

Indicators of Discontinuous Change in the Development of Analogical Reasoning

BETTINA HOSENFELD

Leiden University, Leiden, The Netherlands

AND

HAN L. J. VAN DER MAAS AND DYMPHNA C. VAN DEN BOOM

University of Amsterdam, Amsterdam, The Netherlands

In a longitudinal study, five indicators of a transition in the development of analogical reasoning were examined in young elementary school children, (a) bimodality and (b) inaccessibility in the frequency distributions of test performance, and in the responses of the transitional subjects, respectively, (c) sudden jumps, (d) anomalous variance, and (e) critical slowing down. An open-ended geometric-analogies test was administered eight times during a period of six months to eighty children in Grades 1 and 2 (six- to eight-year-olds). Strong evidence for bimodality was found in the distribution of the test scores and weaker evidence for inaccessibility. In the performance curves of the transitional subjects sudden jumps were demonstrated. Furthermore, the transitional subjects displayed a temporary increase of inconsistent solution behavior and solution time near the sudden jump. The characteristic changes in the analogy performance of the transitional subjects were interpreted as a strategy shift.

© 1997 Academic Press

Analogical reasoning, that is, establishing a correspondence between two or more sets of relations, has been widely acknowledged to be an important skill for learning from instruction (Vosniadou & Ortony, 1989). The majority of researchers (see Goswami, 1991, for a review) suppose that the ability to reason by analogy is already present in 3-year-old children. Change in analogy

We gratefully acknowledge the assistance of Paulien Brouwer, Alice van Dijk, Nienke Graatsma, Monique Harperink, Sharon Hené, Kitty van Houten, Angela Janssens, Mathilde de Jong, Suzanne Spaans, and Liesbeth de Tombe in data collection and of Max Goosen in scoring the data. We also thank Hoben Thomas and two anonymous reviewers for their thoughtful suggestions. Address correspondence and reprint requests to: Bettina Hosenfeld, Department of Developmental and Educational Psychology, Leiden University, P. O. Box 9555, 2300 RB Leiden, The Netherlands.

performance during childhood is generally assumed to be continuous and quantitative and is explained by growing domain knowledge (Gentner, 1977; Vosniadou, 1989) and increasing metaconceptual skills (Brown, 1989).

Only a few authors reported evidence of qualitative change in analogical reasoning during childhood. Within an information-processing approach Sternberg and Rifkin (1979) compared the solution processes of poor and successful analogical reasoners and found that the solution process of poor analogical reasoners lacked the mapping component, that is, the part of the solution process in which the equivalence of two relations is considered. This result was supported by the findings of Goldman, Pellegrino, Parseghian, and Sallis (1982) who reported that young and low-performing analogical reasoners had difficulties verbalizing the parallel relations in analogy problems and, moreover, were seriously distracted by associative alternatives. In a similar way, Alexander, Willson, White, and Fuqua (1987) found that the erroneous performance of nonanalogical reasoners was guided by a hierarchy of matching-by-similarity rules. Hence, developmental change in analogical reasoning during childhood might be described as a genuine shift from using free associations to considering the constraint of parallel relations.

This shift in analogical reasoning seems to match the shift from unidimensional to multidimensional thinking in 5- to 7-year-old children, as described by Siegler (1994) for several Piagetian tasks. When given a multidimensional task, younger children tend to focus on only one dimension, while older children take into account the multidimensionality of the task and, therefore, have a better chance to succeed in task solution. Siegler (1994) suggested three mechanisms that might explain the shift from one- to multidimensional thinking: growing process capacity, growing encoding skills, and decreasing expectations of one-dimensionality.

The majority of the studies on analogical reasoning are cross-sectional in nature. From such studies no definitive conclusions about the issue of continuity versus discontinuity in the development of analogical reasoning can be drawn. One notable exception, a longitudinal training study by Alexander et al. (1989) revealed that 4- and 5-year-olds benefitted from explicit analogy training. The performance of some control group children, however, also improved during a six-month period. The control group children whose performance improved were significantly older than those control group children who remained poor performers. This result might provide indirect evidence for a discontinuity in the spontaneous development of analogical reasoning. The purpose of the longitudinal study reported in this article was to investigate whether the development of analogical reasoning is a discontinuous process.

The model that was used in the present study for uncovering discontinuous change in analogical reasoning was the model of a transition in catastrophe theory (Thom, 1975). Catastrophe theory is a mathematical theory by which phase shifts in dynamic systems, that is, systems changing over time, can be detected and modeled. Up to now, catastrophe theory has been mainly applied

to a diversity of discontinuity phenomena in physics, chemistry, and biology (Poston & Stewart, 1978). An everyday example of a system undergoing a phase shift is that of ice, which is heated to room temperature. During the melting process, two stable behavioral states of the water molecules occur, on the one hand ice, where the water molecules are bound together in a compact crystal pattern, and on the other hand liquid, where the water molecules move randomly. During the last couple of years, some applications of dynamic systems models have emerged in several disciplines of psychology, too. A survey of recent applications in social psychology, for example, was provided by Vallacher and Nowak (1994). Likewise, in developmental psychology, dynamic systems models have been applied to some issues of motor, cognitive, and language development (Thelen & Ulrich, 1991; van der Maas & Molenaar, 1992; van Geert, 1991). These models, however, do not represent completely new theories of the domains they are applied to, but provide a new view on developmental changes within these domains (Van der Maas, 1995). Likewise, the application of catastrophe theory to the domain of analogical reasoning, which is reported in this study, dealt with only the developmental aspects of analogical reasoning.

In catastrophe theory, a transition is defined as an abrupt change in a behavioral variable, which can be explained by continuously changing independent variables, the control variables. In the melting-ice example above, the state of the water molecules represents the behavioral variable, while the added energy represents the control variable. Although the temperature in the melting pot increases continuously, the state of the molecules changes abruptly. When discontinuous change is studied in psychological phenomena, often there is a lack of agreement about possible control variables. But even if the control variables of a particular type of behavior are unknown, transitions can be detected on the basis of their formal characteristics. Gilmore (1981) proposed eight necessary, mathematically defined indicators, the catastrophe flags, for the demonstration of a genuine transition. Van der Maas and Molenaar (1992) translated these eight indicators of a transition from mathematical theory into the domain of cognitive development and suggested to start any investigation of discontinuities with catastrophe detection.

In the present study, we examined those five catastrophe flags that can be identified in a plain longitudinal data set: bimodality, inaccessibility, sudden jump, anomalous variance, and critical slowing down. In the following paragraphs, we will illustrate each of these five flags and present our hypotheses with reference to a transition in the development of analogical reasoning.

Multimodality and Inaccessibility

In the melting-ice example above, at each measurement occasion the great majority of the water molecules can be found either to be moving randomly or to be bound in the crystal state. An intermediate state, with half of the molecules moving and the other half motionless, is rare and highly instable.

The fact that certain sets of behavioral values can possess different probabilities of occurring defines multimodality. Sets of values with high probabilities are called the modes, while sets of values with low probabilities are called inaccessible regions. Inaccessibility is accompanied by metastability, which means that little disturbance will already push behavior from the repelling set to one of the adjoining stable modes. The combination of multimodality and inaccessibility, that is, of attracting and repelling modes, leads to strong multimodality in the score distributions (Van der Maas & Molenaar, 1992).

Multimodality in performance data is given, if in a frequency distribution two or more latent classes of performers are found, which can be sufficiently discriminated by their probabilities to solve a set of tasks correctly, that is, by their success probabilities. Two distinct classes of performers, for example, with success probabilities near chance level and near one, respectively, were consistently demonstrated, when the performance distributions of several age groups solving horizontality and verticality tasks were analyzed (Thomas, 1989; Thomas & Lohaus, 1993; Thomas & Turner, 1991). As stated above, for the analogical reasoning performance of elementary school children the existence of two behavioral modes, free association and analogical reasoning, was assumed. Therefore, we expected to find two distinct classes, inaccurate performers, who mainly apply free association, and accurate performers, who generally apply analogical reasoning. Hence, we looked for strong bimodality in the frequency distribution of analogy performance.

In longitudinal data, the number of modes on a developmental scale together with the growth trajectories of the success probabilities and the proportions of the members of each class can reflect the underlying growth processes (Thomas & Lohaus, 1993). If, for example, unimodality is found for a certain skill on several measurement occasions, while the success probability increases steadily, a gradual growth process can be assumed. If, on the other hand, strong bimodality with stable success probabilities for both latent classes is found and if only the proportion of accurate performers increases over time, discrete-stage growth processes can be assumed. Those subjects who change class membership over time represent the transitional subjects. In a longitudinal study by van der Maas, Walma van der Molen and Molenaar (1993), evidence for stage-like growth processes was found. At each of eleven measurement occasions strong bimodality was detected in the frequency distribution of the conservation performance of 6- to 10-year-old children. Two performance classes, nonconservers with success probabilities near chance level and conservers with success probabilities near one were distinguished. Over a period of eight months, the proportion of the conservers increased steadily, while the success probabilities of each of the latent classes did not change.

