Developmental dynamics of declarative memory from infancy to childhood

Publikationsbasierte Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt dem
Fachbereich Psychologie und Sportwissenschaften
der Johann Wolfgang Goethe-Universität, Frankfurt am Main

von
Dipl.-Psych. Thorsten Kolling
geboren am 17.12.1977
in Neunkirchen (Saar)

18. Juni 2007

Vom Fachbereich Psychologie und Sportwissenschaften
der Johann Wolfgang Goethe- Universität, Frankfurt am Main
als Dissertation angenommen.

Dekan:
Prof. Dr. H. Moosbrugger

GutachterInnen:

Prof. Dr. Monika Knopf
Prof. Dr. Volker Hodapp

Datum der Disputation:
19.12.2007

Dedicated to my family

Acknowledgements

First and foremost, I like to thank my supervisor Prof. Dr. Monika Knopf for constantly criticising and encouraging me while working on the present dissertation. Although sometimes demanding, her persistent drive helped in all stages (designing, analysing and writing) of the present work. Furthermore, I would like to thank my colleagues, Dipl.-Psych. Stefanie Frahsek and Dr. Claudia Goertz for good (and hard) times in the FRAMES (Frankfurt Memory Study) project as well as Dipl.-Psych. Christina Mack for her very careful and empathic assistance during data acquisition. In addition, I would like to thank Margit Feyler for her technical assistance during data acquisition. I would like to thank all student research assistants of the FRAMES project for their constant engagement during data acquisition, data coding, and statistical analyses as well as for assistance in the publication process. Without all the parents and infants participating in the FRAMES study, this project would not have been possible. For proofreading several previous versions of the papers and the dissertation overview, I thank Dr. Regina Kressley-Mba, Dipl.-Psych. Judith Vogel, Dipl.-Psych. Tanja Schatz and M.Sc. Tina Spranger. I like to thank Prof. Dr. Volker Hodapp for reviewing the present dissertation. Further, I would like to thank the German Research Foundation for funding the present study (KN 275/3-1 and KN 275/3-3). Last but not least, many thanks to my partner Maria and my parents Hans-Jürgen and Yvonne Kolling for their emotional support.

June 18th, 2007 Dipl.-Psych. Thorsten Kolling

Table of Contents

Abstract

Deferred imitations assess declarative memory in infants. Many cross-sectional and a few longitudinal studies revealed that, with development, infants learn faster, and retain more target actions over longer retention intervals. Longitudinal stabilities are modest and increase through the second year. To date, there are only few multivariate deferred imitation studies pointing to interactions between declarative memory, language and self-development. However, as these studies applied variable-centered data analysis approaches, the individual stance was not taken into account.

Therefore, the present dissertation focuses on the explanation of inter-individual differences of deferred imitation through the second year. In the multivariate, longitudinal Frankfurt Memory Study (FRAMES), declarative memory (deferred imitation), non-declarative memory (train task), as well as cognitive, language, motor, social, emotional and body self-awareness development (Developmental Test for 6-month- to 6-year-olds, ET6-6) were assessed on three measurement occasions (12-, 18- and 24-month-olds).

From a psychometric perspective, sound tests for the assessment of deferred imitation in the respective age groups were developed (*Paper 1 & 2*). Reliability analyses (*Paper 3*) indicated relatively high short-term-stability for the deferred imitation test (12-month-olds). The co-development of declarative and non-declarative memory in 12- and 18-month-olds provided evidence for discriminative validity (*Paper 4*). Longitudinally, deferred imitation performance tremendously increased throughout the second year, and performance was moderately stable between 12 and 18 months and stability increased between 18 and 24 months.

Using a person-centered analysis approach (relative difference scores; cluster analysis), developmental subgroups were extracted out of the total sample. These groups differed in terms of mean growth and stability. However, between the first and second measurement occasion, the groups did not differ with respect to motor, cognitive and language development (*Paper 5*). Using the data of three measurement occasions, subgroups were extracted showing significant differences with respect to language, motor and body self-awareness development (*Paper 6*). The results are discussed against the background of infancy development theories.

I. Introduction

The art of remembering is the art of thinking.

When we wish to fix a new thing
in either our own mind or a pupil's,
our conscious effort
should not be so much to impress and retain it
as to connect it with something else already there.

The connecting is the thinking; and,
if we attend clearly to the connection,
the connected thing will certainly be likely
to remain within recall.

(*William James* on memory)

1. Learning and memory in infancy

Trying to resolve the beginnings of memory is one of the key issues in developmental psychology (Howe & Courage, 2004). In seminal studies, using the paired-comparison procedure, Fagan (1970) illustrated that (1) infant memory can be scientifically examined, (2) even very young infants are able to retain information over a considerable amount of time, and (3) many memory processes and variables of adult memory research (e.g., stimulus complexity, familiarization time) can be found in infants as well. Since then, other research groups have documented the impressive memory capacity of very young infants, primarily focusing on content-specific memory models, i.e. non-declarative and declarative memory (Tulving, 1985). Following this stream of research, the present publication-based dissertation deals with the appropriate measurement and inter-individual differences of intra-individual change of declarative memory in infancy.

In the following, the content-specific memory systems are introduced before a review of empirical cross-sectional and longitudinal findings of the respective memory tasks for non-declarative (train task) and declarative (deferred imitation) memory is undertaken. It is concluded that the understanding of individual differences of declarative memory performance is in need of improvement. Then, the argument that multivariate, longitudinal studies using person-centered statistical analysis approaches are necessary for understanding individual differences in the development of declarative memory is deduced. Potential correlates of declarative memory and variable- and person-centered approaches are introduced. A description of the present dissertation study's design and the participant characteristics follows, before psychometric properties of the deferred imitation tests and longitudinal findings of inter-individual differences of deferred imitation are reviewed. Finally, the results are discussed and integrated in recent research.

2. Content-specific memory systems, measurement methods, and development

Since the groundbreaking studies with the severely amnesic patient, H.M., it is argued that memory is dissociable into two (or even more) distinct systems with different functions and different operating principles (see Schacter & Tulving, 1994). Figure 1 depicts the memory systems as proposed by Tulving (1985).

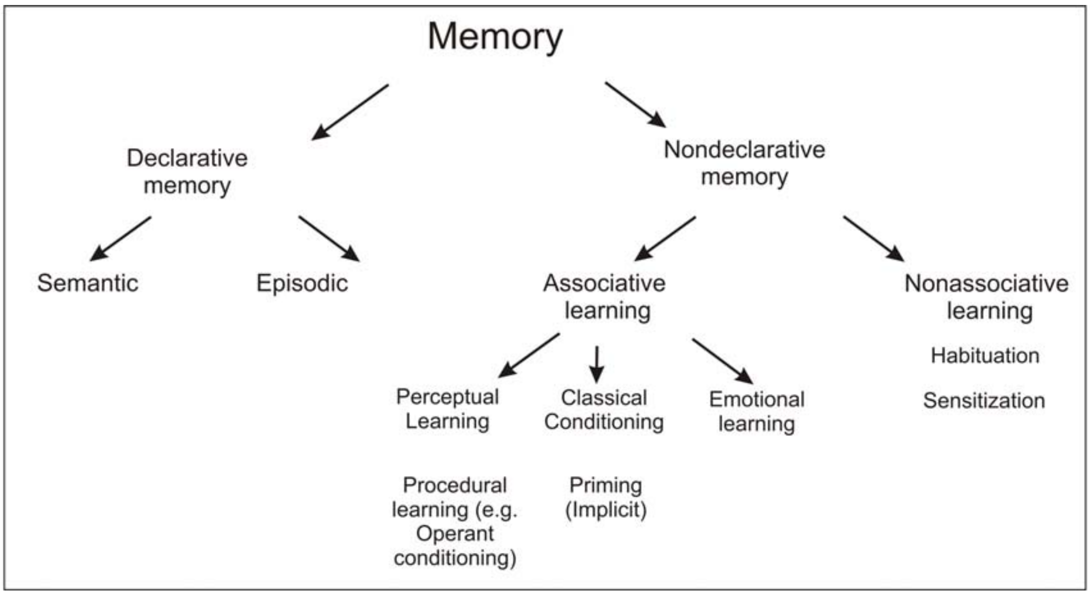


Figure 1. Content-specific memory systems

2.1. Non-declarative memory and its assessment in infancy

The non-declarative memory system is non-consciously accessible and stores information acquired via priming and operant conditioning processes, among others. Non-declarative memory in infants is assessed with the mobile task (up to 6 months of age) and the train task (from 6 months onwards). In these operant conditioning paradigms, the infant is confronted with a stimulus (mobile or train), and a specified response (leg kicking or lever pressing) is expected. In the mobile task, the infant is exposed to a mobile over her bed, which is attached to the infant’s leg via a ribbon. In the train task, the infant sits on her parent’s lap and sees a train, which is activated by pressing the lever in front of the infant. In a first phase, the baseline phase, there is no stimulus– response contingency and pre-experimental activity is recorded. In a second phase, the acquisition phase, each lever pressing leads to a movement of the train. In a third phase, the experimental phase, the stimulus-response-contingency is again deactivated. Depending on the experimental design, in a subsequent “yes”/”no” recognition test (Rovee-Collier, & Barr, 2001), infants are shown either the original stimulus or a different one.

With respect to the developmental timetable, infants as young as 8 to 10 months are able to learn the operant conditioning procedure and remember the stimulus (mobile) for one up to three days. In a wide range of cross-sectional studies (e.g., Rovee-Collier & Barr, 2001; Rovee-Collier, Hayne, & Colombo, 2001 for overviews) it was demonstrated that (1) there are age-related changes, (2) the retention interval increases linearly with age (Hartshorn, Rovee-Collier, Gerhardstein, Bhatt, Wondoloski, & Klein, 1998), (3) memory is increasingly context-independent (Borovsky & Rovee-Collier, 1990), (4) study time decreases with age, and (5) levels of processing affect memory performance in the mobile procedure (Adler, Gerhardstein, & Rovee-Collier, 1998). In studies with reminder procedures, i.e. reinstatement and reactivation, it was shown that reminders protract infants’ memories tremendously (Hartshorn et al., 1998).

2.2 Declarative memory and imitation learning in infancy

In contrast to the non-declarative memory system, the declarative memory system stores information acquired through processes which require conscious awareness. Declarative memory is divided further into two subsystems, namely the semantic and episodic memory systems. Whereas the encoding of factual knowledge about the world defines the semantic subsystem, the episodic subsystem encodes personal experiences and events. One of the most intensively discussed learning mechanisms related to declarative memory in infancy, is *imitation learning*.

In studies of direct facial imitations, Meltzoff and Moore (1977) revealed that infants' imitation is not reflex-driven or automatic but a congruent behaviour of the infant in relation to the human model. With development, imitative acts become more and more decoupled from the presence of the model; hence deferred imitations are observed.

2.2.1 Deferred imitation as declarative memory assessment

Based on earlier work by Piaget (1962), who described deferred imitations of his daughter Jacqueline in natural contexts, Meltzoff (1985) developed the deferred imitation procedure as a nonverbal measure of action-based declarative memory capacity. Within the deferred imitation task, an adult model acts with a series of novel objects with the infant just observing the actions. After a delay of several minutes, hours or even days, the adult gives the objects to the infant and imitative behaviour is observed. Depending on the experimental procedure and the aim of the study, the infant is given these unknown objects prior to the action demonstrations to study baseline behavior (Kressley & Knopf, 2006). A substantial increase in infants' target behavior in the experimental as compared to the control group indicates evidence of declarative memory.

Different arguments have been brought forward to strengthen the assumption that deferred imitation assesses declarative memory (Mandler, 2004; McDonough, Mandler, McKee, & Squire, 1995): (1) the cross-modal character of the deferred imitation task makes it improbable that infants' memory performance is due to priming processes, which are sensitive to modality changes, (2) in contrast to incremental learning processes, no motor exercise is possible, (3) within deferred imitation tests, new, unknown actions are used as memory items which were not available in the infants' motor repertoire before, (4) amnesiacs are unable to show deferred imitation, (5) infants do not only produce single actions but encode structural elements between actions, i.e. they encode item-relational information (Knopf, Kraus, & Kressley-Mba, 2006).

2.2.2 Cross-sectional and longitudinal studies of deferred imitation

There is evidence from several studies that 6-month-old infants show deferred imitation after short retention intervals (e.g., Collie & Hayne, 1999; Hayne, Boniface, & Barr, 2000; Kressley-Mba, Lurg, & Knopf, 2005), and that 9-month-old infants retrieve formerly seen actions after longer retention intervals (e.g., Barr & Hayne, 1996; Hayne et al., 2000; Kressley-Mba, Lurg, & Knopf, 2005; Meltzoff & Moore, 1999). There are age-related changes in deferred imitation parameters, namely that with increasing age infants need less exposure to the target actions; hence they learn faster than younger ones (Barr, Dowden, & Hayne, 1996). Furthermore, in comparison to younger infants, older infants imitate actions after longer retention intervals (e.g., Barr & Hayne, 2000).

Recent longitudinal studies, in which deferred imitation memory performance was repeatedly tested, demonstrated that declarative memory performance increases, and that the individual stability is modest in early infancy and increases throughout the second year (Heimann & Meltzoff, 1996; Nielsen & Dissanayake, 2004). This age-related increase in stability of individual memory differences may be due to an increase in reliability of assessment, or may indicate the successive stabilization of one memory ability. A third possible interpretation of this finding is that imitation memory performance may mirror qualitatively different memory performances in different age groups. While imitation memory performance may indicate semantic memory in young infants ("this object is for doing such an action"), the same memory task may express episodic memory in older children ("I have seen before that this object can be used for such an action"). The last interpretation relates to the theory of episodic memory development in infants (Knopf, Mack, & Kressley-Mba, 2005; Howe & Courage, 1997), which focuses on the conjoint development of the self and the emergence of episodic remembering in infants and children. Knopf, Mack, and Kressley-Mba (2005) argue that the emergence of the categorical self, which develops around 18 months, constitutes a lower boundary for the development of episodic memory.

Taken together, the studies using the deferred imitation paradigm reviewed above indicate that 6-month-olds are capable of deferred imitation. Memory capacity and retention intervals increase with age, and longitudinal stabilities of deferred imitation are modest in infancy. Furthermore, the self is an important moderator of declarative memory development.

However, most of these studies investigated deferred imitation development with univariate, cross-sectional designs. Multivariate analyses focusing on individual differences of deferred imitation are still rare in the research literature. For this reason and following related recent claims (Jones & Herbert, 2006), the analysis of individual differences of deferred imitation is a necessary next research step for understanding relationships between declarative memory and other important developmental correlates.

2.2.3 Multivariate (longitudinal) studies of deferred imitation

Individual differences research of deferred imitation requires longitudinal, multivariate studies as well as adequate statistical techniques to understand the co-varying factors of deferred imitation development. So far, only some studies have reported correlations and predictive regressions between deferred imitation performance and other important developmental variables.

Both cross-sectional and longitudinal studies have demonstrated that declarative memory relates to language development. In a cross-sectional study combining deferred imitation items with verbal cues, infants at the age of 24 months used language cues effectively in a context change condition, whereas 18-month-olds were not able to do so (Herbert & Hayne, 2000). In a longitudinal study, Heimann, Strid, Smith, Tjus, Ulvund, and Meltzoff (2006) investigated the relationship between recall memory, visual recognition memory, social communication, and the emergence of language skills. They reported that visual recognition memory (at 6 months), deferred imitation (at 9 months) and turn-taking skills (at 14 months) predicted language skills at the age of 14 months, with deferred imitation accounting for the highest amount of variance in their regression model. Furthermore, the authors found that deferred imitation and joint attention tested at the age of 9 and 14 months predicted cognitive abilities at the age of 4 years (Strid, Tjus, Smith, Meltzoff, & Heimann, 2006). These studies clearly demonstrate that language and declarative memory are related in the second year.

There are also important co-variations between declarative memory performance and self-development in the second year. Mirror self-recognition, considered as the benchmark test for the development of a self-concept in infancy, is one of the important milestones in the second year of life, typically solved around 18 months. Deferred imitation performance co-varied with the development of a self-concept (mirror self-recognition) in 20-month-old children (Prudhomme, 2005). By combining deferred imitation and mirror self-recognition, Prudhomme demonstrated that children who passed the mirror test were less affected by a change of colours of the relevant target props given within the deferred imitation test than those who did not pass the mirror test. This result is assumed to indicate that the more personal features infants are able to embed in the memory traces, the more memory processing is elaborative and hence rather episodic-like. Furthermore, she concludes that her findings demonstrate the role of the cognitive self as a factor of differentiation between semantic and episodic memory.

Using a multivariate, longitudinal design, Nielsen and Dissayanake (2004) assessed deferred imitation, synchronic imitation, cognitive development (pretend play) and self-development (mirror self-recognition) through the second year on four measurement occasions. The authors found that deferred imitation develops prior to synchronic imitation, pretend play and mirror self-recognition, and that the development of the latter three skills followed a similar developmental trajectory between 18 and 21 months. In this study, using a variable-centered approach (e.g., Spiel, 1998), Pearson’s product–moment correlation coefficients between measurement occasions and constructs as well as mean performance scores were reported. As variable-centered approaches analyze group stabilities at aggregate levels, they do not allow for a differential analysis of individual stability or lability of memory performance over time (Asendorpf, 1990; Ghiselli, 1956, 1960).

2.3. Variable- and person-centered analysis approaches

The variable-centered approach focuses on normative stability, which is defined as "the preservation of a set of individual ranks on a quality within a constant population over a specified amount of time" (Alwin, 1994, p. 139). Thereby, individual stability reflects the relative position of an individual within a reference group across time. Following this view, stability is independent of intra-individual change, but rather reflects the absence of inter-individual differences of intra-individual change. In the variable-centered approach, the focus of interest is on single or combinations of variables, and on correlations between these variables.

In contrast, the person-centered approach focuses on the person and therefore on inter-individual differences of intra-individual change. There are exploratory and confirmatory person-centered analysis approaches, i.e. longitudinal cluster analyses and longitudinal configural frequency analyses. Generally, these approaches allocate the subjects under study into different developmental groups. In a next step of the analysis, intra-individual change is explained by important correlating variables.

The evidence reviewed above demonstrates that language, self-development and declarative memory development are substantially related implying that these constructs are both theoretically and empirically important. However, it is far from clear, how exactly these domains (language, memory, and self-development) interact and how these interactions lead to inter-individual differences of intra-individual change in infancy.

3. *The present dissertation: A multivariate, longitudinal study with person-centered analyses*

To promote differential developmental psychological research in the infancy period, the present dissertation analyzes inter-individual differences of intra-individual change of deferred imitation and its correlates in infancy. Figure 2 depicts the multivariate, 3-wave longitudinal design of the present study.

Declarative memory (deferred imitation) was assessed on three measurement occasions, at 12 (Frankfurt Imitation Test for 12-month-olds, *FIT 12)*, 18 (Frankfurt Imitation Test for 18-month-olds, *FIT 18*) and 24 months (Frankfurt Imitation Test for 24-month-olds, *FIT 24)*. Non-declarative memory (train task) was assessed at the ages of 12 (*TT 12*) and 18 months (*TT 18*). Furthermore, the Developmental Test for 6-month to 6-year olds (*ET6-6;* Petermann & Stein, 2000), comprising six developmental factors (motor, cognitive, language, social, emotional development and body self-awareness) was administered at all three measurement occasions (denoted *ET 12*, *ET 18, and ET 24* in the following).

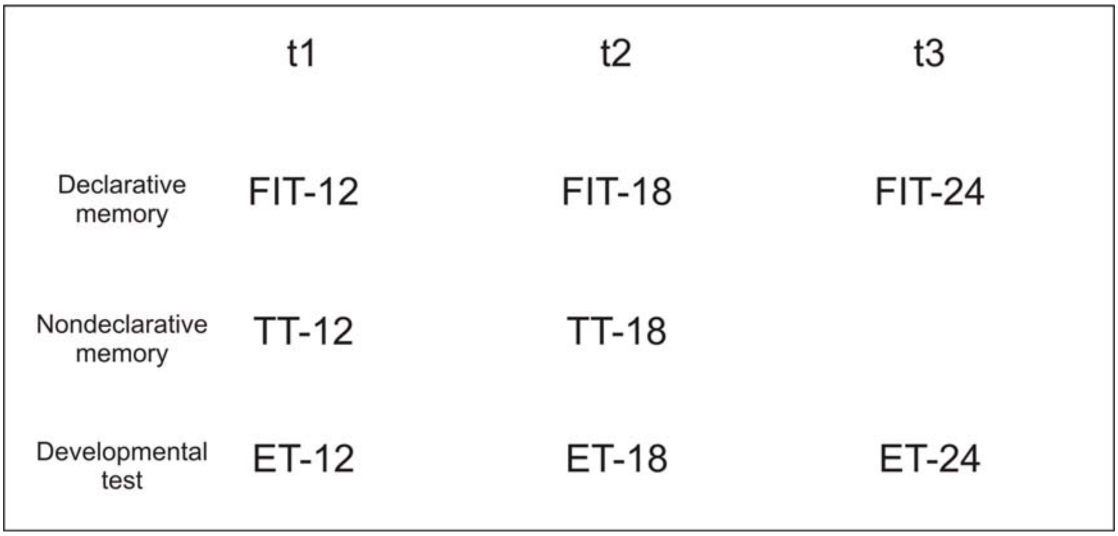


Figure 2. Study design

4. Participants

The sample of the longitudinal Frankfurt Memory Study (FRAMES) consisted of N = 92 infants (N = 48 boys) who were recruited via radio announcement and advertisements in childcare centres and local paediatricians, and by word of mouth. Four subjects did not continue the study to the second testing because of relocation and one infant was excluded from data analysis because of fuzziness. The criteria for admission into the study were no known physical, sensory or mental handicaps, normal length of gestation (over 37 weeks) and normal birth weight (2500 – 4500 grams). The mean age of the subjects at the first measurement occasion was *M* = 362 days (*SD* = 9) and *M* = 551 days (*SD* = 8) at the second testing. At the third measurement occasion children were *M* = 731 (*SD* = 10) days old. Mean birth weight was *M* = 3393 grams (*SD* = 507). Parents reported an *APGAR index* of *M* = 9.8/9.9 with a minimum value of seven.

II. Integrative review of dissertation relevant papers

Most of the studies presented above did not take into account test-theoretical measurement issues with respect to, for example, reliability and validity issues. However, highlighting questions of reliability and validity in measurement issues bridges the gap between experimental psychology and individual differences research. Therefore, in a *psychometric section*, test construction, reliability and validity issues of deferred imitation are discussed, and test construction and test-theoretical refining of the newly developed deferred imitation tests are presented. *Paper 1* and *paper 2* describe the test construction of deferred imitation tests for 12-, 18- and 24-month-olds (Frankfurt Imitation Tests; FIT 12, FIT 18, FIT 24) and report the test characteristics. Moreover, *paper 3* describes the FIT 12 reliability, assessed with a 1-week-test-retest interval. Both FIT 12 and FIT 18 are analyzed concurrently with the train task (non-declarative memory) to obtain a measure of discriminant validity (*paper 4*).

In a subsequent *individual differences section,* inter-individual differences of intra-individual change of deferred imitation through the second year are investigated longitudinally. In *paper 5,* two measurement occasions are analyzed with respect to longitudinal mean growth and stability (variable-centered), and group solutions are used to analyze inter-individual differences of intra-individual change (person-centered). These developmental groups are compared with respect to language, self- and cognitive development in order to explain individual differences.

Paper 6, building up on *paper 5*, uses both variable- and person-centered analyses to extract developmental groups through the second year (three measurement occasions). These developmental groups are then compared with respect to cognitive, language, social, emotional, motor and self-development.

Overall, the present dissertation provides a more complete picture of deferred imitation development in the second year.

III. Dissertation relevant papers– Overview

(see Appendix for papers)

Paper 1

Goertz, C., Knopf, M., **Kolling, T**., Frahsek, S., & Kressley-Mba, R. A. (2006). Entwicklung und Erprobung eines Messinstruments zur Erfassung des deklarativen Gedächtnisses Einjähriger: Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12). *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38,* 88-96.

Paper 2

Goertz, C., **Kolling, T.**, Frahsek, S., & Knopf, M. (subm.). Der Frankfurter Imitations-Test für 18 Monate und 24 Monate alte Kinder (FIT 18-24). *Kindheit und Entwicklung.*

Paper 3

Goertz, C., **Kolling, T**., Frahsek, S., Stanisch, A., & Knopf, M. (2007). Assessing declarative memory in 12-month-old infants: A test-retest reliability study based on the deferred imitation task. *European Journal of Developmental Psychology*, in press. Epub ahead of print retrieved May, 28, 2007, from http://www.informaworld.com/smpp/content~content=a778058220~jumptype=rss

Paper 4

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (subm.). Declarative and non-declarative memory in 12- and 18- month-old infants. *Infancy.*

Paper 5

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (in press). Stability of deferred imitation in 12- to 18-month-old infants: A closer look into developmental dynamics. *European Journal of Developmental Psychology*.

Paper 6

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (subm.). Dynamics of declarative memory from infancy to childhood. *Child Development.*

IV. Psychometrics

1. Construction principles of deferred imitation tests

The construction and pilot testing of the deferred imitation tests applied in the present longitudinal study took into account the most important test development requirements. Generally, structural and differential continuity, the relationship between item difficulty and differential validity as well as the longitudinal shrinkage of variance due to ceiling effects are very important pre-considerations. Structural continuity (Bates & Nowosad, 2005) concerns the degree of constancy in the operational definition of a trait over time, hence developmental measurement equivalence (Hartmann, 2005). Differential continuity (Bates & Novosad, 2005) reflects stability insofar as a person scoring high (low) on a trait remains high (low) in further measurements. It is well known that a test composed of 50% difficulty yields the highest potential differential validity. Furthermore, the item difficulties have to be uniformly distributed over the whole item difficulty distribution, ideally from P = .05 to P = .95 (Gulliksen, 1950). In addition, the problem of the shrinkage of variance had to be taken into account, namely floor and ceiling effects. It can be shown statistically that because of the shrinkage of variance of floor as well as ceiling effect, a lower reliability and lower longitudinal stability correlations (test-criterion correlations) result.

To obtain sound tests for deferred imitation, items for both 12-month-olds and 18-month-olds were pilot tested in several independent studies. Furthermore, control groups (12- and 18- month-olds) were assessed to obtain mean baseline performance and mean test performance. The deferred imitation items finally adopted were chosen among the potential items in the pre-tests using the criteria that (1) they yield good inter-scorer reliability, (2) they involve uniformly distributed item difficulties with a total mean item difficulty of about 50 percent, (3) the items reflect several facets of deferred imitation (number of steps, causal item constraints, goal-relevant *vs*. goal-irrelevant steps), and (4) they comprise formerly unknown actions that infants in the different age groups are able to perform in terms of motor skills. The tests finally applied consisted of these items adjusted for 12-, 18- and 24-month-olds, respectively.

2. *Paper 1*: *Test construction and test characteristics of the FIT 12 (co-author)*

In *paper 1* of the present dissertation (Goertz, Knopf, **Kolling**, Frahsek, & Kressley-Mba, 2006), the test characteristics of the *FIT 12* were analyzed. It was found that the 12-month-olds remembered M = 2.38 (SD = 1.3)[1] target actions. The test characteristics indicated that item difficulties were normally distributed, and no sex differences were found. The analysis of test-item-correlations revealed good indices for all target actions (r_{it} = .43 to r_{it} = .71). Furthermore, no relationship was found between latency of target action imitation and memory performance. The analysis of a high- vs. a low-imitation group provided evidence that higher imitation relates to higher expressive language development. There were no differences for motor development, indicating that the deferred imitation items used are independent of motor development, but that language might play a significant role in declarative memory performance. Finally, data of the FIT 12 indicated that memory capacity might be around three items in this age range.

The author of the dissertation contributed to this paper by analyzing the test statistics and writing several portions (methods & results) of the paper. This paper was printed in *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* (*impact factor*: 0.53).

3. *Paper 2: Test construction and characteristics of the FIT 18 and the FIT 24 (co-author)*

In *paper 2* (Goertz, **Kolling**, Frahsek, & Knopf, submitted), the test characteristics of the *FIT 18* and the *FIT 24* were analyzed. The 18-month-olds remembered M = 6.9 (SD = 1.85) target actions and the 24-month-olds remembered

[1] Note: In the first analysis of target action imitation of the FIT 12, the target actions drum and cup & knife were scored for total action completion. In the successive analysis (starting with paper 3), these actions were coded with 2 points as these actions are 2-stepped actions. Therefore, the maximum sum score of FIT 12 was 7 in successive analysis.

M = 17.9 (*SD* = 3.78) target actions. The results indicated that both FIT 18 and FIT 24 are sound tests for the assessment of declarative memory in infants as they yielded acceptable item difficulties (normally distributed) and item total correlations and test sum scores (normally distributed). Moreover, control and baseline studies demonstrate that both FIT 18 and FIT 24 assess declarative memory performance.

Furthermore, both FIT 18 and FIT 24 correlate significantly with the ET6-6 factors cognitive development and social development. The substantial correlation between declarative memory and social development at the age of 18 months is due to the test dimension “interaction with peers” (i.e. synchronic play). The correlation with the factor cognitive development (24 months) traces back to items, which include immediate imitation performance (e.g., imitating to build a tower with three blocks). In addition, the FIT 24 correlated significantly with receptive language development.

The author of the dissertation contributed to this paper by analyzing the test statistics and writing several portions of all parts of the paper (introduction, method, results and discussion). This paper is *under review* in the journal *Kindheit und Entwicklung* (*impact factor*: 1.80).

4. Paper 3: Reliability of the FIT 12 (co-author)

In paper 3 (Goertz, **Kolling**, Frahsek, Stanisch, & Knopf, in press), the test-retest-reliability of the deferred imitation task for 12-month-olds was assessed. Using a test-retest-control group design, it was found that the short-term stability of the FIT 12 was good (r = .52*). After the exclusion of two outliers, identified with the individual consistency approach by Asendorpf (1990), the stability increased to r = .62*. Considering the high developmental dynamics in infants at the end of their first year, this is a reliability appropriate to that age group. Furthermore, it is comparable to the reliabilities reported in studies with other different indicators of information processing in infants varying between r = .30 and r = .45 (e.g., McCall & Carriger, 1993). In addition, memory performance increased from test 1 to test 2, which indicates a memory benefit, similarly found in older children and adults. This result further supports that deferred imitation assesses declarative memory.

The author of the dissertation contributed to this paper by analyzing the test statistics and writing several portions of all parts of the paper (introduction, method, results and discussion). This paper is *pre-published online* in the *European Journal of Developmental Psychology* (*impact factor*: not yet obtainable).

5. Paper 4: Discriminant validity of the FIT 12 and FIT 18 (first author)

In *paper 4* (**Kolling**, Goertz, Frahsek, & Knopf, under review), the dissociation between declarative and non-declarative memory was analyzed using a cross-lagged panel design. As many infants (n = 41) did not tackle the train task procedure for several reasons (crying, fear, problems with the procedure) at both measurement occasions, the analyses of this paper relate to a subsample of n = 51 infants. With respect to deferred imitation, it was shown that performance increased with age. Furthermore, the subsample showed a moderate stability correlation for deferred imitation performance (r = .31*). Regarding the train task procedure (TT 12, TT 18), the raw analysis of mean lever pressing rates and the immediate retention ratio did not yield interpretable results because the baseline pressing rate for the infants who learned the task was too high. Therefore, qualitative analyses were done afterwards. Showing a satisfying inter-rater reliability, qualitative analyses demonstrated that n = 30 infants learned the train task procedure (59 %) at the first measurement occasion (TT 12), whereas n = 21 infants (41 %) did not learn the task. At the second measurement occasion (TT 18), n = 8 infants (35 %) learned the task whereas n = 15 infants (65 %) did not learn the task. The correlation between measurement occasions for the train task was comparatively low (r = .23).

In sum, the autocorrelations within the cross-lagged panel design illustrate that stability was rather high for deferred imitation (FIT 12, FIT 18) and somewhat lower for operant conditioning. Furthermore, deferred imitation did not correlate with the operant conditioning performance in the train task (synchronous correlation). The calculation of the synchronous correlations revealed that the relationship between the different memory tasks (declarative and non-declarative memory) was rather low (r = -.15 for t_1 and r = .06 for t_2). Therefore, both memory components differentiate early in development. Besides theoretical implications, this result provides further evidence in terms of discriminant validity and provides further support that FIT 12 and FIT 18 assess declarative memory.

The author of the dissertation wrote this paper as the main author. This paper is still *under review* in the journal *Infancy* (*impact factor*: 1.303).

V. Individual differences of deferred imitation

1. Paper 5: Two measurement occasions (12- and 18-month-olds) (first author)

In *paper 5* (**Kolling**, Goertz, Frahsek, & Knopf, in press), memory performance of the first two measurement occasions (12- and 18-month-old infants) was analyzed with both a variable- and a person-centered approach. As deferred imitation performance has primarily been analyzed with variable-centered approaches in longitudinal studies (e.g., stability correlation for the total group), this was the first study which took into account a person-centered perspective. Therefore, the paper aimed at the explanation of inter-individual differences of intra-individual change in deferred imitation. To further explain differences between developmental groups, the analyses included developmental data of the Developmental Test for 6-month to 6-year olds (language, cognitive, and motor development).

Regarding memory performance development, the 12-month-olds remembered $M = 4.0$ ($SD = 1.56$) target actions and the 18-month-olds remembered $M = 6.86$ ($SD = 1.86$) target actions. Although the reliability for FIT 12 was rather high (Goertz et al., in press, *paper 3*), the variable-centered analysis demonstrated that the longitudinal stability correlation for the total sample was rather low ($r = .23$; corrected for attenuation $r = .32$*). For a detailed analysis of the deferred imitation data, a person-centered approach was used in successive statistical analyses. First, the repeated measurement data were scaled with a usual z-transformation. Second, relative z-difference scores (RD scores) were calculated. A categorization of these RD scores (median-split vs. trichotomization) yielded a two and a three group solution.

In the two group solution, a moderate growth group showed high performance at the first measurement occasion and low performance at the second measurement occasion – relative to the total sample. This group remembered *M* = 4.97 (*SD* = 1.22) target actions on the first, and *M* = 6.10 (*SD* = 1.60) on the second measurement occasion. A high growth group showed low performance at the first measurement occasion and high performance at the second measurement occasion. This group remembered *M* = 3.19 (*SD* = 1.26) target actions on the first, and *M* = 7.60 (*SD* = 1.83) on the second measurement occasion. Stability correlations were higher for the subgroups (r_{MG} = .57*; r_{HG} = .69*) than for the total sample.

In a three group solution, a low, a moderate and a high growth group were separated. The low growth group remembered *M* = 5.37 (*SD* = 0.88) target actions on the first, and *M* = 5.70 (*SD* = 1.35) on the second measurement occasion. The moderate growth group remembered *M* = 3.80 (*SD* = 1.27) target actions on the first, and *M* = 6.50 (*SD* = 1.41) on the second measurement occasion. The high growth group remembered *M* = 3.17 (*SD* = 1.47) target actions on the first, and *M* = 8.28 (*SD* = 1.83) on the second measurement occasion. Stability correlations were higher for the subgroups (r_{LG} = .64*; r_{MG} = .95*; r_{HG} = .81*) than for the total sample. This person-centered analysis showed that groups with differential mean growth and stability could be extracted from the total sample. In a last step of analysis, the comparison of these groups with respect to developmental differences in cognitive, language and motor development revealed no significant differences. In paper 6 (see below), however, group differences were further analyzed with data of three measurement occasions.

The author of the dissertation wrote this paper as the main author. This paper is *in press* in the *European Journal of Developmental Psychology* (*impact factor*: not yet obtainable).

2. Paper 6: Three measurement occasions (12-, 18- and 24-month-olds) (first author)

In paper 6 (**Kolling**, Goertz, Frahsek, & Knopf, under review), three measurement occasions (12-, 18- and 24-month-olds) were analyzed with both variable- and person-centered analyses. The longitudinal analysis of this paper revealed that memory performance throughout the second year increases from M_{t1} =

4.2 (*SD* = 1.5) over M_{t2} = 6.9 (*SD* = 1.9) to M_{t3} = 17.0 (*SD* = 3.8) target actions, replicating results of cross-sectional (Barr & Hayne, 1996; Hayne et al., 2000) and longitudinal studies (Heimann & Meltzoff, 1996; Nielsen & Dissanayake, 2004). However, as those studies used deferred imitation tests with fewer items, the present study showed more clearly than before that declarative memory performance tremendously increases through the second year.

Furthermore, the stability indices increased with age (*r* = .17 between 12 and 18 months and *r* = .37* between 18 and 24 months) also replicating results found in other studies (Heimann & Meltzoff, 1996; Nielsen & Dissanayake, 2004). The stability indices were almost identical to those reported by Nielsen and Dissanayake (2004) providing a cross-validation of the former results. Therefore, the conclusion that, from an empirical perspective, inter-individual differences of intra-individual change (developmental dynamics) are high in the first two years of life is reasonable.

A cluster analysis (relative z-difference scores scaled) yielded two cluster groups. Thereby, in terms of mean growth, the first cluster group (*n* = 45) increased from M_{t1} = 3.4 (SD_{t1} = 1.4) over M_{t2} = 7.7 (SD_{t2} = 1.9) to M_{t3} = 17.2 (SD_{t3} = 3.5) target actions. The second cluster group (*n* = 33) increased from M_{t1} = 5.1 (SD_{t1} = 1.0) over M_{t2} = 5.8 (SD_{t2} = 1.3) to M_{t3} = 17.2 (SD_{t3} = 4.3) target actions. An analysis of variance for repeated measures revealed a significant linear trend and a significant interaction, demonstrating that both cluster groups are increasing with respect to declarative memory performance. Furthermore, the significant interaction effect shows that both groups developed differentially. Whereas the first cluster group had a lower memory performance of approx. 1 standard deviation at the first measurement occasion than the second cluster group, the first cluster group outperforms the second cluster group by approx. 1 standard deviation at the second measurement occasion. At the third measurement occasion differences between the cluster groups equalled out.

With respect to stabilities, the first cluster group (*n* = 45) showed high, significant correlations between t_1 and t_2 (r = .66, $p < 0.01$) and between t_2 and t_3 (r = .54, $p < 0.01$). For the second cluster group (*n* = 33) the correlation was high and significant between t_1 and t_2 (r = .53, $p < 0.01$) and moderate between t_2 and t_3 (r = .29, ns).

The multivariate analysis of the developmental factors (ET6-6) yielded a significant difference for the total score of the developmental test at the second measurement occasion, such that those infants of cluster group 2 showed a higher mean developmental sum score at the age of 18 months than cluster group 1. This difference is due to the differences in the developmental factors language, motor and body self-awareness development. This result is, on the one hand, in line with theories of episodic memory development (Howe & Courage, 1997; Knopf, Mack, & Kressley-Mba, 2005) stating that episodic memory co-develops with the self. On the other hand, the difference in language development also corresponds to recent findings. Studies by Herbert and Hayne (2000) and Heimann et al. (2006) demonstrated that language is an important variable for declarative memory performance. How exactly language is related to declarative memory needs to be investigated in more detail in future research.

Furthermore, the cluster groups differed with respect to motor development. However, although it is certainly arguable that motor components are somewhat important in the action-based deferred imitation task, the relatively small difference between groups does not stand in contrast to the importance of language and self-development highlighted above.

The author of the dissertation wrote this paper as the main author. This paper is *under review* in the journal *Child Development* (*impact factor*: 3.021).

VI. General discussion

Using a multivariate, longitudinal (three waves) design, the present dissertation focused on the analysis of inter-individual differences of intra-individual change in memory development. To analyze these individual differences, reliable and valid tests were required. The test parameters, i.e. mean item difficulties, normal distribution of item difficulties, item-total correlations, and control group performance (baseline testing) illustrate that the construction of the Frankfurt Imitation Test (FIT 12, FIT 18, FIT 24) was sound (*paper 1* and *paper 2*). In addition, it was demonstrated that memory performance can be reliably assessed using the FIT 12 (*paper 3)*. A one week retest interval yielded a high reliability (r = .52*; corrected for attenuation r = .62*). The reliability analyses of FIT 18 and FIT 24 are currently in progress. Moreover, the cross-lagged panel analysis of FIT 12 and FIT 18 with the data of the operant conditioning procedure, the train task (TT 12, TT 18, respectively), provide evidence for the discriminative validity of both FIT 12 and FIT 18 (*paper 4*). Taken together, the newly developed Frankfurt Imitation Tests were appropriate to analyze inter-individual differences of intra-individual change of deferred imitation. In the core analysis of the present dissertation, inter-individual differences of intra-individual change were analyzed longitudinally (*paper 5* & *paper 6)*. These analyses revealed that declarative memory performance increases tremendously through the second year, and declarative memory performance was moderately stable between 12 and 18 months (r = .23) and stability increased between 18 and 24 months (r = .37*). Using person-centered approaches (relative z-differences scores, cluster analysis), inter-individual differences of intra-individual change were analyzed by extracting developmental groups. These groups were compared with respect to important developmental correlates in multivariate analyses. Between the first (12 months) and second (18 months) measurement occasion, no significant differences between the developmental groups were found, neither for the two nor for the three group solution (*Paper 5*). However, when taking into account three measurement occasions, significant differences were found at the second measurement occasion (18 months) with respect to three developmental factors, namely language, motor and body self-awareness development (*Paper 6*).

