected frequency was < 5) was used to test our predictions against the null hypothesis of a random distribution. In the case of the *Means of residuals Method,* the differences in means were compared to fractions of the standard deviations, as a way of assessing the importance of the differences.

Finally, it is important to add that the use of various methods (instead of one) is justified by the fact that each method has its particular advantages and disadvantages and that the advantages of one compensated for the disadvantages of another.

Age correction

One of the female infants was born 17 days later than expected and the other 16 days earlier than expected (based on the date of the last menstruation). Because the objective of this study was to corroborate the findings of Van de Rijt-Plooij & Plooij (1992a), and these authors based their conclusions on subjects whose gestational ages had been used in the data analysis, the same age correction was applied to the female subjects of the present study. Thus, the ages of the two infants were corrected by using the dates they were expected to be born instead of the dates they were actually born. The postnatal ages of the infants in question were nevertheless included in an exploratory analysis of the effect of age modifications on the results (see Results, *Observational Data: Age modifications*).

Illness

The infants did not suffer from important illnesses during the study period and were consequently observed every week notwithstanding minor illnesses. The observations during which, according to the mother (and confirmed by the observer), the infant's behavior was clearly affected by fever, the flu, diarrhoea, etc., were the observations pertaining to weeks 30, 32 and 53 for infant E, weeks 20 and 56 for infant J, week 65 for infant F and weeks 44, 47 and 59 for infant S. Because these minor ailments involved a small number of observations and did not systematically affect the studied behaviors, the observations were included in the analyses. Also, in the figure legends the weeks mentioned above have been noted as weeks in which the infant was ill.

Regression periods and criteria

In order to investigate whether the results of the present study were consistent with those of the study by Van de Rijt-Plooij & Plooij (1992a), different measures were derived from the data of their study. 'Strong regression (r) weeks' were defined as those in which as much as 80-100% of the mothers of the Plooij study had reported that their infants showed regressive behaviors. The opposite measure, 'Non-regression (nr) weeks', were defined as those in which only 0-20% of their mothers had reported regressive behaviors in their infants. '60+ regression weeks' were defined as those in which 60-100% of their mothers had reported regressive behaviors (this measure naturally included the strong regression weeks). This last measure was chosen because it represented the majority of the infants of the Plooij study, gave better results than an alternatively tested 70-100% measure, and did not differ greatly from a 50-100% measure.

The R variable (see Peak method) could be calculated between weeks 3-64 for the infants in the present study, so these were the weeks in the Plooij study which were searched for regressions. In this way, 14 strong regression weeks were determined (5, 8, 12, 16-18, 26, 43-44, 51-53, 61-62; note that the periods around weeks 30 and 36 are not represented by this measure), 17 non-regression weeks (6, 7, 10, 11, 13, 14, 20-22, 28, 32, 48, 49, 56-59) and 29 60+ regression weeks (5, 8, 9, 12, 15-19, 24-26, 29, 30, 34-37, 42-45, 51-54, 61-63).

Finally, a continuous 'reported regression' variable was defined as the percentage of the infants from the 1992 study who were reported by their mothers to show regression behaviors for each of the age-weeks. This variable was correlated with the residuals obtained from the polynomial fitting (see below).

Determination of regression weeks in the present study: Observational data

Several criteria were used for determining the regression weeks in this study. First of all and most important, the criterium based on the one used in the 1992 study by Van de Rijt-Plooij & Plooij, namely a peak in both body contact and in crying (duration and/or frequency) or fretting/fussing. This meant that in the data from the same weekly observation there had to be peaks in both body contact and at least one type of negative vocalization. Further, the following criteria were also used independently to determine regressions: peaks in body contact, peaks in crying (duration and frequency) and peaks in fretting/fussing.

Determination of regression weeks in the present study. Maternal report
The criteria used for establishing regression weeks in the data obtained through the mothers were again based on those used by Van de Rijt-Plooij & Plooij (1992a). Thus, two variables were defined as follows. The most important was the one used by the Plooijs to determine the weekly percentage of infants which were in a regression, namely, the mother reported increased crying *and* increased body contact *and* (after the age of 2 months) one or more of the other regressive behaviors (decrease in amount of sleep, fear of other people, childish behaviors, decrease in amount of food intake in one meal, problems with changing/dressing, decrease in activity, peak in cuddling mother and peak in cuddling objects). This variable was called ‘concrete regression’, meaning that the regression was based on the mother’s report of the concrete (i.e. specific, actual) regression behaviors just mentioned.

Furthermore, based on the mothers’ statements in the questionnaire and interview, a general maternal impression of the infant’s regression behavior that week was also established. This variable was called ‘general regression’, meaning that the regression was based on the mother’s report of the infant being more difficult in general, without her necessarily reporting concrete regression behavior. Because it was not always possible to obtain a general impression of the week from the maternal data, this variable suffers from missing data (especially in infants E and J, in which the mothers were less inclined to pass general judgements on their infant’s behavior). Nonetheless, 28 weeks were coded for infant E, 27 for inf.J, 56 for inf.S and 49 for inf.F.

In the analysis of whether the maternal reports supported the earlier findings of Van de Rijt-Plooij & Plooij (1992a) and in that investigating the correspondence between the different infants, weeks in which the mother reported regression behaviors clearly being due to illness/teething (0-3 wks/infant) were treated as missing data and therefore not included. Weeks in which the criteria were fulfilled, but at the same time illness was present and according to the mother could also be affecting the infant’s behavior, were coded as unclear regression weeks and treated as non-regression weeks in the analysis.

In the analysis of whether the maternal reports could be confirmed by the ethological observations, both the weeks in which the infant clearly displayed regression behavior due to illness and the unclear weeks, were included in the analysis.

In this way ‘concrete regression weeks’ were established for all four infants: 21 for infant E, 8 for J (plus 2 unclear weeks), 18 for S (plus 6 unclear wks.) and 15 for F (plus 1 unclear wk.). The number of ‘general regression weeks’ reported were: 11 for inf. E (plus 2 unclear weeks), 9 for inf. J, 19 for inf. S (plus 6 unclear wks.) and 19 for inf. F (plus 1 unclear wk.).

Procedures: the Peak Method

As a preliminary analysis of the data, the four infants' ages in weeks were plotted against the following variables: the percentage of the observation time spent crying and that spent in body contact, the frequency of crying, and the percentage of the total number of one-minute periods (maximum of 180 per observation) in which fretting/fussing had occurred.

A procedure was developed in order to determine the most important peaks in the 65 weeks' time series x(t) of each displayed behavior. The basic line of reasoning was that a peak occurred in week t if x(t) was clearly higher than expected when compared to the values of the surrounding weeks. A variable (R) was created which took into account all four surrounding weeks but gave more weight to the two closest neighboring weeks. It was calculated as follows:

$$R_t = x_t - (x_{t-2} + 2x_{t-1} + 2x_{t+1} + x_{t+2})/6$$

The R-value specifies the difference in level between the behavior(s) in one particular observation week in comparison to the same behaviors during the surrounding observation weeks. Thus a high R-value refers to a single "spike" in the observed behaviors. The R-procedure is particularly suited for those forms of regression where the regressive behaviors occur in a peak-wise fashion and presumably "color" the mother's perception of the child's behavior before and after the peak. For instance, if a regression period is characterized by surges of regressive behavior that occur every three or four days across a period of several weeks, the mother will report all these weeks as regression weeks. This means that the probability of capturing all regressive behavior surges by observing the child one day a week is relatively small. However, if these surges occur during several weeks, there is a significant chance that at least one of them will be covered by an observation session. That is, there is a significant probability that one such observed surge will fall somewhere within the regression period.

The Peak Method is not suited for determining regressive periods that are characterized by prolonged periods of increased regressive behavior. If such type of regression occurs, it will be better captured by the Residuals method (see further).

Linear interpolation was used to fill in the gaps due to missing observations. These interpolated values were only used to obtain the values for R in which either x_{t-2} **or** x_{t+2} were missing. If x_{t-1} and/or x_{t+1} were missing, R_t was also scored as missing. In this way the problem of giving too much weight to an interpolated value was avoided. R was not calculated for the first and last points of the series. In the case of the second observed week, R was calculated as follows: $R_t = x_t - (3x_{t-1} + 2x_{t+1} + x_{t+2})/6$, and in that of the penultimate observed week as: $R_t = x_t - (x_{t-2} + 2x_{t-1} + 3x_{t+1})/6$.