In a previous study (Hosenfeld, van der Maas, & van den Boom, in press), we already found bimodality in analogy performance in cross-sectional data, but we needed to confirm this in a longitudinal inquiry. In the present study, again, two classes of children, accurate and inaccurate performers, were ex-

pected to produce a strong bimodal distribution of analogy performance at several measurement occasions. Because we assumed that the children of each class responded consistently to one of the two strategies, we expected that the two modes remained stable over time. Furthermore, because we assumed that the development of analogical reasoning is a discrete growth process, we expected that in the long run some of the inaccurate performers switch strategies, so that the proportion of accurate performers increases over time.

Sudden Jump

Sudden jumps in individual growth curves point to developmental spurts. Such spurts must take place during a relatively short age interval, but do not necessarily occur at a specific point on the chronological age scale. Therefore, the dimension of chronological age can be replaced by the items of a developmental scale, which provides a relatively continuous measure of the behavior under observation (Fischer, Pipp, & Bullock, 1984). The simultaneous occurrence of sudden jumps in several behavioral measures, as was demonstrated, for example, for infants assessed on a battery of sensorimotor tasks (Lewis & Ash, 1992), provides even stronger evidence of a genuine transition. In the present study, we expected the transitional subjects to display a rapid increase in analogy performance within a few weeks. This increase in the performance of the transitional subjects should be larger than any increase in the performance of the nontransitional subjects.

Anomalous Variance

Another indicator of a transition, anomalous variance, reflects the idea that structural change is linked with a conflict between different modes of behavior. Increased variance in cognitive performance near a transition points to the fact that a tightly organized knowledge base becomes loose and susceptible to random as well as to systematic influences. An example of the operationalization of anomalous variance was given by Graham and Perry (1993), who examined a transition in addition performance. Vagueness in verbal explanations and a discordance between gesture and speech discriminated successfully between low-performing and transitional individuals. In a similar way, van der Maas et al. (1993) studying conservation behavior demonstrated that the responses of transitional subjects were less consistent than the responses of conservers and nonconservers. If analogy performance exists in two qualitatively different modes, free association and analogical reasoning, wavering and alternately applying the two strategies might be characteristic of children during the transition. Therefore, we expected the transitional subjects to display more inconsistency in their analogy solutions near the transition than before and after the transition.

Critical Slowing Down

This indicator of a transition refers to the fact that stable systems recover quickly from perturbations, whereas systems in transition require a longer

period of recovery. Each challenge to the system may be seen as a small perturbation. Therefore, each behavioral manifestation in a critical domain will last longer during the transition than before or after the transition, when the same behavior is more or less automatized. The results of a microgenetic study on addition strategies by Siegler and Jenkins (1989) constitutes an example of the phenomenon of critical slowing down. The response times of individual children were strikingly lengthened shortly before and on the trial on which a new addition strategy was applied for the first time. In our study, we also expected the transitional subjects to require more time for producing an analogy solution near the transition than before or after the transition.

Outline of the Study

In order to examine the catastrophe flags bimodality, inaccessibility, sudden jump, anomalous variance, and critical slowing down in analogical reasoning, we set up a longitudinal study covering the age range in which the transition was likely to occur (Hosenfeld, van der Maas, & van den Boom, in press). The study was a compromise between an inquiry with a sample large enough for group comparisons and a microgenetic study (Siegler & Crowley, 1991) with a high density of observations during a short period of transition. The sample consisted of 6- to 8-year-old elementary school children who were tested every three weeks over a period of six months. Fischer et al. (1984) demonstrated in a study on arithmetic skills that the provision of familiar material, the opportunity to practice, and environmental support induced clear performance spurts, whereas ordinary testing was related to a slow gradual increase in performance. Hence, discontinuities need not occur under all environmental conditions, but seem to be most likely under conditions encouraging optimal performance. To measure analogical reasoning without demanding extended domain knowledge or vocabulary, we administered geometric analogy problems consisting of simple and well-known geometric shapes (Hosenfeld, van den Boom, & Resing, in press). Furthermore, we tried to create optimal circumstances by testing the subjects individually, by providing practice items, and by repeating the instruction several times during each test session.

METHOD

Subjects

Eighty children from ten elementary schools in the Netherlands participated in the study. In every school, eight children, two girls and two boys from Grades 1 and 2, respectively, were selected. In order to have a sample with a uniform age distribution and equal numbers of boys and girls, we divided each grade into four subgroups, younger and older girls, and younger and older boys. Then from each subgroup one child was randomly selected. Children older than their regular age group were not included in the sample. In this way, our sample comprised four cohorts, 6- and 6.5-year-olds in Grade 1 and 7- and 7.5-year-olds in Grade 2.

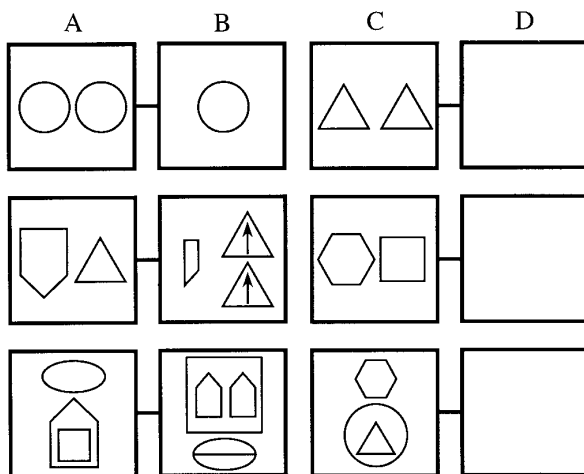


FIG. 1. Examples of open-ended geometric analogy items of (1) low, (2) intermediate, and (3) high level of difficulty. The D-term has to be filled in by the subjects.

The sample was drawn in November 1993. Eight test sessions took place between November 1993 and May 1994. The period of testing ranged from 140 to 161 days, $M = 153$, $SD = 7.32$ and the interval between the test sessions ranged from 13 to 35 days, $M = 21.9$ days, $SD = 4.78$ days. The longest time lag, the one between Test Session 2 and Test Session 3, included the Christmas holidays.

Material

Twenty geometric analogy items, which represented a selection from a highly homogeneous scale of 36 items (Hosenfeld, van den Boom, & Resing, in press), constituted the testing material. The items were designed out of six basic geometric shapes and five transformations by means of a facet design. Three examples of the items are presented in Fig. 1. The level of difficulty of each item could satisfactorily be predicted by the number of elements and the number of transformations the item contained.

A Mokken scale analysis (Mokken, 1971) revealed monotone homogeneity and double monotonicity. Hence, the items and the subjects could reliably be ordered on a common dimension. Furthermore, genuine parallel test items were constructed by means of the same construction rules as were the original items. The level of difficulty of the original and the parallel test items corresponded highly. The same was true for the interrelations of both test versions with external variables.

In Table 1, for the sample of this study, the item difficulties for the twenty items and Cronbach's alpha for the complete scale are displayed for the eight test sessions. The difficulty levels of the items ranged from $p = .00$ to $p =$

TABLE 1
Item Difficulties of the 20 Geometric Analogies at Eight Test Sessions

Item	Test							
	1	2	3	4	5	6	7	8
1	.78	.85	.90	.99	.97	.99	.97	1.00
2	.65	.87	.89	.94	.92	.99	.93	.97
3	.16	.66	.69	.66	.82	.75	.86	.70
4	.09	.39	.51	.68	.59	.72	.73	.73
5	.10	.28	.30	.51	.54	.45	.59	.49
6	.07	.34	.47	.49	.54	.48	.65	.48
7	.41	.51	.72	.73	.75	.83	.82	.86
8	.28	.39	.49	.55	.59	.52	.70	.58
9	.06	.10	.25	.30	.38	.39	.49	.45
10	.35	.63	.72	.86	.66	.96	.82	.97
11	.06	.10	.07	.04	.03	.04	.04	.09
12	.04	.23	.28	.25	.34	.38	.41	.44
13	.14	.35	.54	.61	.68	.62	.72	.70
14	.61	.80	.92	.90	.92	.93	.90	.96
15	.25	.52	.68	.70	.79	.78	.82	.82
16	.00	.13	.25	.20	.28	.21	.41	.34
17	.24	.45	.42	.49	.55	.54	.54	.66
18	.52	.68	.87	.78	.78	.83	.93	.85
19	.10	.10	.07	.10	.10	.04	.13	.06
20	.07	.24	.34	.30	.35	.34	.41	.39
α^a	.88	.91	.90	.89	.89	.87	.89	.89

^a Cronbach's α .