Taken together, declarative memory performance was related to language development both cross-sectionally *(paper 1 & paper 2)* and longitudinally *(paper 5)*. Therefore, the results of the present dissertation clearly demonstrate that language and deferred imitation interact in infant development (e.g., Heimann et al., 2006; Herbert & Hayne, 2000b; Strid et al., 2006).

Moreover, cross-sectional relations between immediate, synchronic and deferred imitation were found (*paper 1 & paper 2*), replicating earlier research findings (e.g., Nielsen & Dissanayake, 2004).

With respect to body self-awareness, the results of the present dissertation show that self-development is an important factor for the developmental dynamics found in the 18-months period. Even if the present results do not yet show unambiguously exactly how the development of the self interacts with declarative memory, the results of the present study indicate that the developmental theory of episodic memory (Howe & Courage, 1997; Knopf, Mack & Kressley-Mba, 2005) is worth to be considered empirically.

Furthermore, the two cluster groups differed with respect to motor development. However, although it is certainly arguable that motor components might be subsidiary important in the action-based deferred imitation task, the relatively small difference between groups does not contrast to the importance of language and self-development highlighted above.

Future research needs more multivariate, longitudinal studies using person-centered analysis approaches (exploratory & confirmatory) to understand the continuities (of memory) from infancy to childhood. If the number of measurement occasions and the sample size is appropriate, the application of newer statistical techniques (e.g. growth curve modelling) for the analysis of the data will provide a better understanding of correlates of intra-individual change of memory. Furthermore, these future studies should take into account exploratory and confirmatory person-centered data analysis approaches to understand the cognitive processes and structures of infant cognition from an individual perspective.

VII. Zusammenfassung

1. Verzögerte Imitation – Quer- und längsschnittliche Befunde

Seit den Studien zur Imitation von Gesichtsgesten von Meltzoff und Moore (1977) gilt Imitationslernen als bedeutender Lernmechanismus bei Säuglingen. Mit fortschreitender Entwicklung, spätestens ab dem 6. Lebensmonat (z.B. Collie & Hayne, 1999), werden zudem verzögerte Imitationen beobachtet. Zur Messung dieser Fähigkeit werden Säuglingen neuartige Objekte und dazugehörige Handlungen präsentiert und nach einer Verzögerungsphase (Stunden, Tage, Wochen) wird das handlungsbezogene Verhalten der Säuglinge beobachtet (Meltzoff, 1985). In Abhängigkeit des Versuchsdesigns werden Basisraten (Baselines) erhoben, um die spontane Ausführungsrate durch die Säuglinge zu bestimmen. Der kreuzmodale Charakter der Aufgabe (Mandler, 2004), der Ausschluss motorischen Lernens, die Nutzung unbekannter Handlungen, klinische Vergleichsstudien (Amnestiker) sowie die Enkodierung Item-relationaler Information (z.B. Knopf, Kraus & Kressley-Mba, 2006) indizieren, dass die Verzögerte Imitation den Leistungen des deklarativen Gedächtnisses zuzurechnen ist. Im Rahmen des inhaltsspezifischen Gedächtnismodells nach Tulving (1985) werden bewusste Abrufprozesse den Leistungen des deklarativen Gedächtnisses zugeordnet, während nicht bewusste Abrufprozesse den Leistungen des nicht-deklarativen Gedächtnisses zugeordnet werden.

Querschnittstudien zur Verzögerten Imitation haben gezeigt, dass Säuglinge und Kleinkinder mit zunehmendem Alter mehr Items enkodieren (z.B. Barr & Hayne, 2000), die Lerngeschwindigkeit zunimmt (z.B. Barr & Hayne, 1996) sowie das Retentionsintervall ansteigt (z.B. Herbert & Hayne, 2000b). In den bisher vorliegenden Längsschnittstudien zeigten sich ansteigende Leistungen des deklarativen Gedächtnisses im Entwicklungsverlauf sowie moderate Stabilitäten (z.B. Heimann & Meltzoff, 1996; Nielsen & Dissanayake, 2004). In den letzten Jahren wurden verstärkt Längsschnittstudien sowie differentielle Analysen (Jones & Herbert, 2006) im Rahmen der Forschung zur verzögerten Imitation gefordert.

Die wenigen multivariaten Längs- sowie Querschnittstudien deuten darauf hin, dass sowohl Sprache (Herbert & Hayne, 2000) als auch das Selbst (Prudhomme, 2006) als wichtige Korrelate deklarativer Gedächtnisleistungen in Frage kommen.

Die vorliegende Arbeit widmet sich daher im Rahmen eines multivariaten Längsschnittdesigns der Frage nach inter-individuellen Differenzen intra-individueller Veränderungen der Verzögerten Imitation im Verlauf des zweiten Lebensjahres.

2. Design und Ziele der Arbeit

In der Längsschnittstudie FRAMES (Frankfurt Memory Study), bei der N = 92 Kinder im Alter von 12, 18 und 24 Monate untersucht wurden, wurden deklarative (Verzögerte Imitation) und nicht-deklarative (Zug-Aufgabe) Gedächtnisleistungen sowie Entwicklungsdaten mit dem Entwicklungstest für 6 Monate bis 6 Jahre alte Kinder (ET6-6; Petermann & Stein, 2000) erhoben. In einem psychometrischen Teil steht die Entwicklung von Messinstrumenten zur Verzögerten Imitation für den untersuchten Altersbereich (12, 18, 24 Monate) im Vordergrund. Im zweiten Teil der Arbeit werden inter-individuelle Differenzen der längsschnittlichen Entwicklungsverläufe dargestellt.

3. Psychometrie

Im ersten Teil der vorliegenden Arbeit wurden Messinstrumente zur Erfassung der Verzögerten Imitation für den Altersbereich von 12 Monaten (Frankfurter Imitations Test für 12 Monate alte Kinder, *FIT 12*, *Artikel 1*) sowie von 18 und 24 Monaten (*FIT 18*, *FIT 24*, *Artikel 2*) entwickelt und hinsichtlich ihrer Eignung getestet. Die testtheoretischen Analysen des FIT 12 (Goertz, Knopf, Kolling, Frahsek & Kressley-Mba, 2006) zeigen, dass die Itemschwierigkeiten normalverteilt sind, keine Geschlechtsunterschiede auftreten, kein Zusammenhang zwischen Latenz und Gedächtnisleistung besteht sowie die mittlere Gedächtnisleistung von Einjährigen bei ca. 3 Items liegt. In einer Reliabilitätsstudie mit einem Test-Retest-Intervall von einer Woche wurde eine für den untersuchten Altersbereich sehr gute Reliabilität (r = .52*, minderungskorrigiert *r* = .62*) ermittelt (*Artikel 3*; Goertz, Kolling, Frahsek, Stanisch & Knopf, in press). Im Hinblick auf FIT 18 und FIT 24 (Goertz, Kolling, Frahsek & Knopf, under review) lässt sich festhalten, dass die Tests bezüglich Umfang und Schwierigkeit altersangepasst sowie die Schwierigkeiten normalverteilt sind. Reliabilitätsanalysen von FIT 18 sowie FIT 24 sind derzeit in Arbeit. Korrelationsanalysen weisen Zusammenhänge zwischen den Leistungen des deklarativen Gedächtnisses (FIT 18; FIT 24) und imitationsbezogenen Items (synchrone sowie unmittelbare Imitation) des ET6-6 auf. Ferner korreliert die Sprachentwicklung mit den deklarativen Gedächtnisleistungen im Alter von 24 Monaten.

Um ein Maß der diskriminativen Validität der FIT 12 und FIT 18 Tests zu erhalten, wurde eine nicht-deklarative Gedächtnisaufgabe (Zug-Aufgabe) im Rahmen eines Cross-Lagged Panel Design eingesetzt (*Artikel 4;* Kolling, Goertz, Frahsek & Knopf, under review). Es zeigte sich, dass die Korrelationen zwischen deklarativen und nicht-deklarativen Gedächtnisleistungen gering sind. Dies steht im Einklang mit Hypothesen bezüglich der frühen Dissoziation beider Gedächtniskomponenten und kann als diskriminative Validität der eingesetzten Tests gewertet werden.

4. Längsschnittliche Befunde

Die Leistungen des deklarativen Gedächtnisses stiegen im Laufe des zweiten Lebensjahres in beachtlicher Weise an. So erinnerten Kinder im Alter von 12 Monaten *M* = 4 (*SD* = 1.55) von 7 Teilhandlungen, im Alter von 18 Monaten *M* = 6.9 (*SD* = 1,85) von 12 Teilhandlungen. Die Zweijährigen erinnerten schließlich *M* = 17.82 (*SD* = 3.81) von 29 Teilhandlungen. Hierbei fanden sich keine Geschlechtsunterschiede. Dieser Leistungszuwachs repliziert die Befunde aus bisherigen Quer- (Barr & Hayne, 2000) sowie Längsschnittstudien (Heimann & Meltzoff, 1996; Nielsen & Dissanayake, 2004).

Im Hinblick auf die interindividuellen Stabilitäten der Leistungen zeigte sich, dass diese zwischen 12 und 18 Monaten relativ niedrig waren (*r* = .23) und zwischen dem 18. und 24. Lebensmonat leicht anstiegen (*r* = .37*). Dieses Befundmuster repliziert die Daten aus der Studie von Nielsen und Dissanayake (2004), wobei in der vorliegenden Studie mehr Items mit mehr Handlungsschritten verwendet wurden. Um die niedrigen Stabilitäten im Rahmen eines differentiellen Ansatzes weiter erklären zu können erfolgten gruppenbasierte Analysen.

5. Inter-individuelle Differenzen

In *Artikel 5* (Kolling, Goertz, Frahsek & Knopf, in press) wurden die Daten der ersten beiden Messzeitpunkte (12 und 18 Monate alte Kinder) analysiert. Es zeigte sich ein signifikanter Leistungsanstieg der Kinder über die Zeit sowie eine niedrige Stabilität. Mit einem personenzentrierten statistischen Verfahren (Medianisierung bzw. Trichotomisierung relativer z-Differenzwerte) wurden zwei bzw. drei Entwicklungsgruppen extrahiert. In einer 2-Gruppen-Lösung wurde eine Gruppen moderaten Wachstums und eine Gruppe hohen Wachstums extrahiert. Durch eine Trichotomisierung der relativen z-Differenzwerte wurde eine 3-Gruppen-Lösung (schwaches, moderates, hohes Wachstum) extrahiert. Die Stabilitäten der Subgruppen waren höher als jene der Gesamtgruppe. Die extrahierten Entwicklungsgruppen wurden im Hinblick auf Differenzen in Sprach-, motorischer sowie kognitiver Entwicklung untersucht. Es fanden sich keine signifikanten Unterschiede zwischen den Gruppen im Hinblick auf diese Variablen. Dennoch konnte in dieser Studie gezeigt werden, dass nicht zwangsläufig Messproblematiken (z.B. niedrige Reliabilitäten) bzw. diskontinuierliche Entwicklungsverläufe für niedrige Stabilitäten im Rahmen der Entwicklungsforschung verantwortlich sind, sondern vielmehr auch bisher noch wenig verstandene individuelle Differenzen niedrige Stabilitäten bedingen können.

Um inter-individuelle Differenzen intra-individueller Entwicklungsverläufe weiter zu verstehen, wurden die Daten von 3 Messzeitpunkten (12, 18, 24 Monate) der Verzögerten Imitation mit Daten des Entwicklungstests ET6-6 in Beziehung gesetzt (*Artikel 6*, Kolling, Goertz, Frahsek & Knopf, under review). Durch personenzentrierte Verfahren (Clusteranalyse absoluter z-Differenzwerte) wurden zwei Gruppen extrahiert, welche differentielle Wachstumskurven sowie relativ hohe Stabilitäten zeigten. Diese beiden Gruppen unterschieden sich im Alter von 18

Monaten in den Entwicklungsdimension *Sprach- sowie motorische Entwicklung* als auch in der Dimension *Körperbewusstsein* signifikant voneinander. Diese Befunde stehen im Einklang mit Theorien zur Rolle des Selbst bei der Entwicklung deklarativer, insbesondere episodischer Erinnerung (vgl. Knopf, Mack & Kressley-Mba, 2005) sowie mit empirischen Ergebnissen zur Rolle der Sprache beim Aufbau deklarativer Erinnerungen (z.B. Herbert & Hayne, 2000).

In weiteren multivariaten Längsschnittstudien sollten die Zusammenhänge zwischen deklarativen Gedächtnisleistungen, Sprache und Selbstentwicklung weiter untersucht werden. In diesen Studien sollte geprüft werden, welche neueren Auswertungs- und Skalierungsmethoden im Bereich der Forschung zu individuellen Differenzen des Säuglingsalters sinnvoll eingesetzt werden können. Mikrogenetische Studien mit akzelerierten Designs, Item-Response-Skalierungen sowie auf linearen Strukturgleichungsmodellen aufbauende Längsschnittverfahren bieten sich für weitere Forschungen an.

VIII. References

Adler, S. A., Gerhardstein, P., & Rovee-Collier, C. (1998). Levels-of-processing effects on infant memory? *Child Development, 69*, 280-294.

Alwin, D. F. (1994). Aging, personality, and social change: The stability of individual differences over the adult life span. In D.L. Featherman, R.M. Lerner, & M. Perlmutter (Eds.), *Life-span development and behavior, Vol. 12* (pp. 135 – 185). Hillsdale, NJ: Lawrence Erlbaum.

Asendorpf, J. B. (1990). The measurement of individual consistency. *Methodika, 4*, 1-13.

Barr, R., & Hayne, H. (1996). The effect of event structure on imitation in infancy: Practice makes perfect? *Infant Behavior & Development, 19*, 253-257.

Barr, R., & Hayne, H. (2000). Age-related changes in imitation: Implications for memory development. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Progress in infancy research* (314 pp). Mahwah, NJ: Lawrence Erlbaum Associates.

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior & Development, 19*, 159-170.

Bates, J. E., & Novosad, C. (2005). Measurement of individual difference constructs in child development, or taking aim at moving targets. In D. M. Teti (Ed.), *Handbook of research methods in developmental science* (pp. 103-122). Oxford: Blackwell Publishing.

Borovsky, D., & Rovee-Collier, C. (1990). Contextual constraints on memory retrieval at 6 months. *Child Development, 61*, 1569–1583.

Collie, R., & Hayne, H. (1999). Deferred imitation by 6- and 9-month-old infants: More evidence for declarative memory. *Developmental Psychobiology, 35*, 83-90.

Fagan, J. F. (1970). Memory in the infant. *Journal of Experimental Child Psychology, 9*, 217-226.

Ghiselli, E. E. (1956). Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology, 40,* 374-377.

Ghiselli, E. E. (1960). The prediction of predictability. *Educational and Psychological Measurement, 20,* 3-8.

Goertz, C., Knopf, M., Kolling, T., Frahsek, S., & Kressley-Mba, R. A. (2006). Entwicklung und Erprobung eines Messinstruments zur Erfassung des deklarativen Gedächtnisses Einjähriger: Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12). [The Frankfurt Imitation Test for 12-month-old infants: Testing a series of items for measuring declarative memory in 1-year-olds.]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38,* 88-96.

Goertz, C., Kolling, T., Frahsek, S., Stanisch, A., & Knopf, M. (2007). Assessing declarative memory in 12-month-old infants: A test-retest reliability study based on the deferred imitation task. *European Journal of Developmental Psychology*, in press. Epub ahead of print retrieved Mai, 28, 2007, from http://www.informaworld.com/smpp/content~content=a778058220~jumptype=rss

Goertz, C., Kolling, T., Frahsek, S., & Knopf, M. (subm.). Der Frankfurter Imitations-Test für 18 Monate und 24 Monate alte Kinder (FIT 18-24). [The Frankfurt Imitation Test for 18-month-old and 24-month-old infants: Testing a series of items for measuring declarative memory in 1.5-2-year-olds.]. *Manuscript submitted for publication.*

Gulliksen, H. (1950). *The theory of mental tests*. New York: Wiley.

Hartmann, D. P. (2005). Assessing growth in longitudinal investigations: Selected measurement and design issues. In D. M. Teti (Ed.), *Blackwell handbook of research methods in developmental science* (pp. 319-339). Oxford: Blackwell Publishing.

Hartshorn, K., Rovee-Collier, C., Gerhardstein, P., Bhatt, R. S., Wondoloski, T. L., & Klein, P. (1998). The ontogeny of long-term memory over the first year-and-a-half of life. *Developmental Psychobiology, 32,* 69-89.

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review*, *24*, 33-73.

Hayne, H., Boniface, J., & Barr, R. (2000). The development of declarative memory in human infants: Age-related changes in deferred imitation. *Behavioral Neuroscience, 114*, 77-83.

Heimann, M., & Meltzoff, A. N. (1996). Deferred imitation in 9- and 14-month-old infants: A longitudinal study of a Swedish sample. *British Journal of Developmental Psychology, 14,* 55-64.

Heimann, M., Strid, K., Smith, L., Tjus, T., Ulvund, S. E., & Meltzoff, A. N. (2006). Exploring the relation between memory, gestural communication, and the emergence of language in infancy: A longitudinal study. *Infant and Child Development, 15,* 233-249.

Herbert, J., & Hayne, H. (2000). Memory retrieval by 18-30-month-olds: Age-related changes in representational flexibility. *Developmental Psychology, 36*, 473-484.

Howe, M. L., & Courage, M. L. (1997). The emergence and early development of autobiographical memory. *Psychological Review, 104*, 499-523.

Howe, M. L., & Courage, M. L. (2004). Demystifying the beginnings of memory. *Developmental Review, 24,* 1-5.

Jones, E. J., & Herbert, J. S. (2006). Using deferred imitation to understand the process of change in infant memory development. *Infant and Child Development, 15,* 215-218.

Knopf, M., Kraus, U., & Kressley-Mba, R. A. (2006). Relational information processing of novel unrelated actions by infants. *Infant Behavior & Development, 29*, 44-53.

Knopf, M., Mack, W., & Kressley-Mba, R. (2005). Wissen und Erinnern. Zur Genese des episodischen Gedächtnisses bei Säuglingen und präverbalen Kindern. [Knowing and remembering: On episodic memory in infants and preverbal children.]. *Psychologische Rundschau, 56,* 113-122.

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (in press). Stability of deferred imitation in 12- to 18-month-old infants: A closer look into developmental dynamics. *European Journal of Developmental Psychology*.

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (subm.). Declarative and non-declarative memory in 12- and 18-month-old infants. *Manuscript submitted for publication*.

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (subm.). Dynamics of declarative memory from infancy to childhood. *Manuscript submitted for publication*.

Kressley, R. A., & Knopf, M. (2006). A comparison of between- and within-subjects imitation design. *Infant Behavior & Development, 29*, 564-573.

Kressley-Mba, R. A., Lurg, S., & Knopf, M. (2005). Testing for deferred imitation of 2- and 3-step action sequences with 6-month-olds. *Infant Behavior & Development, 28*, 82-86.

Mandler, J. M. (2004). Thought before language. *Trends in Cognitive Sciences, 8*, 508-513.

McCall, R. B., & Carriger, M.S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development*, *64*, 57-79.

McDonough, L., Mandler, J., McKee, R., & Squire, L. (1995). The deferred imitation task as a nonverbal measure of declarative memory. *Proceedings of the National Academy of Science, 92*, 7580-7584.

Meltzoff, A. N. (1985). Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. *Child Development*, *56*, 62-72.

Meltzoff, A. N., & Moore, M. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*, 75-78.

Meltzoff, A. N., & Moore, M. (1999). Persons and representation: Why infant imitation is important for theories of human development. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (pp. 9-35). NY, US: Cambridge University Press.

Nielsen, M., & Dissanayake, C. (2004). Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year. *Infant Behavior & Development*, *27*, 342-365.

Petermann, F., & Stein, I. A. (2000). *Entwicklungstest ET 6-6. [Developmental test for 6-month to 6-year-olds.]*. Lisse, Netherlands: Swets & Zeitlinger Publishers.

Piaget, J. (1962). The role of imitation in the formation of representation. *L'Evolution Psychiatrique, 27, 141-150*.

Prudhomme, N. (2005). Early declarative memory and self-concept. *Infant Behavior & Development*, *28*, 132-144.

Rovee-Collier, C., & Barr, R. (2001). Infant learning and memory. In G. Bremner & A. Fogel, (Eds.), *Blackwell handbook of infant development* (pp. 139-168). MA, US: Blackwell Publishing.

Rovee-Collier, C., Hayne, H., & Colombo, M. (2001). *The development of implicit and explicit memory.* Amsterdam: John Benjamins.

Schacter, D. L., & Tulving, E. (1994). *Memory systems* (pp. 1-38). MA, US: The MIT Press.

Spiel, C. (1998). Four methodological approaches to the study of stability and change in development. *Methods of Psychological Research Online, 3,* 7-22.

Strid, K., Tjus, T., Smith, L., Meltzoff, A. N., & Heimann, M. (2006). Infant recall memory and communication predicts later cognitive development. *Infant Behavior & Development*, 29, 545-553.

Squire, L. R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems* (pp. 203-232). Cambridge, MIT Press.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40*, 385-398.

IX. Akademischer Lebenslauf

Dipl.- Psych. Thorsten Kolling

Geburtstag und - ort

Geboren am 17.12.1977 in Neunkirchen (Saar)

Schulausbildung

1984–1988	Grundschule Friedrich von Schiller, Wiebelskirchen (Saar)
1988–1997	Gymnasium am Krebsberg, Neunkirchen (Saar)

Zivildienst

1997–1998	Diakonisches Werk an der Saar

Akademische Ausbildung

1998 - 2003	Studium Psychologie an der Johann Wolfgang Goethe-Universität, Frankfurt am Main
2003	Diplom in Psychologie
2004-2007	Wissenschaftlicher Mitarbeiter im DFG-Projekt „Zur Ontogenese des Gedächtnissystems“, Johann Wolfgang Goethe-Universität Frankfurt (Betreuerin: Prof. Dr. M. Knopf)

Praktika

10/2000–08/2001	Studienbegleitendes Praktikum Abteilung Sozialpsychologie (Prof. Dr. Enno Schwanenberg, Universität Frankfurt am Main)

06/2001–03/2002	Praktikum International Communication Consultants, Nürnberg
08/2001–10/2001	Praktikum Marktforschungsinstitut Global Dynamics, Kronberg (Prof. Dirk Frank)
10/2001 –08/2002	Praktikum in der Schlafambulanz des Universitätsklinikums der J.W. Goethe-Universität, Frankfurt am Main (Prof. Dr. Thomas Heidenreich)

Frankfurt am Main, den 18. Juni 2007 Thorsten Kolling

X. Appendix

Paper 1

Goertz, C., Knopf, M., **Kolling, T**., Frahsek, S., & Kressley-Mba, R. A. (2006). Entwicklung und Erprobung eines Messinstruments zur Erfassung des deklarativen Gedächtnisses Einjähriger: Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12). *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *38,* 88-96.

Paper 2

Goertz, C., **Kolling, T.**, Frahsek, S., & Knopf, M. (subm.). Der Frankfurter Imitations-Test für 18 Monate und 24 Monate alte Kinder (FIT 18-24). *Kindheit und Entwicklung.*

Paper 3

Goertz, C., **Kolling, T**., Frahsek, S., Stanisch, A., & Knopf, M. (2007). Assessing declarative memory in 12-month-old infants: A test-retest reliability study based on the deferred imitation task. *European Journal of Developmental Psychology*, in press. Epub ahead of print retrieved May, 28, 2007, from http://www.informaworld.com/smpp/content~content=a778058220~jumptype=rss

Paper 4

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (subm.). Declarative and non-declarative memory in 12- and 18- month-old infants. *Infancy.*

Paper 5

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (in press). Stability of deferred imitation in 12- to 18-month-old infants: A closer look into developmental dynamics. *European Journal of Developmental Psychology*.

Paper 6

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (subm.). Dynamics of declarative memory from infancy to childhood. *Child Development.*

Paper 1

Entwicklung und Erprobung eines Messinstruments zur Erfassung des deklarativen Gedächtnisses Einjähriger:
Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12)

Goertz, C., Knopf, M., **Kolling, T**., Frahsek, S., & Kressley-Mba, R. A. (2006). Entwicklung und Erprobung eines Messinstruments zur Erfassung des deklarativen Gedächtnisses Einjähriger: Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12). *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38,* 88-96.

Sonderdruck aus: Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38 (2), 88–96

Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12)

Erprobung einer Aufgabenserie zur Erfassung des deklarativen Gedächtnisses Einjähriger

Claudia Goertz, Monika Knopf, Thorsten Kolling, Stefanie Frahsek und Regina Kressley

Johann Wolfgang Goethe-Universität Frankfurt am Main

Zusammenfassung. Die Erforschung des Gedächtnisses von präverbalen Kindern nimmt in der modernen Entwicklungspsychologie einen breiten Raum ein. Es besteht dabei Konsens dahingehend, dass die Methode der Verzögerten Imitation (VI) das Verfahren ist, das deklaratives Gedächtnis im Verlauf der Ontogenese am frühesten erfassen kann. Im deutschen Sprachraum gibt es bislang kein erprobtes Instrumentarium, das für die Messung deklarativer Gedächtnisleistungen mittels VI erprobt worden wäre. Im Anschluss an die Darstellung der Grundlagen des Messverfahrens wird über eine Studie mit *n* = 92 Einjährigen berichtet, in der eine Aufgabenserie zur VI erprobt (Frankfurter Imitations-Test für 12 Monate alte Kinder, FIT 12) wurde. Dieses neu entwickelte Instrument erwies sich als altersadäquat. Die zwischen den Einjährigen gemessenen Leistungsunterschiede erwiesen sich zudem in einer Test-Retest-Reliabilitätsstudie nach einem Zeitintervall von einer Woche als stabil (Goertz et al., under revision). Schließlich konnten Belege für die Konstruktvalidität dieses Gedächtnistests erbracht werden.
Schlüsselwörter: Verzögerte Imitation, deklaratives Gedächtnis, Frankfurter Imitations-Test (FIT 12), Gedächtnistest

The Frankfurt Imitation Test for 12-Month-Old Infants: Testing a series of items for measuring declarative memory in 1-year-olds

Abstract. Research on memory development among pre-verbal infants has grown in significance in the field of developmental psychology. A consensus exists that the method of deferred imitation delivers the earliest measurement of declarative memory in the course of development. In German-speaking areas there has to date been no standardized inventory for assessing declarative memory in one-year-olds via the deferred imitation technique. Following an explanation of the method of deferred imitation, a study is described here with *n* = 92 one-year-olds in which a newly developed inventory was examined (Frankfurt Imitation Test for 12-Month-Old Infants, FIT 12). This newly constructed inventory proved to be age-appropriate. Furthermore, significant performance variability was found among the one-year-olds. In an additional study, a high test-retest stability of interindividual differences over a 1-week interval has been found (Goertz et al., under revision). Finally, evidence for the construct validity of this memory test is presented.
Key words: deferred imitation, declarative memory, Frankfurt Imitation Test for 12-Month-Old Infants (FIT 12), memory task

Die Analyse der Entwicklung des Gedächtnisses im Kindesalter wird zu den bevorzugten Themen kognitiver Entwicklungspsychologen gerechnet (vgl. Schneider, 2002). In den letzten Jahren, in denen u.a. eine intensive Forschungstätigkeit zur Gedächtnisentwicklung bei Säuglingen und präverbalen Kindern einsetzte, hat sich ein Gedächtnismodell durchgesetzt, das auch ontogenetisch frühe Gedächtnisleistungen erfassen kann. Es handelt sich dabei um eine Taxonomie, die auf oberster Ebene zwischen den beiden Bereichen deklaratives und nicht-deklaratives Gedächtnis trennt. Auf einer zweiten Ebene sind das semantische und das episodische Gedächtnis bekanntlich die zentralen Konstituenten des deklarativen Gedächtnisses (Tulving, 2002). Dem nicht-deklarativen Gedächtnis werden demgegenüber eine ganze Sammlung unterschiedlicher Phänomene zugerechnet, die auf eine vorausgehende Beschäftigung mit dem entsprechenden Material zurückgehen. Allerdings können die Inhalte des nicht-deklarativen Gedächtnisses in aller Regel nicht bewusst abgerufen werden (z.B. Verhaltensgewohnheiten, motorische Routinen, Priming-Prozesse, nicht-assoziatives Lernen, klassische Konditionierung).

Während das in der Humanforschung entwickelte Begriffspaar deklaratives und nicht-deklaratives Gedächtnis

Die vorliegende Studie ist Teil der von der DFG geförderten Längsschnittstudie „Zur Ontogenese des Gedächtnissystems bei Säuglingen und Kleinkindern“ (Kn 275/3-1). Wir danken Christina Mack, Heike Sänger, Denise Ginzburg, Mathis Stemmildt und Cornelia Plöger für die Mithilfe bei der Datenerhebung und -auswertung sowie den Gutachtern für wertvolle Hinweise.

DOI: 10.1026/0049-8637.38.2.88

ausdrückt, dass die Einträge des deklarativen Gedächtnisses sprachlich geäußert werden können („deklariert“), was für die Einträge des nicht-deklarativen Gedächtnisses nicht gilt, wird dieses Begriffspaar in jüngster Zeit mit einem aus der Tierforschung stammenden Begriffspaar synonym benutzt, das explizites und implizites Gedächtnis unterscheidet (vgl. z.B. Rovee-Collier, Hayne & Colombo, 2000). Mit der Übernahme der Begriffe explizites versus implizites Gedächtnis in die Humanforschung wurde auch die Vorstellung aufgegeben, dass die Sprache für bewusstes Gedächtnis wesentlich wäre. Für die entwicklungspsychologische Forschung bedeutet dies, dass (aktives) Sprechen keineswegs eine Bedingung für einen bewussten Zugriff auf Gedächtniseinträge ist. Es stellt sich damit aber die Frage, wie deklaratives Gedächtnis von präverbalen Kindern erfasst werden kann.

Nach einer gründlichen Analyse der unterschiedlichen, in der Säuglings- und Kleinkindforschung nachweisbaren Gedächtnisleistungen (z.B. Bauer, 2002; Mandler, 2004; Natour, 2001), die schon in den ersten Lebenstagen erkennbar sind, geht derzeit die Mehrzahl der Forscherinnen davon aus, dass sich in der Fähigkeit zur Verzögerten Imitation von zuvor gesehenen, neuartigen Handlungen deklaratives Gedächtnis bereits früh in der Ontogenese manifestiert. Diese Vorstellung hatten bereits Piaget und Inhelder (1975) geäußert, die in einem ebenfalls zwei Komponenten umfassenden Gedächtnismodell darlegten, dass das „Gedächtnis im strengen Sinne“ – was als weitgehend synonym mit deklarativem Gedächtnis angesehen werden kann – sich gegen Ende der sensumotorischen Phase in der Fähigkeit zur aufgeschobenen Imitation von zuvor gesehenen, neuen Handlungen zeigt. Als klassisches Beispiel gilt dabei ein Wutanfall von Piagets Tochter Jacqueline im Alter von 1;4 Jahren in ihrem Laufställchen, den sie bei einem kleinen Jungen am gleichem Ort gesehen hatte und zuvor ein vergleichbares Verhalten nicht kannte, um ihn tags darauf mit frappierender Genauigkeit nachzuahmen (Piaget, 1975). Für Piaget manifestiert sich in der aufgeschobenen Nachahmung, dass das äußerlich wahrnehmbare Vorbild durch ein „inneres Modell“ (Piaget, 1975) ersetzt worden ist und damit eine Gedächtnisleistung auf der Basis von bildhaften Vorstellungen bzw. symbolischen oder bildhaften Repräsentationen erfolgt. Auch für Piaget sind damit Gedächtnisleistungen im strengen Sinne nicht an sprachliche Repräsentationen geknüpft.

In der experimentellen Gedächtnisforschung bei Säuglingen wurden diese Überlegungen Piagets zunächst von Meltzoff (1985, 1988a, 1988b) aufgegriffen und analoge Aufgaben konstruiert, um die Entwicklung des deklarativen Gedächtnisses im Labor analysieren zu können. In der Standardaufgabe zur verzögerten Imitation sensu Meltzoff werden den präverbalen Kindern durch ein Modell mehrere einfache, (möglichst) neuartige, objektbezogene Handlungen nacheinander demonstriert, die die Kinder nur beobachten können. Nach einer Verzögerung erhalten die Kinder die Objekte sukzessive vorgelegt und ihr Spielverhalten wird beobachtet. Während in Piagets Beobachtungen das Verhalten spontan erfolgte, jedoch in aller Regel ebenfalls bei Wahrnehmung eines handlungsbezogenen Objekts (z.B. Laufställchen), wird in der experimentellen Variante der Aufgabe die Gedächtnisleistung durch Objektvorgabe gezielt ausgelöst. In beiden Fällen handelt es sich also bei den frühesten deklarativen Gedächtnisleistungen um einen unterstützten Gedächtnisabruf.

Um spontanes Ausführen der Zielhandlungen von gedächtnisbasiertem Verhalten unterscheiden zu können, wird das Spiel von Kindern, die das Modell beobachtet hatten, mit dem Spiel von Kindern einer Basis-Kontrollgruppe in Beziehung gesetzt. Während in den Studien von Meltzoff mit unabhängigen Stichproben gearbeitet wird, wird in der Tradition von Bauer und Mandler (1989) die spontane Ausführungsrate der Zielhandlungen vor der Demonstration in einem Intra-Person-Design erfasst. Eine Studie von Kressley-Mba und Knopf (submitted) untersuchte den Einfluss der durch die Basisratenerhebung vermittelten Vorerfahrung mit den Objekten auf die Imitationsleistung und konnte nachweisen, dass sich die Gedächtnisleistungen nicht veränderten, wenn die Kinder bereits vor der Demonstration, eben in der Basisratenerhebung, Kontakt mit den Spielzeugen hatten, die dann in der Verzögerten Imitation verwendet wurden. Dies wurde realisiert über einen Vergleich der Designs mit abhängigen (Mandler-Bauer-Tradition) und unabhängigen Stichproben (Meltzoff-Tradition). Somit ist es gerechtfertigt, Gedächtnisleistungen aus Studien dieser Designs als vergleichbar anzusehen.

Die experimentellen Aufgaben der aktuellen Forschung sind weniger komplex als die von Piaget beschriebenen Beobachtungen; zudem haben die demonstrierten Handlungen in aller Regel ein besonderes Resultat, um die Aufmerksamkeit der Kinder auf die Demonstrationen zu richten. Dies wird häufig durch ein Geräusch oder einen Lichteffekt erreicht, aber auch durch eine neuartige, überraschende Handlung, die sich nicht zwangsläufig aus den Eigenschaften der Objekte ableiten lässt. Entscheidend ist, dass die Kinder während der Demonstration die Handlungen aufmerksam beobachten. Die Demonstrationen erfolgen zudem nicht lediglich einmal, sondern werden mehrfach, je nach Altergruppe bis zu sechs Mal, nacheinander vorgenommen.

Folgende Argumente sprechen dafür, dass die experimentelle Aufgabe zur Verzögerten Imitation explizites und nicht implizites Gedächtnis erfasst. Zum einen wird auf den kreuzmodalen Charakter dieser Aufgabe hingewiesen, wodurch es äußerst unwahrscheinlich wird, dass die kindlichen Gedächtnisleistungen allein auf Primingvorgänge rückführbar sind. Bekanntlich sind Primingphänomene gegenüber einem Modalitätswechsel sensibel. Zum anderen wird ausgeführt, dass motorische Übung beim Erwerb des Verhaltens nicht gegeben ist. Damit unterscheidet sich der Aufbau des Gedächtniseintrags sehr deutlich von inkrementellen Lernprozessen, wie sie vor allem für prozedurale Lernaufgaben typisch sind (z.B. Mobile-Aufgabe von Rovee-Collier). Ein drittes Argument ist, dass es um die Nachahmung neuartiger Verhaltensweisen geht, also nicht um Handlungen, die bereits im

Verhaltensrepertoire der Kinder vorhanden sind. Diese Forderung wird mittels der Bestimmung der Basisrate zu erreichen versucht. Eine vierte Evidenz liefert eine Studie mit Amnesie-Patienten: McDonough, Mandler, McKee und Squire (1995) demonstrierten, dass Amnesiepatienten, deren explizites Gedächtnis im Unterschied zum impliziten geschädigt ist, zur Verzögerten Imitation nicht imstande sind. Fünftens weisen Gehirnforscher darauf hin, dass kortikale Gehirnareale, auf die deklaratives Gedächtnis angewiesen ist, frühestens in der zweiten Hälfte des ersten Lebensjahrs funktionstüchtig werden (Johnson, 1997). Einen sechsten Beleg schließlich liefert eine eigene Studie (Knopf, Kraus & Kressley-Mba, 2006), die zeigt, dass 10- und 11-Monate alte Kinder bei der Verzögerten Imitation nicht nur die Einzelhandlungen reproduzieren, sondern zusätzlich die Struktur zwischen diesen Handlungen kennen, so dass die Vorgabe der Objekte in der identischen Reihenfolge in Enkodier- wie Abrufphase zu besseren Gedächtnisleistungen führt als variierte Reihenfolgen. Die Bedeutung von Reihenfolgeinformation zwischen a-priori unverknüpften Items für die Gedächtnisleistung wird als Indiz dafür gesehen, dass es sich bei der Verzögerten Imitation um ein tatsächliches Analogon zu expliziten verbalen Gedächtnistests handelt, mit Merkmalen nicht nur des Unterstützten Erinnerns sondern sogar des Freien Erinnerns (Bauer, Hertsgaard & Dow, 1994; Bauer, Hertsgaard & Wewerka, 1995; Mandler, 1990, 2004).

Aus den vorstehenden Überlegungen folgt, dass zur Messung der Verzögerten Imitation bei präverbalen Kindern ein Instrument aus einfachen, objektbezogenen Handlungen benötigt wird, die von den Kindern motorisch realisiert werden können. Ferner sollten die Zielhandlungen nicht bereits bekannt sein oder beim Anblick des handlungsbezogenen Objekts automatisch ausgelöst werden, d.h. eine geringe spontane Ausführungsrate haben. Zudem sollten die verschiedenen Zielhandlungen vergleichsweise unabhängig voneinander sein, also keinen a-priori-Bezug zueinander aufweisen; es soll sich also nicht um eine zusammenhängende Sequenz von Handlungen handeln. Bislang existiert im deutschsprachigen Raum kein Instrumentarium erprobter Aufgaben zur Erfassung der Verzögerten Imitation und damit des deklarativen Gedächtnisses von präverbalen Kindern. Eine einfache Übertragung der von Meltzoff entwickelten Items ließ sich nicht realisieren, da im amerikanischen Sprachraum bewährte Aufgaben von deutschen Kindern gleichen Alters nicht durchgängig imitiert wurden (Natour, 2001). Im Verlaufe von einer Reihe von Studien in Frankfurt (Natour, 2001; Knopf, Kraus & Kressley-Mba, 2006; Kressley-Mba & Knopf, submitted) stellten sich fünf Aufgaben für Einjährige als besonders geeignet heraus, mit der Verzögerten Imitation die Gedächtnisleistung auf dieser Altersstufe zu erfassen. Sie wurden zu einer Aufgabenserie von fünf unabhängigen Handlungen, dem FIT 12, zusammengestellt und sollen im Folgenden genauer vorgestellt werden. Neu daran ist, dass die verschiedenen Items meist aus einem Handlungsschritt bestehen und voneinander unabhängig sind. Andere Forscher haben Verzögerte Imitation entweder mit einem Item bestehend aus mehreren Handlungsschritten (Hayne, Boniface & Barr, 2000) oder mit mehreren Items durchgeführt, teils auch aus mehreren Handlungsschritten bestehend (Bauer, Wiebe, Waters & Bangston, 2001; Hayne & Herbert, 2004), die dann zu einem Summenscore zusammengefasst wurden. Die hier entwickelte Prozedur umgeht die Schwierigkeit, dass die Imitation späterer Handlungsschritte möglicherweise auf Grund des Vergessens früherer nicht mehr möglich ist.

Methode

Stichprobe

An der Studie nahmen 92 Kinder (48 Jungen) teil. Ihr mittleres Alter betrug 362,4 Tage ($SD = 8,7$). Alle Kinder wurden nach der 37. Schwangerschaftswoche geboren, ihr mittleres Geburtsgewicht betrug 3393g ($SD = 507,15$) und der APGAR-Index (5/10) betrug im Mittel 9,78/9,94 (Minimum 7). Da alle Kinder mit dem Entwicklungstest ET6-6 von Petermann und Stein (2000) untersucht wurden, wobei die Items für 9 bis 15 Monate alte Kinder zur Anwendung kamen, konnte sichergestellt werden, dass sämtliche Vpn innerhalb des Normbereichs lagen (Dimensionen des ET6-6: Körpermotorik, Handmotorik, Kognitive Entwicklung, Sprachentwicklung, Sozialentwicklung, Emotionale Entwicklung). Die Daten eines Kindes konnten auf Grund technischer Probleme nur teilweise ausgewertet werden.