Because the values of R for each data set went smoothly from high to low (i.e. there was no natural "break" between high and low values), the ten highest R values were chosen for each data set. This was done in the following way. First, the infants' data were divided in two parts: weeks 1-20 and weeks 21-65. The decision to divide the data was based on the fact that the characteristics of the regressions are different in one period as compared to the other. In the first period, 4 regression periods follow each other closely (mean inter-regression period of 2 weeks), while in the second period 6 regressions with longer "rest periods" in between occur (mean inter-regression period of 4 weeks). Also, the values of the data of the first 20 weeks were often dramatically higher and showed important downward trends with relatively small reversals, as compared to that of weeks 21 to 65. By dividing the data in this way, the risk of missing regressive peaks solely due to the interaction of the analysis method with the general developmental trends was being diminished. R was calculated in the same way in both periods.

The next step was to calculate the R values for a given displayed behavior for both periods, and then to select the 4 weeks with the highest R values in the first period and the 6 weeks with the highest R values in the second period as those showing the highest peaks.

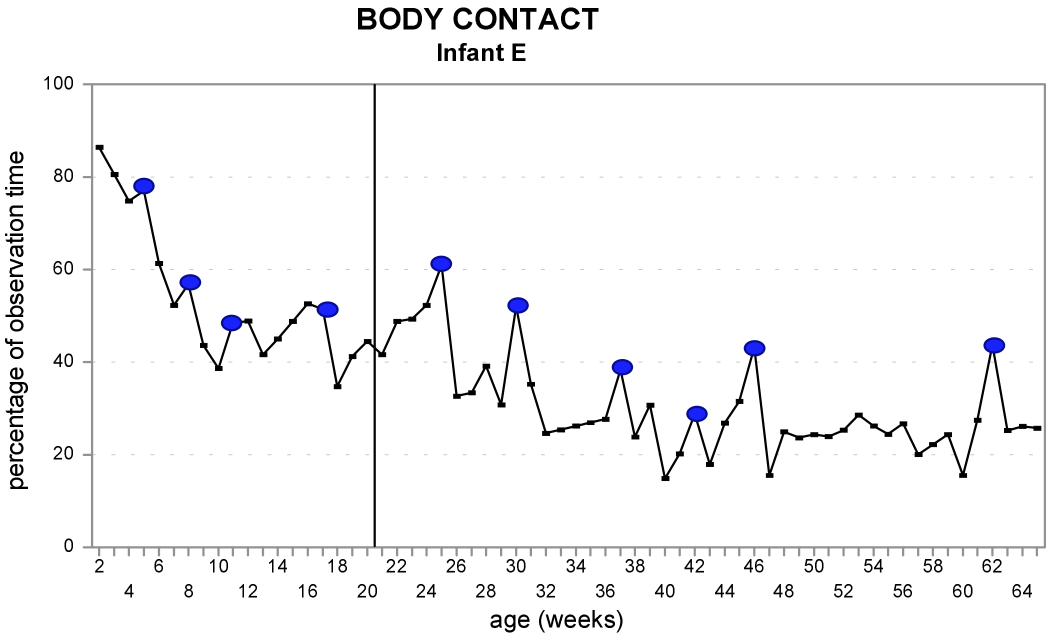


Figure 1. Example of the application of the peak method on the body contact data of infant E. The circles represent the 10 most important peaks, obtained by selecting the four highest Rs from weeks 1-20 and the six highest from weeks 21-65. Weeks 15, 33-35 and 58 contain linearly interpolated values due to missing data: therefore, the R values for weeks 14-16, 32-36 and 57-59 could not be calculated.

When two neighboring weeks were chosen by this method, they were considered to be representing the same regression period. Therefore, the one with the highest R value was selected and its neighbor was replaced by the week with the next highest R value of the list. Also, if the R values turned negative before the predetermined number of weeks had been chosen, only the weeks with positive R values were selected. Figure 1 serves as an example of the results obtained when applying this method of determining the weeks with the highest peaks. As can be seen in the graph, infant E displayed peaks in body contact during the observations in weeks 5, 8, 11, 17, 25, 30, 37, 42, 46 and 62.

The peaks found by calculating R in the regressive behavior graphs of the four infants followed in the present study, were expected to fall in weeks which matched those of the '60+' measure and those of the 'strong r' measure (see *Regression periods and criteria*), but not in weeks which matched those of the 'nr' measure.

Procedures: the Residuals Method

Most of the graphs of the observed behaviors displayed important non-linear trends in the data, for instance exponential decreases. This, for example, made the calculation of correlations between the behavioral variables unreliable as indicator of *local* correspondence between those variables. That is, with such major (non)linear trends, it is impossible to detect whether a higher than average incidence of body contact during a particular week corresponds with a higher than average frequency of fretting and fussing. A good way of extracting prolonged deviations from a local average is to determine the difference between the actual observation point and the value of a corresponding polynomial regression function. The polynomial estimates the best fitting curve for a sequence of observations. If O_t is the value of an observed variable at time t, its corresponding polynomial value P_t equals $P_t = a + b\,t + c\,t^2 + d\,t^3 + e\,t^4 + ...$

The polynomial curve of the second degree (i.e. one with 2 as its highest exponent) was chosen because such curve eliminates linear and eventually also quadratic trends. The residuals for the total data series were then calculated. These residuals were then used to calculate Pearson's correlations between the behavioral variables and the 'reported regression' variable, between the infants and between the different behavioral variables (see Results, *Observational Data: Total series of observed behavior*). The residuals of the second degree polynomials were also used to study the difference in means of behaviors in 60+ regression weeks and in the rest of the observed weeks (see Results, *Observational Data: Means of residuals*). The fifth-order polynomial curves follow the original data more closely and their residuals were therefore specifically chosen for one analysis, which gives much weight to whether the polynomial curve lies above or under each data point. This analysis is namely the study of the amount of positive residuals in 60+ regression weeks and in the rest of the observed weeks (see Results, *Observational Data: Positive residuals*).

Because the data of the Van de Rijt-Plooij & Plooij (1992a) study were sampled into a group curve which shows fairly age-specific characteristics, the data of the four infants of this study were summed over the same-age observations, for all the behavioral variables' residuals (of the second degree polynomials). These summed variables were then used to calculate Pearson's correlations with the 'reported regression' variable.

Finally, the expected regression weeks were analyzed in order to see if they had on average more regressive behavior than the rest of the weeks. If the summed scores of regressive behaviors in the regression weeks did not positively differ from the sum scores in the remaining weeks, there was no use in further pursuing the search for correspondence in the regression patterns, since this would mean that there exists no difference between alleged regression and non-regression weeks. Two methods for comparing average regression indicators were used, one based on the number of positive versus negative residuals, the other on the average of the residuals in the regression versus non-regression weeks.

RESULTS

Consistency of the data with the expected 10-regression period pattern

Observational data: Peaks in the observed behavior

The age of the four infants in weeks was plotted against the variables indicative of regression: the percentage of the observation time spent crying and that spent in body contact, the frequency of crying and the percentage of the total number of one-minute periods (maximum of 180 per observation) in which fretting had occurred. Figures 2-5 show the graphs obtained.

R values (see Method) were then calculated for the data of all the graphs and thereafter used to determine the weeks with the most important peaks in behavior for each one. Note that in the case of infant J, due to missing data no R values could be calculated up till week 6, for weeks 28-30, and for week 61 onwards. Therefore, because three of the regression periods were being missed completely, only 3 weeks were selected in the first period and 4 in the second period for this infant.

Table 1Consistency with regression periods of the earlier study. The body of each infant's section presents the number of peaks which matched the regression measures. These measures represent the percentage of infants from the 1992 study that showed regressive behavior during each week. Because of missing data, the number of weeks for which it was possible to calculate R is different for each infant. This number is displayed in the heading of the second column. The column then presents the number of these weeks belonging to each of the separate measures. The numbers do not add up to the total number of weeks because some weeks do not belong to any of the used measures (i.e. weeks in which 21-59% of the infants of the 1992 study showed regressive behavior) and because the 80-100% weeks are of course included in the 60-100% measure.

regression measures	nr. R(NR) weeks in 49 obs. weeks	Infant J: regression criteria used + resulting number of peak-weeks				
		bc+cry/fr	bc	cry-dur	cry-freq	fretting
		2	7	7	7	7
0-20%	15	1	3 (1)	4 (1)	4 (1)	4 (1)
80-100%	11	0	2	2	2	1
60-100%	23	0	3	3	4	3