.78 at Test Session 1 and from $p = .06$ to $p = 1.00$ at Test Session 8, while the internal consistency coefficients ranged between $\alpha = .87$ and $\alpha = .91$.

All items were administered in an open-ended format, so that the subjects had to come up with their own solutions. Each item was printed on a sheet of paper and was presented separately. Children drew their solutions with a pencil in an empty box. Difficult and easy items were presented in a mixed order, which remained the same over subjects and over test sessions. Four of the items were presented a second time with different geometric shapes so that an inconsistency measure, which will be presented below, could be derived.

Procedure

Ten developmental psychology students were trained as test instructors. The training focused on proper test administration without providing any feed-back. No information about the hypotheses of the study was given to the test instructors. The students were provided with the testing material and the written instruction and had to practice the test instruction with children

who did not participate in the present study. Then, the adequacy of their instruction behavior was checked individually in a role play.

Each of the test instructors was responsible for the test administration to the eight subjects of one school. Once every three weeks, each child was tested individually outside the classroom. At the beginning of every test session the subjects were asked to draw and name the basic geometric shapes so that they got familiarized with the testing material. Then the instruction was given and three practice items were presented and solved with the support of the test instructor.

The core of the instruction ran as follows: "Look at this puzzle. It consists of four boxes. In three of the boxes there is a drawing (point to A, B, and C), the fourth one is empty (point to D). In this box you are to draw the solution of the puzzle. What you need to know is: These two boxes belong together (point to A and B), and those two boxes belong together (point to C and D). These two ones (A and B) belong together in the same way as those ones (C and D) do. Do you know what the solution is?"

Next, the child filled out the 20 test items, while the test instructor recorded the solution time in seconds for each item. Solution time was defined as the time lag between the presentation and the completion of an item. To make sure that the child still knew the instruction, twice during the test session the instructor introduced another practice item and repeated the standard instruction. To prevent the child from recognizing the individual items immediately from one to the next test session, the same items with different geometric shapes were administered alternately.

Scoring

To discriminate between the correct and several types of erroneous item solutions, a scoring system consisting of 15 categories was constructed, which is displayed in Table 2. Three main types of solutions were defined: (a) correct and (b) incomplete solutions, both indicating that analogical reasoning was present, and (c) incorrect solutions indicating that the item was solved by an association strategy. Furthermore, the variable *test score* was defined as the number of correct item solutions (Category 1), which were added up regardless of the level of item difficulty.

Scoring was done by a rater who was trained by the first author according to explicit coding rules. Interrater-reliability computed for ten percent of the data (1600 items), was high, Cohen's $\kappa = .90$ (Cohen, 1960). When the rater repeated the coding for the same 1600 items two months later, consistency was still high, $\kappa = .85$.

In sum, the data set of the longitudinal study reported here contained the data from eight measurement occasions, on which eighty children from Grades 1 and 2 repeatedly filled out 20 open-ended geometric analogies. Three outcome measures were obtained for each child at each test session: (a) test score, (b) solution pattern, and (c) overall solution time.

TABLE 2
Category System for the Solutions of the Geometric Analogies

No.	Description	Solution
1	Correct transformations applied, complete solution	correct
2	Correct transformations applied, elements confused	incomplete
3	Correct transformations applied, incomplete solution	incomplete
4	Elements from A, B, C, copied	associative
5	Undefined transformation applied	associative
6	A copied with horizontal or vertical position change	associative
7	B copied with horizontal or vertical position change	associative
8	C copied with horizontal or vertical position change	associative
9	A partially copied	associative
10	B partially copied	associative
11	C partially copied	associative
12	A completely copied	associative
13	B completely copied	associative
14	C completely copied	associative
15	No classification possible	associative

RESULTS

In order to examine the hypothesis that the development of analogical reasoning is a discontinuous process, we investigated those catastrophe flags that can be detected in a longitudinal data set, bimodality, inaccessibility, sudden jump, anomalous variance, and critical slowing down. We followed the order of analyses that was suggested by van der Maas et al. (1993). Because bimodality and inaccessibility in the cross-sectional frequency distributions constitute the first necessary indicators of a transition, they have to be examined first. If bimodality cannot be detected, the discontinuity hypothesis has already to be rejected. Next, the detection of sudden jumps in longitudinal data helps to reject the hypothesis of linear growth, but is not sufficient for the discrimination between a rapid acceleration and a genuine discontinuity. Finally, the detection of the catastrophe flags anomalous variance and critical slowing down adds strong support to the confirmation of the discontinuity hypothesis.

Reduction of the Data Set

Before running any analyses we discarded seven records that contained incomplete data. In addition, we discarded the records of (a) one subject who was detected as a multivariate outlier and (b) one subject who formed her own class with $n = 1$, when the subjects were assigned to transition classes. The final data set comprised 71 complete records of 19 girls and 16 boys from Grade 1, and 17 girls and 19 boys from Grade 2. The age of the children ranged from

six years and two months to eight years and one month, $M = 86$ months, $SD = 7$ months. On average, the cohorts were 77, 82, 90, and 95 months old.

Bimodality and Inaccessibility

The first analyses addressed the following question: Are there distinct modes of analogical reasoning and can these modes be confirmed statistically? Because we assumed two different strategies for analogy solution in young elementary school children, we expected to find two distinct modes in the frequency distributions of the test scores.

Testing for bimodality and inaccessibility can be done by parametric (Everitt & Hand, 1981) as well as by nonparametric (Silverman, 1981) techniques. Since catastrophe theory does not predict the type of distribution for each mode, nonparametric techniques are preferable. In catastrophe theory all types of nonlinear scalings are allowed, which can change the form of the modes strongly. On the other hand, statistical theory is much more developed for parametric techniques (see also Titterton, Smith, & Makov, 1985). Moreover, if sum scores of binary test scores are investigated, binomial distributions can be assumed (Thomas & Lohaus, 1993). Therefore, we chose to use the finite mixture approach of Everitt and Hand (1981), which delivers a set of statistics to assess the fit of several models and to compare models with different numbers of modes. The parameters to be estimated for each model were the success probabilities of the latent classes ($\theta_1, \theta_2, \dots, \theta_n$) and the proportions of subjects belonging to each class ($\pi_1, \pi_2, \dots, \pi_n$).

Because bimodality was the focus of model testing, we fitted a one-, two-, and three-component binomial mixture model to the frequency distributions of the test scores of the total sample for each of the eight test sessions. Next, we compared the respective goodness-of-fit indices of the three models for each test session. The two-component model was selected, if it fitted the data significantly better than the one-component model and if the three-component model did not enhance the model fit of the two-component model substantially.

Three evaluation statistics were computed for each model in order to find the most parsimonious model with the best fit (Thomas, 1989; Thomas & Turner, 1991). First, Akaike's Information Criterion (AIC), discussed by Thomas and Lohaus (1993), served as a selection criterion within the family of models fitted to the same data set. It is defined as "minus two times the loglikelihood function plus a 'penalty factor' equal to twice the number of parameters estimated from the data" (Thomas & Turner, 1991, p. 182). Hence, the smallest AIC points out the best fitting model in connection with parameter parsimony. Second, the Pearson χ^2 statistic gives an indication of the goodness of fit of the model to the observations. Third, the proportion of variance accounted for (VAF) by the model seems intuitively the clearest indicator of model fit. It can be used for the comparison of different model solutions for the same data set as well as for the comparison of the model solutions of different data sets, but it ignores the issue of parameter parsimony.

TABLE 3
 Model Estimates for the Analogy Performance (20 Items) of 71 Children at Eight
 Measurement Occasions

t^a	M	SD	θ_1	θ_2	θ_3	π_1	π_2	π_3	VAF	χ^2	df^b	AIC	
1	4.96	4.13	.25			1.00			.22	2709	12	539	
			(.011)										
			.09	.44		.55	.45			.87	27	10	396
			(.015)	(.026)		(.070)	(.070)						
			.02	.19	.50	.22	.46	.32	.94	9	8	386	
			(.025)	(.045)	(.041)	(.107)	(.093)	(.088)					
2	8.62	5.16	.43			1.00			.18	4551	15	645	
			(.013)										
			.20	.65		.49	.51			.89	22	13	438
			(.018)	(.020)		(.063)	(.063)						
			.19	.55	.80	.45	.38	.17	.98	8	11	427	
			(.019)	(.040)	(.042)	(.066)	(.077)	(.071)					
3	10.37	4.98	.52			1.00			.20	1244	13	609	
			(.013)										
			.27	.71		.44	.56			.91	12	11	427
			(.022)	(.019)		(.065)	(.065)						
			.24	.51	.75	.34	.24	.42	.97	6	9	425	
			(.038)	(.137)	(.039)	(.111)	(.098)	(.144)					
4	11.07	4.81	.55			1.00			.21	11663	14	580	
			(.013)										
			.27	.71		.36	.64			.94	28	12	409
			(.023)	(.016)		(.060)	(.060)						
			.26	.61	.77	.43	.28	.38	1.00	20	10	410	
			(.025)	(.078)	(.042)	(.065)	(.177)	(.190)					
5	11.55	4.87	.58			1.00			.21	1433	13	599	
			(.013)										
			.31	.74		.37	.63			.89	21	11	433
			(.025)	(.018)		(.064)	(.064)						
			.02	.36	.74	.04	.35	.61	.96	16	9	416	
			(.018)	(.026)	(.017)	(.024)	(.062)	(.063)					
6	11.77	4.35	.59			1.00			.26	417	11	520	
			(.013)										
			.36	.73		.38	.62			.93	13	9	405
			(.026)	(.018)		(.067)	(.067)						
			.14	.41	.74	.05	.37	.58	.98	13	7	403	
			(.061)	(.036)	(.020)	(.036)	(.072)	(.074)					
7	12.86	4.61	.64			1.00			.22	1500	12	576	
			(.013)										
			.43	.82		.44	.56			.89	15	10	418
			(.024)	(.018)		(.067)	(.067)						
			.20	.51	.83	.09	.40	.51	.99	4	8	405	
			(.052)	(.034)	(.017)	(.046)	(.070)	(.069)					