Materialien

Für die Items des FIT 12 wurden Spielzeuge verwendet, die im freien Handel erworben und teilweise etwas abgeändert wurden. Diese Handlungen sind bereits in dieser oder leicht abgewandelter Form in Vorgängerstudien eingesetzt und erprobt worden (z.B. Natour, 2001; Knopf et al., 2006; Kressley-Mba & Knopf, submitted).

1. Handlung: Dose schütteln. Eine hellblaue Dose (Durchmesser = 4,5 cm), gefüllt mit Reis und sicher verschlossen, wurde dreimal auf und ab geschüttelt. Die Kinder erhielten in der Imitationsphase eine identische Dose, die jedoch leer war, um ein zufälliges Ausführen der Zielhandlung zu vermeiden.

2. Handlung: Hut des Schweins abziehen. Einem rosafarbenen Plüschschwein (Größe: 20 cm × 15 cm × 7 cm) wurde der braune, mit Klettband befestigte Hut vom Kopf abgezogen und neben das Schwein auf den Tisch gelegt. Um die Schwierigkeit der Aufgabe etwas zu erhöhen, bekamen die Kinder das Schwein mit der Schnauze nach vorn überreicht, dadurch war der Hut für sie vergleichsweise wenig herausgehoben.

3. Handlung: Mit einem Holzmesser in einer Dose rühren. Ein Holzmesser (Länge = 16,5 cm) wird in einem roten Plastikbecher (Durchmesser = 8 cm) hin und her bewegt. Um die Schwierigkeit der Aufgabe zu erhöhen und

Zufallsausführungen zu minimieren, wurde den Kindern der Becher umgekehrt, mit dem Boden nach oben, präsentiert.

4. Handlung: Maus drücken. Eine Holzmaus (Größe: 8 cm × 5,5 cm), bestehend aus einem gelben Oberteil, einem holzfarbenen Unterteil und einem weißen Gummiband (Länge = 6 cm) mit einer grünen Holzkugel, wurde durch das Herunterdrücken des Oberteils zusammengeklappt.

5. Handlung: Schlagzeug drücken. Bei einem bunten Plastikschlagzeug (Größe = 11 cm × 9 cm) wird mit dem blauen Schlägel (Länge = 8 cm) ein großer roter Knopf (Durchmesser = 5 cm) auf der Oberseite gedrückt. Das erzeugt ein Geräusch ähnlich einem Klingelton. In der Imitationsphase erhielten die Kinder ein Schlagzeug, das bei Knopfdruck kein Geräusch erzeugte, damit die Kinder nicht durch zufälliges Hantieren die Zielhandlung erzeugen konnten.

Die Spielobjekte der fünf Handlungen sind in der Abbildung 1 während der Ausführung der Handlung dargestellt.

In einer früheren Studie, an der 24 Kinder im Alter von 12 Monaten teilnahmen, wurde die Basisrate der Ausführung der fünf Handlungen ermittelt (Goertz, Kolling, Frahsek, Stanisch & Knopf, under revision). Bei der Festlegung der Vorgabereihenfolge der Items im FIT 12 wurde darauf geachtet, dass die Zielhandlungen mit sehr niedriger (Schwein: 8,3 % und Maus: 4,2 %) und niedriger Basisrate (Dose und Schlagzeug je 20,8 % und Becher/Messer: 29,2 %) im Wechsel aufeinander folgten. Zudem sollten einerseits die beiden Geräusch erzeugenden Handlungen Dose und Schlagzeug sowie die Spielzeugtiere Schwein und Maus nicht direkt aufeinanderfolgen. Dieses führte zu der oben beschriebenen Itemreihenfolge.

Versuchsdurchführung

Die Kinder wurden mit ihrer Begleitperson zu der Tageszeit bestellt, zu der sie üblicherweise aktiv waren. Im Labor saß das Kind auf dem Schoß der Begleitperson, die es von hinten stützte, sonst aber angehalten war, sich am Geschehen nicht zu beteiligen. Der Versuchleiter saß dem Kind gegenüber und war über kleine, unauffällige Kopfhörer mit dem zweiten Versuchsleiter verbunden, der vom Aufnahmeraum aus die Videogeräte bediente und die Präsentationszeiten durchsagte.

Demonstrationsphase (5 × 30 Sekunden): Nach einer Aufwärmphase von maximal 180 Sekunden, in der mit einer Ringpyramide gespielt wurde, begann die Demonstrationsphase. Jede Handlung wurde innerhalb von 30 Sekunden viermal demonstriert. Dabei wurde darauf geachtet, dass das Kind die Demonstration aufmerksam beobachtete. Die Handlungen wurden nicht verbal kommentiert, die Interaktion beschränkte sich auf Bemerkungen wie: „Schau mal her, <Name>!“ oder „Sieh mal, <Name>, was ich hier habe.“

Verzögerungsphase (30 Minuten): Das Labor wurde nach der Demonstrationsphase verlassen. In einem anderen Raum wurden nun den Eltern Fragen zur Entwicklung des Kindes gestellt. Nach 30 Minuten wurde das Labor wieder aufgesucht und der Gedächtnistest durchgeführt. In den aktuellen Gedächtnismodellen wird angenommen, dass die aktive Verarbeitung neuer Information im Arbeitsgedächtnis erfolgt, bevor das Erlernte im Langzeitgedächtnis gespeichert wird. Die aktive Bearbeitung von Information erfolgt zügig, d.h. im Sekundenbereich, um dem kontinuierlich eintreffenden Strom neuer Information gerecht zu werden (Baddeley, 1986). Deshalb wird davon ausgegangen, dass ein Behaltenszeitraum von 30 Minuten sicher ausreicht, um Langzeitgedächtnisleistungen zu erfassen.

Imitationsphase (5 × 30 Sekunden): Nach einer erneuten Aufwärmphase von maximal 60 Sekunden begann die Imitationsphase. Dem Kind wurde das erste Spielzeug der Serie überreicht und für 30 Sekunden überlassen. Dann wurde das Spielzeug von ihm zurückerbeten und ihm das nächste überreicht. Dies wurde für alle Objekte wiederholt. Gab ein Kind das Spielzeug vor Ablauf der Phase dem Versuchsleiter zurück, bekam es es wieder mit der Bemerkung: „<Name>, du darfst jetzt damit spielen.“

Abbildung 1. Die Spielobjekte und Handlungen.

Datenerfassung und -auswertung

Die Sitzungen wurden von zwei Kameras aufgezeichnet. Eine Perspektive erfasste das Kind von vorn, die andere den Versuchsleiter von der Decke aus. Jede Aufzeichnung wurde von zwei unabhängigen, naiven Auswertern, die über die Hypothesen der Studie und die tatsächliche Ausführung der Handlungen durch den Versuchsleiter im Unklaren waren, nach definierten Kriterien ausgewertet. Dabei wurde entschieden, ob ein Kind die Handlung ausgeführt hat (Ja- oder Nein-Entscheidung) und zu welchem Zeitpunkt dies der Fall war (Latenzzeit).

Auswertungskonkordanz

Von den drei beteiligten Auswertern erreichten Auswerter 1 und Auswerter 2 eine Übereinstimmung von 92,7% (Cohen's κ = 0,83) und Auswerter 1 und Auswerter 3 eine Übereinstimmung von 89,0% (Cohen's κ = 0,78), wobei Auswerter 1 die Aufzeichnungen aller Kinder auswertete, Auswerter 2 die Videos von 72 Kindern und Auswerter 3 die von 20 Kindern. Nichtübereinstimmungen der Entscheidungen wurden durch einen „forced consent" bereinigt.

Ergebnisse

Imitationsleistung

Die mittlere Reproduktionsleistung lag bei 2,38 von fünf Handlungen (*SD* = 1,3), wobei die Summe der imitierten Handlungen von 0 bis 5 reichte und sich wie folgt verteilte: 0 Handlungen = 10,9%, 1 Handlung = 15,2%, 2 Handlungen = 19,6%, 3 Handlungen = 35,9%, 4 Handlungen = 16,3%, 5 Handlungen = 2,2% (vgl. Abbildung 2). Ein breites Spektrum von Gedächtnisleistungen wurde beobachtet; das Instrument ist damit gut geeignet, die Gedächtnisleistungen von einjährigen Kindern zu erfassen.

Geschlechterunterschiede

Die mittlere Anzahl imitierter Handlungen lag bei den Jungen bei 2,79 (*SD* = 1,01) und bei den Mädchen 2,54 (*SD* = 1,07). Der Unterschied ist nicht signifikant: Z = –0.832, $p = 0.406$. Auch die Analyse der Summe der imitierten Handlungen (T(90) = –0.93, $p = 0.356$) und der Vergleich der Imitationsleistung je Handlung ($F_{\text{Art der Handlung} \times \text{Jungen/Mädchen}}(4,85) = 0.53$, $p = 0.715$) ergab ebenfalls keinen signifikanten Unterschied zwischen den Geschlechtern. Damit ist gezeigt, dass die hier verwendeten handlungsbezogenen Items imstande sind, Gedächtnisleistungen von Jungen und Mädchen vergleichbar gut zu erfassen.

Analyse der Einzelhandlungen

Die einzelnen Handlungen wurden nach der Häufigkeit ihrer Imitation analysiert, die sich signifikant unterscheidet (Cochran-Q-Test (*df* = 4) = 94,029, $p < 0.001$). Das Item Dose wurde von 59,3% der Kinder imitiert, das Item Schwein von 78%, das Item Becher/Messer von 58,2%, das Item Maus von 20,9% und das Item Schlagzeug wurde von 24,2% der Kinder nachgeahmt (siehe dazu Abbildung 3). Es wurde jeweils geprüft, ob die Imitationsleistung die in der Studie von Goertz et al. (under revision) ermittelten Basisrate übersteigt. Ein Mann-Whitney-U-Test ergab einen deutlichen Effekt für die Summe der imitierten Handlungen gegenüber der Summe in der Basisrate (U-Test = 384,0; Z = –4,96, $p < 0{,}001$), für das Item Dose (U-Test = 683,0; Z = –3,21, $p = 0{,}001$), für das Item Schwein (U-Test = 331,0; Z = –6,28, $p < 0{,}001$) und das Item Becher/Messer (U-Test = 774,5; Z = –2,52, $p = 0{,}012$). Das vierte Item Maus verfehlt knapp das Signifikanzniveau von 5% (U-Test = 909,5; Z = –1,91, p = 0,056) und bei fünften Item Schlagzeug zeigt sich kein Unterschied zur Basisrate (U-Test = 1055,5; Z = –0,342, $p = 0{,}732$). Die Basisraten je Handlung sind ebenfalls der Abbildung 3 zu entnehmen.

Beim Zusammenhang zwischen Item- und Gesamtausführung (Trennschärfe) ergaben sich nach einer punkt-biserialen Korrelation folgende Werte für die einzelnen Handlungen (geordnet von der höchsten zur niedrigsten Trennschärfe): Becher/Messer = 0,70, Schwein = 0,58, Dose = 0,55, Schlagzeug = 0,50 und Maus = 0,42.

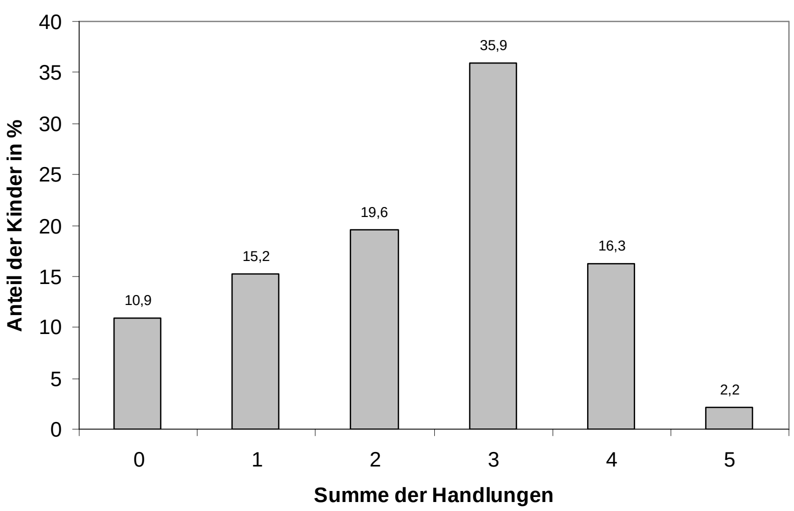
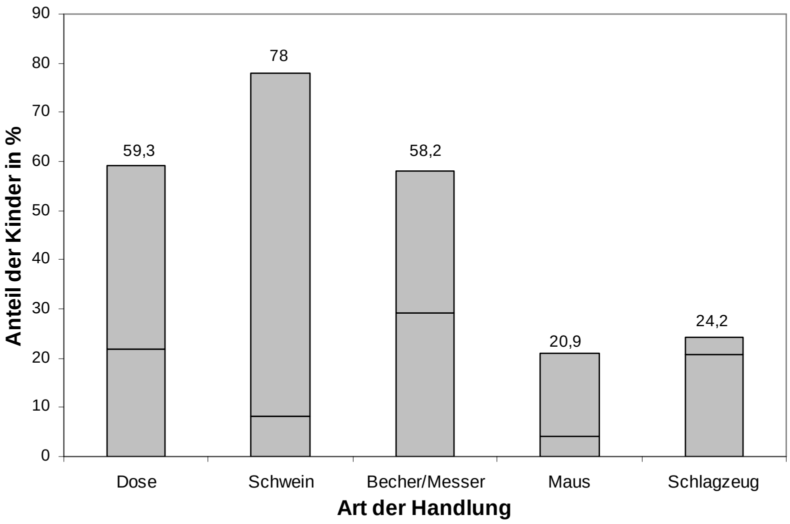


Abbildung 2. Verteilung der Summe der imitierten Handlungen.

Latenzzeiten bis zur Reproduktion der Handlungen

Die Latenzzeit wurde ermittelt, indem die Zeit vom Moment der Übergabe der Spielobjekte an die Kinder bis zum Beginn der Handlungsausführung gemessen wurde. Bei den beiden zweistufigen Items Becher/Messer und Schlagzeug zählte erst der Beginn des zweiten Handlungsschrittes als Zeitpunkt der Ausführung. Die mittlere Latenzzeit betrug: 7,5s (*SD* = 5,68; *N* = 82), wobei auf Grund der geringen Anzahl der Latenz-



Anmerkungen: Imitationsrate ja nach Art der Handlung. Die Basisrate (entnommen Goertz et al., under revision) der jeweiligen Items ist durch eine Linie gekennzeichnet.

Abbildung 3. Analyse der einzelnen Handlungen.

zeiten pro Kind die Mediane in die Analyse eingingen. Die mittlere Latenzzeit betrug bei den Jungen 7s (*SD* = 5,28) und bei den Mädchen 8s (*SD* = 6,16), dieser Unterschied ist nicht signifikant (T(80) = 0,815, *p* = 0,418). Die mittleren Latenzen für die einzelnen Handlungen sind der Tabelle 1 zu entnehmen. Eine ANOVA dieser Zeiten ist nicht möglich, da lediglich zwei Kinder alle fünf Handlungen ausführten und nur sie in die Analyse eingehen würden. Paarweise Vergleiche der in Rangreihe gebrachten Latenzzeiten mittels T-Tests erbrachten jedoch eine Gruppierung von Item Dose (4,9 s), Maus (6,4 s) und Schwein (6,9), die sich jeweils nicht unterschieden ($T_{Dose\text{-}Maus}(8) = –0{,}028$, $p = 0{,}978$, $T_{Maus\text{-}Schwein}(15) = –0{,}471$, $p = 0{,}644$, $T_{Dose\text{-}Schwein}(45) = –1{,}899$, $p = 0{,}064$), gegenüber Becher/Messer (11,7 s) und Schwein (17,8 s), die sich nicht nur gegenüber diesen drei Items sondern auch untereinander unterschieden ($T_{Becher/Messer\text{-}Dose}(37) = 6{,}743$, $p < 0{,}001$, $T_{Becher/Messer\text{-}Schwein}(45) = 2{,}355$, $p = 0{,}023$, $T_{Becher/Messer\text{-}Maus}(13) = 2{,}636$, $p = 0{,}021$, $T_{Becher/Messer\text{-}Schlagzeug}(15) = –2{,}808$, $p = 0{,}013$; $T_{Schlagzeug\text{-}Dose}(12) = 5{,}989$, $p < 0{,}001$, $T_{Schlagzeug\text{-}Maus}(7) = 5{,}063$, $p = 0{,}001$). Die Items Schlagzeug und Schwein unterscheiden sich nicht signifikant ($T_{Schlagzeug\text{-}Schwein}(18) = 1{,}811$, $p = 0{,}087$), obwohl ihre Mittelwerte mit 17,8 und 6,9 s recht weit auseinander liegen. Dies ist aber wohl auf die hohe Variabilität der Latenzzeiten gerade bei diesen beiden Items zurückzuführen ($SD_{Schlagzeug} = 8{,}8$ und $SD_{Schwein} = 9{,}3$). Die Gruppierung ergibt sich aus der Tatsache, dass die Items Dose, Schwein und Maus aus einem Handlungsschritt bestehen, während die Handlungen Becher/Messer und Schlagzeug aus zwei Handlungsschritten bestehen und ihre Ausführung entsprechend länger dauert. Problematisch bleibt dabei die wiederholte Anwendung dieser paarweisen T-Tests, die durch eine Herabsetzung des α auf 1 % teilweise korrigiert werden kann. Ein Vergleich der Latenzzeiten dieser Studie mit denen aus der Basisratenerhebung bei Goertz et al. (under revision) ist auf Grund der geringen Anzahl der Daten in der Basisratenstudie (*N* = 24) nicht möglich. Die mittleren Latenzzeiten der einzelnen Handlungen sind der Tabelle 1 zu entnehmen.

Es zeigte sich kein signifikanter Zusammenhang zwischen der Gedächtnisleistung und der Schnelligkeit der Imitation (*r* = 0,037, *p* = 0,743). Auch die Analyse der einstufigen Handlungen Dose, Schwein und Maus, die im Mittel schneller ausgeführt wurden (mittlerer Median: 5,53 s, *SD* = 4,9), ergab keine signifikante Korrelation zwischen der Summe der imitierten Handlungen und der Schnelligkeit bei der Ausführung (*r* = –0,138, *p* = 0,227).

Test-Retest-Reliabilität des FIT 12

In einer weiteren Studie wurde die Reliabilität des Messinstruments ermittelt (Goertz et al., under revision). In einer Test-Retest-Studie wurde bei 24 Kindern im Alter von 12 Monaten im Abstand von sieben Tagen der Gedächtnistest FIT 12 zweimal durchgeführt. Der Ablauf der Studie und die Reihenfolge der Handlungen entsprachen dem Vorgehen der vorliegenden Studie. Im $Test_1$ (der Abrufphase der ersten Messung) wurden im Mittel 1,92 Handlungen ausgeführt, im $Test_2$ lag die mittlere Ausführung bei 3,17 von fünf Handlungen. Die Imitationsraten in $Test_1$ und $Test_2$ korrelierten mit *r* = 0.52 (*p* = 0.009) und wiesen damit das Instrument FIT 12 als reliablen Gedächtnistest für einjährige Kinder aus. Durch Berechnung der individuellen Konsistenzscores nach Asendorpf (1990) und nach Ausschluss zweier Kinder mit Konsistenzwerten unter –0.5 ergab die anschließende Test-Retest-Korrelation eine Reliabilität von *r* = 0.65 (*p* = 0.001).

Tabelle 1. Mittlere Latenzzeiten je Handlung in s

	Dose	**Schwein**	**Becher/Messer**	**Maus**	**Schlagzeug**
Latenzen in s	4,9	6,9	11,7	6,4	17,8
SD	5,7	9,3	7,0	7,3	8,8
N	53	71	53	19	22

Konstruktvalidität des FIT 12

Alle Kinder der vorliegenden Studie wurden auch mit dem Entwicklungstest ET6-6 von Petermann und Stein (2000) untersucht. Wurden in Abhängigkeit von der Güte der Gedächtnisleistung drei Gruppen gebildet (Gruppe „niedrige Imitationsleistung“ = weniger als 2 Handlungen reproduziert ($N = 24$), Gruppe „mittlere Imitationsleistung“ = 2 oder 3 Handlungen reproduziert ($N = 51$), Gruppe „hohe Imitationsleistung“ = mehr als drei Items reproduziert ($N = 17$)), so zeigte sich, dass die erreichten Werte im ET6-6 in allen Entwicklungsbereichen mit zunehmender Gedächtnisleistung tendenziell anstiegen, ein signifikanter Unterschied war aber nur im Bereich der expressiven Sprachentwicklung zugunsten der Gruppe mit hoher Gedächtnisleistung zu verzeichnen, hingegen ergab sich kein statistisch bedeutsamer Unterschied zwischen den Gruppen in den Bereichen motorische Entwicklung, allgemeine kognitive Entwicklung sowie rezeptive Sprachentwicklung (Kruskal-Wallis-H-Test: Körpermotorik: $\chi^2 = 3{,}083$, $p = 0{,}214$, Handmotorik: $\chi^2 = 4{,}88$, $p = 0.087$, Kognitive Entwicklung: $\chi^2 = 3{,}246$, $p = 0{,}197$, Sprachentwicklung rezeptiv: $\chi^2 = 0{,}634$, $p = 0{,}729$, Sprachentwicklung expressiv: $\chi^2 = 10{,}318$, $p = 0{,}006$). Der Einfluss des Entwicklungsstandes der expressiven Sprache machte sich nur im Unterschied zwischen den Gruppen mit mittlerer und hoher Imitationsleistung bemerkbar (Mann-Whitney-U-Test: $Z_{(\text{Gruppe niedrig-mittel})} = -1{,}858$, $p = 0.063$, $Z_{(\text{Gruppe mittel-hoch})} = -3{,}040$, $p = 0{,}002$). Damit ist einerseits sichergestellt, dass der FIT 12 die Behaltensleistungen unabhängig vom motorischen Entwicklungsstand erfasst und andererseits ein Hinweis auf einen möglichen Zusammenhang des deklarativen Gedächtnisses mit der expressiven Sprachentwicklung gegeben.

Diskussion

Der FIT 12 wurde als Aufgabenserie zur Erfassung der deklarativen Gedächtnisleistung 12 Monate alter Kinder entwickelt. Die 92 an dieser Studie teilnehmenden Kinder mit einem mittleren Alter von 362,4 Tagen zeigten eine durchschnittliche Reproduktion von 2,38 von fünf Handlungen. Dabei ergab sich über die möglichen Summen der imitierten Handlungen von 0 bis 5 eine an eine Normalverteilung erinnernde Häufigkeitsverteilung mit einem deutlichen Gipfel bei drei Items. Damit erfasst der FIT 12 ein breites Spektrum von Imitationsleistungen Einjähriger. Es zeigten sich zudem keine Geschlechterunterschiede, was zeigt, dass die Handlungen des FIT 12 nicht nur altersgerecht, sondern für beide Geschlechter gleichermaßen geeignet sind.

Für die meisten Kinder ist mit drei Items der Umfang dessen erschöpft, was sie durch das serielle Lernen von Handlungen erfassen und/oder behalten können. Nur 17 Kinder reproduzieren vier oder fünf Items. Dabei fällt auf, dass die meisten Kinder bei der Reproduktion mit dem ersten Item beginnen und der vorgegebenen Struktur der Items folgen. Sie verstehen also von Beginn an die Regel des Ablaufs, enkodieren Verknüpfungen zwischen den einzelnen Items und nutzen diese itemrelationale Information für den Abruf, wie es bereits die Studie von Knopf et al. (2006) zeigen konnte. Dort konnte durch Variation der Vorgabereihenfolge der Items in der Abruf- gegenüber der Präsentationsphase nachgewiesen werden, dass die Imitationsleistung bei Wechsel der Vorgabereihenfolge bedeutsam abfällt.

Der Schwierigkeitsgrad der Einzelitems, hier nicht im testtheoretischen Sinne Itemschwierigkeit, ergibt sich aus dem Zuwachs der Ausführungsrate in der Imitationsphase gegenüber der Basisrate, die im Rahmen der Studie von Goertz et al. (under revision) erhoben wurde. Die Ausführungshäufigkeit einer Handlung in der Basisratenerfassung gibt Auskunft darüber, wie sehr sich die Handlung aus den Objekteigenschaften ableiten lässt. Bei der Festlegung der Reihenfolge der Items für den FIT 12 wurde darauf geachtet, dass lediglich die Items aufgenommen wurden, die eine sehr niedrige oder niedrige a-priori-Ausführungsrate hatten. Zudem erfolgte die Anordnung gemischt, indem jeweils Items mit sehr geringer und geringer Basisrate abwechselnd dargeboten. Das letzte Item Schlagzeug wurde nur von wenigen Kindern reproduziert, es unterscheidet sich in seiner Ausführungshäufigkeit in der Imitationsphase nicht vom Basisratenniveau. In früheren Studien (Natour, 2001) fand das Item Schlagzeug auch bei 12 bis 15 Monate alten Kindern Anwendung, wobei Ausführungsraten von bis zu 78% erreicht wurden, dort allerdings im Kontext von nur insgesamt vier Items und anderer Reihenfolge. Es ist also zu vermuten, dass im vorliegenden Fall dieses Item auf Grund seiner Position in der Aufgabenserie und damit aus Gründen der begrenzten Aufmerksamkeit und/oder Gedächtniskapazität eine so niedrige Ausführungsrate erreichte. Weitere Studien mit veränderter Darbietungsreihenfolge werden hier Aufschluss geben.

Die Zeit vom Überreichen der Spielobjekte an die Kinder bis zum Ausführen der Handlungen wurde festgehalten. Diese Latenzzeiten hängen auch von den Ausführungszeiten der Handlungen ab: So werden Handlungen, die aus nur einem Handlungsschritt bestehen (Dose, Schwein und Maus) schneller ausgeführt als zweistufige Handlungen (Becher/Messer und Schlagzeug). Es wurde jedoch kein Zusammenhang zwischen der Summe der nachgeahmten Handlungen und der Schnelligkeit ihrer Ausführung ermittelt, das heißt, Kinder, die viele Handlungen nachahmten, waren nicht zwangsläufig schneller als Kinder mit einer geringeren Gedächtnisleistung. Andererseits zeigte sich auch kein Effekt im Sinne eines Geschwindigkeits-Genauigkeits-trade-offs, dass also Kinder, die sich Zeit bis zur Ausführung der Handlung ließen, nicht dadurch mehr Handlungen imitierten als Kinder, die schnell mit der Nachahmung begannen.

Die Reliabilität der Aufgabenserie wurde in einer Studie von Goertz et al. (under revision) mit einem Test-Retest-Design nachgewiesen. Im Abstand von sieben Tagen wurde der FIT 12 mit 24 Kindern im Alter von 12 Monaten durchgeführt und es zeigte sich eine hohe Kurzzeitstabilität der Imitationsleistungen der Kinder ($r = 0.52$) ähnlich zu Reliabilitätswerten bei Habituations- oder Wiedererkennensaufgaben, die in diesem Altersbereich üblicherweise gefunden werden (McCall & Carriger, 1993). Durch eine individuelle Konsistenzanalyse nach Asendorpf (1990) wurden zwei Kinder identifiziert, die eine sehr inkonsistente Gedächtnisleistung gezeigt hatten; nach ihrem Ausschluss lag die Test-Retest-Reliabilität sogar bei $r = 0{,}65$. Eine hohe Auswertungskonkordanz des FIT 12 ist mit Werten von $\kappa = 0{,}78$ und $\kappa = 0{,}83$ ebenfalls gegeben.

Beim Vergleich der nach ihrer Gedächtnisleistung unterschiedenen Gruppen wurde deutlich, dass die Kinder mit zunehmender Gedächtnisleistung auch tendenziell höhere Werte im ET6-6 (Petermann & Stein, 2000) erreichen, statistisch bedeutsam unterschied sich aber nur die leistungsstärkste Gruppe im Entwicklungsstand ihres expressiven Sprachverhaltens, nicht aber in ihrer motorischen oder kognitiven Entwicklung. Das verweist darauf, dass die sprachlich weiterentwickelten Kinder ihre schon vorhandenen sprachlichen Fähigkeiten beim Enkodieren, Speichern und/oder Abrufen der Handlungen des FIT 12 nutzen können und so zu besseren Gedächtnisleistungen in der Lage sind. Da der FIT 12 ein nahezu sprachfreier Test ist, kann die Überlegenheit der sprachlich weiterentwickelten Kinder nicht an ihrem besseren Verständnis der Instruktion insgesamt liegen. Gegen eine solche Alternativerklärung spricht auch die Beobachtung, dass auch die schwächeren Kinder beim Gedächtnisabruf bevorzugt mit dem ersten Item beginnen, ihnen also der Ablauf des Tests von Anfang an verständlich ist. Ferner sind die Kinder mit den besseren Gedächtnisleistungen nicht diejenigen, die bessere motorische Kompetenzen aufweisen oder einen besseren allgemeinen kognitiven Entwicklungsstand zeigen würden.

Mit zunehmendem Lebensalter wird tatsächlich erwartet, dass die Sprache für die Bewältigung von Lern- und Gedächtnisaufgaben eine Rolle spielt, wie dies in der Tradition der verbalen Gedächtnisforschung mit weiter entwickelten Kindern untersucht und belegt wurde (Nelson, 1984). Anzumerken ist jedoch, dass der ET6-6 noch ein recht neuer Test ist und die Normierungsstichprobe für diesen Altersbereich kleiner war ($N = 61$, Petermann & Stein, 2000) als die der vorliegenden Studie. Für dieses Alter werden die Summenscores der einzelnen Dimensionen teilweise aus sehr wenigen Items gebildet, so dass insgesamt die Ergebnisse des ET6-6 und die daraus abgeleiteten Bezüge zum FIT 12 mit Vorsicht zu interpretieren sind.

Während die Kapazität des Kurzzeitgedächtnisses bei Zweijährigen zwischen zwei und drei Items liegt (Schneider & Pressley, 1989), erfasst der FIT 12 die Anfänge des seriellen Lernens bei Einjährigen und zeigt, dass in diesem Altersbereich die durchschnittliche Lern- und Gedächtnisleistung bei drei Handlungen liegt. Der Imitationstest ist somit mit hoher Wahrscheinlichkeit ein Vorläufer von seriellen Lern- und Gedächtnisaufgaben mit Bildern oder Wörtern, wie sie bei Kindern ab drei Jahren eingesetzt werden. Damit liegt ein Verfahren vor, das es erlaubt, serielles Lernen und Erinnern bereits bei Einjährigen zu analysieren und somit die Gedächtnisforschung bei präverbalen und verbalen Kindern mehr und mehr ineinander zu überführen.

Anders als in anderen Studien zur Verzögerten Imitation, die vor allem an der Erfassung des Behaltensintervalls interessiert waren (z. B. die Arbeiten von Meltzoff, Bauer und Hayne), bietet die vorliegende Studie genauere Informationen über eine Aufgabenserie von fünf Items. Die detaillierte Information über einzelne Items erlaubt in der Zukunft zum Beispiel Strukturanalysen der Reproduktionsleistung der Kinder, für die in der vorliegenden Studie bereits vielversprechende Hinweise gefunden wurden. Da diese Studie Teil der Längsschnittstudie FRAMES ist, bei der Kinder in ihrer Gedächtnisentwicklung bis zum Alter von 36 Monaten studiert werden und neben dem Gedächtnis auch Maße zur Entwicklung des Selbst sowie die Entwicklung sekundärer Repräsentationen über Als-ob-Spiel und Invisible-Displacement-Aufgaben erfasst werden, wird sowohl der prädiktive Wert des FIT 12 abzuschätzen sein als auch Zusammenhänge des deklarativen Gedächtnisses mit anderen Bereichen der kognitiven Entwicklung darstellbar.

Durch weitere Studien mit angrenzenden Altersgruppen und einer Normierung für den deutschen Sprachraum ist die Entwicklung eines standardisierten Tests zur Messung des deklarativen Gedächtnisses von Einjährigen denkbar, den es für dieses Alter noch nicht gibt. Die vorliegende Aufgabenserie FIT 12 kann aber bereits jetzt erfolgreich im Bereich der Gedächtnisforschung bei Säuglingen Anwendung finden.

Literatur

Asendorpf, J. (1990). The measurement of individual consistency. *Methodika, 4,* 1–13.

Baddeley, A. (1986). *Working memory.* Oxford, Claredon Press.

Bauer, P. J. (2002). Early memory development. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 127–150). Oxford: Blackwell.

Bauer, P. J., Hertsgaard, L. A. & Dow, G. A. (1994). After 8 months have passed: Long-term recall of events by 1-to 2-year-old children. *Memory, 2,* 353–382.

Bauer, P. J., Hertsgaard, L. A. & Wewerka, S. S. (1995). Effects of experience and reminding on long-term recall in infancy: Remembering not to forget. *Journal of Experimental Child Psychology, 59,* 260–298.

Bauer, P. J. & Mandler, J. M. (1989). One thing follows another: Effects of temporal srtucture on 1- to 2-year olds' recall of events. *Developmental Psychology, 25,* 197–206.

Bauer, P. J., Wiebe, S. A., Waters, J. M. & Bangston, S. K. (2001). Reexposure breeds recall: Effects of experience on 9-month-old's ordered recall. *Journal of Experimental Child Psychology, 80,* 174–200.

Goertz, C., Kolling, T., Frahsek, S., Stanisch, A. & Knopf, M. (2005). Assessing declarative memory in 12-month-old infants: A test-retest reliability study based on the deferred imitation task. *European Journal of Developmental Psychology* (under revision).

Hayne, H., Boniface, J. & Barr, R. (2000). The development of declarative memory in human infants: Age-related changes in deferred imitation. *Behavioral Neuroscience, 114,* 77–83.

Hayne, H. & Herbert, J. (2004). Verbal cues facilitate memory retrieval during infancy. *Journal of Experimental Child Psychology, 89*, 127–139

Johnson, M. H. (1997). *Developmental cognitive neuroscience*. Oxford: Blackwell.

Knopf, M., Kraus, U. & Kressley-Mba, R. A. (2006). Relational information processing of novel unrelated actions by infants. *Infant Behavior & Development, 29*, 44–53.

Kressley-Mba, R. A. & Knopf, M. (2005). A comparison of between- and within-subjects design for deferred imitation. *Infancy*. Manuscript submitted for publication.

Mandler, J. M. (1990). Recall of events by preverbal children. In A. Diamond (Ed.), *Development and neural bases of higher cognitive functions* (pp. 485–516). New York: New York Academy of Sciences.

Mandler, J. M. (2004). *The foundations of mind.* Oxford: Oxford University Press.

McCall, R. B. & Carriger, M. S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development, 64*, 57–79.

McDonough, L., Mandler, J. M., McKee, R. D. & Squire, L. R. (1995). The deferred imitation task as a nonverbal measure of declarative memory. *Proceedings of the National Academy of Sciences, 92*, 7580–7584.

Meltzoff, A. (1985). Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. *Child Development, 56*, 62–72.

Meltzoff, A. (1988 a). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology, 24*, 470–476.

Meltzoff, A. (1988b). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development, 59*, 217–225.

Natour, N. (2001). *Entwicklung des Gedächtnisses in den ersten zwei Lebensjahren.* Münster: Waxmann.

Nelson, K. (1984). The transition from infant to child memory. In M. Moscovitch (Ed.), *Infant memory: Its relation to normal and pathological memory in humans and other animals* (pp. 103–130). New York, London: Plenum Press.

Petermann, F. & Stein, I. A. (2000). *Entwicklungstest ET 6-6*. Lisse: Swets & Zeitlinger Publishers.

Piaget, J. (1975). *Nachahmung, Spiel und Traum*. Stuttgart: Klett-Cotta (Original 1959 „Imitation, jeu et rêve – Image et représentation“).

Piaget, J. & Inhelder, B. (1975). *Gedächtnis und Intelligenz*. Olten: Walter-Verlag (Original 1968 „Mémoire et intelligence“).

Rovee-Collier, C., Hayne, H. & Colombo, M. (2000). *The development of implicit and explicit memory*. Amsterdam: John Benjamins Publishing Company.

Schneider, W. (2002). Memory development in childhood. In U. Goswami (Ed.), *Blackwell Handbook of Child Cognitive Development* (pp. 236–256). Oxford: Blackwell.

Schneider, W. & Pressley, M. (1989). *Memory development between 2 and 20*. New York: Springer.

Tulving, E. (2002). Episodic memory and common sense: how far apart? In A. Baddeley, M. Conway & J. Aggleton (Eds.), *Episodic memory* (pp. 269–287). Oxford: Oxford University Press.

Dr. Claudia Goertz

Johann Wolfgang Goethe-Universität Frankfurt am Main
Institut für Psychologie – Arbeitseinheit
Entwicklungspsychologie
Georg-Voigt-Straße 8
60054 Frankfurt am Main
E-Mail: <cl.goertz@psych.uni-frankfurt.de>

Paper 2

Der Frankfurter Imitations-Test für 18 Monate und 24 Monate alte Kinder (FIT18-24)

Goertz, C., **Kolling, T.**, Frahsek, S., & Knopf, M. (subm.). Der Frankfurter Imitations-Test für 18 Monate und 24 Monate alte Kinder (FIT18-24). *Kindheit und Entwicklung*.

Kolumnentitel: FRANKFURTER IMITATIONSTESTS FIT 18 UND FIT 24

Die Frankfurter Imitationstests für 18 Monate alte Kinder (FIT 18) und für 24 Monate alte Kinder (FIT 24): Entwicklung altersangepasster Instrumente für die Verzögerte Imitation

Claudia Goertz, Thorsten Kolling, Stefanie Frahsek und Monika Knopf

Johann Wolfgang Goethe-Universität Frankfurt am Main

Zusammenfassung

Das Verfahren der Verzögerten Imitation gilt als eine Methode, durch die deklaratives Gedächtnis bei vorsprachlichen Kindern erfasst werden kann. Dabei werden den Kindern neuartige, objektbezogene Handlungen gezeigt, die sie sich lediglich ansehen dürfen. Nach einem Behaltensintervall werden die Objekte sukzessive präsentiert, das Spiel der Kinder wird analysiert und mit dem Spontanspiel einer Kontrollgruppe verglichen, die das Modell zuvor nicht beobachten konnte. Während wenig umstritten ist, dass dieses Verfahren deklaratives Gedächtnis erfasst, fehlt es an altersangepassten Testinstrumenten. Vorgestellt werden neu entwickelte Testverfahren für 18 und 24 Monate alte Kinder (Frankfurter Imitationstests FIT 18 bzw. FIT 24) und mit einem Test für jüngere Kinder in Beziehung gesetzt (FIT 12). Für diese Instrumente lässt sich nachweisen, dass das Spielverhalten nicht spontan erfolgt sondern zuvor erlernt wurde. Ferner erweisen sich die Tests als objektiv auswertbar, hinsichtlich ihrer Schwierigkeit als altersangemessen sowie als imstande, den alterskorrelierten Zuwachs der Gedächtnisleistung wie die interindividuellen Unterschiede innerhalb der Altersstufen abzubilden. Ein Vergleich mit den Ergebnissen im ET6-6 zeigt an, dass vor allem Dimensionen, die auch Imitationsverhalten abbilden, mit dem Summenwert der FIT-Tests korrelieren.

Schlüsselwörter: deklaratives Gedächtnis, Verzögerte Imitation, FIT 18, FIT 24, ET6-6

Abstract

Deferred imitation is accepted as a method to assess declarative memory in preverbal children. New object-related actions are demonstrated and, at first, the children can only observe them. After a retention interval the objects are then given successively to the children, their play is analyzed and compared to the spontanous play of a control group that did not observe the model. While it is widely accepted that deferred imitation assesses declarative memory, age-adapted instruments are lacking. New testing procedures for 18- and 24-month-olds are presented (Frankfurt Imitation Tests FIT 18 and FIT 24) and related to a test for 12-month-olds (FIT 12). There is evidence for these tests that the behavior children show, in fact, is learned and not spontanous play. Moreover, the tests can be analyzed objectively, are age-adapted in their difficulty and manage to demonstrate the age-related increase in memory performance as well as interindividual differences within one age-group. A comparison with results of the ET6-6 indicates that dimensions assessing imitative behavior are related to the FIT-test score.

Key words: declarative memory, deferred imitation, FIT 18, FIT 24, ET6-6

Einleitung

Die Entwicklung des kindlichen Gedächtnisses steht von jeher im Mittelpunkt des Interesses der kognitiven Entwicklungspsychologie. Dabei hat sich vor allem in den letzten beiden Jahrzehnten viel für das Verständnis des frühen kindlichen Gedächtnisses getan. Gegenwärtig hat sich dabei vor allem ein Gedächtnismodell durchgesetzt, das sich einer Taxonomie bedient, die deklarative und nicht-deklarative Gedächtnisleistungen von einander unterscheidet (Tulving, 2002). Das deklarative Gedächtnis wird weiterhin in ein semantisches und ein episodisches Gedächtnis unterteilt, während dem nicht-deklarativen Gedächtnis zum Beispiel motorische Routinen, Priming-Prozesse, nicht-assoziatives Lernen und klassische Konditionierung zugerechnet werden, Phänomene, die gemeinsam haben, dass ihre Inhalte in aller Regel nicht bewusst abgerufen werden können.