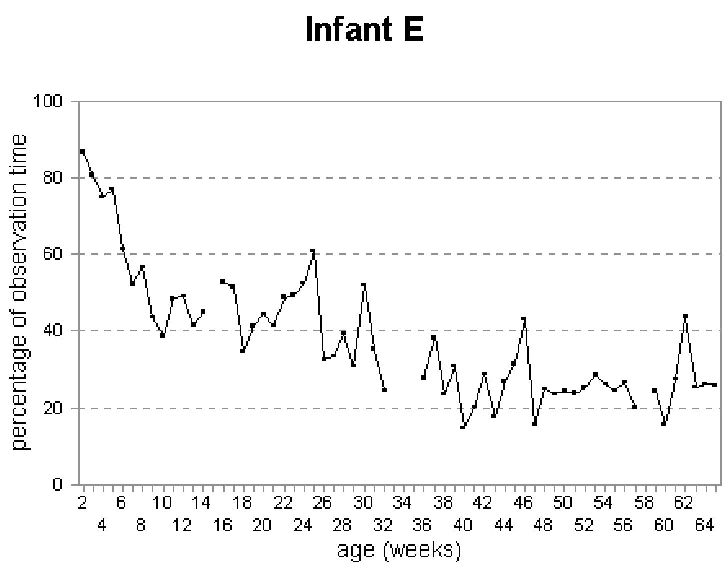
regression measures	nr. R(NR) weeks in 51 obs. weeks	Infant E: regression criteria used + resulting number of peak-weeks				
		bc+cry/fr	bc	cry-dur	cry-freq	fretting
		6	10	10	10	10
0-20%	12	0	1	3	3	4
80-100%	13	3	4	3 (1)	2 (1)	2
60-100%	24	6 **	8 *	7 (1)	6 (1)	5

regression measures	nr. R(NR) weeks in 58 obs. weeks	Infant S: regression criteria used + resulting number of peak-weeks				
		bc+cry/fr	bc	cry-dur	cry-freq	fretting
		5	10	10	10	10
0-20%	15	1	4 (1)	3	4	4
80-100%	14	2	2	4	3	2
60-100%	28	3	4	6	4	5

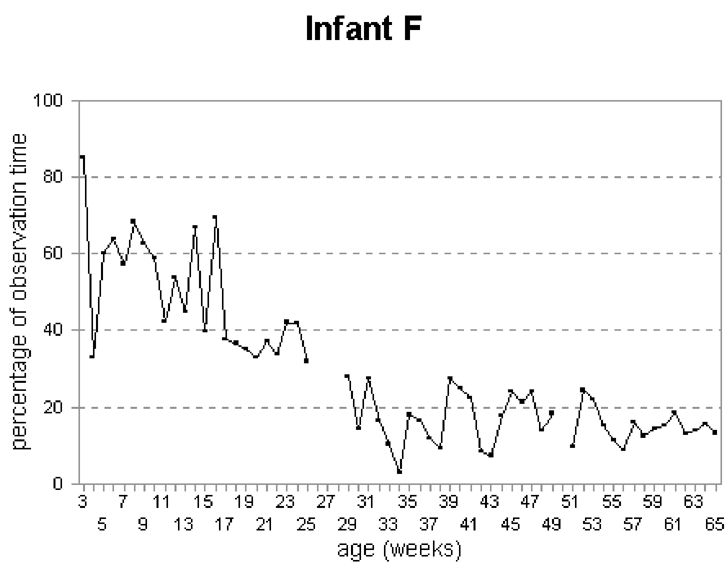
regression measures	nr. R(NR) weeks in 53 obs. weeks	Infant F: regression criteria used + resulting number of peak-weeks				
		bc+cry/fr	bc	cry-dur	cry-freq	fretting
		7	10	10	10	10
0-20%	15	2	2	4	4	4
80-100%	12	3	3	2	3	1
60-100%	25	4	6	3	4	2

Note: nr. R(NR) weeks = number of regression (or non-regression weeks); bc+cry/fr = peak in both body contact and crying (duration or frequency) and/or fretting/fussing; bc = peak in body contact; cry-dur = peak in crying (duration cry-freq = peak in crying (frequency); fretting = peak in fretting/fussing; ( ) = number of matching peak weeks in which the infant was unwell during the observation; * p<.05; ** p<.01 (Fisher's exact tes

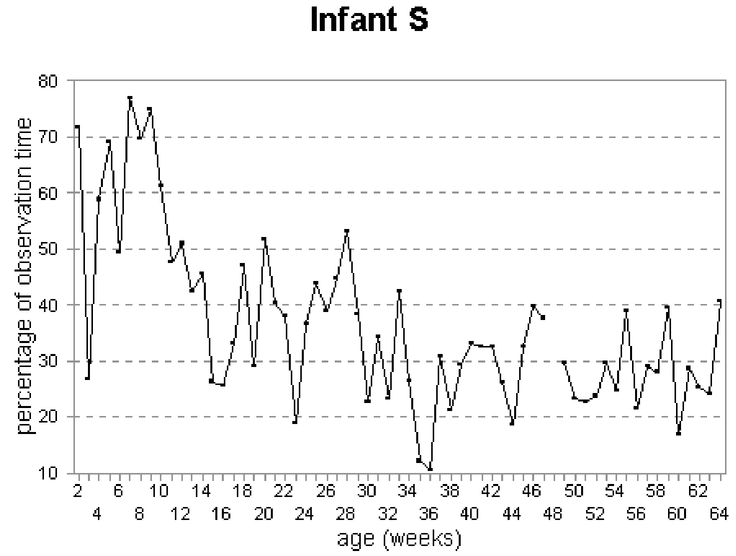
a)


Infant E
percentage of observation time
age (weeks)
100
80
60
40
20
0
2 6 10 14 18 22 26 30 34 38 42 46 50 54 58 62
4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64

b)


Infant F
percentage of observation time
age (weeks)
100
80
60
40
20
0
3 7 11 15 19 23 27 31 35 39 43 47 51 55 59 63
5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65

c)



d)

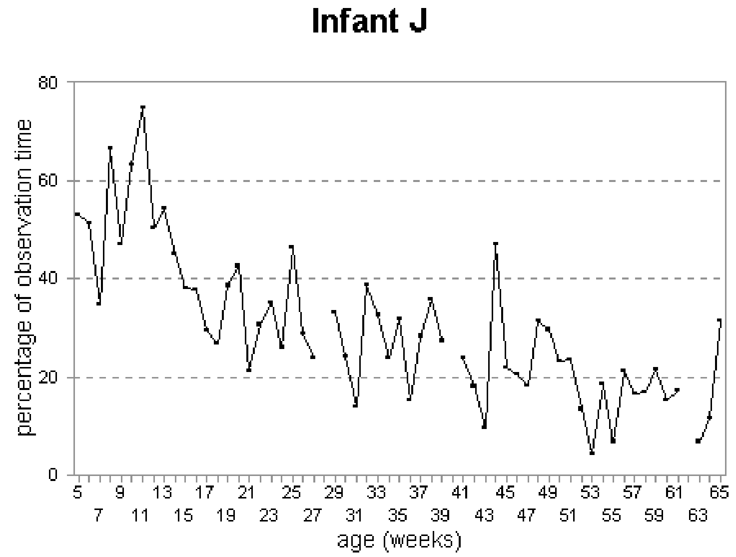
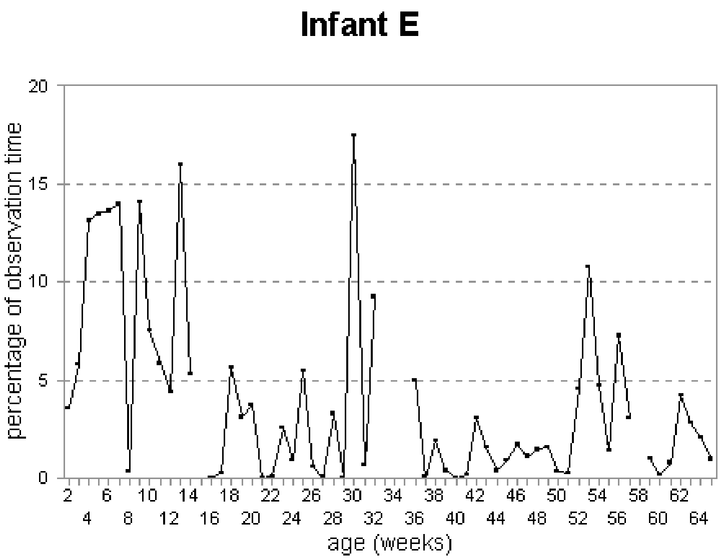
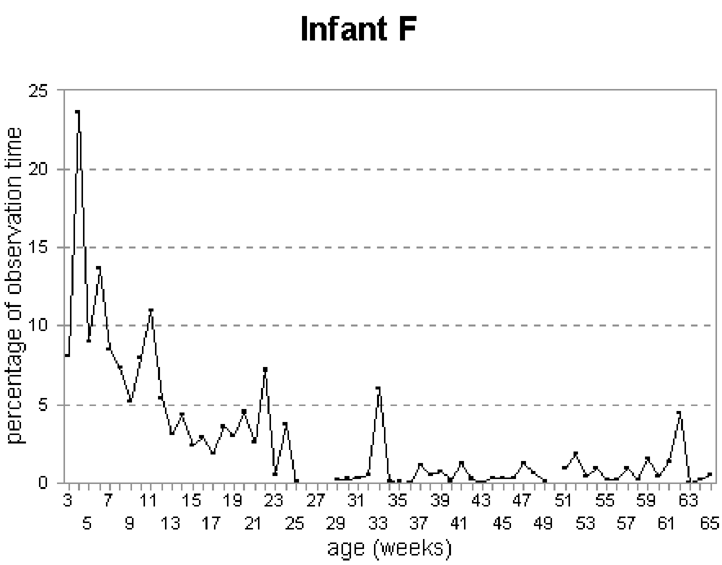


Figure 2a-d. Body contact: percentage of the observation time mother and infant spent in body contact over infant's age in weeks. Missing points on the curves indicate the observations did not take place. Weeks in which the infant was unwell during the observation: infant E, weeks 30, 32 and 53; infant J, weeks 20 and 56; infant S, weeks 44, 47 and 59; infant F, week 65.