TABLE 3—Continued

8	12.54	4.54	.63 (.013)		1.00		.23	1258	12	549
			.34 (.030)	.76 (.017)	.31 (.063)	.69 (.063)	.91	14	10	406
			.28 (.091)	.56 (.206)	.79 (.052)	.20 (.151)	.27 (.146)	.53 (.259)	.97	7 8 404
1-8	10.47	4.99	.52 (.005)		1.00		.18	1441074	18	5283
			.26 (.009)	.71 (.007)	.42 (.023)	.58 (.023)	.88	592	16	3554
			.11 (.019)	.39 (.022)	.75 (.009)	.16 (.029)	.35 (.027)	.49 (.029)	.97	56 14 3392

Note. θ_1 , θ_2 , and θ_3 are estimations of the success probabilities of each component; π_1 , π_2 , and π_3 are the estimated proportions of the sample under each component; VAF is the variance accounted for and AIC is Akaike's information criterion. The numbers in parentheses are the estimations of the standard errors.

^a Test session.

^b The degrees of freedom vary because at each test session the χ^2 test is conducted after removing cells with frequencies less than five (Van der Pol, Langeheine, & de Jong, 1991).

Table 3 displays the means and the standard deviations of the test scores and the results of the model estimations for each of the eight measurement occasions separately and for all eight measurement occasions together. The mean test scores increased almost monotonously from $M = 4.96$ at Test Session 1 to $M = 12.54$ at Test Session 8. The estimates of the success probabilities at Test Session 1, for instance, were $\theta_1 = .25$; $\theta_1 = .09$, and $\theta_2 = .44$; $\theta_1 = .02$, $\theta_2 = .19$, and $\theta_3 = .50$, for the one-, two-, and three-component model, respectively, while the corresponding proportions of subjects in the latent classes were $\pi_1 = 1.00$; $\pi_1 = .55$, and $\pi_2 = .45$; $\pi_1 = .22$, $\pi_2 = .46$, and $\pi_3 = .32$.

For each test session, we first compared the AICs, the χ^2 statistics, and the VAFs of the one- and two-component models (see Table 3). The AICs indicated that at every test session the two-component binomial mixture model described the form of the distribution significantly more parsimoniously than the one-component model. In five of the eight test sessions (Test Sessions 2, 3, 6, 7, and 8), the χ^2 statistic showed that no significant difference existed between the data and the two-component model. The VAF increased from about 20% for the one-component model to about 90% for the two-component model at every test session. When we compared the goodness-of-fit indices of the two- and three-component models, we discovered that the AICs for all test sessions, except for Test Session 4, were slightly reduced for the three-component model. Furthermore, the χ^2 indicated that the three-component model fitted the data satisfactorily for those three test sessions (1, 4, and 5), at which the two-

component model failed to fit the data. Finally, at none of the test sessions did the VAF increase by more than 10% from the two- to three-component model.

Because the geometric analogy items covered a range of difficulty levels, and thus did not meet the assumption of equal item probabilities, we repeated the binomial mixture analyses with subsets of 4, 8, 10, and 16 items covering smaller ranges of item probabilities. For all analyses, the two-component model outperformed the one-component model, and in the majority of the analyses, the three-component model did not add substantially to the variance accounted for by the two-component model. Therefore, the two-component solution, which discriminated two latent classes consisting of inaccurate and accurate performers at every measurement occasion, was retained as robust enough for further interpretation.

Two-component mixture models, however, do not necessarily imply bimodality in the overall distributions. A valid two-component mixture model, for example, can contain two normal distributions with equal means and highly different variances (Thomas & Lohaus, 1993). In that case, the resulting distribution has only one mode. Bimodality and inaccessibility, however, can only be assumed, if there are two clearly separated modes in the modeled frequency distribution. Visual inspection of Fig. 2, which displays the observed frequency distributions of the test scores for each of the eight test sessions together with the estimated distribution curves for the model with the best fit, that is, the two-component model for Test Sessions 2, 3, 6, 7, and 8, and the three-component model for Test Sessions 1, 4, and 5, confirmed bimodality. In the observed as well as in the modeled distributions, two separate modes were clearly visible for each test session. While the majority of the subjects belonged to the first latent class at the first test session, the majority of the subjects belonged to the second latent class at the eighth test session.

As a matter of fact, the two modes of analogy performance were assumed to remain stable over time, while a part of the sample was expected to change class membership. Hence, we examined the longitudinal patterns formed by the estimated parameters of the two-component model, the success probabilities of the two latent classes (θ_1 and θ_2) and the proportion of accurate analogical reasoners (π_2). Linear regression with test session as predictor indicated linear trends for all three parameters (see Fig. 3). Not only did the proportion of members in the high-performing class (π_2) rise significantly from the first to the eighth test session, $R^2 = .62$, $F(1, 6) = 9.78$, $p < .05$, but the success probabilities of the two latent classes, the low-performing (θ_1) and the high-performing (θ_2) individuals, also increased continuously over time, $R^2 = .66$, $F(1, 6) = 11.88$, $p < .05$, and $R^2 = .81$, $F(1, 6) = 25.11$, $p < .01$, respectively. So, in both classes, analogy performance was enhanced to a similar degree over the eight test sessions. This result is in contrast to our expectation that the two modes in the frequency distribution and the inaccessible region in between would remain stable over the eight test sessions.

To explore the inaccessible region on the developmental scale further, we

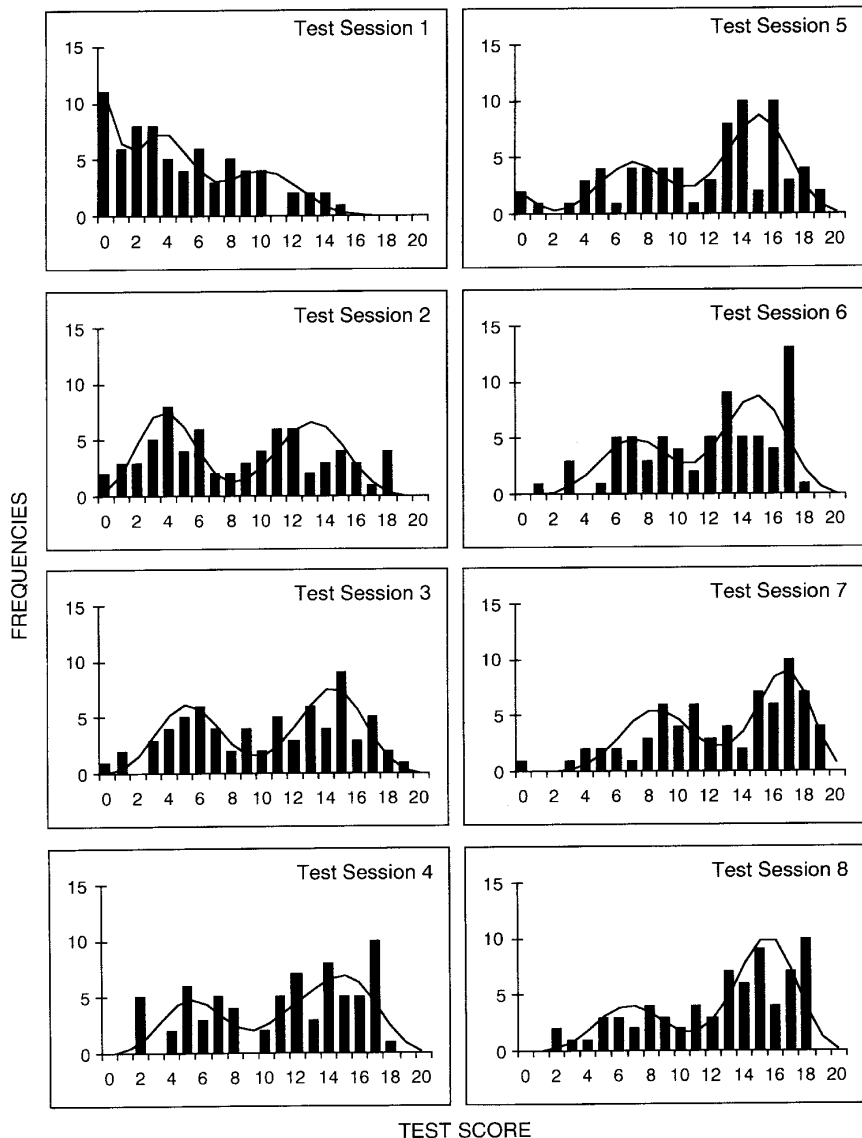


FIG. 2. Frequency distributions of the test score for the eight test sessions. The histogram represents the observed data; the curve represents the model with the best fit (a two-component model for Test Sessions 2, 3, 5, 6, and 7, a three-component model for Test Sessions 1, 4, and 5).