Neben den Begriffen deklaratives und nicht-deklaratives Gedächtnis, die ausdrücken, dass die Inhalte des deklarativen Gedächtnisses sprachlich geäußert werden können, was für die Einträge des nicht-deklarativen Gedächtnisses nicht gilt, werden zunehmend und häufig synonym auch die Begriffe explizites und implizites Gedächtnis verwendet. Dieses Begriffspaar stammt aus der Tierforschung und erweist sich in der Hinsicht als angemessen, als dass ohnehin die Vorstellung, dass das bewusste Gedächtnis Sprache erfordert, aufgegeben wurde. Somit wird auch plausibel, dass schon bei vorsprachlichen Kinder explizites Gedächtnis nachgewiesen werden kann.

Imitation

Imitationsverhalten ist bereits bei Neugeborenen zu beobachten und findet als Lernmechanismus in der Forschung gerade in neuester Zeit viel Beachtung. Während wenige Stunden alte Kinder bereits Gesichtsgesten Erwachsener imitieren, können ältere Kinder bereits Sequenzen von Handlungen nachahmen, die ihnen eine Modellperson demonstriert hat. Meltzoff (2005) stellte in diesem Zusammenhang ein neues Modell der Grundlagen der kognitiven und sozialen Entwicklung vor, das die Repräsentation von Handlungen in den Mittelpunkt stellt und postuliert, dass Säuglinge beobachtete Handlungen speichern, als hätten sie sie selbst ausgeführt. Dies wird gestützt durch die Annahme, dass die Verarbeitung beobachteter und ausgeführter Handlungen vermutlich identische neuronale Netzwerke nutzt.

Mit zunehmendem Alter der Kinder können Demonstration und Imitation auch zeitlich entkoppelt werden. Diese Fähigkeit zur aufgeschobenen Nachahmung, die bereits von Piaget (1975) beschrieben wurde, wird heute mit Hilfe der Methode der Verzögerten Imitation von zahlreichen Forschergruppen untersucht (Meltzoff, 1985, 1988a, 1988b; Bauer & Mandler, 1989; Hayne, MacDonald & Barr, 1997). Mit ihrer Hilfe konnten Belege für deklaratives Gedächtnis bereits im frühen Kindesalter erbracht werden. Während die frühesten Befunde zeigen, dass schon im Alter von sechs Monaten die Fähigkeit zur Verzögerten Imitation vorhanden ist (Collie & Hayne, 1999; Hayne, Boniface & Barr, 2000), zeigen ältere Kinder längere Behaltenszeiten, der Umfang des Behaltenen nimmt zu, sie sind kontextunabhängiger und können Handlungen auf ähnliche Objekte übertragen, was als auch „repräsentationale Flexibilität“ bezeichnet wird (Hayne, Boniface & Barr, 2000).

In vielen Studien steht die Länge der Behaltenszeit im Zentrum des Interesses. So konnten Hayne und Herbert (2004) zeigen, dass sich 18 Monate alte Kinder über 4 Wochen hinweg etwa die Hälfte der demonstrierten drei Handlungen behalten konnten. In einer Studie von Klein und Meltzoff (1999) wurde nachgewiesen, dass einjährige Kinder von fünf demonstrierten Handlungen dreieinhalb Handlungen nach 3 Minuten imitierten, knapp 3 nach einer Woche und nach vier Wochen noch zweieinhalb Handlungen nachahmten, wobei sich statistisch kein Unterschied zwischen der Gedächtnisleistung nach einer und nach vier Wochen zeigte, der größere Teil des Erlernten also innerhalb der ersten Woche vergessen wurde. Eine andere Studie brachte den Nachweis dafür, dass sich 14 und 16 Monate alte Kinder Handlungen über 4 Monate hinweg merken können (Meltzoff, 1995). Die

Arbeitsgruppe um Harlene Hayne erprobte mit Dreijährigen bereits ein Behaltensintervall von einem Jahr (mündliche Kommunikation, 2005). In derartigen Studien werden meist nur sehr wenige Items verwendet, so zum Beispiel 4 Handlungen bei Meltzoff (1995), bei Hayne und Herbert (2004) zwei Items mit je drei Handlungsschritten.

Eine andere Gruppe von Studien widmet sich der Erforschung der Bedingungen des Behaltens, so wird zum Beispiel der Einfluss des Kontextes auf die Gedächtnisleistung näher beleuchtet, wie beispielsweise Veränderungen der Itemdetails (Hayne, Boniface & Barr, 2000; Natour, 2001), der Wechsel des Versuchsleiters (Natour, 2001), die Wirkung sprachlicher Hinweise während der Demonstrations- und Abrufphase (Hayne & Herbert, 2004), der Variation der Reihenfolge der Items in der Abrufphase (Knopf, Kraus & Kressley-Mba, 2006). Auch in diesen Studien werden eher wenige Items bzw. Handlungen verwendet. Seltener kommen mehr als vier Items zur Anwendung (sechs Items bei Meltzoff, 1988a; fünf Items bei Klein & Meltzoff, 1999; bis zu sechs Items bei Collie & Hayne, 1999).

Die in der vorliegenden Arbeit vorgestellten Tests sind entwickelt worden, um den Umfang des Behaltens von Kindern im Alter von 18 und 24 Monaten näher zu untersuchen. Dabei wurde stets mit einem Behaltensintervall von 30 Minuten gearbeitet, dass sich bereits in zahlreichen Studien als praktikabel erwiesen hat (Natour, 2001; Knopf, Kraus & Kressley-Mba, 2006; Kressley & Knopf, 2006; Goertz, Kolling, Frahsek, Stanisch & Knopf, im Druck). Auch im Frankfurter Imitationstest für 12 Monate alte Kinder (FIT 12, Goertz, Knopf, Kolling, Frahsek & Kressley, 2006) wurden altersangemessen fünf Items mit jeweils ein oder zwei Handlungsschritten zu einem Gesamttest zusammengefasst und ein Behaltensintervall von 30 Minuten verwendet. Es zeigte sich, dass die 12 Monate alten Kinder im Mittel vier Handlungsschritte imitieren können, wobei ein Spektrum von ein bis zur maximalen Anzahl von sieben Handlungsschritten ausgeschöpft wurde. Das Instrument für Kinder im Alter von 12 Monaten (FIT 12) wurde in einer Test-Reststudie auch auf seine Reliabilität hin geprüft, die nach einem Zeitintervall von einer Woche zwischen den zwei

Testungen bei r=0,52 lag (Goertz et al., im Druck).

Prinzipien der Testkonstruktion

In den hier vorgestellten Tests für 18 bzw. 24 Monate alte Kinder (FIT 18 bzw. FIT 24) wurden viele Handlungen integriert, so umfasst der Test für das Alter von 18 Monaten 12 Handlungsschritte und der für 24 Monate 28 Handlungsschritte. Es ging darum, gesunde, normal entwickelte Kinder dieses Alters hinsichtlich ihres Gedächtnisses differenzieren zu können. Bei der Entwicklung der Items und ihrer Zusammenstellung zu einem Imitationstest wurde darauf geachtet, (1) dass die mittlere Lösungshäufigkeit des Gesamttests bei etwa 50% lag, (2) eine geeignete Anzahl von Items unterschiedlicher Schwierigkeit dafür sorgte, dass weder Boden- noch Deckeneffekte auftreten, (3) dass die Reihenfolge der Items so gestaltet wurde, dass sowohl motorisch ähnliche als auch visuell ähnliche Handlungen nicht direkt aufeinander folgten, (4) dass sich Items in ihrer Aufeinanderfolge hinsichtlich ihrer spontanen Ausführungshäufigkeit (erfasst über die Basisraten in Kontrollstudien) und ihrer Schwierigkeit abwechselten, (5) weiterhin wurde bei der Konstruktion der Tests dafür Sorge getragen, dass die Handlungen in dem entsprechenden Alter motorisch leicht auszuführen sind, damit sichergestellt wurde, dass mit dem Instrument tatsächlich Gedächtnis und nicht motorische Geschicklichkeit erfasst wird. Eine weitere Besonderheit der einzelnen Items der Tests ist, dass sowohl kausal verknüpfte Handlungsschritte als auch nicht-zielführende Handlungsschritte enthalten sind. Diese lassen über eine einfache Analyse der Lösungshäufigkeiten (Itemschwierigkeit) hinaus auch eine differenzierte Betrachtung der Position dieser Handlungsschritte in der Abrufreihenfolge zu.

Herkömmlich benutzte Tests des deklarativen Gedächtnisses für Kinder

Da die meisten Gedächtnistests, selbst wenn sie mit Bildmaterial arbeiten, sowohl eine sprachliche Instruktion als auch eine verbale Antwort der Kinder erfordern, sind diese erst ab einem bestimmten Alter bzw. einem gewissen Stand der Sprachentwicklung verwendbar. In experimentellen Tests zur Kurzzeitgedächtnisspanne zeigte sich, dass die Kapazität von Kindern im Alter von zwei Jahren etwas mehr als zwei Items beträgt, ob nun die Wortspanne, die Buchstabenspanne oder die Zahlenspanne geprüft wurde (Dempster, 1981). In den Langzeitgedächtnistests für Kinder bis zu drei Jahren, die heute zum Beispiel im Rahmen von Entwicklungstests zur Anwendung kommen, beträgt die Anzahl zwischen 3 (z.B. im ET6-6 von Petermann, Stein & Macha, 2004: 3 Bilder Wiedererkennen und 3 Bilder freie Reproduktion für Dreijährige) und 6 Items („Schatzkästchen" aus dem WET, Kastner-Koller & Deimann, 1998). Diese Tests sind auch überwiegend dafür entwickelt worden, um niedrige Gedächtnisleistungen aufzuzeigen, sie sind hingegen nicht unbedingt geeignet, die Leistungsgrenze nach oben hin zu erfassen. Auch darum stellt die Entwicklung der Frankfurter Imitationstests eine Möglichkeit dar, das deklarative Gedächtnis von Kindern unabhängig vom Stand ihrer Sprachentwicklung mittels der Imitation neuartiger, beobachteter Handlungen zu erfassen. Durch die konzeptionelle Ähnlichkeit der Tests FIT 12, FIT 18 und FIT 24 ist sowohl ein prinzipieller Vergleich zwischen den verschiedenen Altersgruppen als auch eine längsschnittliche Anwendung möglich.

Methode

Stichprobe

18 Monate: Mit 87 Kinder (davon 47 Jungen) wurde der FIT 18 durchgeführt. Das mittlere Alter betrug 551,2 Tage (SD=7,9). Alle Kinder gehörten zur FRAnkfurter MEmory Study-Stichprobe (FRAMES), wiesen bei ihrer Geburt ein normales Geburtsgewicht und hohe APGAR-Werte auf und zeigten bis zum Alter von 18 Monaten keine Entwicklungsauffälligkeiten. Alle Kinder hatten im Alter von 12 Monaten bereits an der ersten Erhebung der FRAMES teilgenommen und somit auch den FIT 12 absolviert (Goertz et al., 2006). Durch technische Probleme konnten die Videoaufnahmen eines Kindes nicht vollständig ausgewertet werden, somit wird über die Ergebnisse von 86 Kindern berichtet.

24 Monate: Mit 87 Kinder (davon 47 Jungen) wurde der FIT 24 durchgeführt. Ihr mittleres Alter betrug 731,3 Tage (SD=10,63). Mit zwei Kindern, die mit 18 Monaten an der Studie teilgenommen hatten, konnte durch Umzug in eine entfernte Stadt der FIT 24 nicht durchgeführt werden. Jedoch nahmen zwei Kinder, die bereits an der ersten Messung von FRAMES teilgenommen und den FIT 12 absolviert hatten, beim FIT 18 jedoch wegen Krankheit bzw. technischer Probleme nicht in die Auswertung eingingen, nun erneut am FIT 24 teil. Durch teilweise Ausfälle einzelner Items blieben die Datensätze von 6 Kindern unvollständig, so dass schließlich die Ergebnisse von 81 Kindern in die Analyse eingingen. Mit allen Kindern wurde zusätzlich sowohl im Alter von 18 als auch von 24 Monaten der Entwicklungstests ET6-6 (Petermann, Stein & Macha, 2004) durchgeführt.

Kontrollstudien

Kontrollstudie 18 Monate: 26 Kindern (davon 18 Jungen) in einem mittleren Alter von 556,5 Tagen (SD=11,65), die ein normales Geburtsgewicht, eine Schwangerschaftsdauer von über 37 Wochen und keine bekannten Entwicklungsauffälligkeiten aufwiesen, nahmen an der Kontrollstudie zum spontanen Spielverhalten mit den Objekten des FIT 18 teil. Durch die Einbeziehung dieser Gruppe wurde die Basisrate des Zielverhaltens bestimmt.

Kontrollstudie 24 Monate: Mit 24 Kindern (davon 13 Jungen), die ein mittleres Alter von 731,4 Tagen (SD=8,57) aufwiesen und ebenfalls allen oben genannten gesundheitlichen Kriterien für die Teilnahme an einer Studie entsprachen, wurde eine Kontrollstudie zum spontanen Spiel mit den Objekten des FIT 24 durchgeführt, um die Basisrate für zweijährige Kinder zu bestimmen.

Versuchsdurchführung

Die Kinder wurden mit ihrer Begleitperson zu der Tageszeit zur Testung eingeladen, zu der sie üblicherweise am aktivsten waren. Im Labor saß das Kind auf dem Schoß der Begleitperson, die es von hinten stützte, sonst aber angehalten war, sich am Geschehen nicht zu beteiligen. Der Versuchleiter saß dem Kind gegenüber und war über kleine, unauffällige Kopfhörer mit dem zweiten Versuchsleiter verbunden, der vom Aufnahmeraum aus die Videogeräte bediente und die Präsentationszeiten durchsagte.

Nach einer Aufwärmphase von maximal 180 Sekunden, in der mit einem neutralen Spielzeug (18 Monate: Ringpyramide; 24 Monate: Holzrampe und Elefant) gespielt wurde, begann die Demonstrationsphase. Jede Handlung wurde innerhalb von 30 Sekunden dreimal (18 Monate) bzw. zweimal (24 Monate) demonstriert. Dabei wurde darauf geachtet, dass das Kind die Demonstration aufmerksam beobachtete. Die Handlungen wurden nicht verbal kommentiert, die Interaktion beschränkte sich auf Bemerkungen wie: „Schau mal her, <Name>!“ oder „Sieh mal, <Name>, was ich hier habe.“

Das Labor wurde nach der Demonstrationsphase verlassen. Nach einem Behaltensintervall von 30 Minuten wurde das Labor wieder aufgesucht und der Gedächtnistest durchgeführt. Im Anschluss an eine erneute Aufwärmphase von maximal 60 Sekunden begann die Imitationsphase. Dem Kind wurden das Objekt des ersten Items überreicht und für 30 Sekunden zum Spiel überlassen. Dann wurde das Objekt von ihm zurückerbeten und ihm das Objekt des nächsten Items überreicht. Dies wurde für alle Items in der gleichen Reihenfolge wiederholt, wie sie in der Demonstrationsphase gezeigt wurden.

Material

FIT 18: Der Imitationstest besteht aus sechs Items, die sich aus Teilhandlungen verschiedener Anzahl zusammensetzen und zu denen jeweils ein oder mehrere Objekte gehören. Die genaue Beschreibung aller Objekte und Handlungsschritte ist der Tabelle 1 zu entnehmen.

FIT 24: Der Imitationstest für 24 Monate alte Kinder besteht aus acht Items mit jeweils drei bis sechs Teilhandlungsschritten. Bei zwei Items gab es neben den Targetobjekten auch Distraktoren. Abbildung 1 enthält die Darstellung eines Beispiel-Items des FIT 24, der „Schildkröte“. Diese Handlung besteht aus drei Handlungsschritten.

hier die Abbildung 1 einfügen

Die Beschreibung aller Objekte und Handlungsschritte des FIT 24 ist der Tabelle 2 zu entnehmen.

Interrater-Reliabilität

FIT 18: Die Ausführung der Zielhandlungen wurde anhand von operationalen Definitionen von jeweils zwei unabhängigen Auswertern entschieden. Insgesamt waren vier Auswerter beteiligt, die folgende Interrater-Reliabilität erreichten: Auswerter 1 und Auswerter 2 (356 Entscheidungen) = 95,8%, κ= 0,912; Auswerter 2 und Auswerter 3 (415 Entscheidungen) = 96,4%, κ= 0,926; Auswerter 3 und Auswerter 4 (261 Entscheidungen) = 96,2%, κ= 0,923). Nichtübereinstimmungen wurden durch einen „forced consent“ bereinigt.

FIT 24: Von drei Auswertern kamen Auswerter 1 und Auswerter 2 auf eine Übereinstimmung von 96,3%, was einem κ von 0,925 entspricht. Sie beurteilten insgesamt

2381 Entscheidungen. Auswerter 2 und Auswerter 3 erreichten bei 684 Entscheidungen eine Übereinstimmung von 93,86%, was einem κ von 0,873 entspricht. Nichtübereinstimmungen wurden wiederum durch einen „forced consent“ bereinigt.

Der Entwicklungstest ET6-6

Der Entwicklungstest ET6-6 von Petermann, Stein und Macha (2004) wurde in einer leicht abgewandelten Form angewandt. Da die Kinder der FRAMES-Studie jeweils in einem Zeitfenster von zwei Wochen vor bis zwei Wochen nach dem Alter von 18 bzw. 24 Monaten getestet wurden, hätten die Kinder eines Messzeitpunkts eigentlich verschiedene Items lösen müssen, da die Altersgrenze jeweils exakt bis zu Alter von 18 bzw. 24 Monaten reicht. So wurde allen Teilnehmern der Studie eine Itemzusammenstellung aus der Altergruppe bis 18 und bis 21 Monate vorgelegt, entsprechend den 24 Monate alten bis zu 30 Monaten. Dies hatte den Vorteil, dass alle Kinder zu einem Messzeitpunkt die gleichen Items zu lösen hatten und dass eventuelle Akzelleration in einzelnen Entwicklungsbereichen abbildbar wurden. Die Berechnung der Testwerte wurde durch eine angepasste Norm vorgenommen.

Ergebnisse

Der Frankfurter Imitationstest für 18 Monate alte Kinder (FIT 18)

Die mittlere Imitationsleistung lag bei $\underline{M}$ = 6,9 Handlungsschritten ($\underline{SD}$ = 1,85), es wurden mindestens 3 und maximal 11 von 12 möglichen Handlungsschritten erreicht, die Schiefe der Verteilung beträgt = 0,19, der Exzess = -0,19. Dem Kolmogorov-Smirnov-Test zufolge kann die Verteilung mit einer Normalverteilung gleichgesetzt werden ($\underline{Z}$ = 1,51, $\underline{p}$ = 0,021). Die Schwierigkeiten und Trennschärfen der Handlungsschritte sind der Tabelle 1 zu entnehmen.

hier Tabelle 1 einfügen

Die Ausführungshäufigkeiten der einzelnen Handlungsschritte erweisen sich als unterschiedlich, die Items Maus und Schlagzeug erreichen Ausführungsraten von 90% und darüber. Sie stellen für dieses Alter sehr leichte Items dar. Sie waren aus dem FIT 12 übernommen worden, in dem sie die schwierigsten Items darstellten mit 20,9% (Maus) und 24,2% (Schlagzeug) Ausführungshäufigkeit (Goertz et al., im Druck). Weiterhin wird deutlich, dass bei den zweischrittigen Handlungen Auto, Gans, Schlagzeug und Ente jeweils der erste Schritt häufiger als der zweite Schritt ausgeführt wird. Allein bei der dreischrittigen Handlung Frosch zeigt sich für den ersten Handlungsschritt „Frosch macht Kopfstand“ die niedrigste Imitationshäufigkeit. Hingegen weisen die beiden folgenden Handlungsschritte „Frosch hüpft in den Ring“ und „Frosch rutscht hin und her“ dann wieder das Muster auf, dass der erste Schritt häufiger ausgeführt wird als der zweite (hier dann der zweite häufiger als der dritte). Eine mögliche Erklärung dafür ist, dass der Kopfstand des Froschs nicht „notwendig“ ist, um dann in den Ring zu hüpfen und zu rutschen. Entweder, dieser Handlungsschritt „Kopfstand“ ist nicht auffällig, salient genug, als dass ihn sich die Kinder merken können, oder er wird als „irrelevant“ für das eigentliche Ziel der Handlung, das Rutschen, erkannt und weggelassen. Das würde implizieren, dass die Kinder ein (vermeintliches) Ziel einer Handlung ausmachen können.

Kontrollstudie 18 Monate: In einer flankierenden Studie mit 26 Kindern wurde anhand des spontanen Spielverhaltens im Umgang mit den Objekten des FIT 18 die Basisrate ermittelt. Die mittlere Basisrate betrug 1,08 (SD = 1,05) von 12 möglichen Handlungsschritten. Das Maximum betrug 4 Handlungsschritte. Je neun Kinder führten keinen oder einen Schritt spontan aus, sechs Kinder zwei Schritte sowie jeweils ein Kind drei oder vier Handlungsschritte. Die Teilschritte der Handlungen Auto und Frosch liegen in ihrer Spontanausführung sämtlich bei Null, ebenso wie die Schritte Gans 2 und Ente 2. Gans 1, Maus und Ente 1 liegen zwischen 10 und 20%. Nur Schlagzeug 1 („Klöppel abnehmen") erreicht eine sehr hohe Basisrate von über 60%. Die Tabelle 1 ermöglicht einen Vergleich zwischen Basisrate, Gedächtnisleistung in der Kontrollstudie sowie der Behaltensleistung in der FRAMES-Studie. Daran wird unter anderem auch deutlich, dass sich die Behaltensleistungen in den Gedächtnistests der Kontrollstudie und der FRAMES-Studie bei den Items Maus und Schlagzeug nicht unterscheiden, was als Hinweis darauf verstanden werden kann, dass sie die Kinder der FRAMES-Studie bei diesen beiden Items, die ja bereits im FIT 12 enthalten waren, keinen Vorteil durch Behaltensreste über die sechs Monate hinweg hatten.

Der Frankfurter Imitationstest für 24 Monate alte Kinder (FIT 24)

Die mittlere Behaltensleistung lag bei 17,09 Handlungsschritten (SD = 3,78) und reichte von minimal 6 bis maximal 25 von 28 möglichen Handlungsschritten. Die Verteilung ist einer Normalverteilung gleichzusetzen (KS-Test: $Z = 1,47$, $p = 0,026$), die Schiefe beträgt -0,67 und der Exzess = 0,29. Die Schwierigkeiten sowie die Trennschärfen der einzelnen Handlungsschritte sind der Tabelle 2 zu entnehmen (der Schritt Boot 3: „Männchen aus der Dose nehmen", wurde von der Analyse ausgeschlossen, da es in den meisten Fällen zu einem Herausfallen des Männchens kam).

hier Tabelle 2 einfügen

Die Items des FIT 24 weisen mit ihren insgesamt 28 Teilhandlungsschritten unterschiedliche Schwierigkeiten auf. Dabei zeigen die Items „Gondel“ und „Schildkröte“ mit ihren jeweils drei Handlungsschritten abnehmende Ausführungshäufigkeiten, die Items „Hase“ und „Magnetteller“ eher homogene Imitationshäufigkeiten der Teilschritte, während bei „Boot“, „Frosch“, „Ball“ und „Kästchen“ die Teilschritte heterogene, von ihrer Reihenfolge unabhängige Schwierigkeiten aufweisen. Einige Teilschritte erreichen sehr hohe Imitationsraten, so liegen die Schritte „Gondel 1“, „Boot 1“ und „Boot 2“, „Ball 1“ und „Ball 2“ jeweils über 90 %. Als die schwersten Handlungsschritte liegen „Schildkröte 3“ und „Magnetteller 6“ unter 10% Imitationshäufigkeit. Das Item „Magnetteller“ erweist sich als relativ problematisch, da die Schritte 2 bis 5 eine etwa gleich hohe Behaltensrate von ca. 80% aufweisen, bei gleichzeitig hoher Basisrate um 40% (siehe unten, Kontrollstudie). Damit liefern diese vier Handlungsschritte für den Gedächtnistest wenig Information. Allein der erste Schritt („Teller umdrehen“) sowie der letzte Schritt („Teller hin- und herdrehen“) weisen einen gewissen Erkenntnisgewinn auf, da sie sich von der Basisrate unterscheiden (Schritt 1) bzw. von nur wenigen Kindern ausgeführt werden (Schritt 6). Trotzdem stellt das Item „Magnetteller“ einen Kontext für die anderen Items dar, so dass es auch nicht vom Test ausgeschlossen werden sollte.

Kontrollstudie 24 Monate: In einer flankierenden Studie, an der 24 Kinder teilnahmen, wurde die Basisrate des FIT 24 bestimmt. Die mittlere Anzahl der spontan ausgeführten Zielhandlungen betrug 3,54 (SD =2,72), wobei 0 bis 9 Handlungsschritte ausgeführt wurden. Im anschließenden Gedächtnistest mit der Kontrollgruppe wurde eine mittlere Gedächtnisleistung von 16,92 (SD =4,0) erreicht, wobei die Behaltensleistung zwischen 8 und 23 Handlungsschritten lag. Zahlreiche Handlungsschritte werden spontan nicht ausgeführt, so beispielsweise alle Schritte des Items „Schildkröte“ überhaupt nicht (Basisrate und Gedächtnistest der Kontrollstudie je Handlungsschritt sind der Tabelle 2 zu entnehmen). Auffällig ist, dass der Handlungsschritt Frosch 2 („Frosch hüpft“) eine höhere Basisrate aufweist als im Gedächtnistest erreicht wird. Im spontanen Spiel wird mit dem Frosch gehüpft, da dies ein Frosch offenbar nahelegt. Im Gedächtnistest, wo das Hüpfen des Frosches der mittlere, vergleichsweise unwichtige Schritt für das eigentliche Ziel, „Frosch rutscht die Rutsche hinab“ ist, wird dieser Handlungsschritt seltener ausgeführt als in der Basisrate (40% versus 25%).

FIT 12, FIT 18 und FIT 24

Zieht man auch die Ergebnisse des FIT 12 (Goertz et al., 2006) hinzu, kann der Leistungsanstieg in der Verzögerten Imitation von 12 bis 24 Monaten nachgewiesen werden. Von einer mittleren Behaltensleistung von 4,02 Teilschritten mit 12 Monaten steigern sich die Kinder mit 18 Monaten auf 6,9 Teilschritte und erreichen mit 24 Monaten schließlich im Mittel 17,09 Teilschritte. Berechnet man für alle drei Tests die T-Werte (siehe dazu Tabelle 3) wird deutlich, dass vor allem die beiden Tests FIT 18 und FIT 24 recht gut differenzieren, der FIT 12 hat mit seinen 7 Handlungsschritten eine zu geringe Anzahl. Dabei differenzieren FIT 18 und FIT 24 vor allem im unteren Bereich sehr gut, während der obere Bereich nicht ganz ausgeschöpft wird.

hier Tabelle 3 einfügen

Zusammenhänge mit dem Entwicklungstest ET6-6

FIT 18: Berücksichtigt man die Ergebnisse in den einzelnen Dimensionen des Entwicklungstests ET6-6 (Petermann, Stein & Macha, 2004), so zeigt sich ein Zusammenhang (siehe dazu Tabelle 4) zwischen der Gedächtnisleistung im FIT 18 und der Dimension „Handlungstrategien“ ($\underline{r} = 0,37$; $\underline{p} = 0,001$), der Interaktion mit Gleichaltrigen ($\underline{r}$

= 0,28; p = 0,031) sowie für die Faktoren „kognitive Entwicklung“ (r = 0,28; p = 0,012) und „soziale Entwicklung“ (r = 0,30; p = 0,007). Weiterhin zeigt sich eine Tendenz zu einem Zusammenhang zwischen Imitationsleistung im FIT 18 und „sozialer Eigenständigkeit“ (r = 0,21; p = 0,053), dem Faktor „emotionale Entwicklung“ (r = 0,21; p = 0,060) sowie dem Gesamtscore des ET6-6 (r = 0,21; p = 0,052).

hier Tabelle 4 einfügen

FIT 24: Für die 24-monatige FRAMES-Gruppe zeigt sich ebenfalls ein starker Zusammenhang zwischen Imitationsleistung im FIT 24 und der Dimension „Handlungsstrategien“ (r = 0,30; p = 0,008) im ET6-6 (Petermann, Stein & Macha, 2004). Weiterhin korrelieren die Imitationsleistungen mit der rezeptiven Sprachentwicklung (r = 0,25; p = 0,027), mit den Faktoren „kognitive Entwicklung“ (r = 0,28; p = 0,015), „Sprachentwicklung“ (r = 0,23; p = 0,040), „soziale Entwicklung“ (r = 0,26; p = 0,021) sowie mit dem Gesamtscore des ET6-6 (r = 0,29; p = 0,009) (siehe auch dazu Tabelle 4).

Diskussion

Mit den vorliegenden Ergebnissen konnte nachgewiesen werden, dass sich die beiden Instrumente FIT 18 und FIT 24 dazu eignen, die Imitationsleistungen bei Kindern im Alter von 18 und 24 Monaten zu erfassen. Die Testinstrumente sind in Umfang und Schwierigkeit altersangepasst und bilden das mögliche Spektrum der deklarativen Gedächtnisleistungen in diesem Altersbereich ab. Anhand des Vergleichs mit den Basisraten aus den Kontrollstudien wird deutlich, dass mit den beiden Instrumenten tatsächlich Gedächtnisleistungen erfasst werden, da das spontane Verhalten im Umgang mit den Spielobjekten in nur geringem Umfang zur Ausführung einzelner Schritte der Zielhandlungen führt.

Das Besondere an diesen beiden Imitationstests ist, dass mit einem Aufwand von insgesamt weniger als einer Stunde bei einer Sitzung die Behaltensleistung des deklarativen Gedächtnisses geprüft werden kann, wobei beide Verfahren durch ihre große Anzahl der Handlungsschritte ermöglichen, auch gesunde, normal entwickelte Kinder hinsichtlich ihrer Gedächtnisleistung zu differenzieren.

Während Tests, die mit freier Reproduktion von zum Beispiel Bildern oder Worten arbeiten, in dem Altersbereich unter drei Jahren nur eine geringe Itemanzahl aufweisen und gleichzeitig Restriktionen durch den noch frühen Stand der Sprachentwicklung unterliegen, konnte mit den beiden Frankfurter Imitationstests durch die nicht-sprachliche Methode der Verzögerten Imitation der erstaunlich große Umfang des Gedächtnisses für Handlungen bei Kindern im Alter von 18 und 24 Monaten nachgewiesen werden. Berücksichtigt man zusätzlich noch die Ergebnisse, die an der gleichen Stichprobe mit dem FIT 12 im Alter von 12 Monaten erhoben wurden (Goertz et al., 2006), so lässt sich der Zuwachs des Gedächtnisumfangs im Laufe des zweiten Lebensjahres anschaulich illustrieren: Von anfänglich durchschnittlich vier Handlungen, die mit 12 Monaten behalten werden, umfasst die mittlere Behaltensleistung mit 18 Monaten bereits knapp sieben Handlungen, mit 24 Monaten schließlich werden im Mittel 17 Handlungsschritte nachgeahmt.

Die Analyse des Zusammenhangs zwischen Imitationsleistung in den beiden Tests FIT 18 und FIT 24 und den Ergebnissen des ET6-6 (Petermann, Stein & Macha, 2004) ergab für beide Alterszeitpunkte ein recht ähnliches Bild. Höhere Imitationsleistungen gingen auch mit höheren Werten in der Itemgruppe „Handlungsstrategien“ einher, was sich auch in einer Korrelation mit dem Faktor „kognitive Entwicklung“ niederschlug. Die Testitems für die Dimension „Handlungsstrategien“ umfassen zahlreiche Items, bei denen dem Kind etwas demonstriert wird, das es dann nachahmen und fortsetzen soll (T55: „Stapelt drei Würfel“, T57 „Aneinanderreihen von mindestens drei Würfeln“, T59 „Pyramide bauen“). Somit wird auch hier die Fähigkeit zur Imitation erfasst und eine Korrelation mit den Imitationstests FIT 18 und FIT 24 ist naheliegend. Es ergab sich kein Zusammenhang zur Dimension „Handmotorik“, was ein Hinweis darauf ist, dass die Tests FIT 18 und FIT 24 so konstruiert sind, dass sie nicht motorische Geschicklichkeit abbilden, sondern vielmehr für Kinder des entsprechenden Alters motorisch zu bewältigen sind. Weiterhin zeigte sich zu beiden Alterszeitpunkten ein Zusammenhang zum Faktor „soziale Entwicklung“, bei den 18 Monate alten Kindern auch in der Dimension „Interaktion mit Gleichaltrigen“. Dabei handelt es sich um Fragebogenitems, die das Zusammenspiel und auch das Imitieren von Gleichaltrigen erfassen (F35 „Es antwortet einem Gleichaltrigen durch eine Geste oder einen Zuruf.“ und F36 „Es spielt in der Nähe eines gleichaltrigen Kindes: beide spielen das Gleiche, aber jedes für sich allein (Parallelspiel)“. Ein Zusammenhang zwischen der Imitationsleistung im FIT 18 und dem Verhalten gegenüber anderen Kindern erscheint hier deshalb durchaus schlüssig, da sich jeweils das Imitationsverhalten auswirkt. Bei den 24 Monate alten Kindern ergab sich weiterhin ein Zusammenhang zwischen der Imitationsleistung im FIT 24 und der Dimension „rezeptive Sprachentwicklung“ des ET6-6, was sich dann auch als Korrelation mit dem Faktor „Sprachentwicklung“ niederschlug. Da die Verzögerte Imitation ein nicht-sprachlicher Test ist, außer „Schau mal, was ich hier habe.“ oder „Jetzt bist du dran mit spielen.“ kaum sprachliche Äußerungen vorgenommen werden, kann es sich um eine

Korrelation handeln, hinter der wiederum eine gemeinsame Fähigkeit steckt. So wäre denkbar, dass Kinder, die besonders häufig in Interaktion mit Erwachsenen stehen, sowohl in ihrer rezeptiven Sprachentwicklung gefördert als auch in ihrem Imitationsverhalten verstärkt werden.

Mit den Frankfurter Imitationstests FIT 12, FIT 18 und FIT 24 lässt sich anhand des Imitationsverhaltens das deklarative Gedächtnis von Kindern unabhängig vom Stand ihrer Sprachentwicklung erfassen, wobei durch die konzeptionelle Ähnlichkeit der Tests sowohl ein prinzipieller Vergleich zwischen den verschiedenen Altersgruppen als auch eine längsschnittliche Anwendung möglich sind. Damit stellen die Frankfurter Imitationstests Vorläufer von seriellen Lern- und Gedächtnisaufgaben mit Bildern oder Wörtern dar, wie sie bei Kindern ab drei Jahren eingesetzt werden und erlauben es, serielles Lernen und Erinnern bereits bei Kindern im Alter von 18 und 24 Monaten zu analysieren und somit die Gedächtnisforschung bei präverbalen und verbalen Kindern mehr und mehr ineinander zu überführen. Die simultane Erfassung von nicht-sprachlichen und sprachlichen deklarativen Gedächtnisleistungen ist eine theoretisch interessante Möglichkeit, die weiter zu klären vermag, in welcher Relation nicht-sprachliche Imitationstests wie die hier vorgelegten und die herkömmlichen sprachlichen Gedächtnistests zueinander stehen. Ferner ist in Nachfolgestudien zu prüfen, inwieweit sich die nicht-sprachlichen Imitationstests für die Analyse von Gedächtnisleistungen von Kindern eignen, die Entwicklungsverzögerungen in der Sprachentwicklung oder sprachliche Defizite insgesamt aufweisen.

Autorenhinweis

Die vorliegende Studie ist Teil der von der DFG geförderten Längsschnittstudie „Zur Ontogenese des Gedächtnissystems bei Säuglingen und Kleinkindern“ (Kn 275/3-1 und Kn 275/3-3). Wir danken Christina Mack, Margit Feyler, Ali Koc, Denise Ginzburg und Frauke Saelz für die Mithilfe bei der Entwicklung und Durchführung der Aufgaben sowie allen Kindern und ihren Familien.

Kontaktadresse:

Dr. Claudia Goertz

Johann Wolfgang Goethe-Universität Frankfurt am Main

Institut für Psychologie - AE Entwicklungspsychologie

Georg-Voigt-Str. 8

60054 Frankfurt am Main

Tel.: (069) 798 25 049

Fax: (069) 798 28 595

email: cl.goertz@psych.uni-frankfurt.de

Literatur

Bauer, P. J. & Mandler, J. M. (1989). One thing follows another: Effects of temporal structure on 1- to 2-year olds' recall of events. Developmental Psychology, 25, 197-206.

Collie, R. & Hayne, H. (1999). Deferred imitation by 6- and 9-month-old infants: More evidence for declarative memory. Developmental Psychobiology, 35, 83-90.

Dempster, F.N. (1981). Memory span: Sources of individual and developmental differences. Psychological Bulletin, 89, 63-100.

Goertz, C., Knopf, M., Kolling, T., Frahsek, S. & Kressley, R. A. (2006). Entwicklung und Erprobung eines Messinstruments zur Erfassung des deklarativen Gedächtnisses Einjähriger: Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12). Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38, 88-96.

Goertz, C., Kolling, T., Frahsek, S., Stanisch, A., & Knopf, M. (in press). Assessing declarative memory in 12-month-old infants: A test-retest reliability study based on the deferred imitation task. European Journal of Developmental Psychology.

Hayne, H., Boniface, J., & Barr, R. (2000). The development of declarative memory in human infants: Age-related changes in deferred imitation. Behavioral Neuroscience, 114, 77-83.

Hayne, H. & Herbert, J. (2004). Verbal cues facilitate memory retrieval during infancy. Journal of Experimental Child Psychology, 89, 127-139.

Hayne, H., MacDonald, S., & Barr, R. (1997). Developmental changes in the specifity of memory over the second year of life. Infant Behavior & Development, 20, 233-245.

Kastner-Koller, U. & Deimann, P. (1998). Der Wiener Entwicklungstest: Ein allgemeines Entwicklungstestverfahren für Kinder von 3 bis 6 Jahren. Göttingen: Hogrefe.

Klein, A. & Meltzoff, A. (1999). Long-term memory, forgetting, and deferred imitation in 12-month-old infants. Developmental Science, 2, 102-113.

Knopf, M., Kraus, U. & Kressley-Mba, R. A. (2006). Relational information processing of novel unrelated actions by infants. Infant Behavior & Development, 29, 44-53.

Kressley-Mba, R. A. & Knopf, M. (2006). A comparison of between- and within-subjects design for deferred imitation. Infant Behavior & Development, 29, 564-573.

Meltzoff, A. (1985). Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. Child Development, 56, 62-72.

Meltzoff, A. (1988a). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. Developmental Psychology, 24, 470-476.

Meltzoff, A. (1988b). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. Child Development, 59, 217-225.

Meltzoff, A.N. (2005). Imitation and other minds: The „like me“ hypothesis. In S. Hurley & N. Chater (Eds.), Perspectives on imitation: From neuroscience to social science (Vol. 2, pp. 55-93). Cambridge, MIT Press.

Natour, N. (2001). Entwicklung des Gedächtnisses in den ersten zwei Lebensjahren. Münster: Waxmann.

Petermann, F., Stein, I.A. & Macha, T. (2004). ET6-6. Frankfurt: Harcourt Test Services.

Piaget, J. (1975). Nachahmung, Spiel und Traum. Stuttgart: Klett-Cotta (Original 1959 „Imitation, jeu et rêve - Image et représentation“).

Schneider, W. & Pressley, M. (1989). Memory development between 2 and 20. New York, Berlin: Springer-Verlag.

Tulving, E. (2002). Episodic memory and common sense: how far apart? In A. Baddeley, M. Conway & J. Aggleton (Eds.), Episodic memory (pp.269-287). Oxford: Oxford University Press.

Tabelle 1

Beschreibung der Handlungsschritte des FIT 18, ihre Basisrate, Itemschwierigkeit und Trennschärfe

Handlungs-schritte	Beschreibung der Handlungsschritte	Basisrate (Test Kontroll-studie)[a]	Item-schwierigkeit[b]	Trenn-schärfe[b]
Auto 1	Das gelbes Auto wird (wie ein Handschuh) über die Hand gezogen.	0 (0,58)	0,75	0,56**
Auto 2	Durch Knicken der Finger wird mit dem Auto gewinkt	0 (0,19)	0,30	0,45**
Gans 1	Der weißen Gans wird die silberne Dose an den Bauch geheftet.	0,1 (0,53)	0,71	0,27*
Gans 2	Die Gans wird auf die Dose gelegt.	0 (0,05)	0,17	0,56**
Maus 1	Die dunkelblaue Maus aus Holz wird zusammengedrückt.	0,19 (0,85)	0,90	0,18
Frosch 1	Der hellgrüne Frosch macht einen Kopfstand.	0 (0)	0,05	0,28**
Frosch 2	Der Frosch hüpft in den dunkelgrünen Ring.	0 (0,6)	0,69	0,51**
Frosch 3	Der Frosch rutscht mit dem Ring auf dem Tisch hin und her.	0 (0,24)	0,26	0,54**
Schlagzeug 1	Von dem bunten Plastikschlagzeug wird der blaue Klöppel abgenommen.	0,62 (0,88)	0,98	0,15
Schlagzeug 2	Mit dem Klöppel wird der große rote Knopf gedrückt (ein Geräusch ertönt).	0,07 (0,92)	0,94	0,28**
Ente 1	Die gelbe Ente hüpft auf die gelb-rote Krake (schaut mit dem Kopf zur Seite).	0,12 (0,68)	0,79	0,39**
Ente 2	Die Ente dreht sich auf der Krake hin zum Kind und zurück.	0 (0,12)	0,33	0,44**

Anmerkungen. [a] Basisrate beruhend auf Kontrollstudie ($N = 26$), Lösungshäufigkeiten im Test sind in Klammern angeben. [b] Beruhend auf FRAMES-Studie ($N = 86$).