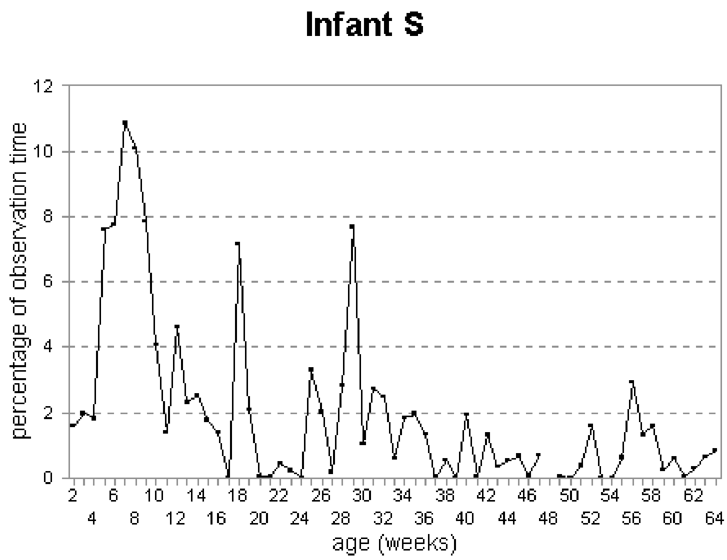
a)


Infant E
percentage of observation time
age (weeks)
0
5
10
15
20
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
56
58
60
62
64

b)


Infant F
percentage of observation time
age (weeks)
0
5
10
15
20
25
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
35
37
39
41
43
45
47
49
51
53
55
57
59
61
63
65

c)



d)

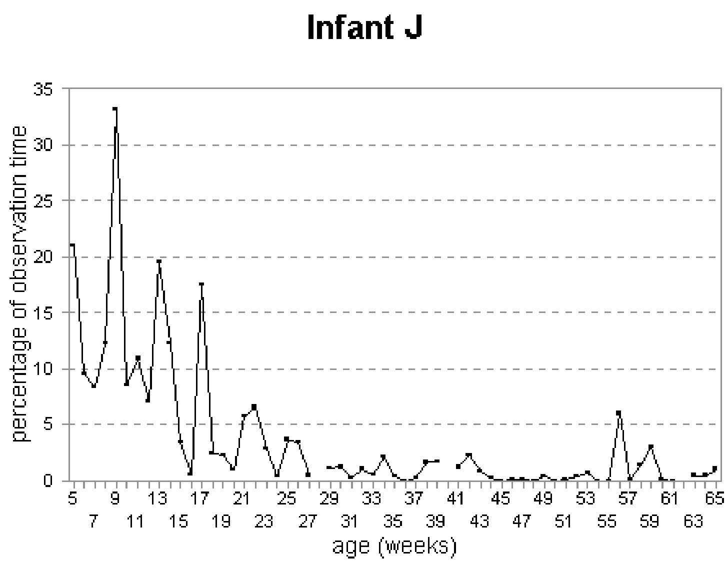
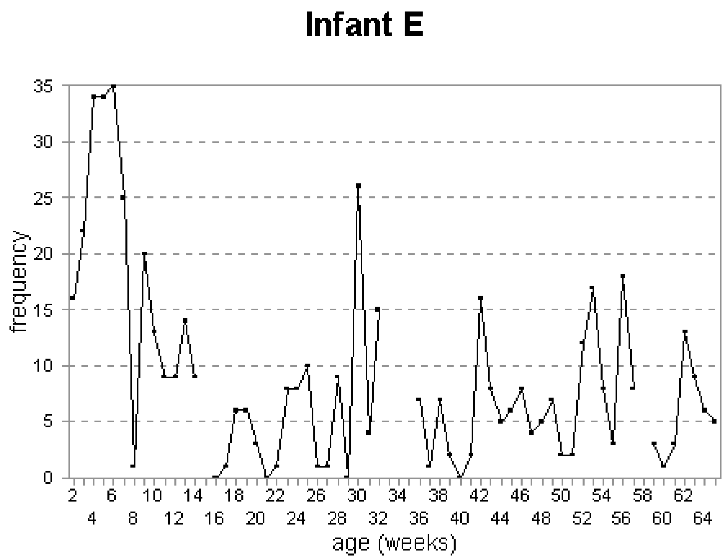
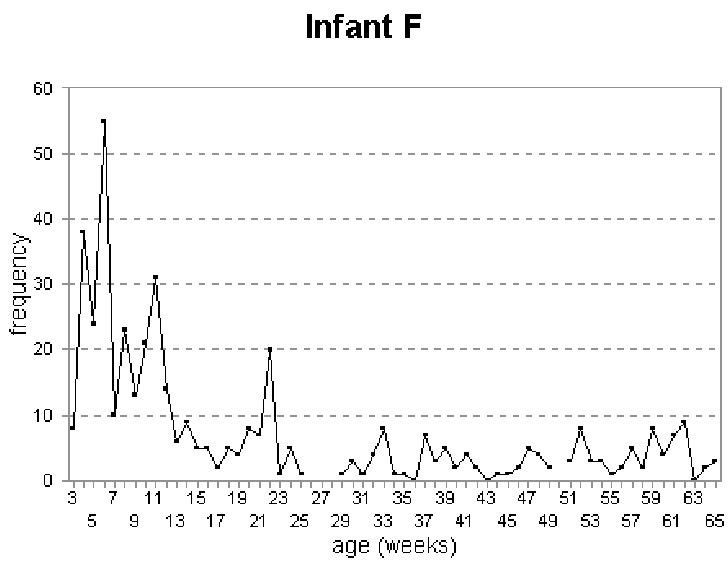


Figure 3a-d. Crying (duration): percentage of the observation time mother and infant spent in body contact over infant's age in weeks. Missing points on the curves indicate the observations did not take place. Weeks in which the infant was unwell during the observation: infant E, weeks 30, 32 and 53; infant J, weeks 20 and 56; infant S, weeks 44, 47 and 59; infant F, week 65.

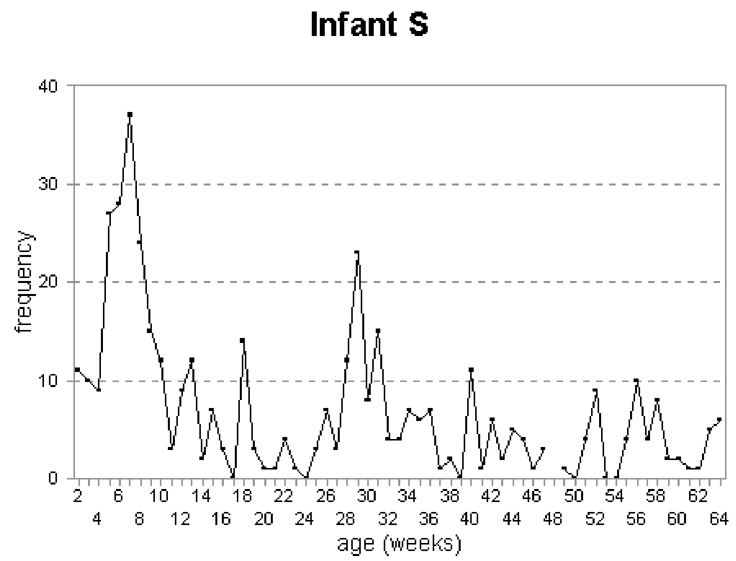
a)


Infant E
frequency
age (weeks)
35
30
25
20
15
10
5
0
2 6 10 14 18 22 26 30 34 38 42 46 50 54 58 62
4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64

b)