partitioned the sample at each measurement occasion into two discrete groups using the procedure described by Thomas and Lohaus (1993). For this purpose, for each individual test score two posterior probabilities were estimated on the basis of the estimated parameters of the two-component model, the probability

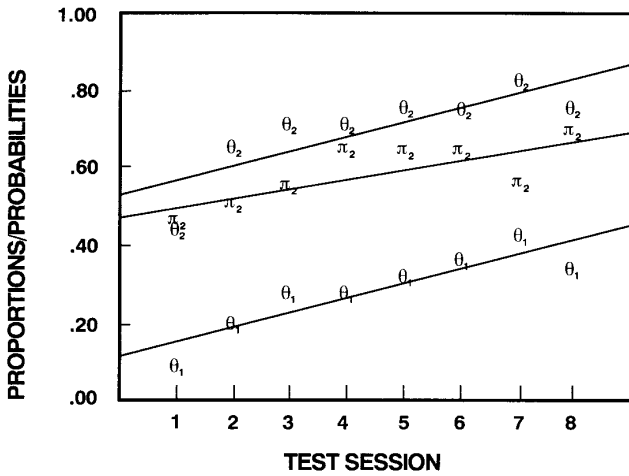


FIG. 3. Parameter estimates of the two-component model as a function of test session. θ_1 and θ_2 represent the success probabilities of the two components; π_2 represents the proportion of high-performing analogical reasoners. The solid lines (—) represent the linear trends for each parameter.

of belonging to the first component and the probability of belonging to the second component. Then each test score was assigned to the component with the largest posterior probability. At Test Session 1, for example, the test scores zero through four were assigned to the first component, while the test scores five through twenty were assigned to the second component. Then the minimum number of correct responses required to be classified as accurate performer was determined, for instance, five for Test Session 1. This minimum number of correct responses, however, increased steadily from 5 to 13 over the eight test sessions (see Table 4), which indicated a shift of the inaccessible region over the eight test sessions in the direction of the higher test scores. When the analogy performance over all eight test sessions was examined, the minimum number of items correct necessary for being classified as belonging to the second component, was ten, irrespective of whether the eight raw scores or the mean score for each child was analyzed. The criterion of ten items correct will be used for the definition of the transitional subjects in subsequent analyses.

The first result, strong bimodality in the majority of the measurement occasions, clearly supported the hypothesis of discontinuous development in analogical reasoning, whereas the second result, instability of the success probabilities and of the inaccessible region, was unexpected and not in support of discontinuous developmental change.

The Effect of Repeated Testing

The overall increase of test performance from the first to the eighth test session might have been an outcome of the repeated administration of almost identical testing material. To examine whether analogy performance was facil-

TABLE 4
 Minimum Number of Items Correct Necessary for Component 2 Membership
 and Its Posterior Probability at Test Sessions 1 through 8 and Overall

Test session	Criterion	p(Comp.21 Crit)
1	5	.61
2	9	.83
3	10	.66
4	10	.73
5	11	.79
6	11	.62
7	13	.64
8	11	.63
total ^a	10	.73
total ^b	10	.68

Note. The definition of the criterion is based on the estimated model parameters of the two-component model.

^a Mixture distribution based on 568 raw scores.

^b On 71 mean scores.

itated by repeated testing, we compared the test performance at the first and the last test session while controlling for age. Because both the age differences between the cohorts and the time lag between the first and the last test session was about half a year, each cohort at its last test session was of the same age as the subsequent cohort at its first test session. Because the test scores from both the first and the last test session were available for only three age levels, the 6.5, the 7-, and the 7.5-year-old children, a Mann-Whitney U-test (Hays, 1981, pp. 587–589) on test score with test session as the independent variable was performed for each of these three levels. At 6.5- and 7.5 years, the test performance in the last test session was significantly higher than in the first test session, $z = 4.44$, $p < .001$, and $z = 4.65$, $p < .001$, respectively, while at 7 years the same trend was obtained, $z = 1.54$, $p = .06$. In general, even after controlling for age, we found that our subjects performed significantly better, when they were tested for the eighth time than when they were tested for the first time. Hence, some part of the effect under observation can be explained by practice.

Solution Patterns of the Nonanalogical, the Transitional, and the Analogical Reasoners

To explore the performance characteristics of the transitional subjects, we divided the sample into three classes, nonanalogical, transitional, and analogical reasoners. For this purpose, we used the criterion of ten or more items correct, which was derived from the estimation of the posterior probabilities

TABLE 5
 Frequencies and Proportions of 15 Types of Item Solutions
 for Three Classes of Analogical Reasoners

Category	Class		
	Nonanalogical	Transitional	Analogical
1	515 (.21)	4054 (.56)	1376 (.78)
2	13 (.01)	156 (.02)	14 (.01)
3	491 (.20)	1830 (.25)	311 (.18)
4	259 (.11)	287 (.04)	32 (.02)
5	104 (.04)	248 (.03)	16 (.01)
6	2 (.00)	0 (.00)	0 (.00)
7	0 (.00)	0 (.00)	0 (.00)
8	54 (.02)	120 (.02)	1 (.00)
9	17 (.01)	4 (.00)	0 (.00)
10	48 (.02)	12 (.00)	0 (.00)
11	197 (.08)	57 (.01)	1 (.00)
12	46 (.02)	13 (.00)	0 (.00)
13	116 (.05)	28 (.00)	3 (.00)
14	487 (.20)	321 (.04)	1 (.00)
15	51 (.02)	70 (.01)	5 (.00)
Total	2400	7200	1760

Note. Category 1 represents correct solutions; Categories 2 and 3 represent incomplete solutions; and Categories 4 through 15 associative solutions. The numbers in brackets are the column percentages.

of the two-component model for the overall frequency distribution (Thomas & Lohaus, 1993). Subjects who consistently solved ten or more analogies were defined as analogical reasoners, $N = 11$, while subjects who consistently solved less than ten items were defined as nonanalogical reasoners, $N = 15$. All 45 subjects whose test scores passed the criterion of ten items correct during the half year of testing were defined as transitional subjects.

Next, the solution patterns of the three classes were examined to determine whether the transitional subjects produced any unique item solutions. Table 5 shows that, in general, the same types of item solutions were produced by all three classes. Large differences between classes were found in the proportions of the solution types. Analogical reasoners produced correct solutions for the greater part, nonanalogical reasoners came up with associative solutions most of the time, and transitional subjects produced correct as well as associative solutions. There was a significant difference in the number of incomplete solutions (Categories 2 and 3) in favor of the transitional class (1986 (27%) vs 504 (21%) and 325 (19%) for the nonanalogical and the analogical reasoners, respectively), $\chi^2(2, N = 2815) = 58, p < .01$, but it

seems to be an overstatement to conclude that the incomplete solutions were typical for the transitional subjects.

To illustrate the change of the prevailing solution over time, we examined the three main solution types separately, that is, the correct, the incomplete, and the associative solutions. For each class a characteristic pattern, as displayed in Fig. 4, was found.

For the nonanalogical reasoners (see Fig. 4 A), the majority of the item solutions were associative solutions at the first test session; at the eighth test session, correct, incomplete, and associative solutions were about equally represented. For the analogical reasoners (see Fig. 4 C), at every test session correct item solutions dominated the incomplete and associative solutions. The latter two decreased slightly in favor of the correct solutions. For the transitional subjects (see Fig. 4 B), a reversal of the prevailing solution took place. Associative solutions, which were dominant at the first test session, almost disappeared, whereas the correct solutions increased rapidly and took the largest part. The proportion of incomplete solutions, finally, remained more or less constant over eight test sessions for all three classes.