* $p < 0{,}05$; ** $p < 0{,}01$

Tabelle 2

Beschreibung der Handlungsschritte des FIT 24, ihre Basisrate, Itemschwierigkeit und Trennschärfe

Handlungs-schritte	Beschreibung der Handlungsschritte	Basisrate (Test Kontroll-studie)[a]	Item-schwierig-keit[b]	Trenn-schärfe[b]
Gondel 1	Das schwarze Männchen wird in die weiße Gondel auf den Dorn gestellt.	0,08 (0,92)	0,92	0,29**
Gondel 2	Der Kochlöffel wird in die Gondel gestellt und an den Fahrer angelehnt.	0 (0,71)	0,41	0,58**
Gondel 3	Die Gondel fährt.	0,12 (0,29)	0,34	0,36**
Boot/Dose 1	Der blaue Filzschlauch wird von der blauen, länglichen Dose gezogen.	0,5 (1,0)	0,94	0,05
Boot/Dose 2	Die Dose wird durch Ziehen geöffnet.	0,28 (1,0)	0,97	0,29**
Boot/Dose 3	Das Männchen wird herausgenommen.	von der Analyse ausgeschlossen		
Boot/Dose 4	Die Beine des Männchens werden geknickt.	0 (0,29)	0,28	0,34**
Boot/Dose 5	Das Männchen wird in das Boot gesetzt.	0,08 (0,54)	0,40	0,32**
Frosch 1	Das Brett wird an den Sockel gelehnt.	0,04 (0,92)	0,80	0,35**
Frosch 2	Der grüne Frosch nimmt Anlauf und hüpft.	0,4 (0,25)	0,35	0,34**
Frosch 3	Der Frosch rutscht die Rutsche hinunter.	0 (54)	0,60	0,28*
Ball 1	Der schwarz markierte Schlitz des roten Schaumstoffballs wird gefunden.	0,44 (1,0)	0,99	0,19
Ball 2	Die Plastikaugen werden in den Schlitz gesteckt.	0,12 (1,0)	0,95	0,32**
Ball 3	Der Ball hüpft über den Tisch.	0 (0,33)	0,35	0,27*
Schildkröte 1	Die in blaues Tüll gehüllte Halbkugel und die Pyramide aus Styropor werden zusammengesetzt.	0 (0,71)	0,75	0,50**
Schildkröte 2	Beide Teile werden auf die Schildkröte gelegt.	0 (0,38)	0,52	0,61**
Schildkröte 3	Die Schildkröte fliegt.	0 (0,08)	0,11	0,36**
Hase 1	Das runde gelbe Kissen wird vorn an den braunen Stoffhasen geheftet.	0,12 (0,63)	0,61	0,37**
Hase 2	Das viereckige grüne Kissen wird auf den Kopf des Hasen geheftet.	0,12 (0,5)	0,57	0,32**
Hase 3	Das dreieckige rosa Kissen wird hinten angeheftet. (drei weitere Kissen =	0,12 (0,54)	0,57	0,35**

	Distraktoren)			
Kästchen 1	Der Pappring wird an den Haken des braunes Holzkästchen gehängt.	0 (0,42)	0,58	0,30**
Kästchen 2	Der Ring wird gedreht.	0,04 (0, 29)	0,35	0,25*
Kästchen 3	Die Schublade wird aufgezogen und ein weißes Vögelchen erscheint.	0,16 (0, 69)	0,52	0,24*
Magnetteller 1	Der beigefarbene Blechteller wird umgedreht.	0,08 (0,65)	0,71	0,26*
Magnetteller 2	Der rote Knopf wird oben aufgesetzt.	0,40 (0,78)	0,84	0,32**
Magnetteller 3	Der gelbe Knopf wird darunter gesetzt.	0,48 (0,70)	0,81	0,49**
Magnetteller 4	Der schwarze Knopf wird darunter geheftet.	0,44 (0,91)	0,86	0,34**
Magnetteller 5	Das Plastikcroissant wird unten angeheftet.	0,48 (0,78)	0,84	0,25*
Magnetteller 6	Der Teller wird hin und her gedreht. (zwei Plastikwürstchen = Distraktoren)	0 (0,13)	0,01	0,17

Anmerkungen. [a] Basisrate beruhend auf Kontrollstudie ($N = 24$), Lösungshäufigkeiten im Test sind in Klammern angeben. [b] Beruhend auf FRAMES-Studie ($N = 81$).

* $p < 0,05$; ** $p < 0,01$

Tabelle 3

T-Werte der drei Tests FIT 12, FIT 18 und FIT 24

Test	N	M	SD	Min	T20	T30	T40	T50	T60	T70	T80	Max
FIT 12	90	4,02	1,55	1		1,53	3,59	5,00	6,14			7
FIT 18	86	6,90	1,85	3	3,00	4,71	6,00	7,50	9,36	11,00		11
FIT 24	81	17,09	3,78	6	7,33	12,00	15,50	18,74	21,57	25,00		25

Tabelle 4

Korrelation zwischen den Gedächtnistests FIT 18 bzw. FIT 24 und Entwicklungstest ET6-6 mit 18 bzw. 24 Monaten

Dimensionen und Faktoren des ET6-6	ET6-6 Norm 18/21 Korrelation mit FIT 18	ET6-6 Norm 24/30 Korrelation mit FIT 24
Körpermotorik	0,09	0,14
Handmotorik	0,13	0,14
Handlungsstrategien	0,37**	0,30**
Kategorisierung	0,14	0,16
Körperbewusstsein	0,01	0,04
Sprachentwicklung rezeptiv	0,05	0,25*
Sprachentwicklung expressiv	0,08	0,17
Interaktion mit Erwachsenen	0,18	0,19
Interaktion mit Gleichaltrigen	0,24*	0,18
soziale Eigenständigkeit	0,21	0,19
Faktor emotionale Entwicklung	0,21	0,13
Faktor motorische Entwicklung	0,14	0,16
Faktor kognitive Entwicklung	0,28*	0,28*
Faktor Sprachentwicklung	0,07	0,23*
Faktor soziale Entwicklung	0,30**	0,26*
Gesamtscore	0,21	0,29**

* $p < 0{,}05$; ** $p < 0{,}01$

Abbildung 1

Beispiel-Item „Schildkröte“ aus dem FIT 24: Ausgangsposition, Schritt 1 „Halbkugel und Pyramide zusammensetzen“, Schritt 2 „beide Teile auf die Schilfkröte legen“, Schritt 3 „Schildkröte fliegt“

Paper 3

Assessing declarative memory in 12-month-old infants:

A test-retest reliability study based on the deferred imitation task.

Goertz, C., **Kolling, T**., Frahsek, S., Stanisch, A., & Knopf, M. (2007). Assessing declarative memory in 12-month-old infants: A test-retest reliability study based on the deferred imitation task. *European Journal of Developmental Psychology*, in press. Epub ahead of print retrieved May, 28, 2007, from http://www.informaworld.com/smpp/content~content=a778058220~jumptype=rss

EUROPEAN JOURNAL OF DEVELOPMENTAL PSYCHOLOGY
0000, 00 (00), 000–000, i*First* article

Assessing declarative memory in 12-month-old infants: A test–retest reliability study of the deferred imitation task

Claudia Goertz, Thorsten Kolling, Stefanie Frahsek, Annett Stanisch and Monika Knopf
Johann Wolfgang Goethe–University Frankfurt am Main, Frankfurt am Main, Germany

This study examined whether declarative memory in infants can be reliably assessed using the deferred imitation task. Twenty-four infants at the age of 12 months were given the same deferred imitation task twice within a short period of time (week-to-week assessment). Replicating the results of former studies the second memory test yielded better memory performances on the group level than the first one, indicating a memory benefit as is typically found in older children as well as in adults. Stability of memory performance level was analysed using two indicators, namely test–retest correlations assessing stability of individual memory performances for the whole sample, as well as corrected test–retest correlations using individual consistency scores. Test–retest reliability was highly significant ($r = .52$, $p = .009$), as well as corrected test–retest reliability ($r = .62$, $p = .001$), thus demonstrating that the individual memory performance level in infants can reliably be assessed using the deferred imitation task.

Deferred imitation has already been studied by Piaget (1975), who described the occurrence of deferred imitation of complex action sequences in natural contexts. Meltzoff (1985, 1988a, 1988b, 1988c) developed a standardized

Correspondence should be addressed to Claudia Goertz, Johann Wolfgang Goethe–University Frankfurt am Main, Department of Developmental Psychology, Georg-Voigt-Str. 8, D-60054 Frankfurt am Main, Germany. E-mail: cl.goertz@psych.uni-frankfurt.de

This study was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft) to Monika Knopf (KN 275/3–1).

We would like to thank Regina Kressley-Mba for her help during the preparation of the study and we gratefully acknowledge the assistance of Denise Ginzburg and Andrea Lotz in scoring the video tapes. Special thanks are due to all the babies and their families who participated in the study.

http://www.psypress.com/edp DOI: 10.1080/17405620600910186

deferred imitation task that has been used in many developmental studies in recent years. In this task, a short series of simple, object-based actions are successively shown in the laboratory context to young infants. The infants only observe the demonstration of actions without touching the action-related props. Both the different objects and the instrumental actions are novel to the infants. After a delay of several minutes, hours, days or even weeks the props are given successively to the infants and it is observed whether they perform the target actions or not. The performance of the target actions is seen as an evidence for declarative memory (Meltzoff, 1990, 1995).

To separate spontaneous behaviour from imitative behaviour the base rate of performing the target actions spontaneously is also assessed and compared with the occurrence of the target actions after they have been demonstrated. This is usually accomplished by either using an independent control group (Meltzoff, 1985, 1988a, 1988b, 1988c) or by assessing the baseline behaviour before the demonstration of the relevant target actions within the same subject sample (e.g., Bauer & Mandler, 1992). In a study by Kressley and Knopf (in press) imitative behaviour of between- and within-subjects imitation designs has been compared. There was no indication of an effect of prior contact with the objects on delayed imitation. Therefore, assessing the baseline behaviour with a within-subjects design seems to have no influence on the memory performance in the deferred imitation test.

Evidence for an early emergence of declarative memory at the age of 6 months was found by Collie and Hayne (1999) and Hayne, Boniface, and Barr (2000) using deferred imitation with a delay of 24 hours. With 9-month-old infants it has been demonstrated that deferred imitation is not only shown after short but also after longer retention intervals (e.g., Barr & Hayne, 1996; Bauer, Heertsgard, & Wewerka, 1995; Hayne et al., 2000; Kressley-Mba, Lurg, & Knopf, 2005; Mandler & McDonough, 1995; Meltzoff & Moore, 1999).

In the deferred imitation task infants reproduce instrumental actions, which they have only observed and which they do not incrementally acquire by performing them motorically. Therefore, it is assumed that the deferred imitation task assesses the ability to acquire, voluntarily retrieve or recall information. McDonough, Mandler, McKee, and Squire (1995) brought forward some additional arguments for the assumption that it is indeed declarative memory that is assessed in the deferred imitation task. They demonstrated that adults suffering from amnesia, where the declarative memory system as opposed to non-declarative memory is disrupted, are unable to show deferred imitation. Electrophysiological studies have yielded further supporting evidence (Carver, Bauer, & Nelson, 2000). Given these different considerations, the deferred imitation task has become an important instrument for assessing declarative memory in infants in recent years.

While the validity of the deferred imitation task as an instrument for assessing declarative memory is currently not under debate, it is still of major importance to test the psychometric adequacy of the deferred imitation task as a reliable memory instrument. The question is whether stable individual memory scores are found with this task after short time intervals, thus reliably testing individual memory performance level in infants.

The necessity of reliably assessing psychological abilities is fundamental for any psychological testing. The need to assess infant competencies and abilities in a reliable manner is even more crucial if the goal of developmental research is not only to describe age-related changes but also to analyse development on an individual level. The goal of describing and explaining individual patterns of development in order to predict later abilities on the basis of earlier assessments can only be reached when reliable assessments of the respective competencies and abilities are available for the different times of measurement (for an overview see Colombo & Fagan, 1990).

It is also useful to quantify the effects of short-term fluctuation in memory performance level and to distinguish it from the effects of true development. This leads to a more differentiated analysis of individual memory development by revealing the different factors that contribute to test outcomes in longitudinal measurement.

It is widely assumed that it is more difficult to get psychometrically reliable assessments for infants than it is for older children or for adults (e.g., McCall & Carriger, 1993), because: (a) infant behaviour is subject to higher instability and therefore each assessment is not as reliable as psychological testing in older children or even adults; and (b) each assessment in infants has to be based on only a few items and has to be realized in short time frames due to limited vigilance and the shorter attention span of infants, thereby greatly limiting the sample of assessed behaviour. The fact that testing in infants comprises a small number of items, typically, has the consequence that some reliability indicators, like internal consistency or split-half test reliability, are less appropriate for infants than test–retest reliability.

Up to now, the test–retest reliability of assessments in infants has most often been studied for habituation measures (e.g., Brian, Landry, Szatmari, Niccols, & Bryson, 2003). The findings in these studies are in line with the aforementioned argumentation. Bornstein, Slater, Brown, Roberts, and Barrett (1997), who provide an overview about this piece of research, state that the test–retest reliability for habituation measures assessed at points close in time (day-to-day and week-to-week assessments) yield good reliability estimates, ranging from $r = .40$ to $r = .60$. The intercorrelations between two testings from points more distant in time (month-to-month assessments) deliver low stability estimates (*r*s around .20). However, one may question whether these low individual stabilities found after longer time intervals do in fact mirror an unreliability of the habituation measure or rather reflect interindividual differences in the onset or speed of development.

McCall and Carriger (1993) provided a meta-analytic review of infant habituation and recognition memory performance as predictors of later IQ. Short term test–retest-reliabilities were rather low, ranging from $r = .30$ to $r = .45$. Different explanations for poor reliabilities of infant measures are discussed by McCall and Carriger (1993), e.g., the issue that different aspects of stimuli are processed at different developmental stages. The poor short-term reliability in these studies may also be due to the small number of items that these memory tests typically consist of. McCall and Carriger therefore suggest aggregating several assessments and thus increasing age-to-age reliability, as did Rose, Feldman, and Wallace (1988) as well as Colombo, Mitchell, and Horowitz (1988).

There are some longitudinal studies that used the deferred imitation task repeatedly to follow memory development in individuals, showing some stability of the individual memory performance over time, modest in younger children and increasing throughout infancy (Heimann & Meltzoff, 1996; Nielsen & Dissanayake, 2004). Since, to our knowledge, a short-term test–retest reliability test for this memory assessment procedure is still lacking, it is unknown whether these findings reflect interindividual variability in developmental patterns in infants and young children or are due to the unreliability of the assessment procedure.

The aim of the present study was to investigate short-term reliability of memory performance level in infants using the deferred imitation task. As has been realized in the Mandler tradition (e.g., Bauer & Mandler, 1992), the baseline was assessed in a within-subjects design. After the demonstration of the target actions and a retention interval of 30 minutes the memory test was assessed. A retention interval of 30 minutes is supposed to assess long-term memory; in addition, this time interval is often used in studies with infants.

The split-half method of assessing reliability was not considered because the respective testing would have been too long to be realized with infants. Similarly, not enough items were available to determine test reliability following the parallel test method. So a test–retest design was used assessing declarative memory twice within a short period of time. Choosing the time interval between the two tests had to be a compromise between two opposing effects. On the one hand ceiling effects should be prevented which might be the result of retention effects. So the test–retest interval had to be long enough to prevent the infants from benefiting too much from the first test. On the other hand, the retest should be given at a point so close in time to the first test that individual memory performance levels should not have changed due to developmental processes. Schneider and Sodian (1997) assessed test–retest reliability of memory performance of 4-year-old children in a hide-and-seek task within a 2-week interval. It is assumed that retention effects are larger in older children than in infants. Moreover it seems reasonable to suppose that the progress in development is more rapid in toddlers and young infants compared to older children, so that individual performance level changes more quickly in younger compared to older children. Therefore, a shorter test–retest interval, as has been used in the study of Schneider and Sodian (1997), seemed to be adequate.

In addition, Klein and Meltzoff (1999) demonstrated in a deferred imitation study with 12-month-old infants that most of the forgetting occurred during the first week after the demonstration of the actions, actually. Three different retention intervals have been used by Klein and Meltzoff (1999). An imitation score of 3.47 out of 5 target actions has been found in a first memory test that was given after a 3-minute-interval, a score of 2.81 target actions in a second memory test that was done after one week, and a score of 2.56 imitated target actions resulted after a 4-week retention interval. Given these different considerations it was concluded that a time interval of one week between the two tests might be optimal.

METHOD

Participants

24 children (9 girls) with a mean birth weight of 3409.2 g ($SD = 425.3$) took part in the study. Their age ranged from 11 months and 15 days to 12 months and 15 days, with a mean age of 363 days ($SD = 9.4$). All children fitted into this age at both test sessions. Participants were recruited by notices sent to mother–child recreational groups and to local paediatricians, and by word of mouth. One infant was excluded because it was born four weeks prior to the calculated date of birth.

Procedure

Appointments were made for the time of day when the parents reported that their children were most alert, usually mid-morning. An interview prior to the experiment was conducted to explain the purpose of the study, details of the procedure, and to obtain informed consent. In the lab, the child was seated on the parent's lap. The child's hips were firmly supported by the accompanying adult so that children could move both arms freely and had optimal access to objects placed on the table. The experimenter was sitting opposite to the child. A warm-up toy was given to the child to accustom it to the test environment as well as to provide a first indicator of the infant's ability to engage in visually guided reaching.

Session 1. The first session consisted of baseline testing, the demonstration phase, the delay, as well as the first memory test (test 1). After up to 180 s in a warm-up phase, as soon as infants appeared comfortable, baseline testing began. The first object was given to the infant for 30 s to assess spontaneous playing. Then the demonstration of the first object-based target action began. It was shown four times within 30 s. This procedure, baseline testing and demonstration of a target action, was repeated with the remaining four props. If an infant became distracted, the experimenter tried to redirect the infant's attention to the task by saying: "[Name], look at this". The experimenter refrained from calling the props by name or describing any part of the target action. After a delay of 30 minutes, a test for deferred imitation was administered to infants (test 1). Infants were given each of the props sequentially for 30 s in the same order as in the demonstration phase.

Session 2. Seven days later session 2 was conducted, consisting of a second demonstration of the same target actions, the delay, and a second deferred imitation test (test 2). After a warm-up phase of up to 180 s all five target actions were demonstrated successively four times by the experimenter within 30 s. Order of objects and target actions were the same as in session 1. Again the experimenter tried to redirect the infant's attention when distracted by saying: "[Name], look at this". After a delay of 30 minutes test 2 was administered. The children were given each of the five objects sequentially for 30 s in the same order as before.

Material and apparatus

The five props used in this study are shown in Figure 1. The objects were commercially available toys and were specially adapted for the experiments. Some of the objects and target actions were used in previous work (Knopf, Kraus, & Kressley-Mba, 2006; Kressley & Knopf, in press).

Objects and target actions (all target actions were repeated four times):

1. *Tin can*: Shaking the can. A blue tin can (diameter = 4.5 cm) filled with rice was shaken up and down three times. The tin can given to the infant during baseline testing and the two memory tests was empty in order to prevent children from shaking it to produce the perceived noise.

Figure 1. Objects: tin can, pig, cup and knife, wooden mouse, drum.

2. *Toy pig with a hat*: Removing the hat. The hat was removed from a light pink stuffed toy pig (20 cm × 15 cm × 7 cm) wearing a brown hat attached by Velcro.
3. *Cup and knife*: Stirring back and forth with the knife in the cup. The upside down red plastic cup (diameter = 8 cm) was turned with the open side up and the wooden knife (length = 16.5 cm) was inserted and moved back and forth in the cup. This action is highly similar to an item used by Klein and Meltzoff (1999), the blue plastic box and the wooden stick, accept that in the present study the cup had to be turned first. This variation was done to diminish the spontaneous realization of the target behaviour in the baseline phase.
4. *Wooden mouse*: Shutting the mouth of the mouse. The upper yellow half of a wooden mouse (8 cm × 5.5 cm) with an attached tail (6 cm) was shut four times.
5. *Drum*: Pressing a red button with a drum stick. A multi-coloured plastic drum (11 cm × 9 cm) had one large red button (diameter = 5 cm) and three smaller buttons on the top as well as two small buttons on the side. The blue drumstick (length = 8 cm) attached with a white string was pressed against the red button twice in order to produce a battery-run noise. In baseline testing and in the test phases infants received a drum without noise in order to prevent children from producing the sound by pressing the button with their fingers.

Data collection and data scoring

The experimental sessions were video-taped by two cameras, one taping the infant and the second the experimenter. Two naïve and independent observers scored target action completion by using operational definitions. They were uninformed about the hypotheses of the study, how exactly the experimenter demonstrated the target actions, and which phase of the study they were actually watching (baseline testing, test 1, or test 2). Target actions were scored as "yes" or "no" responses.

The operational definitions for the target actions were:

1. *Tin can*: A "yes" was coded if the tin can was shaken with one or both hands. The movement up and down or back and forth should have been done more than once.
2. *Toy pig*: A "yes" was coded if at least an attempt to draw the hat from the pigs head was observed. Simple touching the hat was not coded as "yes".
3. *Cup and knife*: A "yes" was coded if an attempt at putting the knife into the cup was observed. The knife had to be inserted at least a bit into the cup.
4. *Mouse*: A "yes" was coded if the child pressed on the top of the mouse with the hand or with single fingers no matter whether the mouse was placed on the table or on one hand of the infant.
5. *Drum*: A "yes" was coded if the drum stick was taken and the red button on the drum top was touched with it (no matter with which side of the stick). An accidental fall of the stick on the button was not coded as "yes".

The two scorers reached an interrater reliability of $r = 89.8\%$ and a Cohen's kappa of $\kappa = .78$. Diverging decisions were equally distributed over baseline phase, test 1, and test 2 (12, 12, and 13) and neither of the scorers showed a response bias (observer 1: 227 "yes" responses/133 "no" responses; observer 2: 226 "yes" responses/134 "no" responses). Discrepancies in scoring were resolved by consensus to 100% agreement.

RESULTS

Target action completion

As in our earlier studies (Knopf et al., 2006; Kressley & Knopf, in press) it was found again that each of the five actions is adequate to assess deferred imitation in 12-month-olds. While two actions, namely toy pig and mouse, have very low spontaneous performance rates in the baseline phase, the three other actions, namely tin can, cup and knife and drum, were spontaneously performed somewhat more often. Irrespective of the base-rate level, however, for all five items target action completion in test 1 was at least more than 10% higher than in the baseline phase and again at least more than 10% higher in test 2 compared to test 1. Therefore all five items were generally imitated in the two different memory tests.

The number of mean target action completions in baseline testing, test 1, and test 2 are given in Table 1. Comparison of mean target action completions yielded a significant difference between the three phases: mean target actions in baseline testing = .83 ($SD = 1.0$), in test 1 = 1.92 ($SD = 1.2$), and in test 2 = 3.17 ($SD = 1.1$), respectively (Friedman-Test: $df = 2$, $\chi^2 = 30.06$, $p < .001$; Wilcoxon tests: baseline – test 1: $Z = 3.056$, $p = .002$; test 1 – test 2: $Z = 3.613$, $p < .001$).

The role of prior contact with target-related objects on test 1

Of the 24 subjects, 12 showed one or more target actions spontaneously in baseline testing. In order to test a possible influence of prior contact with target-related objects on mean deferred imitation level two subgroups were distinguished. While the 12 infants in one group did not show any target behaviour in the baseline phase, the 12 subjects in another group spontaneously realized at least one target action in the baseline phase. The mean number of actions completed in test 1 M of $\text{group}_{\text{baseline0}} = 1.67$ ($SD = 1.16$, min – max: 0 – 4), M of $\text{group}_{\text{baseline} \geq 1} = 2.17$ ($SD = 3.75$, min – max: 1 – 4). There was no difference in mean memory performance level between these two groups in test 1 (Mann – Whitney-U: $Z = -0.782$; $p = .239$). The finding that no differences between these two groups occurred in the memory test, which was administered 30 minutes after the baseline phase, indicates that there seems to be no effect of the prior contact with the objects on delayed imitation, thus replicating the findings of Kressley and Knopf (in press).

Realization of target behaviour at the three different assessment procedures

To further test to what extent imitation rate in both memory tests may be regarded as an effect of acquisition of the target actions in the demonstration phases respectively as an effect of retention benefit from a previous phase, the proportion of "old" and "new" target actions performed was calculated for the two memory tests. These findings are

TABLE 1
Mean action completion in the three phases baseline testing, test 1, and test 2

	M	*SD*	*min – max*
Baseline	.83	1.0	0 – 3
Test 1	1.92	1.21	0 – 4
Test 2	3.17	1.13	0 – 5

given in Figure 2. Of the target actions completed in test 1, 30.2% had already been shown in baseline, whereas 71.8% of the actions were newly acquired. Similarly, 47.3% of the actions performed in test 2 had already been completed in test 1, whereas 44.8% of the target actions completed in test 2 had not been performed before. Further 7.9% of the target actions in test 2 had been performed in baseline but not in test 1.

In addition, Figure 2 reveals that not all target actions completed in baseline testing were recalled in test 1. Moreover, actions imitated in test 1 were not inevitably imitated in test 2. Given the fact that most of the actions performed in both tests were not shown in earlier test phases, recall of actions in the deferred imitation task seems to be mostly due to encoding the respective actions in the demonstration phase and less due to pre-experimental knowledge or retention effects of earlier phases of the study.

Stability of individual memory performance levels

Though there is a significant increase of memory performance on the group level from test 1 to test 2, the stability of individual memory performance

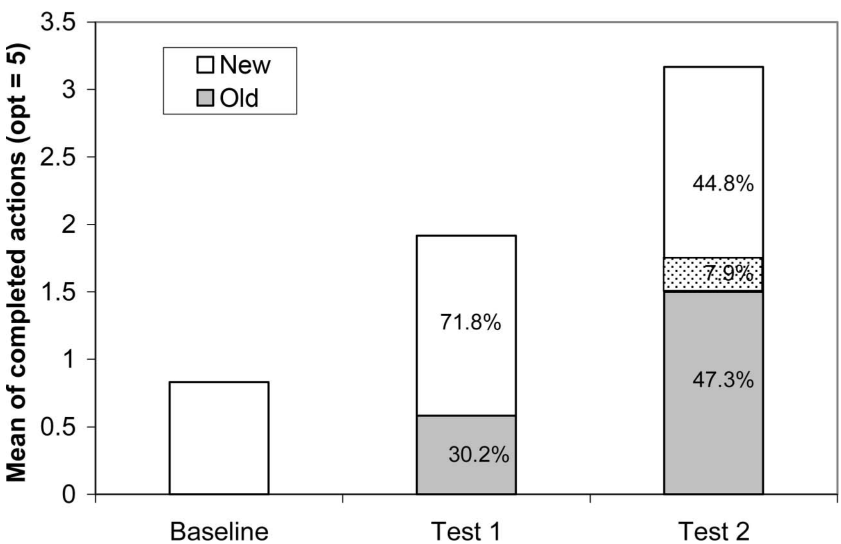


Figure 2. Learning effect or retention benefit from the previous phase. *Note*: Target actions completed in test 1 are classified as "old", if they were already shown in baseline testing, otherwise as "new". In correspondence in test 2 target actions already completed in test 1 are classified as "old", extent of increase during the test 2 is classified as "new". The dotted area identifies the 7.9% of target actions completed in baseline but not in test 1.

level is high: The test – retest correlation between deferred imitation scores in test 1 and test 2 is $r = .52$ ($p = .009$). It is therefore concluded that the deferred imitation task is a reliable instrument for assessing memory differences in infants.

Test – retest correlations express mean group stabilities at aggregate levels. While test – retest correlations indicate stability of memory performance for all individuals of the sample, this analysis does not allow a differential analysis of individual stability or lability of memory performance over time (Asendorpf, 1989, 1990). Individual stability reflects the relative position of an individual within a reference group across time. From a more differential perspective the stability of individual development pattern might be interesting. To take a more differential perspective on the test – retest data in a second analysis the individual consistency (IC) scores separately for each infant were computed using the formula $1 - [(Z_x - Z_y)^2]/2$ as proposed by Asendorpf (1990). As can be seen from the formula the relative position of an individual (*z*-score) on one measurement occasion is subtracted from its relative position on the other measurement occasion. The higher the difference between the *z*-scores, i.e., the more different the relative position of an individual at the two measurement occasions, the lower the IC-index. As the index may vary between $-\infty$ and 1 Asendorpf (1989) proposed a transformation for the IC scores, which normalizes the skewed distribution. A comparison of the IC-scores before and after transformation revealed negligible statistical differences for the subjects studied here. The untransformed individual consistency scores are given in Figure 3. As is shown, two infants, specifically, show low consistency scores. Via the procedure proposed by Asendorpf these two cases are identified as subjects who performed differently in the two memory tasks compared with the other subjects and therefore they were treated as outliers. The threshold was arbitrarily set at $-.50$.

After the exclusion of these two subjects having scores lower than $-.50$ from data analysis, the test – retest correlation based on the individual performance level was $r = .65$, $p = .001$ ($n = 22$). This finding shows that the test – retest reliability of this instrument for assessing declarative memory increases when those infants having very low stability scores are identified and excluded from data analysis.

DISCUSSION

The present study assessed the reliability of a deferred imitation task, which is often used to study declarative memory in preverbal infants with a test – retest design. A sample of 24 12-month-old infants was given the same two demonstration phases and memory tests within a 1-week interval. Five target actions were demonstrated to them twice and their imitative

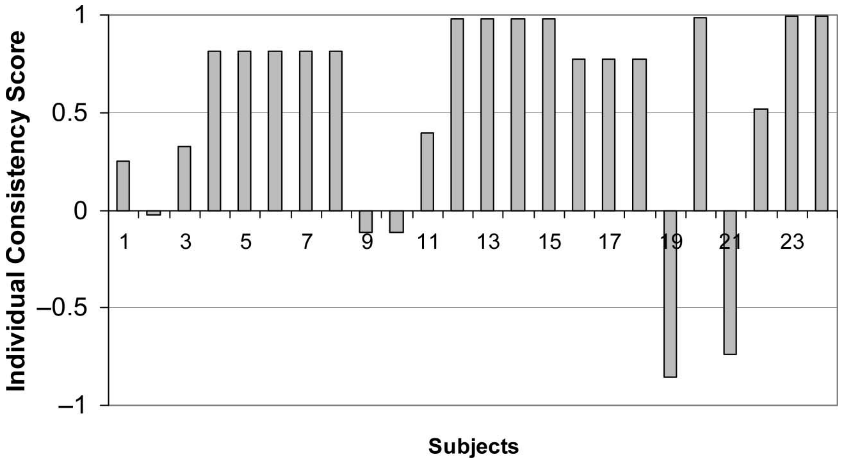


Figure 3. Individual consistency scores. *Note*: Individual consistencies were computed with the formula $1-[(Z_x-Z_y)^2]/2$ proposed by Asendorpf (1990). The individual consistency index varies between $-\infty$ and 1.

behaviour was also assessed twice. The baseline of the spontaneous demonstration of the target actions was assessed in a within-subjects design.

Whereas the mean number of target action completions in test 1 was 1.92 out of 5 actions, 3.17 out of 5 actions were recalled in test 2. The fact that a second demonstration of the same actions resulted in a higher memory performance level compared to a first demonstration, as is the case in older children and adults in similar memory tasks (e.g., serial word learning tasks), may be seen as another argument for the assumption that it is indeed declarative memory that is assessed in the deferred imitation task.

The intraindividual stability of memory performance level was estimated with a test–retest procedure. Test–retest correlations for the completion of target actions during test 1 and test 2 were highly stable ($r = .52$, $p = .009$). A similar result was found in a second analysis, which considered the individual consistency scores, as proposed by Asendorpf (1990). In this analysis only two infants were highly inconsistent in the number of target actions performed in the two memory tests. Excluding these two outliers from computing the test–retest reliability yielded a reliability score of $r = .65$ ($p = .001$). Considering the high developmental dynamics in infants at the end of their first year this is a reliability appropriate to that age and comparable to the reliabilities reported in studies with other different indicators of information processing in infants varying between $r = .3$ and $r = .45$ (e.g., McCall & Carriger, 1993). It therefore seems that the development of interindividual differences in declarative memory in infants can be assessed using this type of memory task.

As is shown in this study, while most infants show stable performance levels with respect to declarative memory at the end of their first year, there are some infants with very inconsistent memory performances. It would be highly interesting to analyse in future longitudinal studies what differentiates these two groups—stable and unstable memory performance level.

Although designing items for deferred imitation for that age group is a demanding task, assessing reliability with a parallel test paradigm would eliminate the influence of retention effects and is therefore desirable. By additionally assessing the imitative behaviour in test 2 before the demonstration the retention effect would be explicitly measured and its contribution to memory performance in test 2 could be identified.

The present study produced evidence that: (a) despite the fact that the behaviour of younger infants is subject to higher instability and variability, declarative memory can reliably be assessed in infants at the end of their first year in a deferred imitation test; (b) though the assessment had to be realized with only a few items, due to the limited attention span, the results yielded reliable results; and (c) an increase in memory performance level from test 1 to test 2 thus speaks for a memory benefit, as is found in similar studies with older children or adults. This finding is in line with the assumption that it is declarative memory that is being assessed with the deferred imitation task.

With the information about the reliability of this deferred imitation task a more differentiated analysis of factors that contribute to test results is possible. For instance, in a longitudinal context the effect of the test characteristics could be distinguished from the individual memory performance.

Manuscript received 15 November 2005
Revised manuscript accepted 14 July 2006
First published online day month year

REFERENCES

Asendorpf, J. B. (1989). Individual, differential, and aggregate stability of social competence. In B. H. Schneider, G. Attili, J. Nadel, & R. Weissberg (Eds.), *Social competence in developmental perspective* (pp. 71–86). Dordrecht, NL: Kluwer.

Asendorpf, J. B. (1990). The measurement of individual consistency. *Methodika*, *4*, 1–13.

Barr, R., & Hayne, H. (1996). The effect of event structure on imitation in infancy: Practice makes perfect? *Infant Behavior and Development*, *19*, 253–257.

Bauer, P. J., Heertsgard, L. A., & Wewerka, S. S. (1995). Effects of experience and reminding on long-term recall in infancy: Remembering not to forget. *Journal of Experimental Child Psychology*, *59*, 260–298.

Bauer, P. J., & Mandler, J. M. (1992). Putting the horse before the cart: The use of temporal order in recall of events by one-year-old children. *Developmental Psychology*, *28*, 441–452.

Bornstein, M. H., Slater, A., Brown, E., Roberts, E., & Barrett, J. (1997). Stability of mental development from infancy to later childhood: Three waves of research. In G. Bremner, A. Slater, & G. Butterworth (Eds.), *Infant development. Recent advances* (pp. 191–215). Hove, UK: Psychology Press.

Brian, J. A., Landry, R., Szatmari, P., Niccols, A., & Bryson, S. (2003). Habituation in high-risk infants: Reliability and patterns of responding. *Infant and Child Development*, *12*, 387–394.

Carver, J. C., Bauer, P. J., & Nelson, C. A. (2000). Associations between infant brain activity and recall memory. *Developmental Science*, *3*, 234–246.

Collie, R., & Hayne, H. (1999). Deferred imitation by 6- and 9-month-old infants: More evidence for declarative memory. *Developmental Psychobiology*, *35*, 83–90.

Colombo, J., & Fagan, J. (Eds.). (1990). *Individual differences in infancy: Reliability, stability, prediction*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Colombo, J., Mitchell, D. W., & Horowitz, F. D. (1988). Infant visual attention in the paired-comparison paradigm: Test–retest and attention-performance relations. *Child Development*, *59*, 1198–1210.

Hayne, H., Boniface, J., & Barr, R. (2000). The development of declarative memory in human infants: Age-related changes in deferred imitation. *Behavioral Neuroscience*, *114*, 77–83.

Heimann, M., & Meltzoff, A. N. (1996). Deferred imitation in 9- and 14-month-old infants: A longitudinal study of a Swedish sample. *British Journal of Developmental Psychology*, *14*, 55–64.

Klein, J. P., & Meltzoff, A. N. (1999). Long-term memory, forgetting, and deferred imitation in 12-month-old infants. *Developmental Science*, *2*, 102–113.

Knopf, M., Kraus, U., & Kressley-Mba, R. A. (2006). Relational information processing of novel unrelated actions by infants. *Infant Behavior & Development*, *29*, 44–53.

Kressley, R. A., & Knopf, M. (in press). A comparison of between- and within-subjects imitation designs. *Infant Behavior & Development*.

Kressley-Mba, R. A., Lurg, S., & Knopf, M. (2005). Testing for deferred imitation of 2- and 3-step action sequences with 6-month-olds. *Infant Behavior & Development*, *28*, 82–86.

Mandler, J. M., & McDonough, L. (1995). Long-term recall of event sequences in infancy. *Journal of Experimental Child Psychology*, *59*, 457–474.

McCall, R. B., & Carriger, M. S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development*, *64*, 57–79.

McDonough, L., Mandler, J., McKee, R., & Squire L. (1995). The deferred imitation task as a nonverbal measure of declarative memory. *Proceedings of the National Academy of Science*, *92*, 7580–7584.

Meltzoff, A. N. (1985). Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. *Child Development*, *56*, 62–72.

Meltzoff, A. N. (1988a). Infant imitation after 1-week delay: Long-term memory for novel actions and multiple stimuli. *Developmental Psychology*, *24*, 470–476.

Meltzoff, A. N. (1988b). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development*, *59*, 217–225.

Meltzoff, A. N. (1988c). Imitation of televised models by infants. *Child Development*, *59*, 1221–1229.

Meltzoff, A. N. (1990). Towards a developmental cognitive science: The implications of cross-modal matching and imitation for the development of representation and memory in infancy. In A. Diamond (Ed.), *The development and neural bases of higher cognitive functions* (pp. 1–31). New York: Annals of the New York Academy of Sciences.

Meltzoff, A. N. (1995). What infant memory tells us about infantile amnesia: Long-term recall and deferred imitation. *Journal of Experimental Child Psychology*, *59*, 497–515.

Meltzoff, A. N., & Moore, M. K. (1999). Persons and representation: Why infant imitation is important for theories of human development. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (pp. 9–35). Cambridge, UK: Cambridge University Press.

Nielsen, M., & Dissanayake, C. (2004). Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year. *Infant Behavior & Development*, *27*, 342–365.

Piaget, J. (1975). *Nachahmung, Spiel und Traum*. Stuttgart, Germany: Klett. (Original: Imitation, jeu et rêve – Image et représentation, published 1959).

Rose, S. A., Feldman, J. F., & Wallace, I. F. (1988). Individual differences in infants' information processing: Reliability, stability, and prediction. *Child Development*, *59*, 1177–1197.

Schneider, W., & Sodian, B. (1997). Memory strategy development: Lessons from longitudinal research. *Developmental Review*, *17*, 442–461.

Paper 4

Declarative and non-declarative memory in 12- and 18-month-old infants

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (subm.). Declarative and non-declarative memory in 12- and 18- month-old infants. *Infancy.*

Declarative and non-declarative memory in 12- and 18-month old infants

Submitted 2007/06/08

Word/Pages count:

Abstract: 131 words   Manuscript 35 pages long, 5 Figures

Abstract

Whereas deferred imitations are used to assess declarative memory, non-declarative memory is assessed with operant conditioning tasks (mobile and train task). Except for some studies, however, both memory research programs are still largely separated. Therefore, the present longitudinal study (12 and 18 months old infants; *N* = 51) analyzed the co-development of these memory tasks. Both declarative memory (deferred imitation task) and non-declarative (train task) memory were assessed in a cross-lagged panel design. Auto-, synchronous and cross-lagged correlations were computed. While autocorrelations were rather high (r = .31* for deferred imitation and r = .23 for the train task), synchronous correlations were low at both measurement occasions (r = -.15 for t_1 and r = .06 for t_2). The findings are in line with the dissociation hypothesis of infant memory development.