Infant F
frequency
age (weeks)
60
50
40
30
20
10
0
3 7 11 15 19 23 27 31 35 39 43 47 51 55 59 63
5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65

c)



d)

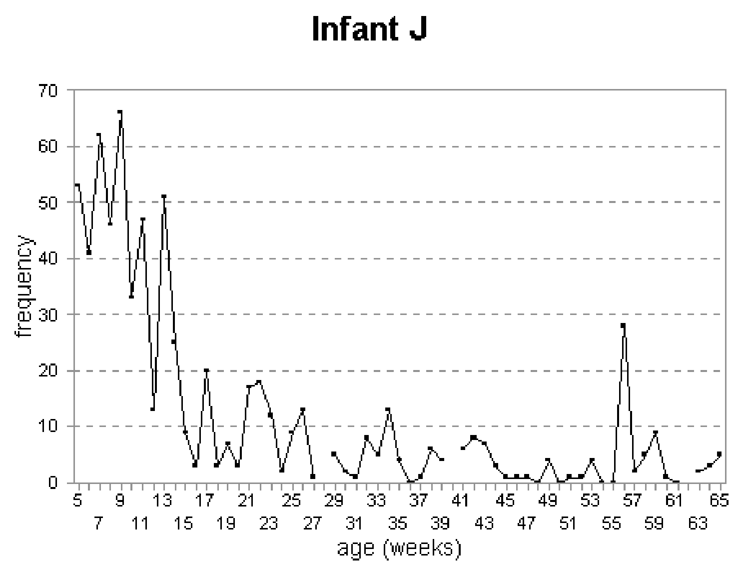
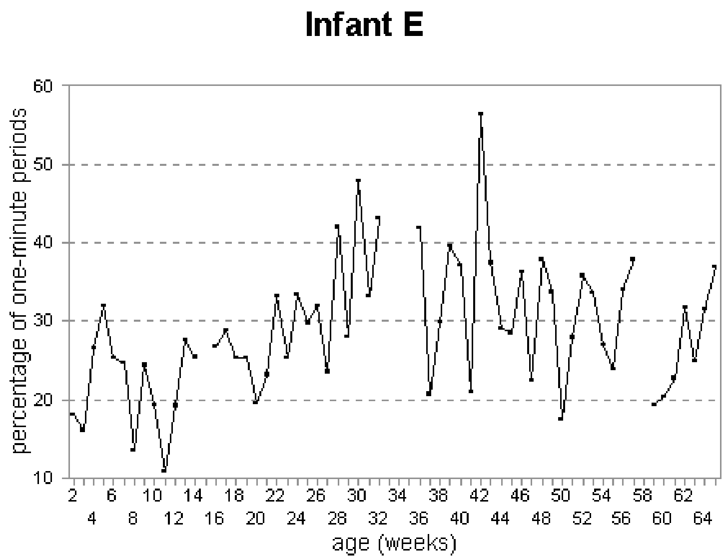
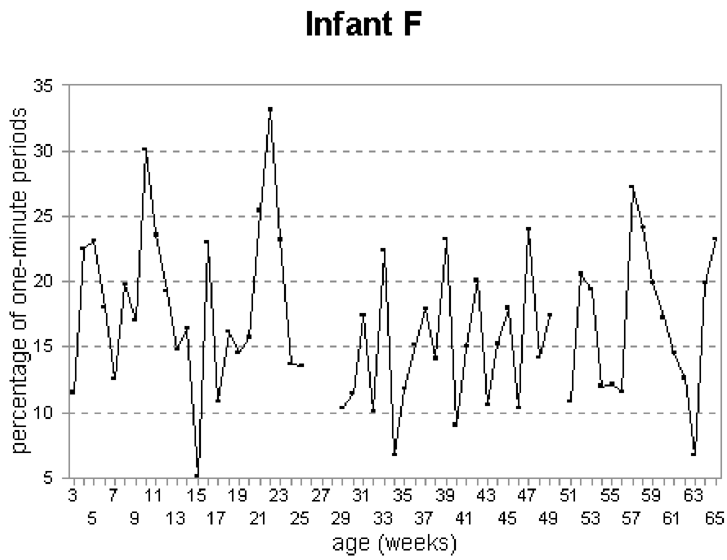


Figure 4a-d. Crying (frequency): absolute frequency of crying episodes during the observation period, over infant's age in weeks. Missing points on the curves indicate the observations did not take place. Weeks in which the infant was unwell during the observation: infant E, weeks 30, 32 and 53; infant J, weeks 20 and 56; infant S, weeks 44, 47 and 59; infant F, week 65.

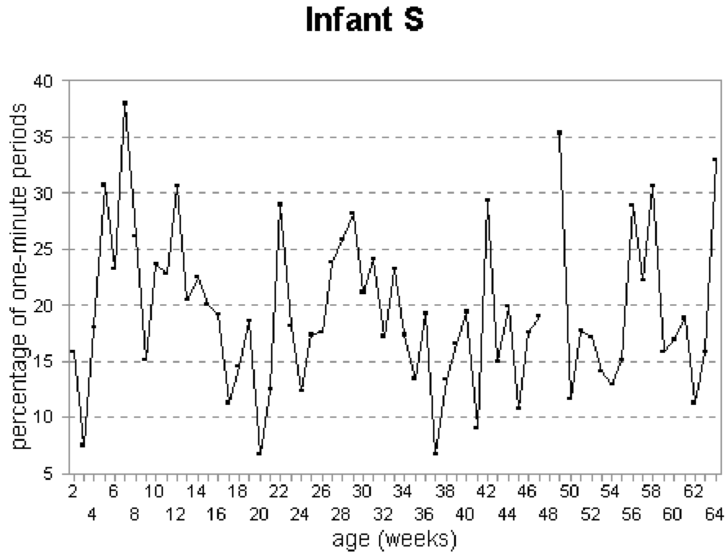
a)


Infant E
percentage of one-minute periods
age (weeks)

b)


Infant F
percentage of one-minute periods
age (weeks)

c)



d)

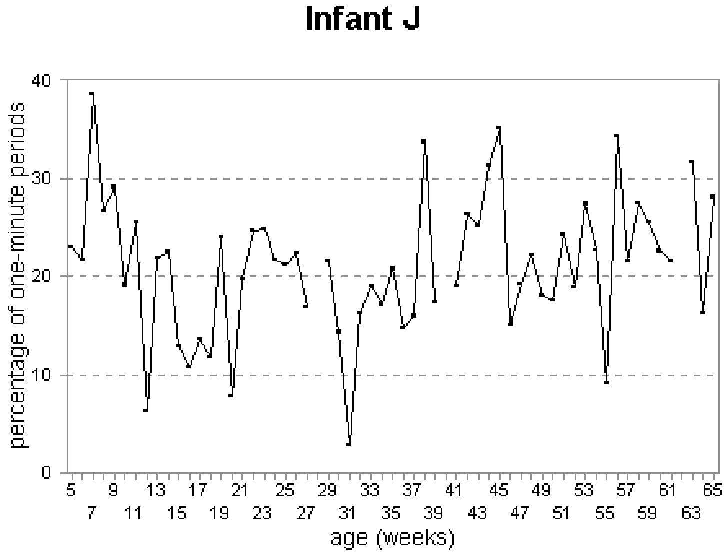


Figure 5a-d. Fretting/fussing: percentage of observed one-minute intervals in which the infant fretted/fussed over infant's age in weeks. Missing points on the curves indicate the observations did not take place. Weeks in which the infant was unwell during the observation: infant E, weeks 30, 32 and 53; infant J, weeks 20 and 56; infant S, weeks 44, 47 and 59; infant F, week 65.

The matches found between the regression weeks obtained for the four infants when using the different regression criteria, and the theoretically strong regression weeks, non-regression weeks and 60+ regression weeks (as derived from the study by Van de Rijt-Plooij & Plooij, 1992a), are presented separately for each subject in Table 1. In general, no single criterion resulted in more matches than the others. Significant results in the number of matches were found for infant E's peak-weeks as defined by the criteria body contact plus crying/fretting (Fisher Exact Test, $p<.01$; peaks in periods around weeks 5, 17, 26, 30, 44 and 62) and by body contact alone (Fisher Exact Test, $p<.05$; peaks in periods around weeks 5, 8, 17, 26, 30, 36, 44 and 62), for the regression weeks defined by the 60+ measure. However, the other 3 infants did not have peak-weeks falling significantly more often in these weeks. Furthermore, peaks were expected to fall in the strong regression weeks and not in the non-regression weeks. This could not be confirmed for any of the four infants. Moreover, a surprisingly high amount of peak-weeks fell in non-regression weeks.