In sum, no item solution specific for the transitional subjects was found. The aspect that discriminated the transitional class from the two other classes was the reversal of the prevailing solutions. This result is in accordance with the expectation that the transitional subjects need not represent a homogeneous group with unique characteristics that can be identified on one measurement occasion alone. It is the longitudinal pattern of their performance and other dynamical characteristics of their solution that define the transitional reasoners. Three of the dynamical characteristics of transitional subjects will be investigated in the next section.

Sudden Jump, Anomalous Variance, and Critical Slowing Down

To demonstrate the sudden jump and to test the hypotheses of anomalous variance and critical slowing down in the performance of the transitional subjects we identified for each transitional subject, the test session at which it passed the criterion of ten items correct and assigned it to one of seven subgroups. All subjects who passed the criterion at the second test session constituted Transition Class 1, all subjects who passed the criterion at the third test session constituted Transition Class 2, etc. The test score patterns of the classes are displayed in Table 6, while the raw test scores of each subject are provided in the Appendix. Because one transition class contained only one subject, the following analyses were conducted on the remaining six transition classes.

Furthermore, for each transition class the measurement occasions were aligned to a new scale with the transition from the state of nonanalogical reasoner to the state of analogical reasoner as an anchor, as represented by the vertical lines in Table 6. Each test session at which a transitional child solved nine or less items, was labeled "before the transition," the remaining

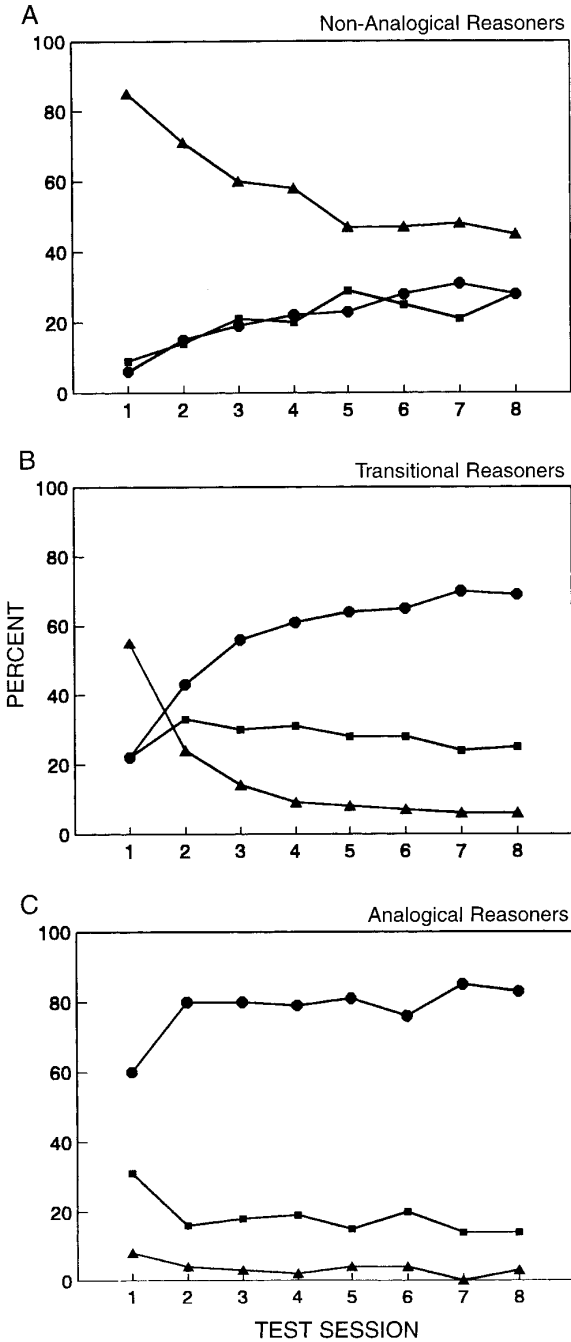


TABLE 6
Division of the Sample into Classes by Longitudinal Response Pattern

Class	N	Test							
		1	2	3	4	5	6	7	8
AR ^a	11	+	+	+	+	+	+	+	+
T1 ^b	22	-	+	+	+	+	+	+	+
T2	8	-	-	+	+	+	+	+	+
T3	6	-	-	-	+	+	+	+	+
T4	3	-	-	-	-	+	+	+	+
T5	2	-	-	-	-	-	+	+	+
T6	4	-	-	-	-	-	-	+	+
T7	0	-	-	-	-	-	-	-	+
NA ^c	15	-	-	-	-	-	-	-	-

Note. “+” = ten or more items correct and “-” = less than ten items correct,” respectively. The vertical lines represent the anchor for aligning the transition classes.

^a AR = Analogical Reasoners.

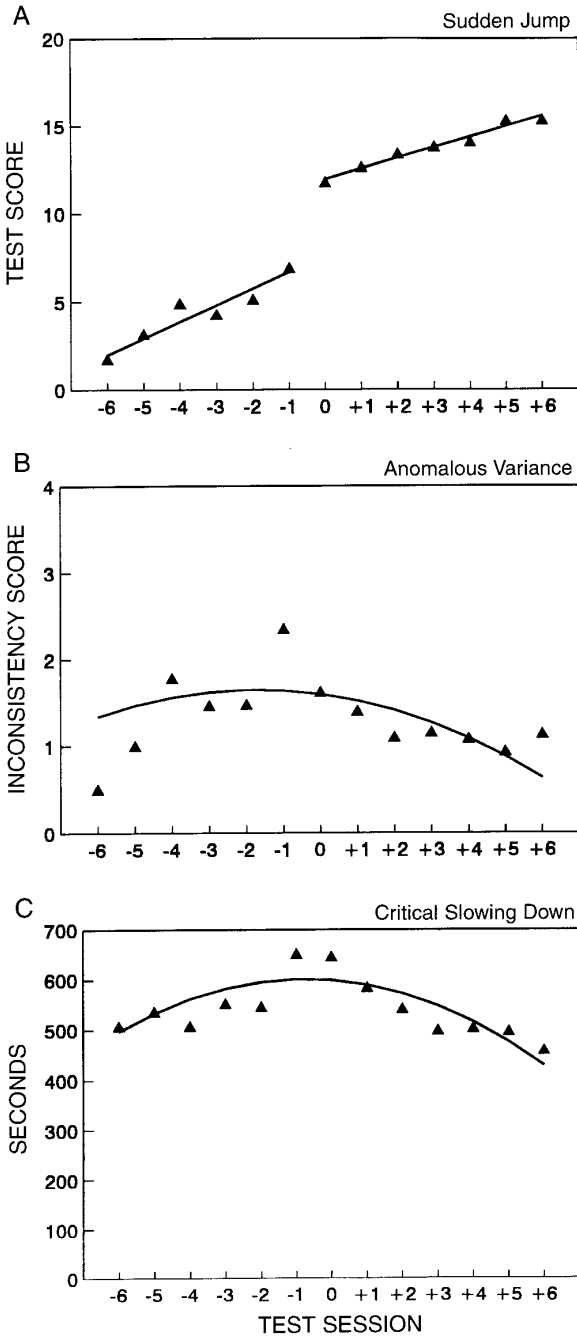
^b T1 to T7 = Transitional Subjects.

^c NA = Nonanalogical Reasoners.

test sessions were labeled “after the transition.” For children in Transition Class 1, for example, one test session took place before and seven test sessions after the transition, for children in Transition Class 2, two test sessions took place before and six after the transition, etc. Hence, the new scale comprised 13 scale points with varying numbers of observations on each scale point.

Sudden jump. For the transitional subjects, the test score was expected to increase rapidly after the transition. Figure 5 A shows the mean test scores for each test session, which indeed form a step-like pattern. To investigate this pattern of a sudden jump we fitted two separate regression lines on test score, one for the test sessions “before the transition” and one for the test sessions “after the transition.” The slopes of the two regression lines were similar, $\beta = .53$, $t = 6.31$, $p < .001$, and $\beta = .41$, $t = 7.36$, $p < .001$, while there was a clear difference in level, $a = 1.03$ and $a = 7.77$. Both regression equations together accounted for 72% of the variance of the variable test score. Thus, the growth curve of the transitional subjects can be satisfactorily modeled by two increasing lines with slightly different slopes on clearly different levels. The regression line, which fits the test scores before the transition, seems to jump to a higher level after the transition, where it is continued. The demonstration of the sudden jump clearly confirmed the classi-

FIG. 4. Solution pattern as a function of test session for (A) nonanalogical reasoners, (B) transitional subjects, and (C) analogical reasoners. Triangles (▲) represent associative solutions, squares (■) incomplete solutions, and circles (●) analogical solutions.



fication criterion, but it was not sufficient to discriminate between rapid acceleration and a genuine discontinuity in the development of analogical reasoning (Van der Maas et al., 1993).