Keywords: Infant memory, deferred imitation task, train task, declarative memory, non-declarative memory

Introduction

The ontogeny of multiple memory systems is one of the most interesting questions in infants' memory research (e.g. Rovee-Collier, Hayne & Colombo, 2001). In the past couple of years, an increasing number of researchers have argued that memory is not a unitary faculty, but rather is comprised of two different systems that serve different functions, operate according to different principles and are based on different neural systems (Eichenbaum, 2002; Milner, Squire & Kandel, 1998; Tulving & Craik, 2000). Several terms are used in this multiple-memory systems view to describe these independent memory systems. The dichotomy declarative versus non-declarative memory (e.g., Squire, 1994) is used very often in human memory research whereas explicit vs. implicit memory is used in non-human memory research (Schacter & Tulving, 1994). Although not exactly identical in meaning, both share the idea that the declarative (explicit) memory system allows the conscious recollection of past experiences, while the non-declarative (implicit) memory system constitutes the basis of priming processes, the retention of habits and skills, of incremental learning, of classical and operant conditioning. For ease of presentation, the terms declarative and non-declarative memory will be adopted throughout the present article.

From the beginning of memory research in infants and young children (e.g., Piaget & Inhelder, 1974; Piaget, 1962) until the late eighties it was assumed that memory in infants is based on a first, more primitive system, functional after birth, the non-declarative memory system. The declarative memory system was assumed a late-maturing system that became functional at the end of the second year. Observing children in natural settings, Piaget found the first evidence of declarative memory between 18 and 24 months of age, marking the beginning of a representational system. Prior to this development, Piaget postulated, children do not have the cognitive prerequisites to store information. Younger children were presumed to rely and act on information available in the here-and-now. This view has been modified due to recent evidence from systematic and intense studies exploring memory in infants and toddlers using the deferred imitation task (for recent reviews see Bauer, 2002, 2006; Jones & Herbert, 2006) as well as reminder/priming tasks (see Rovee-Collier, 1997, for overview).

In a first research program, based on seminal work by Piaget (1962), Meltzoff (1985) developed an experimental deferred imitation task that has been used in different labs over the past 20 years. In the deferred imitation paradigm, a model demonstrates a series of novel object-related actions to infants. After a retention interval of hours, days or even weeks the infants are allowed to act with the respective objects. Playing behavior is analyzed. It has repeatedly been found that infants in the demonstration group produce significantly more target actions than infants in the control group that never saw the target actions demonstrated, i.e. declarative memory capacity.

Several arguments speak for the claim that declarative memory is assessed in the deferred imitation task, i.e. (1) the cross-modal character of the deferred imitation task makes it improbable that infants´ memory performance is due to priming processes which are sensitive to modality changes, (2) in contrast to incremental learning processes no motor exercise is possible, (3) within deferred imitation tests, new, unknown actions are used as memory items which are not yet available in the infants´ motor repertoire, (4) amnesiacs are unable to show deferred imitation, (5) infants do not only produce single actions but do encode structural elements between actions, i.e. they encode item-relational information (Knopf, Kraus, & Kressley-Mba, 2006).

There is plenty of evidence that 6 month-old infants are able to imitate actions formerly observed after short retention intervals (Barr, Dowden & Hayne, 1996; Barr, Vieira & Rovee-Collier, 2001; Collie & Hayne, 1999; Hayne, Boniface & Barr, 2000; Kressley, Lurg & Knopf, 2005), older infants demonstrate formerly modelled actions after longer retention intervals (e.g., Barr & Hayne, 1996; Bauer, 2006; Hayne et al., 2000). Moreover, older infants acquire novel actions faster than younger ones (Barr, Dowden & Hayne, 1996; Meltzoff, 1995). Additionally, older infants are able to store the observed actions longer than younger infants (Barr & Hayne, 2000).

Whereas most of the research with the deferred imitation paradigm was cross-sectional, several studies assessed memory development longitudinally. These studies demonstrated that the individual stability of deferred imitation is modest in early infancy and increases throughout the second year (Heimann & Meltzoff, 1996; Kolling, Goertz, Frahsek & Knopf, under revision; Nielsen & Dissayanake, 2004). In addition, in a psychometrically oriented short-term longitudinal study (one week-interval) with 12-month-old infants, Goertz, Kolling, Frahsek, Stanisch and Knopf (2007) found a retest-reliability of $r = .52$. After the exclusion of two outliers with very low stability indicators, following the individual consistency approach by Asendorpf (1990), the short-term test-retest reliability even increased to $r = .63$, indicating a good reliability of the deferred imitation task. In a further longitudinal study, Kolling, Goertz, Frahsek and Knopf (under review) found that developmental groups can be extracted which show differential mean growth and stabilities.

To sum up, the present findings using the deferred imitation paradigm indicate that 6-month-olds are capable of deferred imitation, thus, demonstrating declarative memory for object-related action events. Declarative memory in infants can therefore be assessed reliably, memory capacity increases with age, longitudinal stabilities are modest in infancy, and individual differences are found through the second year.

In a second memory research program lead by Rovee-Collier (1997), infants are tested either with the mobile or the train task, which are used as memory task of non-declarative memory of infants. Whereas infants up to 6 months of age are tested with the mobile task, the train task is used with infants between 6 and 18 months of age. Within the mobile conjugate reinforcement procedure, the infant is shown a mobile over her bed that can be moved by leg kicking via a ribbon. In the train task procedure, infants are exposed to a miniature train which moves around a circular track by pressing a lever in front of the infants. In a *baseline phase*, motor activity does not lead to a movement of the mobile or the train, i.e. leg (lever) and mobile (train) are disconnected. Thereby pre-experimental individual activity level is assessed in that the target motor activities (infants' kicking base rate or infants' pressing base rate) are measured. In the *acquisition phase*, kicking (pressing) behavior results in a movement of the mobile (train). During the *immediate retention test* the linkage between ankle (lever) and mobile (train) is interrupted, i.e. the infants´ target behavior does not lead to a movement of the mobile (train). In a *delayed recognition test* infants are shown either the original mobile or a different one. This test is taken as a simple "yes"/"no" recognition test (Rovee-Collier, & Barr, 2001).

With respect to the developmental timetable, infants as young as 8-10 months are able to learn the operant conditioning procedure and remember the mobile for one up to three days. In a wide range of cross-sectional studies, it was demonstrated (e.g., Rovee-Collier et al., 2001; Rovee-Collier & Barr, 2001 for overviews), that (1) there are age-related changes, (2) the retention interval increases linearly with age (Hartshorn, Rovee-Collier, Gerhardstein, Bhatt, Wondoloski, & Klein, 1998), (3) memory is increasingly context-independent (Borovsky & Rovee-Collier, 1990), (4) study time decreases with age, and (5) levels of processing affect memory performance in the mobile procedure (Adler, Gerhardstein & Rovee-Collier, 1998).

In studies with reminder procedures, i.e. reinstatement and reactivation, it was shown that reminders protract infants´ memories tremendously. With a constant reinstatement procedure, Hartshorn (1998), for example, demonstrated that 6 month-old infants exhibit significant retention with 18 months after brief reinstatements at 7, 8, 9 and 12 months. In addition, the critical time window for a reminder to enlarge the retention interval with both reactivation and reinstatement is age-dependent. As almost all studies with operant conditioning procedures were cross-sectional, longitudinal studies with both mobile and train task are very rare.

To sum up, the findings reviewed above of both mobile as well as train task indicate that both operationalizations are adequate for the different age groups, developmental trends are reported and memories are long lasting if constantly reinstated or reactivated, respectively. Furthermore, longitudinal studies are still missing in this research program.

Although the cross-sectional, developmental findings reviewed above are uncontroversial, there is considerable debate about what exactly is assessed by these operant conditioning tasks. There are two lines of experimental approaches to assign the different memory tasks to the respective memory systems. Squire and colleagues apply an *amnesia filter* (Hayne, 2004) as the benchmark test for the assignment of tasks to the specific memory system. By comparing temporal-lobe amnesia patients with healthy controls, tasks are assigned to the prevailing memory system. As human adults cannot be tested with operant conditioning procedures resembling the train task, the amnesia filter cannot be applied to this task. Contrary, Rovee-Collier and colleagues apply a *parameter filter* within their studies. By studying the influence of a wide range of independent variable (age, retention interval, context change, interference, levels of processing and alike) memory tasks are assigned to the prevailing memory systems.

The amnesia filter indicates that whereas long-term retention of the mobile and train task is taken as declarative memory, both acquisition of the operant task and reminder/priming tasks are associated with non-declarative memory. Furthermore, according to Rovee-Collier and colleagues (e.g., Rovee-Collier et al., 2001), the experimental application of a parameter filter shows that both non-declarative as well as declarative memory may even be available before the age of 6 months.

As the study of the development of declarative memory (deferred imitation task), and non-declarative memory (operant conditioning paradigms) are currently realized largely independent from each other, research that assesses simultaneously both memory aspects and their longitudinal co-development is still missing.

Several studies assessing simultaneously declarative memory (puppet imitation task) and non-declarative memory (train task) exist (e.g. Barr, 1996; Barr et al., 2001; Barr et al., 2002). The question in this research, however, was if a primed association between these two tasks would be built up when presented together in a close time window (Barr et al., 2001; Barr et al., 2002).

The major research question in the present study was the charting of the co-development of declarative and non-declarative memory and the investigation of longitudinal relationships. In terms of the dissociation assumption brought forward by Rovee-Collier, it was expected that both memory systems are dissociated very early in development; hence low correlations between memory tasks are expected. To our knowledge, such an experimental design assessing the dissociation of the two memory systems, declarative and non-declarative memory, is still lacking for infants.

Therefore, in the present study, both a deferred imitation task and a train task were assessed both cross-sectionally and longitudinally in the age range between 12 and 18 months with a cross-lagged panel design.

Method

Participants

The total sample consisted of fifty one (27 male) infants who were recruited via radio announcement, advertisements in child care centres and local paediatricians and by word of mouth. Criteria for admission into the study were no known physical, sensory or mental handicaps, normal length of gestation (over 37 weeks) and normal birth weight (2300 – 4400 grams). The mean age at the first measurement occasion was M = 361.8 days (SD = 8.4) and at the second measurement occasion M = 550.7 days (SD = 7.5). Mean birth weight was M = 3373 grams (SD = 490) and parents reported an *APGAR index* of M = 9.79/9.91 with a minimum value of 7. Furthermore, the Developmental Test of Petermann and Stein (2000) which aims to assess several developmental factors in the age range between 6 month and 6 years (e.g., motor development, language development, and cognitive development) showed that all children of the sample were within normal range in all developmental dimensions.

The second assessment of declarative memory, using a second deferred imitation test, as well as non-declarative memory with the train task was administered half a year later. Out of the sample assessed on t1 n = 28 subjects (55 percent) were excluded, because (1) they pressed the lever of the train task rather seldom, namely less than 10 times over the total baseline, acquisition and retention phase (n = 9), (2) were too agitated during testing (n = 6), (3) were uninterested in the task and testing procedure (n

= 3), or (4) cried excessively (*n* = 9). No infant was excluded due to problems with the deferred imitation task.

A drop-out control comparison analysis showed that (1) there were no significant differences in the different developmental dimensions assessed with the Developmental Test of Petermann and Stein (2000), and (2) no sex differences with respect to drop-out were found.

Testing environment

All testing took place in a baby lab that was unfurnished expect for the experimental apparatus. During the test, the infant was seated on his or her parent's lap at a small rectangular table just opposite to the experimenter. Behind and to the left of the experimenter was a video camera that was focused on the child's head and torso and most of the tabletop. A second camera behind the infant on the right recorded the experimenter. The recording apparatus was housed in an adjacent viewing room to prevent auditory distractions. The experiment was electronically timed by a computer that mixed elapsed time in 0.10 sec increments directly onto the videotaped records.

Materials and Apparatus

Deferred imitation test for 12-month-olds (DI-12). The deferred imitation test for 12 months olds (DI-12) consisted of 5 actions, of which two were two-step actions and the three were one-step actions. The objects were commercially available toys and were specially adapted for the experiments. Some of the objects and target actions were used in previous work (Knopf, Kraus & Kressley-Mba, 2006; Kressley & Knopf, 2006; Goertz, Knopf, Kolling, Frahsek & Kressley–Mba, 2006; Goertz et al., in press) Baseline behavior for the test items was assessed with *N* = 24 infants[1]. The actions were shown successively to the child three times each within a 30 seconds interval.

Following the delay, infants acted with the objects successively for 30 seconds each. The retention delay was 30 minutes. The maximal sum score of the DI-12 is 7. Mean item performance analysis for a sample of *N* = 92 infants showed that the items were uniformly distributed. Item difficulties ranged from .21 to .78 (mean item difficulty = .48). The item-total correlations ranged from .43 to .71 (mean item-total correlations = .56). The three scorers of the DI-12 reached an inter-rater reliability of *r* =. 89.0 - 92.7 % and a Cohen's kappa of κ = .78 – .83. Discrepancies in scoring were resolved by consensus to 100% agreement. A short term (2 weeks) test-retest reliability study of the DI-12 showed a reliability correlation of *r* = .52 (Goertz et al., in press).

Deferred imitation test for 18- month-olds (DI-18). At the second measurement occasion, the deferred imitation test for 18-month-olds (DI-18) consisted of 6 actions (one three-step action, four two-step actions and one one-step action). All the objects were commercially available and specially adapted for the experiments. Objects and target actions were used in previous work (Kolling et al., in press; Goertz, Kolling, Frahsek & Knopf, under review). The actions were shown successively to the child three times each within a given time interval (30 to 45 seconds). The retention delay was 30 minutes. Following the delay, infants acted with the objects successively for 30 to 45 seconds each. Among the new items, the two most difficult objects (*mouse* and *drum*) for the 12-month-olds were taken over for the 18-month-olds test. Baseline behavior for the test items for the 18-month-olds was assessed with *N* = 26 infants[2]. The maximal sum score of the DI-18 is 12. Mean item performance analysis for a sample of *N* = 88 infants showed that the items were uniformly distributed. Item difficulties ranged from .33 to .96 (mean item difficulty = .62). The item-total correlations ranged from .18 to .67 (mean item-total correlations = .45). The four scorers of the DI-18 reached an inter-rater reliability of *r* =. 95.7- 96.4 % and a Cohen's kappa of κ = .91 – .92 [3]. Discrepancies in scoring were resolved by consensus to 100% agreement. A short term test-retest reliability study of the DI-18 is currently in progress.

Data scoring. All sessions were videotaped by two cameras, one taping the infant and the second the experimenter. Two naïve and independent observers scored target action completion using operational definitions. The scorers were uninformed about the hypotheses of the study and how exactly the experimenter demonstrated the target actions. Target actions were scored as yes- or no-responses.

Train task. The train set used consisted of a wooden board (110x 100 cm) on which a circulating miniature train (HO-scale) was placed. The miniature train consisted of an engine and two rail cars (blue-coloured). Placed around the track were several houses and trees to make the surroundings interesting enough for the infants but to keep the infants´ mind on the train. The train task was continuously illuminated by a light (60 W) above the wooden board. At the beginning of each experimental procedure the train was placed directly in front of the front window of the circular track (88 cm diameter; see Figure 1).

Insert Figure 1 about here

During the test, the infant was seated on his or her parent's lap directly in front of a wooden lever directly in front of the front window. The wooden lever (26 x 19 x 9 cm) activated a micro switch when pressed. The micro switch was connected to an interface box and an IBM computer. The IBM computer used a visual basic program which recorded lever presses (10 second bins) and activated the train at reinforcement periods. Furthermore, the train task session was video-recorded with one video camera that was focused on the infant and the wooden lever.

During reinforcement periods, each lever press activated the train for 3 s. Lever presses that occurred when the train was in motion were recorded on the computer but did not lead to another motion of the train. During the non-reinforcement periods (baseline as well as retention test), the lever was deactivated, so that each lever press was recorded in the control file but did not lead to a motion of the train.

Retention measure. In the experimental studies by Rovee-Collier, three response rates were assessed: the baseline response rate (B), the response rate immediately after acquisition (I) and the response rate at a long-term retention test (L). Two indicators to assess non-declarative learning and memory, i.e. the *long-term retention ratio* and the *baseline retention ratio* were used by Rovee-Collier and her coworkers, typically. The *long-term retention ratio* is defined by the ratio of the lever presses in the long-term retention test and the immediate retention test (LRT = L/I). If LRT is not significantly different from 1 then significant retention is assumed. Retention decreases linearly with the LRT ratio with no retention assumed if LRT is significantly lower than one. The *baseline retention ratio* is defined by the ratio of the lever presses in the baseline test and the long-term retention test (RR = L/B). There exist two different interpretations of RR. The first interpretation states that if RR > 1.0 then some retention exists. The other interpretation (Rovee-Collier & Shyi, 1992) states that if RR => 1.5 then a conservative retention criterion is reached, such as the infant is at least responding during the test at the rate required to meet the learning criterion. The learning criterion is defined as I/B => 1.5. In her studies, Rovee-Collier included only infants in their final samples if the learning criterion I/B > 1.5 was reached. This means that she systematically excluded all infants from future analyses, which could not learn the operant procedure properly. In the present study, we did not analyze the priming aspect of non-declarative memory and the retention interval but primarily the operant conditioning aspect of non-declarative memory. Therefore, we intended to use an *operant learning index (OL = I/B)* as our primary dependent variable.

Design

Figure 2 shows the statistical design in line with the operant conditioning procedure used in the present study.

Insert Figure 2 about here

The design is a cross-lagged panel design with two autocorrelations, two synchronous correlations and two cross-lagged correlations. The cross-lagged correlations are only cautiously interpretable in the present design for several reasons, i.e., (1) the analysis of cross-lagged correlation differences presupposes a large sample size, and (2) ideally continuous variables. Therefore, an analysis of the cross-lagged correlations in terms of the direction of causation is not the aim of the present study. The main focus lies on the analysis and interpretation of the auto- and synchronous correlations. In the present design, the two autocorrelations reflect the longitudinal stability of the measures analyzed. The synchronous correlations reflect the dissociation hypothesis.

Procedure

Measurement occasion 1. Appointments were made for the time of day when the parents reported that their children were most alert, usually midmorning. In a waiting room, an interview prior to the experiment was conducted to explain the purpose of the study, details of the procedure, and to obtain informed consent. In the lab, the child was seated on the parent's lap. The first testing day started with the demonstration phase of the deferred imitation task (DI-12). In the 30 minute retention interval, the Developmental Test of Petermann and Stein (2000) was administered. After the 30 minutes retention interval, the retention phase of the deferred imitation task took place. At the second testing day, most often the next day, the train task (1 min. baseline phase, 6 min. acquisition phase, 1 min. retention phase) was conducted.

Measurement occasion 2. At the second measurement occasion, the following testing was realized. After a short warm-up (1) the demonstration of deferred imitation (DI-18) was realized, (2) after a break of approx. 10 min., the (3) train task was implemented (1 min. baseline phase, 4 min. acquisition phase, 1 min. retention phase), followed by an (4) approx. 14 min. long break which was succeeded by the (5) retention phase of deferred imitation. Finally (6) the Developmental Test of Petermann and Stein (2000) was administered.

Results

Analysis of operant learning index. The analysis of the primary dependent variable, the operant learning index (OL = I/B) yielded that infants at time 1 showed a mean lever pressing rate in the baseline phase of *M* = 18.24 (*SD* = 12.54) and a mean lever pressing rate in the acquisition phase of *M* = 12.51 (*SD* = 8.76). In the retention phase the mean lever pressing rate was *M* = 12.9 (*SD* = 9.8). As can be seen from the descriptive statistics, the calculation of the operant learning index was not a useful index in the present study because almost all of the infants had a very high spontaneous baseline lever pressing rate. At time 2, mean lever pressing in the baseline phase was *M* = 11.29 (*SD* = 10.75) and a mean lever pressing rate in the acquisition phase of *M* = 9.1 (*SD* = 9.7). In the retention phase the mean lever pressing rate was *M* = 11.77 (*SD* = 12.6).

As at time 1, baseline behavior was too high for the application of the operant learning index for the data of the study. Therefore, as the results of the operant learning index for time 1 and time 2 did not lead to a classification of successful learners (denoted TT+ in the following) and non-successful learners (denoted TT- in the following) a second statistical measure was used. As it was expected that successful learners (TT+) would respond with high lever pressing in the acquisition phase and non-successful learners (TT-) would show low lever pressing, the mean sum of lever pressing across the acquisition phase was z-transformed. This procedure lead to a group of high lever pressing infants and a group of low lever pressing infants for both measurement occasions. This analysis showed that at time 1 (12 months), 39 percent (*n* = 19) infants of the sample acquired the lever press (stimulus) – train activation (response) relation and 61 percent (*n* = 30) were not able to acquire the relationship. At time 2 (18 months) of the remaining *n* = 23 infants 36 percent (*n* = 8) infants acquired the relationship between lever and train activation, whereas 64 percent (*n* = 14) did not acquire the relationship. However, as we were aware of the fact, that even a high lever pressing rate of the infants might not obligatory indicate a successful operant learning, we developed a *qualitative analysis of the video tapes* to obtain a second criterion for learning.

Video analysis train task. The *qualitative analysis of the video tapes* took into account important behaviors of the infants, i.e. (1) goal-oriented (for example, clear signs of voluntarily pressing the lever) vs. accidental lever pressing (for example, accidental lever pressing with the elbow), (2) train-oriented gazing while lever pressing or (3) astonishment of train movement following lever pressing. Furthermore, the time at which infants showed signs of understanding the stimulus-response relationship was assessed. The analysis included all three phases of the train task procedure. The two scorers of the TT-12 task reached an inter-rater reliability of $\kappa = .57$ (7 out of 50 misclassifications). The two scorers of the TT-18 task reached an inter-rater reliability of $\kappa = 1.0$ (no misclassifications). Discrepancies in scoring were resolved by consensus to 100% agreement. The comparison of the inter-rater reliabilities indicates that the rating for the 12-month-olds was more difficult than the rating for the 18-month-olds.

The qualitative analysis of the train task procedure showed that at time 1 (12 months), 59 percent ($n = 30$) infants in the sample acquired the lever press (stimulus) – train activation (response) relation and 41 percent ($n = 21$) were not able to acquire the relationship, following these acquisition criteria. At time 2 (18 months) 35 percent ($n = 8$) of the remaining $n = 23$ infants acquired the relationship between lever and train activation, whereas 65 percent ($n = 15$) did not. Figure 3 depicts the mean lever pressing rates of both t_1 and t_2 for TT+ and TT- (according to the qualitative analysis) and the total group.

Insert Figure 3 about here

Figure 3 demonstrates that at time 1 the TT+ group shows a higher lever pressing rate at baseline testing than the total group and the TT- group. Whereas the TT+ groups´ lever pressing decreases within the baseline phase and increases immediately after the acquisition phase started, the TT – group does not increase their lever pressing in the acquisition phase. At time 2, the TT+ group starts with high lever pressing and remains constantly high through the acquisition phase, whereas the TT- group demonstrates a low lever pressing behavior through both baseline and acquisition. This analysis of the lever pressing rates of the TT+ vs. TT- group indicates that the calculation of the operant learning index did not yield useful results for a grouping of infants into TT+ vs. TT-. Therefore, in the following statistical analyzes, the grouping of the infants into TT+ (learners) vs. TT- (non-learners) according to the qualitative analysis at both time 1 and time 2 is taken as the *dependent variable* in further analysis.

Comparison of quantitative and qualitative analysis. A comparison of the percentage of TT+ vs. TT- at time 1 calculated with a z-transformed mean lever pressing rate in the acquisition phase and the qualitative analysis of the video tapes showed that there is no perfect consistency between the statistical measure and the qualitative analysis. 11 infants were classified as TT- according to z-transformed mean lever pressing rate but classified as learners according to qualitative analysis. An analysis of the time when infants made signs of acquiring the stimulus-response relationships showed that those 11 misclassified infants learned the train task later in the acquisition phase than the remaining n = 19. Whereas the n = 19 infants acquired the relationship after approx. M = 197 (SD = 95) seconds, the late learning infants learned the relationship after approx. M = 267 (SD = 63) seconds. An unpaired two sample t-test showed a significant difference for time ($t = 2.1$, $df = 31$, $p < 0.05$). Therefore the analysis of mean lever presses misclassified those infants because they had not enough time left in the 6 minute acquisition time frame to increase their lever pressing.

A comparison of the percentage of TT+ vs. TT- at time 2 showed that only n = 3 did not match the mean lever pressing rate analysis and the qualitative analysis. Two of those three infants (classified as TT+ according to lever pressing and as TT- according to qualitative analysis) were so agitated that they hit the lever or beat against the front window and thereby pressed the lever accidentally with their forearms or elbows. They therefore had high lever pressing rates but showed no signs of learning the lever-train-relationship. The other infant (classified as TT- according to lever pressing and as TT+ according to qualitative analysis) lost interest in the train task very quickly but showed a very high goal-oriented lever pressing in the first acquisition minute.

Deferred imitation performance

Mean Memory Performance (N=52). The analysis of the mean number of target actions performed by the 12-month and 18-month-olds showed that 12-month-olds imitated M = 3.96 (SD = 1.44) target actions, while the 18-month-old infants imitated M = 6.8 (SD = 1.99) target actions. A paired samples t-test revealed a significant difference between the two measurement occasions ($t = 9.6$, $df = 48$, $p < .01$).

Interrelation between declarative and non-declarative memory

Mean performance analysis. In a mean performance analysis of the sum of imitated actions the TT + groups were contrasted with the TT – groups in the two measurement occasions. Figure 4 depicts the mean number of target actions for the 12-olds and 18-month-olds dependent on the TT+ and TT- groups.

Insert Figure 4 about here

Figure 4 shows that at the first measurement occasion there was no significant difference between the two groups (TT+ vs. TT-) according to deferred imitation performance for both time 1 (DI-12) and time 2 (DI-18). At the second measurement occasion, a comparison of TT+ vs. TT- showed that non-learners imitated significantly more actions at time 1 (DI-12) (t = 2.31, df = 21, p < .05) but not at time 2 (DI-18).

Cross-lagged panel correlations. The correlations of the cross-lagged panel design are given in Figure 5.

Insert Figure 5 about here

The autocorrelation for the deferred imitation performance was $r = .31$ ($p < .05$) and the autocorrelation for the train task was $r = .23$, demonstrating a rather high relationship between same memory tasks over time. Whereas the synchronous correlation at time 1 was $r = -.15$, the synchronous correlation at time 2 was $r = .06$, showing that the correlation between different memory tasks is close to zero. While the cross-lagged correlation between DI-12 and TT-18 is $r = .02$, the cross-lagged correlation between DI-18 and TT-12 is $r = -.45$ ($p < .05$). The analysis of stationarity ($Z = -.75$, $p > .25$) for the synchronous correlations revealed that the cross-sectional correlations between acquisition of the train task and deferred imitation performance were not significantly different across measurement occasions. As only one variable in the cross-lagged panel design was tested for its short-term-reliability (DI-12), an analysis corrected for unreliability of the measurement (analysis of quasi-stationarity) was not considered useful in the present analysis.

Discussion

In the present study, both declarative and non-declarative memory was assessed simultaneously and repeatedly in 12- and 18-month-old infants. The operant conditioning (train task) and declarative memory performance (deferred imitation) was used to assess both non-declarative and declarative memory. The main objectives of the study, however, were to (1) chart the co-development of declarative and non-declarative memory, and (2) to test the hypothesis that declarative and non-declarative memory are dissociated very early in development.

The test parameters of the two tests DI-12 (five target actions) and DI-18 (six target actions) indicate the adequacy of the declarative memory procedure applied in the study. The test difficulty of both tests is adequate, the inter-rater reliabilities are high, and the tests are sufficiently motivating, i.e. all children successfully passed the deferred imitation tests twice. This means that a suitable procedure was applied to follow early declarative memory development.

The analysis of mean deferred imitation memory performance showed a significant increase between 12 and 18 months, thus replicating findings of other research groups (Heimann & Meltzoff, 1996; Nielsen & Dissayanake, 2004). As the relationship between the two declarative memory assessments (autocorrelation of DI tests) of 12- and 18-month-olds is r =. 31 (p < .05) and the short-term test-retest stability is close to r = .60 even in 12 month-olds, the developmental dynamics on the individual level are high in the first half of the second year. This high developmental dynamic of the total sample is also in line with earlier longitudinal studies (Heimann & Meltzoff, 1996; Nielsen & Dissayanake, 2004). However, fine-grained person-centred analysis (Heimann & Meltzoff, 1996; Kolling et al., in press) showed that if groups of infants with differential growth are separated, stabilities are higher.

The autocorrelation for the operant learning procedure (train task) again reflects a high amount of developmental dynamics for this memory task. The autocorrelation between the first and second measurement occasion is r = .23, which is high but not significant. As there are no other studies, which assessed the operant learning aspect of the train task within a longitudinal design with the age group used, comparison data are not obtainable. Further research is needed to analyze the short- and long-term stability of operant learning capacity (non-declarative memory) of infants. In future research, it would be supportive for longitudinal studies to analyze both short-term stability, i.e., reliability and long-term stability in 12- and 18-month-old infants.

The analysis of the synchronous correlations of the cross-lagged panel design reflected the dissociation hypothesis. A low correlation between the two different memory aspects assessed in 12- and 18-month-olds was expected. The analysis of the synchronous correlations yielded that whereas the operant learning procedure (TT) and the declarative memory assessment (DI) are weakly negatively correlated at time 1 (r = -.15), the correlation further decreases to r = .06 at time 2. The analysis of stationarity yielded that the cross-lagged correlations do not differ between measurement occasions reflecting the fact that there are no significant differences between the synchronous correlations of time 1 and time 2. The comparison of the mean performance in the declarative memory task yielded that at time 1 neither the TT+ group nor the TT- group differed in their deferred imitation performance for both time 1 (DI-12) and time 2 (DI-18). At time 2, non-learners (TT-) imitated significantly more actions of the DI-12 assessment (t = 2.31, df = 21, $p < .05$).

From a multiple memory systems view, the results are therefore in accordance with the assumption that non-declarative (implicit) and declarative (explicit) (Rovee-Collier et al., 2001) memory dissociate very early in life. Furthermore, as the correlations between declarative and non-declarative memory do not differ between measurement occasions, both memory systems seem to mature at the same rate. The present study therefore presents further evidence that both memory systems develop simultaneously with 12- and 18-month-old infants.

As the present study is the first one to analyze the longitudinal development of operant learning and declarative memory performance simultaneously within a cross-lagged panel longitudinal design further replications are doubtlessly needed. Besides replication purposes, future (longitudinal) studies could combine the study of priming oriented aspects (reinstatement and reactivation) and operant learning aspects of non-declarative memory with declarative memory performance (deferred imitation) in the first half of the second year to provide further evidence for an early dissociation of non-declarative and declarative memory. Furthermore, individual differences in deferred imitation (Jones & Herbert, 2006; Kolling et al., in press) should be taken more and more into account.

Overall, future research needs to (1) combine non-declarative memory indicators (operant learning; priming) and declarative memory indicators simultaneously in clever experimental designs, (2) take into account individual differences both within the train task procedure and within deferred imitation analysis, and (3) study the short-term vs. long-term stability of the train task assessment.

References

Adler, S. A., Gerhardstein, P., & Rovee-Collier, C. (1998). Levels-of-processing effects on infant memory? *Child Development, 69*, 280-294.

Asendorpf, J. (1990). The measurement of individual consistency. *Methodika, 4*, 1-13.

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior & Development*, *19*, 159-170.

Barr, R., & Hayne, H. (1996). The effect of event structure on imitation in infancy: practice makes perfect? *Infant Behavior & Development*, *19*, 253-257.

Barr, R. & Hayne, H. (2000). Age-related changes in imitation: Implications for memory development. In C. Rovee-Collier, L. P. Lipsitt & H. Hayne (Eds.), *Progress in infancy research* (Vol. 1, pp. 21-67). Hillsdale, NJ: Laurence Erlbaum Associates.

Barr, R., Vieira, A., & Rovee-Collier, C. (2001). Mediated imitation at 6 months of age: remembering by association. *Journal of Experimental Psychology, 79*, 229-252.

Barr, R., Vieira, A., & Rovee-Collier, C. (2002). Bidirectional priming in infants. *Memory & Cognition, 30,* 246-255.

Bauer, P. J. (2002). Early memory development. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 126 - 146). Malden, MA: Blackwell Publishers.

Bauer, P. J. (2006). Event memory. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol 2. Cognition, perception, and language* (6th ed., pp. 373-425). New York: Wiley.

Borovsky, D., & Rovee-Collier, C. (1990). Contextual constraints on memory retrieval at 6 months. *Child Development, 61*, 1569–1583.

Collie, R. & Hayne, H. (1999). Deferred imitation by 6- and 9-month-old infants: More evidence for declarative memory. *Developmental Psychobiology*, *35,* 83-90.

Eichenbaum, H. (2002). *The cognitive neuroscience of memory*. New York: Oxford University Press.

Goertz, C., Knopf, M., Kolling, T., Frahsek, S., & Kressley-Mba, R. A. (2006). Entwicklung und Erprobung eines Messinstruments zur Erfassung des deklarativen Gedächtnisses Einjähriger: Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12). [The Frankfurt Imitation Test for 12-Month-Old Infants: Testing a series of items for measuring declarative memory in 1-year-olds.]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38,* 88-96.

Goertz, C., Kolling, T., Frahsek, S., Stanisch, A., & Knopf, M. (2007). Assessing declarative memory in 12-month-old infants: A test-retest reliability study based on the deferred imitation task. *European Journal of Developmental Psychology*, in press. Epub ahead of print retrieved Mai, 28, 2007, from http://www.informaworld.com/smpp/content~content=a778058220~jumptype=rss

Goertz, C., Kolling, T., Frahsek, S., & Knopf, M. (subm.). Der Frankfurter Imitations-Test für 18 Monate und 24 Monate alte Kinder (FIT18-24). [The Frankfurt Imitation Test for 18-Month-Old and 24-month-Old Infants: Testing a series of items for measuring declarative memory in 1,5-2-year-olds.]. *Manuscript submitted for publication.*

Hartshorn, K., Rovee-Collier, C., Gerhardstein, P., Bhatt, R. S., Wondoloski, T. L., & Klein, P. (1998). The ontogeny of long-term memory over the first year-and-a-half of life. *Developmental Psychobiology, 32*, 69-89.

Hayne, H., Boniface, J., & Barr, R. (2000). The development of declarative memory in human infants: Age-related changes in deferred imitation. *Behavioral Neuroscience*, *114* , 77-83.

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review, 24*, 33-73.

Heimann, M., & Meltzoff, A. N. (1996). Deferred imitation in 9- and 14-month-old infants: A longitudinal study of a Swedish sample. *British Journal of Developmental Psychology*, *14*, 55-64.

Hsu, V. C., & Rovee-Collier, C. (2006). Memory reactivation in the second year of life. *Infant Behavior & Development*, *29*, 91-107.

Jones, E. J. H., & Herbert, J. S. (2006). Exploring memory in infancy: Deferred imitation and the development of declarative memory. *Infant and Child Development, 15*, 195-205.

Knopf, M., Kraus, U. & Kressley-Mba, R. (2006). Relational information processing of novel unrelated actions by infants. *Infant Behavior & Development*, *29*, 44-53.

Kolling, T., Frahsek, S., Goertz, C. & Knopf, M. (in press). Stability of deferred imitation in 12- to 18-month-old infants: a closer look into developmental dynamics. *European Journal of Developmental Psychology.*

Kressley, R. A. & Knopf, M. (2006). A comparison of between-and within-subjects imitation designs. *Infant Behavior & Development*, *29*, 564-573.

Kressley-Mba, R. A., Lurg, S. & Knopf, M. (2005). Testing for deferred imitation of 2- and 3-step action sequences with 6-month-olds. *Infant Behavior & Development*, *28*, 82-86.

Meltzoff, A. N. (1985). Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. *Child Development*, *56*, 62-72.

Meltzoff, A. N. (1995). What infant memory tells us about infantile amnesia: long-term recall and deferred imitation. *Journal of Experimental Child Psychology, 59,* 497-515.

Milner, S., Squire, L. R., & Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. *Neuron*, *20*, 445-468.

Nielsen, M., & Dissanayake, C. (2004). Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year. *Infant Behavior & Development*, *26*, 213-226.

Petermann, F., & Stein, I. A. (2000). *Entwicklungstest ET 6-6. [Developmental test for 6-month to 6-year-olds.].* Lisse, Netherlands: Swets & Zeitlinger Publishers.

Piaget, J. (1962). *Play, dreams and imitation in childhood.* New York: Norton. (Original work published 1945)

Piaget, J. & Inhelder, B. (1974). *Gedächtnis und Intelligenz (Memory and intelligence)*. Olten, Switzerland: Walter Verlag (Original: Mémoire et intelligence, 1968).

Rovee-Collier, C. (1997). Dissociations in infant memory: Rethinking the development of implicit and explicit memory. *Psychological Review, 104*, 467-498.

Rovee-Collier, C., & Barr, R. (2001). Infant learning and memory. In G. Bremner & A. Fogel (Eds.), *Blackwell handbook of infant development* (pp. 139 – 168). Malden, MA: Blackwell Publication Company.

Rovee-Collier, C., Hayne, H., & Colombo, M. (2001). *The development of implicit and explicit memory.* Amsterdam: John Benjamin's Publication Company.

Schacter, D. L. & Tulving, E. (1994). What are the memory systems of 1994? In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (p. 1-38). Cambridge, MA: MIT Press.

Squire, L. R. (1994). Declarative and non-declarative memory: Multiple brain systems supporting learning and memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 203-232). Cambridge, MA: MIT Press.

Tulving, E. & Craik, F. I. M. (Eds.), (2000). *The Oxford handbook of memory*. Oxford: Oxford University Press.

Footnotes

1. The target action *tin can* had a baseline performance rate of 20.8% the target action *toy pig with hat* of 8.3 %, the target action *cup and knife* of 29.2 %, and the target action *drum* of 20.8 % and the target action *mouse* of 4.2 %.

2. The target actions *car, goose, frog, octopus* had a baseline performance rate of 0%. The target action *mouse* had a baseline performance rate of 19.2 percent and the target action *drum* of 7.7 %.

Figure Captions

Figure 1

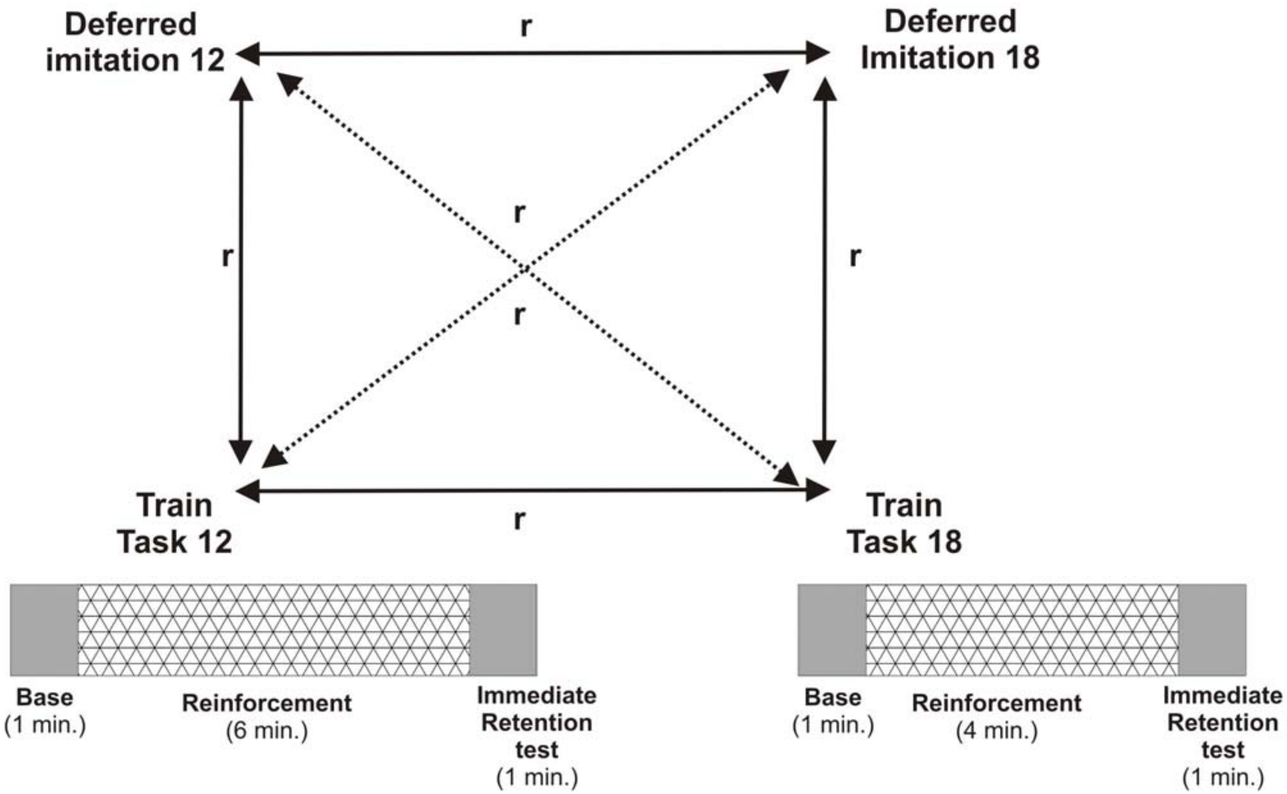

Deferred
imitation 12
r
Deferred
Imitation 18
r
r
r
r
Train
Task 12
r
Train
Task 18
Base
(1 min.)
Reinforcement
(6 min.)
Immediate
Retention
test
(1 min.)
Base
(1 min.)
Reinforcement
(4 min.)
Immediate
Retention
test
(1 min.)