Observational Data: Total series of observed behavior

The results of the correlations of the 'reported regression' variable and the three observed behaviors were disappointing. No significant positive correlations were found between the variable and body contact, crying (duration and frequency) and fretting/fussing. Moreover, the variable was negatively correlated to fretting/fussing in infants S ($r=-.21$, $p=.05$) and F ($r=-.29$, $p<.01$), and to frequency of crying in infant J ($r=-.23$, $p<.05$).

The correlations for the summed variables were still low, and also negative. For body contact: $r=-.06$, $p=.62$; for crying (duration): $r=-.12$, $p=.34$; for crying (frequency): $r=-.27$, $p<.05$, and for fretting/fussing: $r=-.26$, $p<.05$.

Observational Data: Age modifications

Although the above analyses were computed with the infants' gestational ages, as in the original paper by Van de Rijt-Plooij & Plooij (1992a), the poor results obtained gave rise to speculations over possible individual differences in the onset of the 10-regression period pattern. In order to investigate this alternative, the data series of all the infants were moved in four different ways with respect to the data of the Plooijs: 2 weeks back, 1 week back, 1 week forward and 2 weeks forward. Note that by modifying the age in these ways infant S's correction for gestational age disappeared (with 2 weeks forward). Infant J's age was also corrected by moving it 3 weeks back, so that also in her case the correction for gestational age disappeared. This means that for these two infants the data now also corresponded with their postnatal age, which eliminates the doubt of whether postnatal age might actually be more important in the phenomena studied than gestational age.

Both analyses presented above were thereafter performed on the modified data series. Only 4 significant positive results were found: 2 in the peak analysis (infant J, 2 wks. back, fretting/fussing, Fisher Exact Test, $p<.01$; infant F, 1 wk. forward, duration of crying, Fisher Exact Test, $p<.05$) and 2 in the correlations (infant E, 1 wk. back, duration of crying, $r=.23$, $p<.05$; infant S, 2 wks. back, body contact, $r=.26$, $p<.05$). Because the data was modified in diverse ways and all 4 variables were tested for each possibility, a total of 153 statistical tests had to be performed. In view of this number, the amount of significant results found is to be expected by chance at the 5% level, and therefore cannot be given any weight. Thus, the remaining analyses were performed without age modifications.

Observational Data: Positive residuals

The residuals obtained after detrending the data were analyzed in order to investigate whether a greater amount of positive residuals fell in 60+ regression weeks than in the rest of the observed weeks. This simple analysis gives less weight to peaks -outliers- and more to general increases in the observed behaviors, which were hypothesized to occur in these weeks. Chi-square and Fisher Exact Tests were used to test for significance. However, none of the four infants showed more positive residuals in 60+ than in non-60+ weeks.

Table 2 Differences in means of observational variables between 60+ and non-60+ weeks. a) all the differences are shown; b) only the differences greater than 1/4 (and 1/2) standard deviation are shown.

a)

INFANT	BEHAVIOR			
	BC	CRY-D	CRY-F	FRET
E	+	+	+	+
J	-	+	-	+
S	-	+	+	-
F	-	-	-	-
ALL	-	+	-	-

b)

INFANT	BEHAVIOR			
	BC	CRY-D	CRY-F	FRET
E	+			
J	-		-	
S	-	+		-
F		-	-	-*
ALL			-	

Note: All = sum of four infants' data; bc = body contact; cry-d = duration of crying; cry-f = frequency of crying; fret = fretting/fussing; + = the mean of the residuals is higher in the non-60+ weeks than in the 60+ weeks; * = differences which are greater than ½ standard deviation.

Observational Data: Means of residuals

The means of the residuals obtained after detrending the data were calculated for the 60+ regression weeks and for the rest of the observed weeks. The goal was to see whether the means were larger in the 60+ weeks as compared to the remaining weeks, which is the result to be expected were the hypothesis to be true. This analysis gives weight both to peaks -outliers- and to general increases in the observed behaviors. Table 2a shows the results obtained for the analysis of the four behavioral variables and the four infants. As can be seen in the table, there are approximately the same number of cases in which the mean of the 60+ weeks is larger than that of the remaining weeks (+'s), as cases in which it is smaller (-'s). In order to assess the importance of the difference between two means, the standard deviations of each behavioral variable were used. Differences which were greater than 1/2 and 1/4 standard deviation were calculated and are presented in Table 2b. The results indicate very small differences between the means of the 60+ regression weeks and the remaining weeks. Furthermore, 8 out of 10 of the differences which are greater than 1/4 standard deviation are of means of 60+ weeks which are *lower* than those of non-60+ weeks.

Observational Data: Checking for False negatives

One may argue that with only a single (albeit lengthy) observation a week, a fair chance exists of missing the regressive behavior. If behavior characteristic of regression occurs only once every *n* days, or, if regression weeks are those with a higher probability that regressive behavior will occur across the *whole* week, it is likely that an observation during a single day not always captures the true regressive character of a particular week. Since the preceding analyses showed that there is no evidence that regressive behavior occurs in the predicted weeks, we should check the probability that the results of our observation studies yield *false negatives* (which occur if the observation data result in the qualification "non-regression week" during a true regression week).

What is the probability that 64 consecutive observation weeks yield a pattern of observed variables in which the average residual scores of the predicted regression weeks are not bigger than, or eventually smaller, than those of the non-regressive weeks? The answer to this question is not trivial, since there exists no fixed criterion for what distinguishes a regression from a non-regression week in terms of the amount and distribution of the regressive behaviors. We decided to solve this problem in two different ways, based on two different ways in which regressive behavior may occur.

To begin with, it is reasonable to assume that both during the regression and the non-regression weeks, the distribution of regressive behaviors per unit time shows a normal distribution (the residuals of all empirically observed variables were normally distributed; Chi^2 test with 95% confidence range). This means that if we observe for a full three hours during one day in the week, our observations are like a randomly drawn, single sample from a normal distribution, the value of which can lie above or below that week's average. As far as the occurrence of regressive behaviors is concerned, there exist two possibilities.

According to the first, it is reasonable to assume that the averages for regressive behaviors are higher during regression weeks than non-regression weeks. The difference can be expressed in terms of percentages standard deviation difference between the averages of the regression and non-regression distributions. We assume that this difference must be big enough for a mother to notice it in a consistent way, given the fact that the predicted regression weeks from the Van de Rijt-Plooij and Plooij study were based on maternal reports. We took a ½ standard deviation of the residuals as a conservative estimation of the required difference. By means of a Monte Carlo simulation, we determined the probability distribution of observation results. We tested the probability that the observed regressive behavior levels of a set of 17 non-regression weeks were in fact ½ standard deviation smaller than those of 29 regression weeks, given the averages and standard deviations of the actual observations. This was done by drawing random observation samples from two normal distributions of scores, one representing the regression weeks, the other the non-regression weeks, with averages that differed a ½ standard deviation from one another. In all cases, the probability that such difference was mistakenly overlooked was 5% or smaller (for simulations based upon 500 cases for each infant and each regression variable). We can thus reject the hypothesis that our data missed the real differences that would have been found if we had observed every day instead of only one day of the week.

There exists also a second possibility, namely that regressive behaviors occur as spikes or surges of regressive behavior among an otherwise normal distribution of regressive behaviors. That is, a regressive week is characterized by behavioral variation that does not differ from the variation in the non-regression weeks, with one exception, namely that every now and then a temporal surge of regressive behaviors occurs in the regression weeks. Since every observation lasted for a long part of a day (recall that an observation consisted of an additive *total* of three hours of behavior) we found it reasonable to assume that the hypothetical regression "spikes" occur once in every three periods of time covered by a single observation. We reasoned that, for an infant to be found really "difficult" by his or her mother, those difficult, i.e. regressive, behaviors must occur relatively frequently (the frequency of 1 out of 3 observation periods is just an initial guess). A "spike" of regressive behavior was defined as one in which the regressive behaviors were 1 standard deviation higher than the average. The technique was applied on the distributions of the behavioral variables of each infant. For each infant, each variable was checked for the probability that the observed results indeed indicated a difference between non-regression weeks and "spiked" regression weeks. The check was based on 500 simulated distributions of weekly observations for each case. The results are as follows: the probability that the observed results are indicative of the occurrence of regression weeks are for infant F: 0.125, 0.004, 0.007, and 0.000 for total contact, cry duration, cry frequency and fretting respectively; for infant S: 0.01, 0.611, 0.269 and 0.003; for infant J: 0.01, 0.205, 0.003 and 0.165; for infant E: 0.55, 0.229, 0.11 and 0.266. Although some of those probabilities are high, the distribution of results does not differ from what could be expected on the basis of chance alone. Given that Van de Rijt-Plooij and Plooij (1992a) defined regressions as weeks in which increased crying *and* increased contact occur, it can be concluded that our results do not support the regression weeks hypothesis, with regression weeks defined as those in which spikes of regressive behavior occur.