Anomalous variance. To detect the presence of anomalous variance in the solution behavior of the transitional subjects, we derived an inconsistency index from the classification of the analogy solutions of the four test items that were presented twice. The three main categories, correct, incomplete, and associative solutions, were used for this comparison. For example, if one correct and one incomplete solution was found within an item pair, the item pair was labeled inconsistent. The same was done for item pairs consisting of one correct and one associative solution or one incomplete and one associative solution. For each subject an inconsistency score was obtained from the number of different-solution pairs for each test session. Inconsistency scores ranged from 0 to 4. While overall inconsistency was also high for both the nonanalogical reasoners, $M = 1.14$, $SD = 1.03$, and the analogical reasoners, $M = 1.15$, $SD = .98$, the transition group gained the highest inconsistency scores, $M = 1.48$, $SD = 1.10$, which means that it produced inconsistent solutions in more than one third of the item pairs.

Next, to test whether inconsistency in the transitional subjects increased near the transition, a polynomial regression analysis was conducted on inconsistency with test session (13 test sessions) as predictor. The regression equation contained a significant quadratic term, $F(1, 347) = 9.04$, $p < .01$, indicating a maximum in the inconsistency curve. Figure 5 B shows that this maximum corresponded clearly with the sudden jump. Inconsistent solution behavior was highest shortly before the sudden jump.

Critical slowing down. The sum of the item solution times for each subject was defined as solution time, which ranged from 170 to 1517 s, $M = 539$ s, $SD = 201$ s. In order to search for evidence of the phenomenon of slowing down, a polynomial regression analysis on solution time with test session aligned with the sudden jump was performed. A significant quadratic trend was detected, $F(1, 347) = 11.20$, $p < .001$, in the solution time curve. As can be seen in Figure 5 C, the maximum of this curve is closely related to the sudden jump. Solution time continuously increased before the sudden jump and decreased thereafter.

In sum, the sudden increase in test performance of the transitional subjects was found to be accompanied by a temporary increase of both inconsistent solution behavior and solution time.

DISCUSSION

The longitudinal study reported here provided support of the hypothesis that there is a discontinuity in the development of analogical reasoning. Four

FIG. 5. (A) Sudden jump, (B) anomalous variance, and (C) critical slowing down in the performance of the transitional group. The test sessions are aligned by the criterion of ten or more items correct. Triangles (▲) represent the mean observations, solid lines (—) the trends.

of the five indicators of a transition as defined by catastrophe theory, bimodality, sudden jump, anomalous variance, and critical slowing down were satisfactorily demonstrated in the analogical reasoning performance of young elementary school children. Two qualitatively different modes of analogical reasoning were observed in 6- to 8-year-old children, free association and systematic analogical reasoning. During half a year of testing, a large part of the sample switched solution mode, progressing from free association to systematic analogical reasoning. This transition was accompanied by a sudden increase of analogy performance, a temporary increase of inconsistent solution behavior, and a temporary slowing down of solution behavior during a period of six months. In particular, the latter two indicators, anomalous variance and critical slowing down, pointed to the fact that a conflict between two strategies might be the cause of the progress.

In contrast to the expectations derived from the model of a transition, not only the number of analogical reasoners increased, while the number of nonanalogical reasoners decreased accordingly, but the success probabilities of both classes also increased steadily. As a consequence, the inaccessible region on the developmental scale moved from the lower to the higher scale points over the eight measurement occasions. Apparently, repeated exposure to analogy problems raised the overall probability of success. Most probably, not only the transitional subjects, but all children became familiarized with the form and the content of the testing material, which helped them to deal economically with the memory load imposed by the analogies. In adults a similar finding was reported by Sternberg (1977). Poor and successful analogy performers allocated their solution time in different ways. Successful performers spent more time encoding the attributes of an analogy and less time processing comparisons than the poor performers did. The strategic allocation of solution time seemed to help them to master the memory load of the analogies. In the longitudinal study reported here, repeated dealing with analogies might have reduced encoding difficulties in all children.

Furthermore, we found evidence that the transition in analogical reasoning we looked for occurred in the first grades of elementary school. It was, however, not the aim of the study to make a definitive statement about the age level at which children master analogical reasoning. We agree with Goswami (1992) that changing the test procedure may facilitate analogy solution so that children become able to solve one or the other type of analogy earlier, but we expect that changing the test procedure only pushes off the phenomenon of discontinuity to a lower age level. Evidence for this assumption may be found in the study of Alexander et al. (1987), where the frequency distribution of the analogy performance of 4- and 5-year-old children also displayed two different modes. The tasks presented in this study were similar, but easier geometric analogy problems than the tasks in the study presented here.

There are three more indicators of a transition, divergence, divergence of linear response, and hysteresis, whose detection could contribute to the confir-

mation of the discontinuity hypothesis (Van der Maas & Molenaar, 1992). They were not investigated in our study, because experimental designs are needed for their detection. In the following we will suggest guidelines for future investigations concerning the detection of the remaining three transition flags in the development of analogical reasoning. First, the clarity of the phenomena of bimodality and sudden jump depends on the conditions under which they are observed (divergence). Hence, the systematic variation of test conditions needs to be studied. Under less optimal test conditions than those provided in our study, for example, when the test is administered in groups and without any instruction, bimodality and sudden jump should weaken or even disappear, while under even more favorable conditions than ours, when for example, feedback or examples of correct solutions are provided, they should emerge even more clearly than demonstrated in this study. Second, if there is a transition in analogical reasoning, specific training in analogical reasoning skills should have an effect on transitional subjects only, that is, those children who have both modes of analogical reasoning at their disposal (divergence of linear response). Training can strengthen the conflict between the two modes and possibly accelerate the process of establishing the more successful strategy, but training cannot induce analogical reasoning in children who master only free association. Hence, training studies are needed with experimental groups consisting of nonanalogical reasoners and transitional subjects. The distinction between these groups can be made on the basis of, for example, inconsistency and solution time measures. Third, the transition from one mode to the other is expected to occur on different values of the control variable depending on the direction in which the control variable changes (hysteresis). If we assume that, for instance, information processing capacity is a control variable responsible for the transition in analogical reasoning, information load might be manipulated systematically. The same analogy items might be administered with increasing information load to one group of subjects and with decreasing information load to another one. For the latter group we would expect the first associative solution at easier items than for the first group.

The framework of catastrophe theory provided the possibility to demonstrate a discontinuity in the development of analogical reasoning, which may be described as a genuine shift from using free association to considering the constraint of parallel relations. Such a discontinuity is a typical characteristic of complex developmental processes. In the last decade, research in physics, chemistry, biology, and other disciplines has shown many examples of discontinuous phase transitions in processes of structural change (Haken, 1977; Prigogine & Stengers, 1984). These phase transitions are generally in accord with the original Piagetian description of discontinuous change, except for the domain specific character of the current approaches. From the point of view of nonlinear dynamic systems theory, of which catastrophe theory is a part, domain-specific transitions can be expected when mastering a domain requires the acquisition of new strategies.

At the moment, we can only speculate about the developmental mechanisms responsible for the transition in analogical reasoning in elementary school children. In our view, an explanation in Piagetian terms, such as competing preoperational and operational structures, is insufficient. What we need are hypotheses that include control variables suitable for operationalization. Growing information processing capacity, for example, or decreasing field dependency might be reasonable candidates. These variables have already been suggested in connection with investigations about cognitive performance more characteristic for the concrete operational stage like volume conservation (Van der Maas & Molenaar, 1992) and the understanding of horizontality and verticality (Thomas & Lohaus, 1993), but await empirical investigation. A fruitful addition to empirical investigations of possible control variables might be simulation research. An example of a computer model, which simulated analogical reasoning as a product of a self-organizing system, was presented by Mitchell and Hofstadter (1990).

Finally, we would like to emphasize our conviction that the idea of discovering structural change with the help of formal indicators may be worthwhile for investigations in other domains of cognitive development as well, for example, number conservation, class inclusion, or balance scale problems.