Figure 2

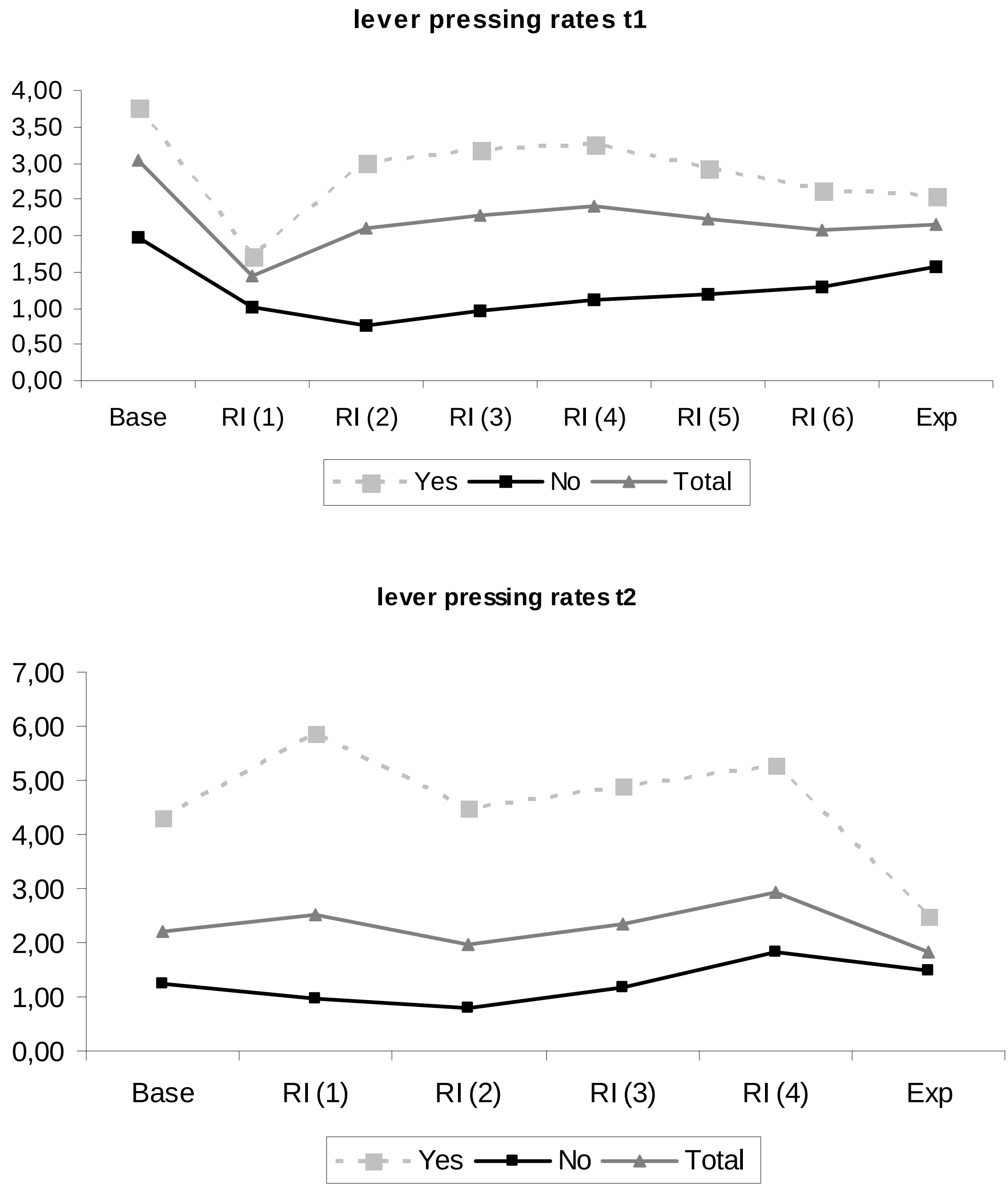


Figure 3

Note: Base denotes the baseline (non-reinforcement) phase, *RI* (x_i) denotes the learning (reinforcement) phase with x_i the i_{th} minute within this phase and *Exp* denotes the experimental (non-reinforcement) phase.

	Time 1			Time 2		
	TT +	**TT -**	**p**	**TT +**	**TT -**	**p**
DI-12	3.8 (1.60)	4.1 (1.34)	ns	3.4 (1.41)	4.5 (0.99)	*
DI-18	6.9 (1.9)	6.7 (2.15)	ns	7.1 (2.0)	6.93 (1.58)	ns

Figure 4

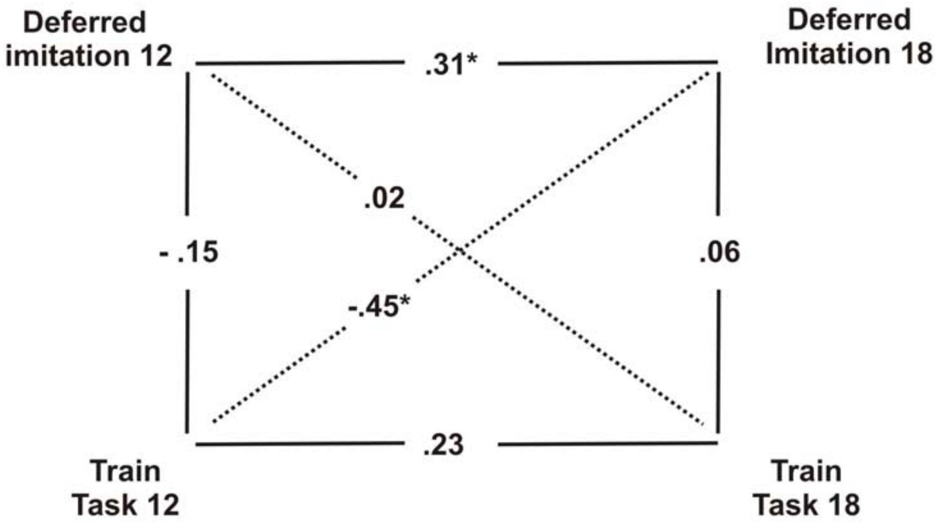

Deferred
imitation 12
.31*
Deferred
Imitation 18
.02
- .15
.06
-.45*
.23
Train
Task 12
Train
Task 18

Figure 5

Paper 5

Stability of deferred imitation in 12- to 18-month-old infants:
A closer look into developmental dynamics

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (in press). Stability of deferred imitation in 12- to 18-month-old infants: A closer look into developmental dynamics. *European Journal of Developmental Psychology*.

Stability of Deferred Imitation in 12- to 18-month-old Infants: A Closer Look into Developmental Dynamics

Thorsten Kolling, Claudia Goertz, Stefanie Frahsek and Monika Knopf

Institute of Psychology, University of Frankfurt/Main

Correspondence address:

Thorsten Kolling
Johann Wolfgang Goethe–University
Department of Developmental Psychology
P.O. Box 11 19 32
D-60054 Frankfurt am Main, Germany
E-mail: T.Kolling@psych.uni-frankfurt.de

Keywords: Deferred imitation, infant memory, stability of early declarative memory, variable-centred approach, person-centred approach

Acknowledgements

This study was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft) to Monika Knopf (KN 275/3-1). We would like to thank Regina Kressley-Mba for her help during the preparation of the study and we gratefully acknowledge the assistance of Christina Mack, Veit Kubik, Micha Best, Ali Koc, Margit Feyler and Eva Riedmueller for scoring the video tapes and helping with data collection. Special thanks are due to all the babies and their families who participated in the study.

Abstract

Deferred imitation is used to assess declarative memory in infants. Although a lot of studies show that infants from 6 months onwards are able to re-enact actions following a delay, only a few studies describe early declarative memory longitudinally. From a variable-centred approach, these studies found a modest relationship between measurement occasions that increases with age. However, no studies have analysed the differential aspect of memory development from a person-centred perspective. The present study analyses memory stabilities both from a variable-centred and person-centred approach within a sample of N=87 infants of a 2-wave longitudinal design (12- and 18-month-olds). From a variable-centred perspective, the results indicate that firstly, there was a significant increase in infants´ memory performance and secondly, that although reliable, the stability of infants´ memory performance was relatively low. From a person-centred perspective two vs. three groups of infants were differentiated showing different developmental growth trajectories and stability correlations but no differences in language, cognitive and motor development. The implications of those results in terms of further test development are discussed. Furthermore, important methodological expansions for the analysis of infants´ memory data are presented and discussed.

Introduction

The deferred imitation task is one of the most often used procedures to assess declarative memory in infants. It was observed in a natural context by Piaget (1962) and later standardized by Meltzoff (1985, 1988a, b, c) and is used today in many developmental studies. In this task, short series of simple, object-based actions are successively shown to young infants who observe modelled actions and are generally not allowed to act with the action-related props. After a delay of minutes, hours or even days the children are given the props and their target behaviour is observed. A substantial increase in infants´ target behaviour in the experimental in comparison to the control group indicates evidence of declarative memory.

Several arguments can be made in favour of the assumption that deferred imitation in fact measures declarative memory, which can be defined as a memory system whose encoding requires conscious awareness. In the deferred imitation task infants reproduce instrumental actions, which they have only observed and which they do not incrementally acquire by performing them motorically. Therefore, it is assumed that the deferred imitation task assesses the ability to acquire, voluntarily retrieve, or recall information. McDonough, Mandler, McKee and Squire (1995) brought forward some additional arguments for this assumption. They demonstrated that adults suffering from temporal lobe amnesia, where the declarative memory system as opposed to non-declarative memory is disrupted, are unable to show deferred imitation. Electrophysiological studies have since then yielded further supporting evidence (Carver, Bauer & Nelson, 2000). Given these different considerations, the deferred imitation task has become an important instrument to assess declarative memory in infants in recent years.

From a developmental psychological viewpoint, up to now, most studies within the deferred imitation paradigm were cross-sectional with age as a factor in the experimental design. In several studies, age (Barr, Dowden & Hayne, 1996; Collie & Hayne, 1999; Hayne, MacDonald & Barr, 1997; Herbert & Hayne, 2000a, b) and retention intervals (Bauer, Heertsgard

& Wewerka, 1995; Hayne et al., 2000; Herbert & Hayne, 2000a,b; Mandler & McDonough, 1995; Meltzoff & Moore, 1999) were varied to investigate the onset of deferred imitation ability and its development in terms of length of retention and absolute performance (number of target actions recalled).

Those studies demonstrated that the earliest manifestation of declarative memory in an imitation task can be found at the age of 6 months (Collie & Hayne, 1999; Hayne, Boniface & Barr, 2000; Kressley-Mba, Lurg & Knopf, 2005), though, to a certain degree, these studies are still controversial. By 9 months of age, infants are reliably capable of imitating actions following a delay across a variety of imitation designs (Collie & Hayne, 1999; Heimann & Meltzoff, 1996; Meltzoff & Moore, 2001). Furthermore, there are age-related changes in deferred imitation parameters, namely that with increasing age infants need less exposure to the target actions; hence they learn faster than younger ones. Furthermore, older infants imitate actions after longer retention intervals (Barr et al., 1996; Barr & Hayne, 1996; Barr & Hayne, 2000; Bauer et al., 1995; Hayne et al., 1997; Hayne et al., 2000; Herbert & Hayne, 2000a, b; Mandler & McDonough, 1995; Meltzoff & Moore, 1999). Moreover, Knopf, Kraus and Kressley-Mba (2006) provided evidence that infants as young as 10 and 11 months encode item-relational information in a deferred imitation task, a finding that parallels results in adult memory studies with item lists. This finding can be taken as a further argument for the assumption that the deferred imitation task in fact measures declarative memory.

Since the deferred imitation task has become an established research instrument to assess infants´ declarative memory, a more developmental stance has emerged in recent years (e.g., Heimann & Meltzoff, 1996; Nielsen & Dissayanake, 2004; Goertz, Kolling, Frahsek, Stanisch & Knopf, in press), as longitudinal studies are increasingly used to study infants´ declarative memory development. In the first longitudinal study of deferred imitation with 9- and 14-month-olds, Heimann and Meltzoff (1996) showed that there was a mean increase of deferred imitation performance from 9- to 14-months. Furthermore, in an exploratory post-hoc analysis of the stability of individual patterns the authors found that after classifying the infants at both ages into being either low or high imitators, about 80 per cent of the infants were stable in their imitative behaviour. This analysis gave evidence that there seems to be substantive stability of the deferred imitation task after disentangling mean growth from longitudinal stability correlations.

In a second longitudinal study, Nielsen and Dissayanake (2004) report longitudinal stabilities of the deferred imitation task. This study found some stability of individual memory performance over time, modest in younger infants and increasing throughout infancy. For the interval between 12 and 18 months, Nielsen and Dissayanake (2004) reported a stability correlation of r =.21. As this study focused primarily at the development of a new, more advanced, representational level (i.e. secondary representations, Perner, 1991) in the second year, the authors did not disentangle mean growth and stability patterns by statistical analyses. They only reported mean growth and stability correlations for deferred imitations and did not analyse the individual developmental trajectories of deferred imitation.

Within a third longitudinal study, investigating whether stable individual memory scores can be found with the deferred imitation task, Goertz et al. (in press) applied a test-retest-design (1 week interval) to assess the short-term stability of memory performance level. In this study, a reliability coefficient of r =.52 was reported. Further exploration of the data demonstrated that only two out of 24 infants showed inconsistent memory performances across the two test sessions. The inconsistency was identified with the individual consistency approach proposed by Asendorpf (1991). The authors discussed whether the two infants were mere outliers or whether a different pattern of stability correlations might arise in a longitudinal study with a larger sample of infants.

Although Heimann and Meltzoff (1996) report moderate stabilities after an exploratory post-hoc analysis of the stability of individual patterns and Goertz et al. (in press) excluded outliers with the individual consistency approach, statistically, all the longitudinal studies reported followed primarily a variable-centred approach to analyse the stability and mean growth of memory performance in infants. The variable-centred approach (e.g., Spiel, 1998) focuses on normative stability, which can be defined as “the preservation of a set of individual ranks on a quality within a constant population over a specified amount of time” (Alwin, 1994, p. 139). Since mere stability correlations calculate mean group stabilities at aggregate levels, they do not allow a differential analysis of individual stability or lability of memory performance over time (Asendorpf, 1990; Ghiselli, 1956, 1960). Thereby, individual stability reflects the relative position of an individual within a reference group across time. Following this definition, stability is independent of intra-individual change, but rather reflects the absence of inter-individual differences in intra-individual change.

To disentangle group-specific mean development and stability correlations, infants´ memory researchers could consider a person-centred approach (e.g., Spiel, 1998) for the analysis of longitudinal memory performance. Whereas in the variable approach the primary interest is on the relationship between variables, in the more differential person-centred approach the individual is at the focus of research. This person-specific perspective becomes especially important if scales of deferred imitation are increasingly integrated into test scales of infant development.

To analyse the individual development within a reference group, the person-centred approach focuses on relationships among individuals by grouping them into categories. From a developmental perspective, this approach identifies groups with different stabilities and growth trajectories within a longitudinal approach. Hence, one main advantage of the person-centred approach is the detection of heterogeneity in developmental trajectories, thereby taking a differential perspective on the research question. Those groups which differ in developmental trajectories can then be further described by other variables. This approach reflects a much more developmental point of view than the variable-centred one as it focuses at the person as the main object of study. There are several methodological pathways to analyse 2-wave longitudinal data with a person-centred approach.

Firstly, researchers may infer a person-specific development from different patterns of standard deviations at different measurement occasions with higher differences in standard deviations indicating more person-specific development. Despite being a highly interesting approach for a first screening of longitudinal data, this analysis does not provide any further information about different developmental groups moderating the statistical relationship. Secondly, simple difference scores between measurements may be calculated, which raises the reliability problem of difference scores (Cronbach & Furby, 1970). Thirdly, Ghiselli (1960) introduced an index which calculates the individual deviation from the regression line as a person-centred approach. As the method proposed by Ghiselli (1956, 1960) is based on *absolute* difference scores (|D|) between Z_{t1} and Z_{t2}, the index merely calculates deviations between measurement occasions without taking into account the direction of the deviation from the regression line. Fourthly, the individual consistency index proposed by Asendorpf (1990) calculates the squared differences of the *z*-scores $(1 – (Z_{t1} - Z_{t2})^2 / 2)$ as a person-specific measure and then logarithmically transforms them. Finally, a *relative* difference score D between Z_{t1} and Z_{t2} (Zedeck, 1971) may be computed. All those indices may potentially be used in a person-specific framework where the main aim is the detection of subgroups of infants with differential developmental stabilities and mean growth. The decision of which index to use, however, depends both on the question under study and on the scaling of the other variables obtained.

Yet, there are still no studies which aim primarily at describing the differential aspect of declarative memory development by identifying groups of children, showing a different pattern of development in terms of mean growth and stability. Therefore, the present study integrates both variable- and person-centred analyses in a 2-wave longitudinal design with 12 and 18 month-olds. Firstly, from a variable-centred perspective, the study aims at a replication and cross-validation of the mean growth and stability correlation of the Nielsen and Dissayanake (2004) study through the second year of life. Secondly, from a person-centred perspective, the aim was to analyse individual infants′ memory performances according to different mean growth and stabilities and thereby identify different groups of infants showing differential development in the deferred imitation task over time. Finally, the present study broadens the results of Heimann and Meltzoff (1996) with a different statistical procedure and within another age group.

Method

Participants

The sample consisted of $N = 92$ infants (48 male) who were recruited via radio announcement and advertisements in child care centres and local paediatricians and by word of mouth. Four subjects did not continue the study to the second testing because of relocation and one infant was excluded because of fussiness. Therefore, the statistical analyses were done for N = 87 infants. Criteria for admission into the study were no known physical, sensory or mental handicaps, normal length of gestation (over 37 weeks) and normal birth weight (2500 – 4500 grams). The mean age at the first measurement occasion was $M = 362.4$ days ($SD = 8.7$) and at the second measurement occasion $M = 551$ days ($SD = 7.9$). Mean birth weight was $M = 3393$ grams ($SD = 507.15$) and parents reported an average *APGAR index* of $M = 9.78/9.94$ with a minimum value of 7. Furthermore, the Developmental Test for 6-month-olds to 6-year-olds (Petermann & Stein, 2000), which assesses several developmental dimensions (e.g., motor development, language development, and cognitive development) showed that the total sample was within normal range in all these developmental parameters.

Testing environment

Testing took place in a small room that was unfurnished except for the experimental apparatus. During the test, the infant was seated on his or her parent's lap at a small rectangular table just opposite to the experimenter. Behind and to the left of the experimenter was a video camera that was focused on the child's head and torso and most of the tabletop. A second ceiling camera behind the infant recorded the experimenter. The recording apparatus was housed in an adjacent viewing room to reduce auditory distractions. The experiment was electronically timed by a computer that mixed elapsed time in 0.10 sec increments directly onto the videotaped records.

Material and Apparatus

Construction principles of tests. The construction and pilot testing of the deferred imitation tests applied in the present longitudinal study took into account most important test development requirements. Generally, structural and differential continuity, the relationship between item difficulty and differential validity as well as the longitudinal shrinkage of variance due to ceiling effects are very important pre-considerations. *Structural continuity* (Bates & Nowosad, 2006) concerns the degree of constancy in the operational definition of a trait over time; hence developmental measurement equivalence (Hartmann, 2006). *Differential continuity* (Bates & Novosad, 2006) reflects stability insofar as a person scoring high on a trait remains high in further measurements and vice versa. Richardson (1936) showed that a test composed of 50 per cent difficulty yields the highest potential differential validity. Furthermore the item difficulties have to be uniformly distributed over the whole item difficulty distribution, ideally from $P = .05$ to $P = .95$ (Gulliksen, 1950).

In addition, the problem of the shrinkage of the variance had to be taken into account, namely floor (too difficult task) and ceiling (too easy task) effects. It can be shown statistically that shrinkage of variance due to floor or ceiling effects results in lower reliability and lower longitudinal stability correlations (test-criterion correlations).

To obtain sound tests for deferred imitation, items for both 12-month-olds and 18-month-olds were pilot tested in several studies. Furthermore, control groups (12- and 18-month-olds) were assessed to obtain mean baseline performance and mean item performance. The finally adopted deferred imitation items were chosen among the potential items in the pre-tests when (1) they yielded good inter-rater reliability, (2) involved uniformly distributed item difficulties with a total mean item difficulty of about 50 percent, (3) comprised unknown actions that infants in the different age groups are able to perform motorically.

Furthermore, the selection of the deferred imitation items represented the several facets of deferred imitation (number of steps, causal item constraints, goal-relevant vs. goal-irrelevant steps). In general, the items for the 12-month-olds were constructed such as 2 out of 5 items consisted of 2 action steps, which were causally constrained (the second step could not be solved without the first). For the 18-month-olds the number of multi-staged items was increased (one 1-stepped, four 2-stepped, one 3-stepped) and one item (frog) included one goal-irrelevant step.

Deferred Imitation Tests. At the first measurement occasion (12-month-olds), the deferred imitation test comprised 5 actions, of which two were two-step actions and the three others one-step actions. At the second measurement occasion (18-month-olds), the deferred imitation test had 6 items consisting of one- up to three-step actions. All actions and the respective operational definitions for the deferred imitation items are described below.

Deferred Imitation Test for 12-month-olds. The objects were commercially available toys and were specially adapted for the experiments. Some of the objects and target actions were used in previous work (Knopf, Kraus & Kressley-Mba, 2006; Kressley & Knopf, 2006; Goertz, Kolling, Frahsek & Knopf, in press). In an independent study, baseline behaviour for the test items was assessed with *N* = 24 infants[1]. Mean item performance analyses showed that the items were uniformly distributed. Detailed test characteristics are given in Goertz, Knopf, Kolling, Frahsek and Kressley-Mba (2006). For all infants the target actions were presented in the same order. The five actions used in the study are shown in Figure 1.

Insert Figure 1 about here

Objects and target actions (all target actions were demonstrated four times):

1. *Tin can*: The blue tin can (diameter=4.5cm) filled with rice was shaken up and down three times. During retrieval, the tin can was empty to prevent infants from shaking it in order to produce the perceived noise.
2. *Toy pig with a hat*: The brown hat was removed from a light pink stuffed toy pig (20cm x 15cm x 7cm).
3. *Cup and knife*: Step 1: The upside down red plastic cup (diameter = 8cm) was turned with the opening on the top. Step 2: The wooden knife (length=16.5cm) was inserted and moved back and forth in the cup. This action is highly similar to an item used by Klein and Meltzoff (1999), the blue plastic box and the wooden stick, except that in the present study the cup had to be turned first.
4. *Wooden mouse*: The upper yellow half of a wooden mouse (8cm x 5.5cm) with an attached tail (6cm) was shut four times.

5. *Drum*: *Step 1*: The blue drumstick (length=8cm) attached with a white string was removed from the multi-coloured plastic drum. *Step 2*: The red button on the top was pressed twice in order to produce a battery-run noise. During retrieval, infants received a drum without noise in order to prevent children from producing the sound by pressing the button with their fingers.

Deferred Imitation Test for 18-month-olds. All the objects were commercially available and specially adapted for the experiments. Among the new items, the two most difficult objects (*mouse* and *drum*) for the 12-month-olds were taken over for the 18-month-olds test. In an independent study, baseline behaviour for the test items for the 18-month-olds was assessed with N = 26 infants[2]. Mean item performance analyses showed that the items were uniformly distributed. Detailed test characteristics are given in Goertz, Kolling, Frahsek and Knopf (under review). For all infants, the target actions were presented in the same order. The six actions for the 18-month-olds used in this study are shown in Figure 2.

Insert Figure 2 about here

Objects and target actions (all target actions were demonstrated four times):

1. *Car*: *Step 1*: The yellow car-resembling glove was placed on the hand. *Step 2*: The hand was folded three times to produce a waving.

2. *Goose and tin box: Step 1*: The silver tin box was clicked on the magnets at the white gooses´ belly. *Step 2*: The goose was lain down on the box.

3. *Wooden mouse*: The upper yellow half of a wooden mouse (8cm x 5.5cm) with an attached tail (6cm) was shut four times.

4. *Frog and ring*: *Step 1*: The green frog was turned upside down and placed with the head on the table. *Step 2*: The frog was turned back and placed into the dark-green ring. *Step 3*: The frog was moved with the ring from right to left and back to the middle.

5. Drum: *Step 1*: The blue drumstick (length=8cm) attached with a white string was removed from the multi-coloured plastic drum. *Step 2*: The red button on the top was pressed twice in order to produce a battery-run noise. During retrieval, infants received a drum without noise in order to prevent children from producing the sound by pressing the button with their fingers.

6. Octopus: *Step 1*: The yellow duck was placed on the red top of the yellow octopus. *Step 2*: The duck was turned back and forth.

Data scoring. The experimental sessions were videotaped by two cameras, one taping the infant and the second the experimenter. Two naïve and independent raters scored target action completion using operational definitions. They were uninformed about the hypotheses of the study and how exactly the experimenter demonstrated the target actions. Target actions were scored as yes- or no-responses.

Operational definitions and inter-rater reliabilities

12-month-olds. The operational definitions for the target actions were:

1. *Tin can*: A “yes” was coded if the tin can was shaken with one or both hands. The movement up and down or back and forth had to be done more than once.

2. *Toy pig*: A “yes” was coded if at least an attempt to draw the hat from the pigs head was observed. Simply touching the hat was not coded as a correct imitation.

3. *Cup and knife*: *Step 1*: A “yes” was coded if the cup was turned. *Step 2*: A “yes” was coded if an attempt of putting the knife into the cup was observed. The knife had to be inserted at least a bit into the cup.
4. *Mouse*: A “yes” was coded if the child pressed on the top of the mouse with the hand or with single fingers no matter whether the mouse was placed on the table or on one hand of the infant.
5. *Drum*: *Step 1*: A “yes” was coded if the infant took the drumstick in her hand. *Step 2*: A “yes” was coded if the red button on the top was touched with the stick (no matter with which side of the stick). An accidental fall of the stick on the button was not coded as a correct imitation.

Inter-rater reliabilities (12-month-olds). Rater 1 and 2 coded 72 of the subjects and reached an inter-rater reliability of r = 92.7% and a Cohen's kappa of κ = .83. Rater 1 and 3 coded 20 subjects and reached an inter-rater reliability of 89.0% and Cohen’s kappa of κ =.78 Discrepancies in scoring were resolved by consensus to 100% agreement.

18-month-olds. The operational definitions for the target actions were:
1. *Car*: *Step 1*: A “yes” was coded if the child inserted a hand or at least one finger into the glove-like car. *Step 2*: A “yes” was coded if the child waved with the car.
2. *Goose and tin box*: *Step 1*: A “yes” was coded if at least the attempt to put the tin box onto the magnets was observed (even if the box was not attached in the end). *Step 2*: A “yes” was coded if the goose with the attached box was turned to lie on the box or the goose alone is lain down with the belly on the box. An accidental fall on the box was not coded as “yes”.
3. *Mouse*: A “yes” was coded if the child pressed on the top of the mouse with her hand or with single fingers no matter whether the mouse was placed on the table or on one hand of the infant.
4. *Frog and ring*: *Step 1*: A “yes” was coded if the frog was placed with the head on the table. *Step 2*: A “yes” was coded if the frog was at least half-seated into the ring. If the infant

accidentally passed the ring while moving the frog over the table, no “yes” was coded. *Step 3*: A “yes” was coded if the frog (with at least 50% of it placed in the ring) was moved with the ring to and fro or back and forth. Even if the frog fell out of the ring, the attempt of moving it together with the ring was coded as “yes”.

5. *Drum*: *Step 1*: A “yes” was coded if the drum stick was taken. *Step 2*: A “yes” was coded if the red button on the top was touched with the stick (no matter with which side of the stick). An accidental fall of the stick on the button was not coded as ”yes”.

6. *Octopus*: *Step 1*: A “yes” was coded if the duck was placed on the octopus or the attempt of doing so was observed. *Step 2*: A “yes” was coded if the duck was turned on the octopus at least a small distance in one direction or back and forth.

Inter-rater reliabilities (18-month-olds). The four raters reached an inter-rater reliability of r = 95.7 - 96.4 % and a Cohen's kappa of κ = .91 – .92. Discrepancies in scoring were resolved by consensus to 100% agreement[3].

Procedure

First measurement occasion. Appointments were made for the time of day when the parents reported that their children were most alert, usually midmorning. In a waiting room, an interview prior to the experiment was conducted to explain the purpose of the study, details of the procedure, and to obtain informed consent. In the lab, the child was seated on the parent's lap. After the child appeared comfortable with the new situation, the demonstration phase started. The five props were shown successively to the child three times each within a 30 second interval. In the 30 minute retention interval, the Developmental Test for 6-month to 6-year-olds (Petermann & Stein, 2000) was administered. After the 30 minutes retention interval, the testing took place during which the infant was again given the five action-related props successively for 30 seconds each. Playing behaviour was video-taped.

Second measurement occasion. Testing at the second measurement occasion equalled almost exactly the first measurement occasion expect for the timing of the action presentation. Five of the actions were shown three times within a 30 second interval, whereas two actions were shown three times within a 45 second interval. The infants acted with the objects successively for 30 to 45 seconds each. Playing behaviour was videotaped.

Results

Variable-centred analysis

Mean developmental pattern for Deferred Imitation performance. The analysis of the mean number of target actions performed by the 12-month and 18-month-olds revealed a developmental trend. 12-month-olds imitated $M = 4.00$ ($SD = 1.56$) target actions and the 18-month-old infants imitated $M = 6.86$ ($SD = 1.86$) target actions. The mean difference test (t-test) between the two measurement occasions was significant ($t = 12.05$, $df = 85$, $p < .01$, $d = 0.62$).

Longitudinal stability. The longitudinal stability between the first and the second measurement occasion was $r = .23$ ($p < .05$, $n = 86$). Furthermore, the computation of the stability index corrected for attenuation with the uncorrected reliability of $r = .52$ yielded a slightly higher stability coefficient of $r = .32$.

Person-centred analysis. In order to gain more insight into this dynamic and to disentangle mean growth and stability within the longitudinal data, a differential, person-centred perspective was taken into account.

Statistical procedure. In a first step of the analysis, deferred imitation data for both measurement occasions were scaled with the common z-transformation, $z = (x – M) / SD$, where x is the individual infants´ value, M is the mean, SD is the standard deviation and z is the z-transformed individual score. As the study consisted of two measurement occasions, one z-score for the first measurement occasion (z_{t1}) and one z-score for the second measurement occasion (z_{t2}) for each infant was obtained. As the z-score transforms the scales with a resulting mean of 0 and a standard deviation of 1, this procedure leads to measurement equivalence at least from a differential continuity stance. If structural continuity had been given, then a centring of the variables would potentially have been a more adequate transformation procedure.

In order to analyse individual developmental trajectories different indicators for the measurement of change were considered. Firstly, *absolute* difference scores (|D|) between z_{t1} and z_{t2} (Ghiselli, 1956, 1960) were calculated with equation 1.

$$|D| = z_{t1} - z_{t2} \quad (1).$$

As the absolute difference score merely calculates deviations between measurement occasions without taking into account the direction of the deviation from the regression line, a further di- vs. trichotomization yielded neither methodologically nor theoretically plausible results. Therefore, this analysis was not considered any further. Secondly, the individual consistency approach proposed by Asendorpf (1990) was used to analyse different trajectories of deferred imitation capacity following equation 2.

$$IC = 1 - \frac{(z_{t1} - z_{t2})^2}{2} \quad (2)$$

As can be seen from equation 2, the procedure proposed by Asendorpf (1990) squares the differences of the z-scores. This means that the relative position of an individual (z-score) at one measurement occasion is subtracted from its relative position at the other measurement occasion. The more different the relative position of an individual to the two measurement occasions, i.e. the higher the difference between the z-scores, the lower the IC-index. As the index may vary between $-\infty$ and 1, Asendorpf (1990) proposed a logarithmic transformation for the IC scores which normalizes the skewed distribution. In contrast to the absolute difference score introduced by Ghiselli (1960), the individual consistency index avoids some statistical problems. However, the analysis of the Asendorpf index with a moderator approach yielded neither interpretable nor statistically meaningful results as it also uses an absolute difference measure for the measurement of change. Therefore, following a proposal by Zedeck (1971), in the present study, *relative* difference scores D between Z_{t1} and Z_{t2} were calculated with equation 3.

$$D = z_{t1} - z_{t2} \quad (3)$$

where D is the relative difference between $z_{t1} - z_{t2}$. Those *relative* difference scores *D* are reported in Figure 3.

Insert Figure 3 about here

In a last step of the statistical procedure, the relative differences D were categorized (*cat D*) by a median split (*two groups solution*) or a trichotomization (*three groups solution*).

The dichotomization by a median split yielded two groups ($D_{HG} \geq 0$; $D_{LG} < 0$). Graphically, in figure 3, the first group (*high growth*) is the group above the thickened, black zero line and the other group is the group (*moderate growth*) below the zero line. Then, stability correlations and means were calculated for the two groups separately, which are reported below.

In the last step of the analysis, motivated by the statistical framework of Zedeck (1969), the sample was divided into three equally sized groups (*n* = about 29) by trichotomizing the *relative* z-difference score (*D*). This procedure yielded three groups ($D_{HG} \leq -.57$; $D_{MG} \leq -.56 \leq .51$, $D_{LG} \geq .52$). Graphically, in figure 3, the first group is the group above the dotted line (*low growth*), the second group (*moderate growth)* is the group between dotted and dashed line and the third group (*high growth)* below the dashed line.

To sum up, the two groups solution was obtained by dichotomizing and the three groups solution by trichotomizing the relative difference scores. This statistical procedure allows to separate mean growth development from individual stability. In the following paragraphs, the mean growth and stabilities in line with the developmental characteristics of the two vs. three groups solution are presented.

Two groups solution

Mean growth of two subgroups. The analysis of the mean growth for the two subgroups revealed that the *moderate growth* group remembered $M = 4.97$ ($SD = 1.22$) target actions on the first measurement occasion and increased to $M = 6.10$ ($SD = 1.60$) target actions on the second measurement occasion ($t = 5.4$, $df = 42$, $p < .01$, $d = 1.67$). The *high growth* group remembered $M = 3.19$ ($SD = 1.26$) target actions on the first measurement occasion and $M = 7.6$ ($SD = 1.83$) target actions on the second measurement occasion ($t = 21.76$, $df = 42$, $p < .01$, $d = 6.72$). For a graphical representation of differential growth patterns, the reader is referred to Figure 4.

Insert Figure 4 about here

Stability correlations of two subgroups. The moderate growth group ($n = 43$) showed a stability correlation of $r = .57$ (*ns*) and the high growth group ($n = 43$) showed a stability correlation of $r = .69$ ($p < .05$). Figure 5 depicts the correlation diagram with the total sample, the two separated groups and their respective regression lines.

Insert Figure 5 about here

Developmental characteristics of the two subgroups. Furthermore, the high and moderate growth groups were analysed according to their motor, cognitive and language development assessed with the Developmental Test for 6-month to 6-year-olds (Petermann & Stein, 2000). There were no differences in motor development (M_{HG} = .40, *SD* = .21; M_{MG} = .33, *SD* = .19; *t* = 1.55, *df* = 84, *ns*), cognitive development (M_{HG} = .58, *SD* = .12; M_{MG} = .54, *SD* = .12; *t* = 1.58, *df* = 84, *ns*) or language development (M_{HG} =0.71, *SD* = .25; M_{MG} = .67, *SD* = .24; *t* = 0.72, *df* = 84, *ns*) between the high vs. moderate growth groups.

Three groups solution

Mean growth of three subgroups. The analysis of the mean growth for the three subgroups showed that the *low growth* group remembered *M* = 5.37 (*SD* = 0.88) target actions on the first measurement occasion and increased to *M* = 5.70 (*SD* = 1.35) target actions on the second measurement occasion (*t* = 1.67, *df* = 26, *ns*). The *moderate growth* group remembered *M* = 3.80 (*SD* = 1.27) target actions on the first measurement occasion and *M* = 6.50 (*SD* = 1.41) target actions on the second measurement occasion (*t* = 31.73, *df* = 29, $p < .01$, *d* = 11.78). The *high growth* group remembered *M* = 3.17 (*SD* = 1.47) target actions on the first measurement occasion and increased to *M* = 8.28 (*SD* = 1.83) target actions on the second measurement occasion (*t* = 25.44, *df* = 28, $p < .01$, *d* = 9.62). For a graphical representation of the differential growth patterns, the reader is referred to Figure 6.

Insert Figure 6 about here

Stabilities of three subgroups. The low growth group (n = 27) showed a stability correlation of r = .64 (p < .01), the moderate growth group (n = 30) a stability correlation of r = .95 (p < .01) and the high growth group (n = 29) a stability correlation of r = .81 (p < .01). Figure 7 depicts the correlation diagram with the total sample and the three separated subgroups and their respective regression lines.

Insert Figure 7 about here

Developmental characteristics of the three subgroups. Furthermore, the three subgroups (low, moderate, and high growth) were analysed according to their motor, cognitive and language development. There were no differences in motor development (M_{HG} = .37, SD = .19; M_{MG} = .36, SD = .22; M_{LG} = .36, SD = .21; F = 0.24, df = 2, *ns*), cognitive development (M_{HG} = .57, SD = .11; M_{MG} = .56, SD = .13; M_{LG} = .55, SD = .12; F = 0.23, df = 2, *ns*), or language development (M_{HG} = .75, SD = .23; M_{MG} = .66, SD = .27; M_{LG} = .67, SD = .22; F = 0.91, df = 2, *ns*).

Long-term retention and item specificity effect as confounds

The two most difficult items (*drum* and *mouse*) at the first measurement occasion were used again in the second measurement occasion. As long-term retention effects might have confounded the results of the growth group analyses, we analysed mean item performances for these two items. However, the calculation of mean item performance showed that, consistent with the analysis of the total items, the moderate growth group imitated more often both the *drum* and the item *mouse* at the first measurement occasion and less at the second measurement occasion than the high growth group. The same pattern was found for the three-group solution.

Furthermore, there might also have been differential group effects because of the different facets of the deferred imitation items (number of steps, causal item constraints, goal-relevant vs. goal-irrelevant steps) integrated into the tests. However, neither the correlation matrix nor a factor analysis showed signs of item-specificity for the different growth groups. Therefore, neither retention effects nor item specificity effects did confound the subgroups analysis of the present study.

Discussion

The results of the present study are twofold. First, infants´ memory performance increased significantly with age, replicating cross-sectional (Barr et al., 1996; Barr & Hayne, 2000) as well as longitudinal results (Heimann & Meltzoff, 1996; Nielsen & Dissayanake, 2004). Second, from a variable-centred stabilities viewpoint, a comparison of the mean stability index in the present study with the longitudinal study by Nielsen and Dissayanake (2004) shows that the stability indices are almost identical. Whereas Nielsen and Dissayanake (2004) report a stability correlation between memory performance of 12- and 18-month-olds of $r = .21$, the present study found a correlation of $r = .23$. This correspondence of outcomes both cross-validates the results of Nielsen and Dissayanake (2004) and shows that the stability correlation seems to be stable across different memory items and different laboratories. Furthermore, with regard to the high short-term stability of memory in infants (r = .52) found by Goertz et al. (in press), the longitudinal stability seems rather low and can therefore be ascribed to high dynamics in memory development occurring in the first half of the second year.

The main aim of this study was the analysis of infants´ declarative memory data at the age of 12- and 18-months with a person-centred, differential perspective. Two and three groups with different mean growth rates and stabilities were extracted. In the two groups solution, one moderate growth group and one high growth group were separated. In the three groups solution, a low growth, a moderate growth and a high growth group were differentiated. The longitudinal stabilities for the two and three groups solutions were much higher than the overall stability of r = .23 (total sample). This analysis of different groups broadens the exploratory analysis of the Heimann and Meltzoff (1996) study because of the larger sample size. Moreover, the deferred imitation test applied here consists of more items. Finally, a different age range was analysed in this study.

Taken together, a differential mean growth pattern for infants can be maintained. Furthermore, the stability of deferred imitation can be considered high if it is taken into account that a highly varying mean growth or a scattered correlation (low stability) does not always point to measurement error or instability but rather to differential development in infants.

From a statistical stance, critics of the present study could point to the statistical algorithm used in this study, which tries to disentangle mean growth from stability by categorizing the relative z-difference scores. The partitioning of the total intra- vs. inter-individual variance in groups of different mean growth lowers intra-individual variance at least in the predicted group. Therefore, it is statistically apparent that the stabilities rise. Refining the statistical properties of the algorithm will be a future methodological research aim. However, in the present study the algorithm showed that a differential development can be assumed in early memory development.