In summary, we may reject the hypothesis that our observations, made once a week, have missed the 10-regression period pattern, both if regressions occur as general increases of regressive behavior or as infrequent spikes.

Maternal report

The 'concrete' and the 'general' measures based on the maternal report of regressive behavior were strongly correlated within individuals for infants E, S and F (Fisher Exact Test, $p=.000$), but not within infant J (Fisher Exact Test, $p=.30$).

The results of the analysis of the consistency of the maternal report with the 10-regression period pattern of Van de Rijt-Plooij & Plooij (1992a) did not give support to this pattern of regressions. The only significant relationship between the regression weeks reported by the mothers (both concrete and general measures) and the 60+ regression weeks showed a negative significance for the concrete regressions for infant E (Chi-square=5.01, $p<.05$).

Relationships among the data of the four infants

Observational Data: Correlations between behavioral variables (detrended data)

Not surprisingly, the amount of fretting/fussing of all four infants was positively correlated to the time spent crying and, even more strongly, to the frequency of crying. Pearson's r values for the duration of crying were: .50 for infant E (p<.001), .25 for infant J (p<.05), .45 for infant S (p<.001) and .32 for infant F (p<.01), and for the frequency of crying: .64 for infant E (p<.001), .55 for infant J (p<.001), .51 for infant S (p<.001) and .38 for infant F (p=.001).

Based on the findings of Van de Rijt-Plooij & Plooij (1992a), an infant's negative vocalizations were expected to be positively correlated to the time spent in body contact. However, in the case of duration of crying this was only true for infant S (r=.39, p=.001). Moreover, infant F showed a negative correlation between the two variables (r=-.46, p<.001). The frequency of crying was positively correlated to body contact for infants S (r=.32, p<.01) and E (r=.22, p<.05), and negatively for infant F (r=-.24, p<.05).

Fretting/fussing was found to be positively correlated to body contact for infant S (r=.30, p<.01) and infant J (r=.21, p<.05), but not for the remaining two infants. In general, it is possible to say that although the correlations were not as strong as expected, 3 of the 4 infants showed a positive correlation between body contact and a negative vocalization.

Observational Data: Correlations between infants

The time spent crying was found to be significantly correlated between infants E-S (r=.24, p<.05) and E-J (r=.24, p<.05). The frequency of crying was significantly correlated between more infants: E-S (r=.31, p<.01), E-J (r=.36, p<.01), E-F (r=.28, p=.01) and J-S (r=.32, p<.01). Fretting/fussing was found to be significantly correlated between E-S (r=.36, p<.01) and S-F (r=.21, p=.05). Body contact was not significantly correlated between any pair of infants. Based on these results, it is possible to say that the negative vocalizations of the infants are apparently related to a certain degree, while their time spent in body contact is not.

Maternal reports

The mothers of the four infants of this study did not report their infants going through regressive periods in the same weeks. Thus, no significantly matching patterns of regression behavior were found (Chi-square, Fisher Exact tests). This was true for both the concrete and the general measures of regression.

Observational data with maternal reports

Several of the approaches used to determine the consistency of the observational data with the 10-regression period pattern, were also used to investigate how well the weekly observations matched the weekly maternal reports (for methodological information refer to Methods section). Each week on which the mother reported began at the morning of one observation and ended at the following observation. Therefore, the maternal reports were matched with the observations of the beginning of the week (begin measure), and also with the observations at the end of the week (end measure), which coincided with the filling of the questionnaire.

Observational data with maternal reports: Peaks in the observed behavior

The peaks obtained by means of the peak method were matched against the regression weeks reported by the mother. Only infant F’s peak-weeks as defined by the criteria body contact plus crying/fretting (Fisher Exact Test, p=.01) and by body contact alone (Fisher Exact Test, p=.001) matched the regression weeks reported by the mother (general measure). These results were obtained by matching the observations of the beginning of the week with the maternal reports. The remaining infants showed no significant relationships, although both infants E and S's matches with the end of the week displayed a tendency for the weeks in which the mother reported regression to have peaks in fretting/fussing (Fisher Exact Tests, p=.07 -gen. and conc. measures- and p=.06 -gen. measure-, respectively).

Observational data with maternal reports: Positive residuals

The residuals obtained after detrending the observational data were analyzed in order to find out whether a greater amount of positive residuals fell in maternally reported regression weeks than in the rest of the observed weeks. However, this was not found to be so for any of the four infants. Moreover, infants S and F's matches with the end of the week showed a significantly greater amount of positive residuals of one regression criterion falling in weeks not reported by the mother as regression weeks. For infant S this was so for body contact (Chi-square=7.88, p=.00, gen. measure, Chi-square=10.69, p=.00, conc. measure) and for F it was true for duration of crying (Chi-square=2.67, p=.05, gen. measure).

Observational data with maternal reports: Means of residuals

The means of the residuals obtained after detrending the data were calculated for the maternally reported regression weeks and for the rest of the observed weeks. This analysis rendered the best matches between the observational and maternal report data. Table 3a shows the results obtained for the four behavioral variables and the four infants, of the measure - general or concrete - which gave the best matches. As can be seen in the table, the means of virtually all the behaviors for all the infants were higher for the weeks in which the mother reported regressive behavior as compared to those of the remaining weeks. The differences between the means which were greater than 1/2 and 1/4 standard deviation were calculated and are presented in Table 3b. The results indicate that for 3 of the infants these differences were relatively large. Thus, it is possible to conclude that in general the maternal report could be confirmed by ethological observations.

Table 3 Differences in means of observational variables between maternally reported regression and non-regression weeks; a) all the differences are shown; b) only the differences greater than 1/4 (and 1/2) standard deviation are shown.

a)

INFANT	BEHAVIOR			
	BC	CRY-D	CRY-F	FRET
E-conc	+	+	+	+
J-conc	+	+	+	+
S-gen	-	-	+	+
F-gen	+	+	+	+

b)

INFANT	BEHAVIOR			
	BC	CRY-D	CRY-F	FRET
E-conc	+		+	+
J-conc	+*		+*	+
S-gen	-			
F-gen	+		+	

Note: In the results presented, for infants E, J and S the observation for the end of the week was matched with the maternal report, while for infant F it was that of the beginning of the week. Conc = concrete measure of regression; gen = general measure of regression; bc = body contact; cry-d = duration of crying; cry-f = frequency of crying; fret = fretting/fussing; + = the mean of the residuals is higher in the maternally reported regression weeks than in the remaining weeks; - = the mean of the residuals is higher in the remaining weeks than in the maternally reported regression weeks; * = differences which are greater than ½ standard deviation.

DISCUSSION

The data of the four infants of this study failed to replicate the findings of Van de Rijt-Plooij & Plooij (1992a). This was true both for the data obtained through ethological home observations as well as for the data obtained through maternal reports. No conclusive evidence was found (that is, evidence above mere chance level) for the model of ten regressions during infancy described in the study by Van de Rijt-Plooij & Plooij (1992a).

These results immediately bring up the question of whether either the selected population, the data, and/or the analysis methods are the factors causing this discrepancy. Taking into account the strict selection procedure of the mother/infant pairs and the further normal development of the infants during the study period, the differences in results are most unlikely to lie with the subjects which participated in the study[2]. The *type* of data collected closely replicated the data obtained in the 1992 study by Van de Rijt-Plooij and Plooij. The

[2] See *Acta Neuropsychiatrica* (no. 3, 1998) for a lengthier discussion on the topic.

quality of the data collection was controlled through inter- and intra-observer reliability checks, both for the ethological observations and for the coding of the maternal questionnaires. It is therefore also extremely unlikely that the differences in results lie in the data *per se.* The remaining possibility is the methodological aspect of the data analysis and as will be explained in detail in the following paragraphs, this point is a weak contender for being the major agent in the failure to replicate the 10-regression period pattern.

The consistency of the data obtained through weekly home observations and the pattern of regressions found by Van de Rijt-Plooij and Plooij (1992a) was analyzed in four different ways. We are of the opinion that the combination of these methods thoroughly covers the possible manifestations of regressive behavior as are to be expected from the article of the Plooijs. First of all, the *peak method* gives more weight to sharp, one-week long peaks than to longer periods in which the values are elevated (i.e. a three-week long 'mountain'). This could have resulted in longer regression periods being underrepresented. However, the study of *positive residuals* underlined increases in regression behavior without giving importance to the magnitude of the increases, and the study of the *means of the residuals* emphasized both subtle and large increases in regression behavior. Both these methods should detect the 10-regression period pattern if instead of being characterized solely by peaks it were characterized by 'mountains' or even by general subtle increases, and also if it consisted of a mixture of 'mountains' and peaks. Finally, the *correlations* between the observed variables and the population regression curve of the Plooij study, served as a means of investigating whether the patterns found followed those of the earlier study in a general way. In this case, even if the infants 'missed' a couple of regression periods, high correlations should have been found between the variables of the two groups. This method therefore controlled for the weaker hypothesis that less than 10 of the hypothesized regression periods took place.