APPENDIX

The Raw Test Scores of 71 Children at Eight Test Sessions

Class	Subject	t1	t2	t3	t4	t5	t6	t7	t8	
AR ^a	507	15	16	18	17	18	17	16	18	
	607	14	18	19	15	19	17	17	15	
	606	14	18	17	17	18	17	18	18	
	708	13	18	15	17	14	14	17	18	
	705	13	17	17	16	14	17	17	18	
	703	12	16	16	16	17	17	19	18	
	706	12	15	15	13	14	16	15	15	
	205	10	18	17	17	16	13	18	16	
	307	10	15	15	15	16	13	17	17	
	002	10	15	13	17	16	13	18	16	
	008	10	11	13	14	16	14	16	14	
	T1 ^b	501	9	15	16	15	15	16	15	18
		508	9	14	14	15	17	15	18	15
		406	9	12	13	14	10	13	16	14
806		9	12	12	14	13	15	16	15	
605		8	14	14	17	14	15	15	16	
108		8	13	15	17	16	17	12	17	
301		8	11	11	17	14	12	14	15	
005		8	10	14	15	12	16	19	15	
206		7	14	15	16	12	15	15	18	
807		7	10	12	8	13	13	15	12	
903		6	16	15	18	16	17	16	17	

APPENDIX—Continued

Class	Subject	t1	t2	t3	t4	t5	t6	t7	t8
	105	6	12	13	16	19	18	19	18
	203	6	11	17	17	14	17	18	14
	804	6	11	16	14	18	16	18	17
	906	6	10	8	10	14	12	11	11
	207	5	13	18	17	16	17	17	18
	208	5	11	15	16	17	14	17	18
	803	4	12	15	14	13	13	15	13
	402	3	12	13	11	7	10	13	13
	901	3	11	11	12	18	14	17	17
	302	3	10	13	14	16	17	18	17
	003	2	12	10	13	13	10	13	9
T2	007	8	9	15	14	13	15	13	14
	905	7	8	17	14	14	17	17	15
	102	5	3	11	12	9	9	8	11
	907	4	9	11	13	14	17	19	15
	106	3	6	12	12	13	10	11	10
	904	0	9	14	11	15	13	15	17
	604	0	7	11	12	13	11	12	14
	902	0	7	10	11	13	13	17	16
T3	001	4	5	9	11	16	17	17	15
	701	4	4	9	10	11	11	10	11
	504	3	6	6	11	7	8	10	8
	303	3	3	6	12	14	14	14	14
	502	2	6	9	12	16	13	16	12
	305	0	3	9	12	12	12	13	13
T4	503	6	8	8	7	10	9	10	13
	801	4	5	7	7	10	8	8	10
	103	2	3	7	7	10	12	9	13
T5	202	2	5	6	6	8	10	11	12
	505	0	2	3	8	9	12	11	11
T6	603	2	5	7	8	8	9	12	13
	308	2	4	7	5	7	6	11	8
	704	2	4	5	5	5	7	11	13
	702	1	4	6	5	8	9	10	9
NA ^c	403	5	6	5	7	5	7	9	5
	104	3	6	5	8	5	9	9	9
	802	3	2	1	2	1	3	3	3
	908	2	1	4	6	4	6	4	2
	107	1	6	4	6	7	7	9	8
	201	1	4	5	2	4	5	9	5
	808	1	4	3	4	4	3	5	4
	408	1	3	6	5	5	6	8	7
	006	1	0	0	2	3	6	5	6
	004	0	4	6	5	9	8	6	8
	204	0	4	3	2	0	3	6	6
	608	0	2	5	5	8	6	4	6
	304	0	1	4	7	9	7	9	5
	601	0	1	4	4	6	7	7	7
	306	0	0	1	2	0	1	0	2

^a AR = Analogical Reasoners.^b T1 to T6 = Transitional Subjects.

REFERENCES

- Alexander, P. A., Willson, V. L., White, C. S., & Fuqua, J. D. (1987). Analogical reasoning in young children. *Journal of Educational Psychology*, **79**, 401–408.
- Alexander, P. A., Willson, V. L., White, C. S., Fuqua, J. D., Clark, G. D., Wilson, A. F., & Kulikowich, J. M. (1989). Development of analogical reasoning in 4- and 5-year-old children. *Cognitive Development*, **4**, 65–88.
- Brown, A. L. (1989). Analogical learning and transfer. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 369–412). Cambridge: Cambridge University Press.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.
- Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman & Hall.
- Fischer, K. W., Pipp, S. L., & Bullock, D. (1984). Detecting developmental discontinuities: Methods and measurement. In R. N. Emde & R. J. Harmon (Eds.), *Continuities and discontinuities in development* (pp. 95–121). New York: Plenum Press.
- Gentner, D. (1977). Children's Performance on a spatial analogies task. *Child Development*, **48**, 1034–1039.
- Gilmore, R. (1981). *Catastrophe theory for scientists and engineers*. New York: Wiley.
- Goldman, S. R., Pellegrino, J. W., Parseghian, P., & Sallis, R. (1982). Developmental and individual differences in verbal analogical reasoning. *Child Development*, **53**, 550–559.
- Goswami, U. (1991). Analogical reasoning: What develops? A review of research and theory. *Child Development*, **62**, 1–22.
- Goswami, U. (1992). *Analogical reasoning in children*. Hove, UK: Erlbaum.
- Graham, T., & Perry, M. (1993). Indexing transitional knowledge. *Developmental Psychology*, **29**, 779–788.
- Haken, H. (1977). *Synergetics: An introduction: Nonequilibrium phase transitions and self-organization in physics, chemistry, and biology*. Heidelberg: Springer.
- Hays, W. L. (1981). *Statistics*. New York: Holt, Rinehart, & Winston.
- Hosenfeld, B., van den Boom, D. C., & Resing, W. C. M. (in press). Constructing geometric analogies for the longitudinal testing of elementary school children. *Journal of Educational Measurement*.
- Hosenfeld, B., van der Maas, H. L. J., & van den Boom, D. C. (in press). Detecting bimodality in the analogical reasoning performance of elementary school children. *International Journal of Behavioral Development*.
- Lewis, M. D. & Ash, A. J. (1992). Evidence for a neopiagetian stage transition in early cognitive development. *International Journal of Behavioral Development*, **15**, 337–358.
- Mitchell, M., & Hofstadter, D. R. (1990). The emergence of understanding in a computer model of concepts and analogy making. *Physica D*, **42**, 322–334.
- Mokken, R. J. (1971). *A theory and procedure of scale analysis with applications in political research*. Paris: Mouton.
- Poston, T., & Stewart, I. N. (1978). *Catastrophe theory and its applications*. London: Pitman.
- Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature*. New York: Bantam.
- Siegler, R. S. (1994). Unidimensional thinking, multidimensional thinking, and characteristic tendencies of thought. Chapter to appear in A. Sameroff & M. Haith (Eds.), *Reason and responsibility: The passage through childhood*. Chicago, IL: University of Chicago Press.
- Siegler, R. S., & Crowley, K. (1991). The microgenetic method. *American Psychologist*, **46**, 606–620.
- Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.
- Silverman, B. W. (1981). Using Kernel density function to investigate multimodality. *Journal of the Royal Statistical Society*, **43**, 97–99.

- Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, **84**, 353–378.
- Sternberg, R. J., & Rifkin, B. (1979). The development of analogical reasoning processes. *Journal of Experimental Child Psychology*, **27**, 195–232.
- Thelen, E., & Ulrich, B. D. (1991). Hidden skills: A dynamic systems analysis of treadmill stepping during the first year. *Monographs of the Society for Research in Child Development*, **56**, (1, Serial No. 223).
- Thom, R. (1975). *Structural stability and morphogenesis*. Reading, MA: Benjamin.
- Thomas, H. (1989). A binomial mixture model for classification performance: A commentary on Waxman, Chambers, Yntema, and Gelman (1989). *Journal of Experimental Child Psychology*, **48**, 423–430.
- Thomas, H., & Lohaus, A. (1993). Modeling growth and individual differences in spatial tasks. *Monographs of the Society for Research in Child Development*, **58**, (9, Serial No. 237).
- Thomas, H., & Turner, G. F. W. (1991). Individual differences and development in water-level task performance. *Journal of Experimental Child Psychology*, **51**, 171–194.
- Titterton, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.
- Vallacher, R. R., & Nowak, A. (1994). *Dynamical systems in social psychology*. San Diego, CA: Academic Press.
- Van der Maas, H. L. J. (1995). Beyond the metaphor? *Cognitive Development*, **10**, 621–642.
- Van der Maas, H. L. J., & Molenaar, P. C. M. (1992). Stagemwise cognitive development: An application of catastrophe theory. *Psychological Review*, **99**, 395–417.
- Van der Maas, H. L. J., Walma van der Molen, J., & Molenaar, P. C. M. (1993). Discontinuity in conservation acquisition: A longitudinal experiment. In H. L. J. van der Maas, *Catastrophe analysis of stagemwise cognitive development. Model, methods and application*. Doctoral dissertation, University of Amsterdam, The Netherlands.
- Van der Pol, F., Langeheine, R., & de Jong, W. (1991). *Panmark User Manual*. Voorburg: CBS.
- Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, **98**, 3–53.
- Vosniadou, S. (1989). Analogical reasoning as a mechanism in knowledge acquisition: A developmental perspective. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 413–437). Cambridge: Cambridge University Press.
- Vosniadou, S., & Ortony, A. (1989). *Similarity and analogical reasoning*. Cambridge: Cambridge University Press.

RECEIVED: February 20, 1995; REVISED: July 15, 1996