In the light of the disappointing results obtained, one might ask oneself whether ethological observations constitute adequate means of replicating the phenomenon reported by Van de Rijt-Plooij & Plooij (1992a). The main reason for believing that they do is that, as stated in the Introduction, the Plooijs themselves used this method, found that it corroborated their maternal reports, and recommended it be used in a weekly fashion in future research on the topic. But aside from this point, what are the actual chances of missing a regression period with observations that took place on only one day of the week, began at 9.00 and went on till any time between 12.00-17.00 (depending on whether the infant slept or not)? One may object that the continuous, spontaneous observations that a mother makes of her own infant are far superior to a (lengthy) observation made only once a week. If this objection holds, ethological observation data are not suited for testing the hypothesis of regression weeks, the signs of which may easily have been missed in the observations. However, the present study failed to replicate the 10-regression period model not only on the basis of ethological observations, but also on the basis of the mothers' reports. It is important to note that the observations are consistent with the mothers' reports if the results of the *means of residuals* analysis are taken into account. This finding suggests that mothers base their impression of regressive behavior and "difficult" weeks upon intuitive estimations of the deviation of *week averages* of regressive behavior as compared with the average of such behaviors across longer periods. A check for the probability of false negatives, based on a simulation, showed that it is highly unlikely (at and below 5% chance level) that our observations have underestimated the differences between the summed scores of the regression and non-regression weeks. We also checked for the possibility that our observations could have missed regressive behaviors that occur in short spikes or surges. We found that if regressions indeed occur in this way, the chance of missing them in single day observations is sometimes considerable. However, we also found that, on the whole, the distribution of those probabilities is not different from what can be expected on the basis of chance alone. In addition, since it is assumed that regression weeks are characterized by a combination of regression indicators, the evidence for the pattern of regression and non-regression weeks becomes (very) weak. Finally, one may also object that correlations between observed data and the predicted regression weeks are likely to strongly underestimate the similarity between regressions in our population on the one hand and the original population in the Plooij study on the other hand. The reason for this objection is that our observations most likely suffer from naturally expected variations in the expression of regressive behaviors. That is, it is likely that our observations give a distorted image of the regressive behaviors that occurred during the week in question. Correlations, however, are very robust indicators of similarity between variables that show high levels of error variance. If one correlates a variable with a strongly distorted copy of itself (the distortion being either plus or minus the standard deviation of the series randomly applied to each value in the correlated series), the Pearson r between series of 30 values each is still .70 on average, which is far above the correlations found in this study.

But for critics who still doubt the value of the observations, perhaps the most important point is that for 3 of the 4 infants of the present study, the maternal reports could be corroborated through the ethological observations. This important fact shows that the relatively long weekly observation period was fairly representative of the infant's behavior that week, as measured by the mothers' reports. Nonetheless, the consistency between the two data types was not perfect and did not occur in one infant. The observational data could of course be subject to a certain amount of chance variation, but they are objective by definition. The situation is different for information that is obtained from a caregiver, which is subject to problems of recollection and subjectivity (Bates & Bayles, 1984; Bernard et al., 1984; Quandt, 1987; St James-Roberts & Wolke, 1984). For example, Vitzthum (1994) studied the frequency and duration of breast feeding and compared daily maternal reports with observational data. Due to lack of accuracy of the mothers, she found virtually no agreement between the two types of data. The author observes that in the light of her results, the conclusions of studies based on mothers' recall should be seriously questioned and re-examined. Given that in the present study mothers were asked to recall a much greater constellation of behaviors and on a weekly basis, the reliability of their reports cannot be directly assumed. This being said, it still does not explain why, reliably or unreliably, the mothers of the Plooij study reported regression behaviors in the *same* weeks.

The analyses discussed above rendered only a few significant relationships between the observational data and the Plooij data. Apart from the fact that both positive *and* negative relationships were found between the two data sets, due to the large number of analyses which were carried out, it is necessary to conclude that the number of significant relationships did not surpass the amount to be expected by chance. In this context it is also important to note that these results were obtained even though the criteria we used to define regressions were much more flexible than those used by Van de Rijt-Plooij and Plooij (1992a): namely, besides looking for regressions in increases/peaks of both body contact *and* crying/fretting/fussing, we also looked for regressions in these behaviors individually. However, this broadening of criteria did not result in a greater consistency with the 10-regression period pattern.

Another attempt at increasing the consistency between the observational data and the regression period pattern of the Plooijs, was made with the age modifications. In this case, the data of the infants were analyzed again after modifying their ages in four different ways (from 2 weeks backwards to 2 weeks forwards). Although the Plooijs' theory consists of regression periods which are very strictly timed, we took the liberty of assuming that individual infants might begin going through their series of regressions at slightly different ages. The point of whether using a time window of 5 possible starting weeks constitutes a grave breach with the Plooij theory could be further pursued, were it not that the poor results obtained through these age modifications make this discussion futile.

Perhaps the most surprising result obtained in this study is that the mothers failed to report a pattern of regression periods in their infants which supported the 10-regression period pattern found by Van de Rijt-Plooij & Plooij (1992a). The mothers *did* report regression weeks (between 8 and 21 'concrete regression' weeks depending on the mother), but they did not follow the pattern proposed by the Plooijs. Moreover, whereas in the Plooij study 29 weeks could be found in which 60-100% of the mothers reported regression behavior, in the present study no significant agreement could be found between the reports of the four mothers. This points in the direction of individuality in patterns of maternally reported regression as compared to the uniformity of the maternal reports of the Plooij study.

Another interesting point of the results is that the correlations between the time spent in body contact with the mother and the infant's negative vocalizations, were not as strong (r's between .21 and .39) and as general as expected. This implies that no general maternal strategy could be determined in which the mothers consistently responded to the infant's increases in crying or fretting/fussing by increasing the time spent in contact. This finding differs from the questionnaire data of the mothers of the Van de Rijt-Plooij & Plooij (1992a) study. The data from the present study's naturalistic observations point at a greater variability in behavior than could be expected from the 1992 study; namely, there are apparently important differences in response between mothers and/or differences in response within mothers over situations.

Finally, the infants of our study went through periods of more crying and fretting/fussing, and of being more in body contact with the mother. This was found both through the observations and the maternal reports. Did these four infants follow an alternative common pattern of 'n' regressions, i.e. did they all go through a fixed number of regression periods, which were not necessarily ten and not necessarily in the hypothesized weeks? If this had been the case, strong positive correlations would have been found between the infants' detrended regression behaviors. Although body contact was not correlated between the infants, a number of positive correlations which included all the infants were indeed found for crying (frequency and duration) and fretting/fussing. Therefore, these negative vocalizations apparently followed their own pattern of 'ups and downs' in these four infants. Future analyses will help reveal the importance of this finding, by comparing the patterns found for the group and for the individual infants with the data of other researchers that have reported major transitions in early infancy (see Introduction). Furthermore, both movements and sounds made by the infant during the time he/she spent in bed were automatically recorded. These data should give more insight into the relationship between the developing infants' sleeping patterns and periods of regression and transition.

CONCLUSION

No support has been found for the model of ten regression periods described by Van de Rijt-Plooij & Plooij (1992a), notwithstanding the fact that the original criteria were at points significantly weakened (e.g. by shifting ages of onset, by relaxing the requirement that all or at least several indicators of regression should coincide, and so forth). This finding does not imply that regressions and transitions do not occur during infancy. It does suggest, however, first, that such regressions and transitions are subject to much more individual variation than the Plooij data indicate, and second, that the number of universally occurring regression periods (if any such exist) is considerably smaller than ten. Further analyses of our data will focus on the structure and dynamics of the observed behaviors and on their relationship with newly emerging skills. We believe that the search for universal stages of regression and transition, especially when as many as ten are assumed, is not only futile in light of the present detailed observational time series, but also that such endeavor distracts our attention from the truly interesting dynamic aspects of infant development. By their very nature, those aspects show their underlying similarity and coherence in the form of characteristic variations and dynamic patterns, both within and between individuals.

ACKNOWLEDGMENTS

This research was funded by the Netherlands Organization for Scientific Research (NWO), grant number 575-63-078. We wish to thank Prof. Dr. B.C.L. Touwen for performing the neurological examinations of the infants and Prof. Dr. W. Molenaar for his important contributions to the development of the peak method. And above all, we wish to thank the families that participated in the study.