The analysis of the developmental characteristics of both the two and three group solution showed no differences in motor, cognitive and language development. Therefore, the explanation of this differential development within the deferred imitation task will be the aim for future research. Although no explaining factors were found, the present study demonstrates that a person-centred perspective on memory development in infancy is both necessary and fruitful. From a test theoretical perspective, infant researchers should be aware of the fact that there might be several groups of infants showing differential memory development due to differential quantity or quality changes. Those changes might be especially due to the large developmental dynamics in the first two years of life.

Theoretically, several explanations for this differential development can be put forward. Firstly, Meltzoff and his colleagues as well as Hayne and her colleagues (e.g. Herbert & Hayne, 2000) argue that infants change the way by which they represent certain facts in the second year. According to Herbert and Hayne (2000), representational flexibility, the ability of infants to increasingly de-contextualize the actions seen in the deferred imitation task, could be the mechanism, by which this qualitative change might be explained. With development, infants´ reliance on literal representations of their experience decreases and the use of hypothetical representations (deductive reasoning) increases.

Secondly, several researchers (McCormack & Hoerl, 2001; Tulving, 2002; Knopf, Mack & Kressley-Mba, 2005) hypothesize that with the development of the categorical self around 18 months of age a qualitative shift from semantic to episodic encoding may take place. Whereas semantic memory is the ability to consciously encode into and retrieve knowledge from declarative memory, episodic memory is the recollection of memories related to the personal self. Possibly, the different growth patterns and stabilities obtained might point to qualitative differences by which the infants encode the deferred imitation items. The obtained groups might indicate that some infants are already in a phase of qualitative change, whereas other infants have not reached this developmental phase yet. To shed further light on these theoretical considerations, a longitudinal study through the second year assessing self-development (e.g. mirror self-recognition), deferred imitation as well as representational development indicators is currently in progress in our lab.

Thirdly, an alternative theoretical hypothesis derived from empirical research proposes that solely quantitative changes, at least on the behavioural level, dominate the developmental process (Courage & Howe, 2002). Since the present study used only a two-wave assessment to analyse the differential stabilities, the data presented here do not rule out this possibility.

Irrespective of the theoretical notion favoured, further research is needed to enlarge the knowledge base of early declarative memory development within narrower and more frequent time intervals. This kind of longitudinal research in combination with both variable- and person-centred analysis approaches would provide further interesting data for declarative memory research. In terms of variable-centred growth trajectories, linear or quadratic trends could be analysed to resolve more clearly whether deferred imitation develops continuously or in stages. If declarative memory performance increased linearly over the infancy period, one might argue that the developmental process would only be quantitative. However, if performance did not increase linearly, a stage-sequential development would potentially be the better theoretical explanation.

Although a variable-centred, age-related growth trend would help to understand the development of declarative memory in infancy, a multivariate context increases the potential explanation of individual growth trends. Further multivariate, multi-wave longitudinal studies, in which infant researchers assess, for example, declarative memory indicators (deferred imitation), non-declarative memory indicators (for overview see Rovee-Collier, 1997), language acquisition, invisible displacement tasks (Haake & Somerville, 1985) and new, nonverbal theory of mind tasks (Onishi & Baillargeon, 2005) could further determine the understanding of individual differences in the second year.

In a longitudinal, multivariate study other statistical procedures to analyse individual trends are indicated, for example, growth curve modelling. Growth curve modelling analyses growth within a structural equation framework. This method is capable of integrating variable- and person-centred perspectives. In a first step, intercept (starting point) and slope (growth) are calculated for the total sample. Then, through a succession of predefined statistical stages, researchers can analyse individual intercepts and slopes and correlate those individual parameters with other developmental variables of interest. Therefore, person-specific intercepts and slope are explained by other variables obtained in the study. The hypothetical model that is finally used depends on the specific longitudinal design, on the focus of the research, scaling of the data, the underlying hypotheses, and other methodological assumptions.

Furthermore, in terms of data scaling, an interesting test-theoretical scaling alternative for the longitudinal analysis of deferred imitation performance would be a probabilistic scaling of the deferred imitation data. Even though this methodological scaling approach has strong prerequisites on the data, it could provide another methodological approach to the problem of variable and person effects in deferred imitation. In the last couple of years, item-response theory (IRT) has become very important for the scaling of test data, especially in scaling adaptive tests. IRT assumes an item-characteristic curve, which allows separating ability and person effects. As this is impossible with classical test theory assumptions, this state-of-the-art approach provides an adequate individual perspective method for scaling deferred imitation items in future studies. However, the application of such an approach necessitates the construction of deferred imitation items, which are item-response scalable (see Embretson & Reise, 2000, for overview).

The integration of the proposed methodological research strategies leads to the proposal for infant researchers to design longitudinal studies, use item-response scalable items and analyse those with newer statistical methods.

In sum, future research needs to (1) explain the developmental characteristics of different growth and stability groups in line with other developmental phenomena, (2) find plausible and methodologically sound combinations of psychological measures to shed further light on the theoretical notions of the development of deferred imitation, and (3) use adequate statistical methods for the analysis of the obtained data.

References

Alwin, D. F. (1994). Aging, personality, and social change: The stability of individual differences over the adult life span. In D.L. Featherman, R.M. Lerner & M. Perlmutter (Eds.), *Life-span development and behavior, Vol. 12* (pp. 135 – 185). Hillsdale, NJ: Lawrence Erlbaum.

Asendorpf, J. B. (1991). The measurement of individual consistency. *Methodika, 4,* 1-13.

Barr, R., & Hayne, H. (2000). Age-related changes in imitation: Implications for memory development. In C. Rovee-Collier, L. Lipsitt, & H. Hayne (Eds.), *Progress in infancy research* (Vol. I, pp. 21-67). Mahwah, NJ: Erlbaum.

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior & Development, 19,* 159-170.

Bates, J. E., & Novosad, C. (2005). Measurement of individual difference constructs in child development, or taking aim at moving targets. In D. M. Teti (Ed.). *Handbook of research methods in developmental science* (pp. 103-122). Oxford: Blackwell Publishing.

Bauer, P. J., Heertsgard, L. A., & Wewerka, S. S. (1995). Effects of experience and reminding on long-term recall in infancy: Remembering not to forget. *Journal of Experimental Child Psychology, 59***,** 260-298.

Carver, J. C., Bauer, P. J., & Nelson, C. A. (2000). Associations between infant brain activity and recall memory. *Developmental Science, 3,* 234-246.

Collie**,** R., & Hayne, H. (1999). Deferred imitation by 6- and 9-month-old infants: More evidence for declarative memory. *Developmental Psychobiology, 35,* 83-90.

Courage, M. L., & Howe, M. L. (2002). From infant to child: The dynamics of cognitive change in the second year of life. *Psychological Bulletin, 128,* 250-277.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change" - or should we? *Psychological Bulletin, 74,* 68-80.

Embretson, S. E., & Reise, S. P. (2000). *Item-response-theory for psychologists*. Mahwah, NJ : Lawrence Erlbaum Publishers.

Goertz, C., Knopf, M., Kolling, T., Frahsek, S., & Kressley-Mba, R. A. (2006). Entwicklung und Erprobung eines Messinstruments zur Erfassung des deklarativen Gedächtnisses Einjähriger: Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12). [The Frankfurt Imitation Test for 12-Month-Old Infants: Testing a series of items for measuring declarative memory in 1-year-olds.]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38,* 88-96.

Goertz, C., Kolling, T., Frahsek, S., Stanisch, A., & Knopf, M. (in press). Assessing declarative memory in 12-month-old infants: A test-retest reliability study based on the deferred imitation task. *European Journal of Developmental Psychology.*

Goertz, C., Kolling, T., Frahsek, S., & Knopf, M. (2007). Der Frankfurter Imitations-Test für 18 Monate und 24 Monate alte Kinder (FIT18-24). [The Frankfurt Imitation Test for 18-Month-Old and 24-month-Old Infants: Testing a series of items for measuring declarative memory in 1.5-2-year-olds.]. *Manuscript submitted for publication.*

Ghiselli, E. E. (1956). Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology, 40,* 374-377.

Ghiselli, E. E. (1960). The prediction of predictability. *Educational and Psychological Measurement, 20*, 3-8.

Gulliksen, H. (1950). *The theory of mental tests*. New York: Wiley.

Haake, R. J., & Somerville, S. C. (1985). Development of logical search skills in infancy. *Developmental Psychology, 21,* 176-186.

Hartmann, D. P. (2006). Assessing growth in longitudinal investigations: Selected measurement and design issues. In D. M. Teti (Ed.). *Handbook of research methods in developmental science* (pp. 340-366). Oxford: Blackwell Publishing.

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review, 24,* 33-73.

Hayne, H., Boniface, J., & Barr, R. (2000). The development of declarative memory in human infants: Age-related changes in deferred imitation. *Behavioral Neuroscience, 114,* 77-83.

Hayne, H., McDonald, S., & Barr, R. (1997). Developmental changes in the specificity of memory over the second year of life. *Infant Behavior & Development, 20,* 233-245.

Heimann, M., & Meltzoff, A. N. (1996). Deferred imitation in 9- and 14-month-old infants: A longitudinal study of a Swedish sample. *British Journal of Developmental Psychology, 14,* 55-64.

Herbert, J., & Hayne, H. (2000a). The ontogeny of long-term retention during the second year of life. *Developmental Science, 3,* 50–56.

Herbert, J., & Hayne, H. (2000b). Memory retrieval by 18-30-month-olds: Age-related changes in representational flexibility. *Developmental Psychology, 36,* 473-484.

Howe, M. L., & Courage, M. L. (2004). Demystifying the beginnings of memory. *Developmental Review, 24,* 1-5.

Knopf, M., Kraus, U., & Kressley-Mba, R.A. (2006). Relational information processing of novel unrelated actions by infants. *Infant Behavior & Development, 29,* 44-53.

Knopf, M., Mack, W., & Kressley-Mba, R. (2005). Wissen und Erinnern. Zur Genese des episodischen Gedächtnisses bei Säuglingen und präverbalen Kindern. [Knowing and remembering: On episodic memory in infants and preverbal children.]. *Psychologische Rundschau, 56,* 113-122.

Kressley R. A., & Knopf, M. (2006). A comparison of between- and within-subjects imitation design. *Infant Behavior & Development, 29*, 564-573.

Kressley-Mba, R. A., Lurg, S., & Knopf, M. (2005). Testing for deferred imitation of 2- and 3-step action sequences with 6-month-olds. *Infant Behavior & Development, 28, 8*2-86.

Mandler, J. M., & McDonough, L. (1995). Long-term recall of event sequences in infancy. *Journal of Experimental Child Psychology, 59,* 457-474.

McCormack, T., & Hoerl, C. (2001). The child in time: Temporal concepts and self-consciousness in the development of episodic memory. In Ch. Moore & K. Lemmon (Eds.), *The self in time: Developmental perspectives* (pp. 203-227). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

McDonough, L., Mandler, J., McKee, R., & Squire L. (1995). The deferred imitation task as a nonverbal measure of declarative memory. *Proceedings of the National Academy of Science, 92,* 7580-7584.

Meltzoff, A. N. (1985). Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. *Child Development, 56,* 62-72.

Meltzoff, A. N. (1988a). Infant imitation after 1-week delay: Long-term memory for novel actions and multiple stimuli. *Developmental Psychology, 24,* 470-476.

Meltzoff, A. N. (1988b). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development, 59,* 217-225.

Meltzoff, A. N. (1988c). Imitation of televised models by infants. *Child Development, 59,* 1221-1229.

Meltzoff, A. N., & Moore, M. K. (1999). Persons and representation: Why infant imitation is important for theories of human development. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (pp. 9-35) Cambridge: Cambridge University Press.

Meltzoff, A. N., & Moore, M. K. (2001). 'Discovery procedures' for people and things: the role of representation and identity. In F. Lacerda, C. von Hofsten, & M. Heimann (Eds.), *Emerging cognitive abilities in early infancy* (pp. 213–230). Mahwah, NJ: Erlbaum.

Nielsen, M., & Dissanayake, C. (2004). Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year. *Infant Behavior & Development, 27*, 342-365.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old-infants understand false beliefs? *Science, 308,* 255-258.

Petermann, F., & Stein, I. A. (2000). *Entwicklungstest ET 6-6. [Developmental Test for 6-month to 6-year-olds.]*. Lisse: Swets & Zeitlinger Publishers.

Perner, J. (1991). *Understanding the representational mind.* Cambridge: MIT Press.

Piaget, J. (1962). *Play, dreams and imitation in childhood.* New York: Norton. (Original work published 1945).

Richardson, M. W. (1936). The relation between the difficulty and the differential validity of a test. *Psychometrika, 1,* 33 – 49.

Rovee-Collier, C. (1997). Dissociations in infant memory: Rethinking the development of implicit and explicit memory. *Psychological Review, 104,* 467-498.

Spiel, C. (1998). Four methodological approaches to the study of stability and change in development. *Methods of Psychological Research Online, 3,* 7-22.

Tulving, E. (2002). Episodic memory and common sense: how far apart? In A. Baddeley, M.

Conway & J. Aggleton (Eds.), *Episodic memory* (pp. 269-287). Oxford: Oxford University Press.

Zedeck, S. (1971). Identification of moderator variables by discriminant analysis in a multipredictable group validation model. *Journal of Applied Psychology, 55,* 364-371.

Footnotes

1. The target action *tin can* had a baseline performance rate of 20.8% the target action *toy pig with hat* of 8.3 %, the target action *cup and knife* of 29.2 %, and the target action *drum* of 20.8 % and the target action *mouse* of 4.2 %.

2. The target actions *car, goose, frog, octopus* had a baseline performance rate of 0%. The target action *mouse* had a baseline performance rate of 19.2 percent and the target action *drum* of 7.7 %.

3. Four raters rated the video tapes. Rater 1 and Rater 2 reached an inter-rater reliability of r = 96.1% (Cohen's kappa κ = .92), Rater 2 and Rater 3 an inter-rater reliability of r = 96.1% (Cohen's kappa κ = .92) and Rater 3 and Rater 4 an inter-rater reliability of r = 95.8% (Cohen's kappa κ = .91).

Figure Captions

Figure 1 Objects and target actions for the 12 month-olds: Tin can, pig, cup and knife, wooden mouse, drum

Figure 2 Objects and target actions for the 18 month-olds: Car, goose, tin box, mouse, frog and ring, drum, octopus

Figure 3 Z-difference scores between the two measurement occasions

Figure 4 Growth for total sample and 2 subgroups

Figure 5 Correlation pattern (total sample) with group stabilities and different mean development for the 2 groups

Figure 6 Growth for total sample and 3 subgroups

Figure 7 Correlation pattern (total sample) with group stabilities and different mean development for the 3 groups





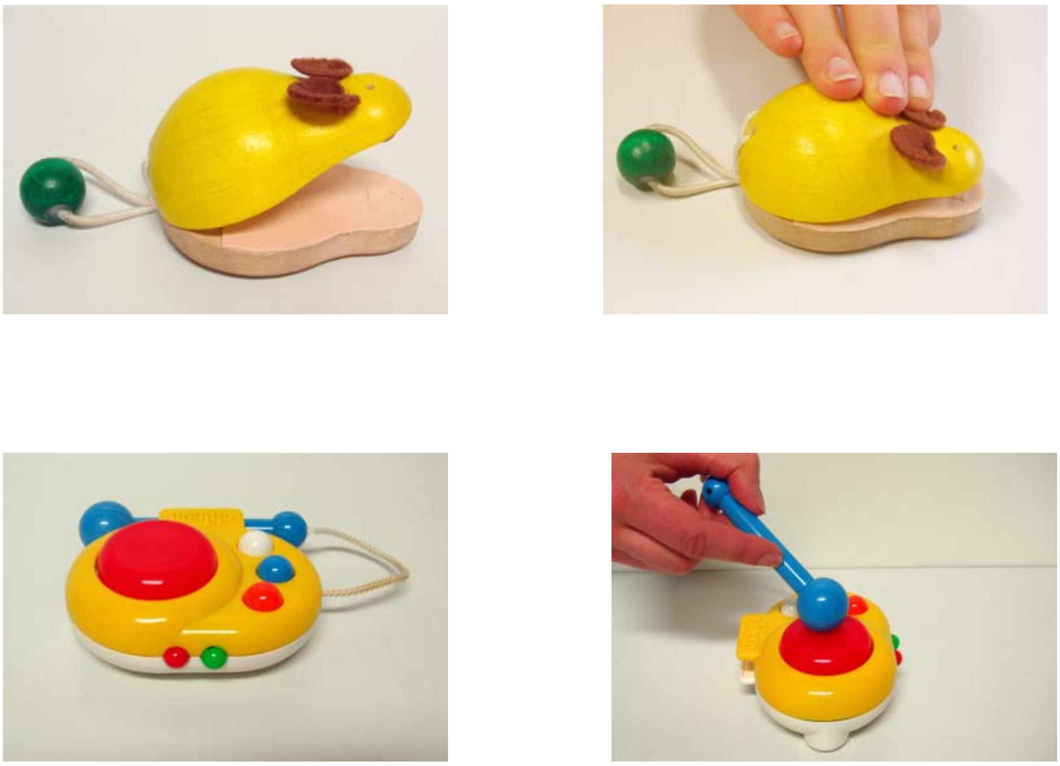



Figure 1




Figure 2

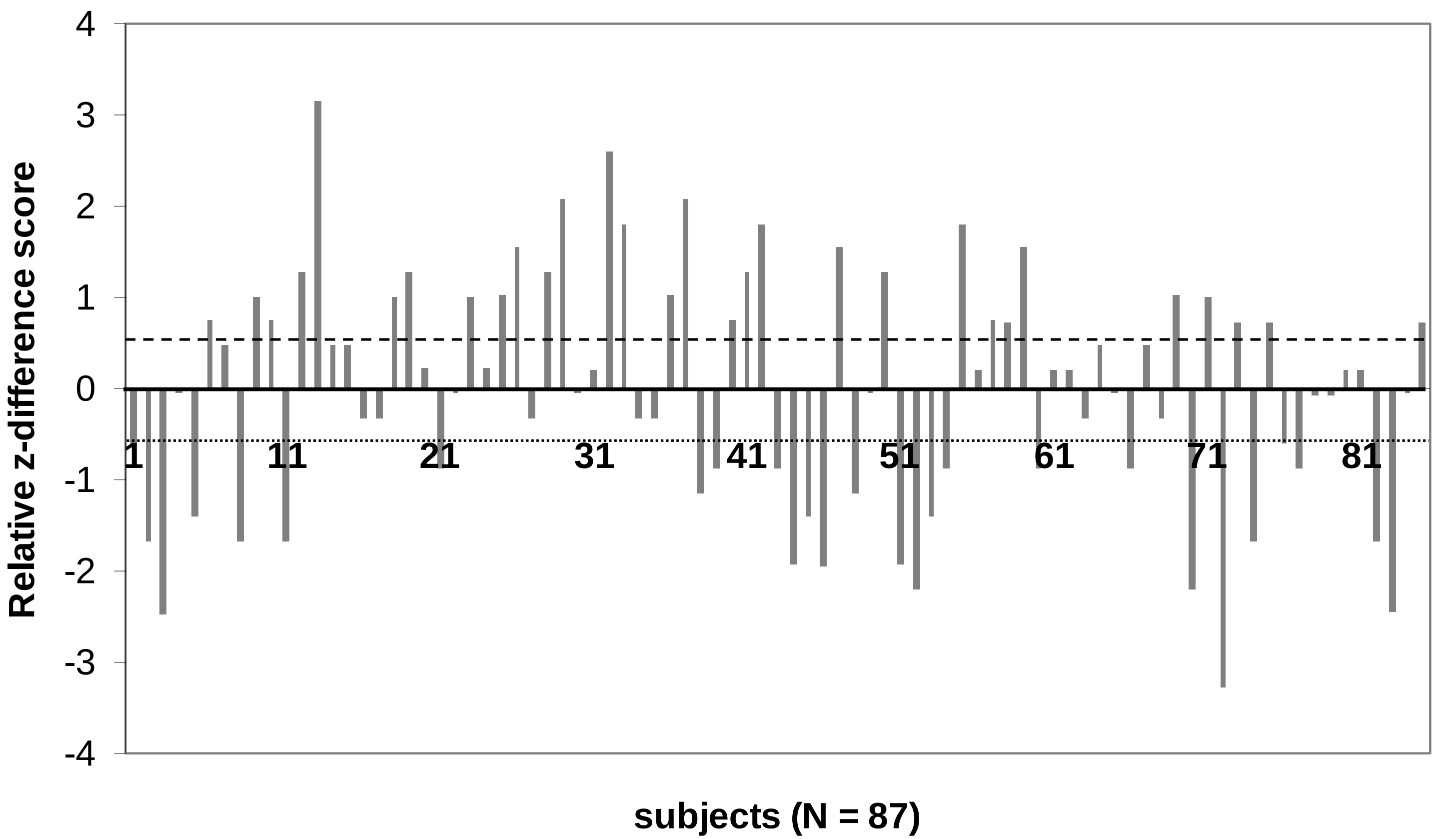


Figure 3

Note: Relative z-difference scores were computed with the formula (z_{t2}-z_{t1}) Relative z-difference scores vary between -∞ and ∞.

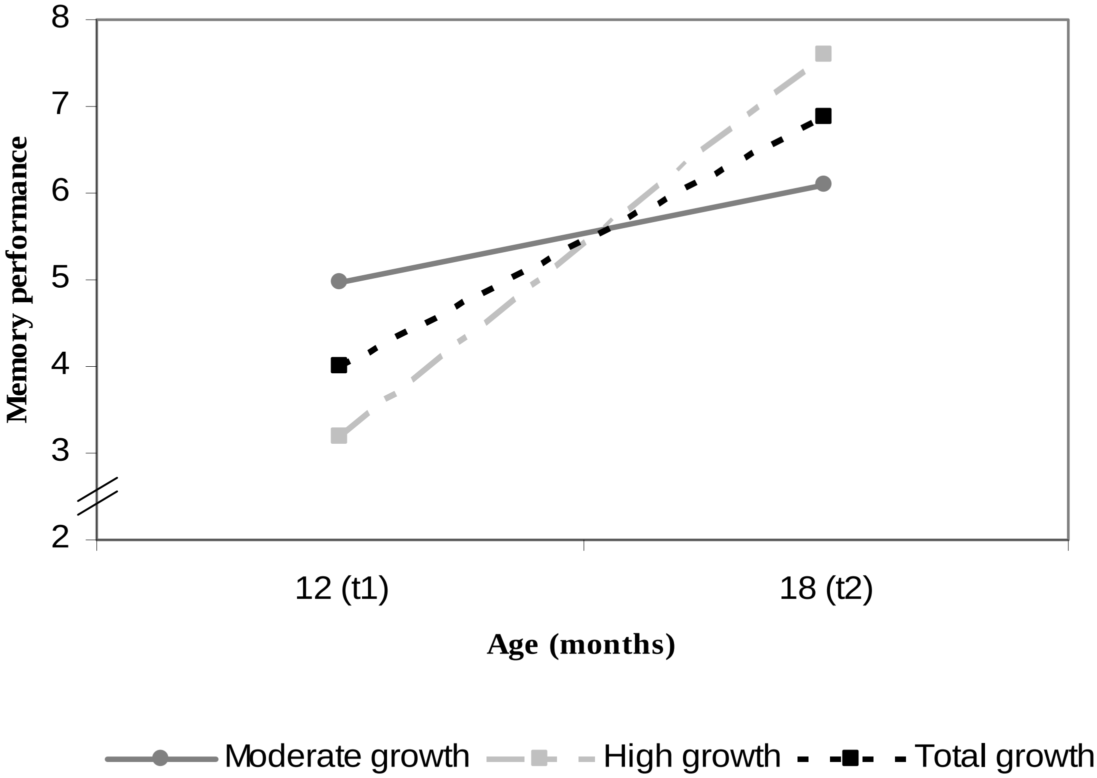

8
7
6
5
4
3
2
Memory performance
12 (t1)
18 (t2)
Age (months)
Moderate growth
High growth
Total growth

Figure 4

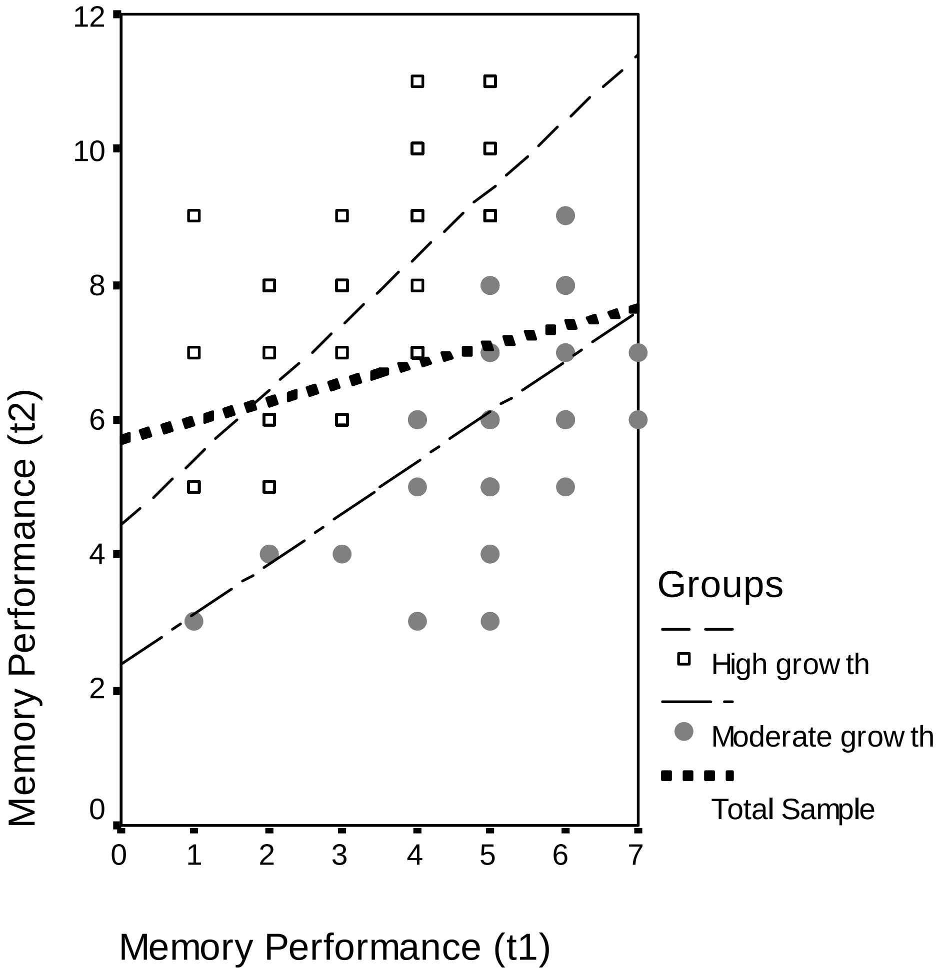

12
10
8
6
4
2
0
Memory Performance (t2)
0
1
2
3
4
5
6
7
Memory Performance (t1)
Groups
High grow th
Moderate grow th
Total Sample

Figure 5

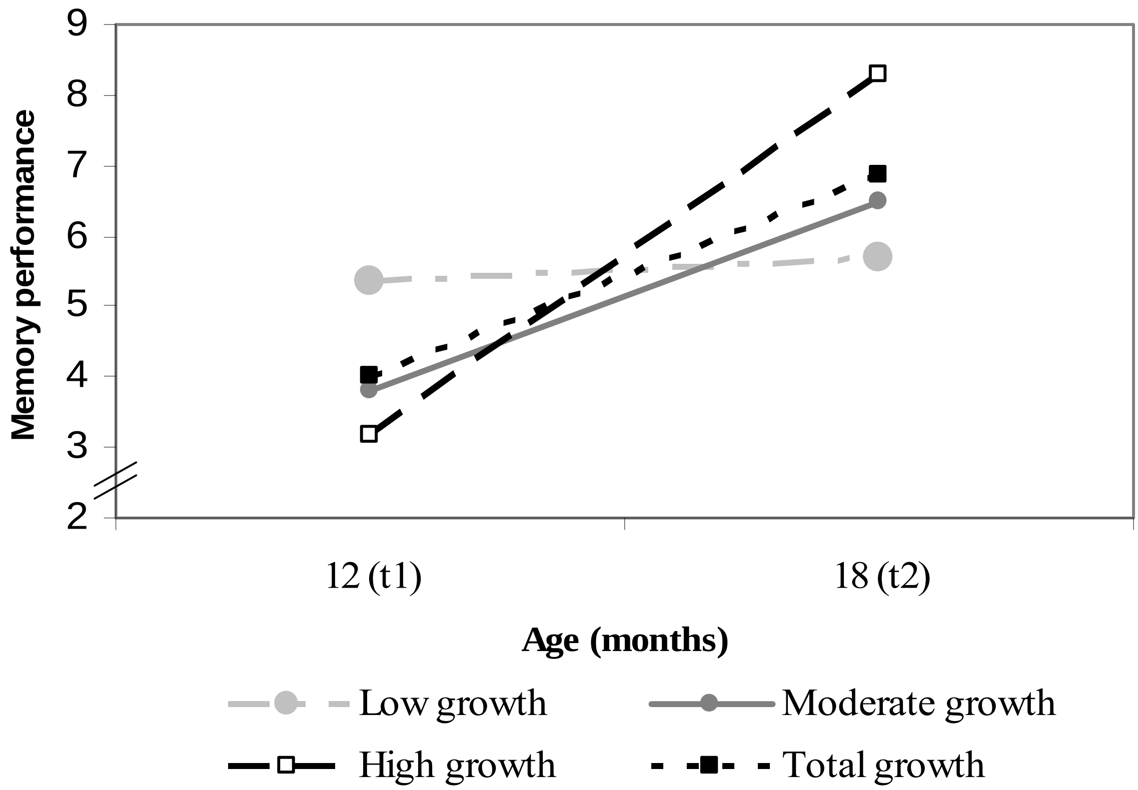

9
8
7
6
5
4
3
2
Memory performance
12 (t1)
18 (t2)
Age (months)
Low growth
Moderate growth
High growth
Total growth

Figure 6

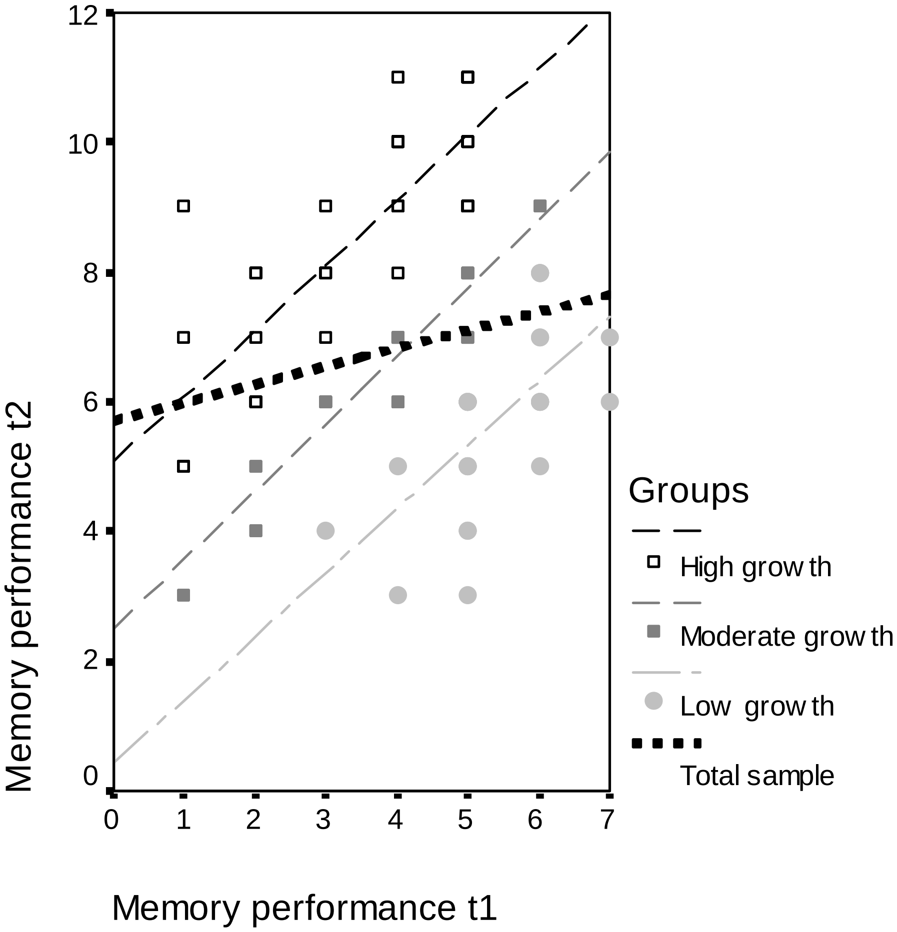

Memory performance t2
12
10
8
6
4
2
0
0
1
2
3
4
5
6
7
Memory performance t1
Groups
High grow th
Moderate grow th
Low grow th
Total sample

Figure 7

Paper 6

Dynamics of declarative memory from infancy to childhood

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (subm.). Dynamics of declarative memory from infancy to childhood. *Child Development.*

Dynamics of declarative memory from infancy to childhood

Submitted 2007/06/15

Word/Pages count:

Abstract: 120 words   Manuscript 34 pages long, 4 Figures, 1 Table

Abstract

Deferred imitations assess declarative memory in infants. Cross-sectional and longitudinal studies demonstrated that with increasing age, infants learn faster, and retain more target actions over longer retention intervals. Furthermore, individual stability correlations are moderate but increase throughout the second year. However, multivariate approaches focusing on inter-individual differences of deferred imitation are largely missing. The present three-wave (12-, 18- and 24-month-old infants), multivariate, longitudinal study analyzed deferred imitation in line with cognitive, motor, emotional, social, language and self-development. A person-centered approach (cluster analysis) demonstrated inter-individual differences in intra-individual change, i.e. a two group solution. Multivariate analyses revealed differences of language, motor and self-development at the second measurement occasion between groups. Results are discussed against the background of contemporary developmental theories.

Keywords: Infant memory, deferred imitation task, declarative memory, individual differences

Introduction

Imitation is a central mechanism for learning and memory in humans and non-human primates (e.g., Meltzoff, 2002, 2005). Barr and Hayne (2003), for example, argue that infants acquire up to 1-2 new behaviors per day through an imitative learning process. With the course of development, the imitative acts of the infants become more and more decoupled from the presence of a model and deferred imitations are observed.

Based on seminal work by Piaget (1962) who observed and described deferred imitations in natural contexts, Meltzoff (1985, 1988a, 1988b, 1988c) introduced an experimental method for assessing declarative memory in preverbal infants. In this task, young infants observe short series of simple, object-based actions performed by an experimenter. After a delay of minutes, hours or even days the children are given the props and their target behavior is observed. Depending on the experimental procedure and the aim of the study, the infant is given these unknown objects prior to the action demonstrations to study baseline behavior (Kressley & Knopf, 2006). A substantial increase in infants' target behavior in the experimental group, in comparison to the control group, indicates evidence of declarative memory. Thereby, in contrast to non-declarative memory, declarative memory is defined as the consciously accessible part of the memory system (Tulving, 1985). Several arguments strengthen the assumption that deferred imitation in fact measures declarative memory (Mandler, 2004; McDonough, Mandler, McKee & Squire, 1995). Firstly, the cross-modal character of the deferred imitation task makes it improbable that infants' memory performance is due to priming processes, which are sensitive to modality changes. Secondly, in contrast to incremental learning processes, no motor exercise is possible. Thirdly, within deferred imitation tests, new, unknown actions are used as memory items, which were not available in the infant's motor repertoire before. Fourthly, amnesic subjects are unable to show deferred imitation. Fifthly, infants not only produce single actions but also encode structural elements between actions, i.e. they encode item-relational information (Knopf, Kraus & Kressley-Mba, 2006).

Since the work by Piaget (1962) and Meltzoff (1985) many cross-sectional studies have been conducted, but only a few longitudinal studies. There is evidence from several studies that 6-month-old infants show deferred imitation after short retention intervals (e.g., Collie & Hayne, 1999; Hayne, Boniface, & Barr, 2000; Kressley-Mba, Lurg, & Knopf, 2005), and that 9-month-old infants retrieve formerly seen actions after longer retention intervals (e.g., Barr & Hayne, 1996; Hayne, Boniface & Barr, 2000; Kressley-Mba, Lurg, & Knopf, 2005; Meltzoff & Moore, 1999). There are age-related changes in deferred imitation parameters, namely that with increasing age infants need less exposure to the target actions; hence they learn faster than younger ones. Furthermore, in comparison to younger infants, older infants imitate actions after longer retention intervals (Barr, Dowden & Hayne, 1996; Barr & Hayne, 1996; Barr & Hayne, 2000; Bauer et al., 1995; Hayne, McDonald & Barr, 1997; Hayne et al., 2000; Herbert & Hayne, 2000a, 2000b; Mandler & McDonough, 1995; Meltzoff & Moore, 1999).

Although most of the deferred imitation studies have been cross-sectional, longitudinal studies have emerged increasingly in recent years. Consistent with cross-sectional studies, longitudinal studies indicate that memory performance increases with age (Heimann & Meltzoff, 1996; Nielsen & Dissayanake, 2004; Goertz, Kolling, Frahsek, Stanisch, & Knopf, 2007; Kolling, Goertz, Frahsek, & Knopf, in press). Furthermore, these studies demonstrated that the individual stability of deferred imitation is modest in early infancy and increases throughout the second year. In a psychometrically oriented short-term longitudinal study, Goertz et al. (2007) report that deferred imitation performance was highly stable (*r* = .52, corrected for attenuation *r* = .63) for a one week test-retest interval. After the exclusion of two outliers with very low stability indicators, following the individual consistency approach by Asendorpf (1990), the short-term test-retest reliability even increased to *r* = .63, indicating a good reliability of the deferred imitation task.

Taken together, the studies using the deferred imitation paradigm reviewed above indicate that 6-month-olds are capable of deferred imitation, thus demonstrating declarative memory for object-related action events. Declarative memory in infants can be assessed reliably, memory capacity increases with age, and longitudinal stabilities are modest in infancy.

Most of these studies, however, investigated deferred imitation using univariate, cross-sectional designs. Multivariate analyses focusing on individual differences of deferred imitation were only marginally, if at all, taken into account. For this reason and following related claims (Jones & Herbert, 2006), the analysis of individual differences of deferred imitation is a necessary next research step for understanding relationships between declarative memory and important developmental correlates. Research of individual differences of deferred imitation requires both cross-sectional and longitudinal, multivariate studies, as well as adequate statistical techniques to understand the co-varying factors of deferred imitation development. So far, only a few multivariate studies have reported correlations and predictive regressions between deferred imitation performance and the development of other cognitive as well as non-cognitive aspects.

In several studies, co-variations between declarative memory performance and language development were found. In one study combining deferred imitation items with verbal cues, Herbert and Hayne (2000) showed that infants at the age of 24 months used language cues effectively in a context change condition, whereas 18-month-olds were not able to do so. Heimann, Strid, Smith, Tjus, Ulvund, and Meltzoff (2006) investigated the relationship between recall memory, visual recognition memory, social communication, and the emergence of language skills. They reported that visual recognition memory (at 6 months), deferred imitation (at 9 months) and turn-taking skills (at 14 months) predicted language skills at the age of 14 months, with deferred imitation accounting for the highest variance in their regression model.

There are also important co-variations between declarative memory performance and self-development in the second year. Mirror self-recognition, considered as the benchmark test for the development of a self-concept in infancy, is one of the important developmental milestones in the second year of life, typically reached around 18 months. Deferred imitation performance co-varied with the development of mirror self-recognition in 20-month-old children (Prudhomme, 2005). By combining deferred imitation and mirror self-recognition, Prudhomme demonstrated that children who passed the mirror test were less affected by a change of colours of the relevant target props given within the deferred imitation test than those who did not pass the mirror test. This result is assumed to indicate that with the more personal features infants are able to embed in the memory traces, the more memory processing is elaborative and hence rather episodic. Furthermore, she concludes that her findings show the role of the cognitive self as a factor of differentiation between semantic and episodic memory. This empirical evidence is in line with episodic memory development theories, which focus on the development of the self and the emergence of episodic remembering in infants and children. Howe and Courage (1997), for example, argue that the emergence of the categorical self, which develops around 18 months, constitutes a lower boundary for the development of episodic memory. Furthermore, Knopf, Mack, and Kressley-

Mba (2005) and Howe (2004) assume that the prerequisites for episodic remembering begin to develop in the second year. This theory indicates that self-development is an important correlate of declarative memory development.

Nielsen and Dissayanake (2004), in a multivariate, longitudinal study, assessed deferred imitation, synchronic imitation, cognitive development (pretend play) and self-development (mirror self-recognition) through the second year with four measurement occasions. The authors found that deferred imitation develops prior to synchronic imitation, pretend play and mirror self-recognition, and that the development of the latter three skills followed a similar developmental trajectory between 18 and 21 months. In this study, using a variable-centred approach (e.g., Spiel, 1998), Pearson's product–moment correlation coefficients between measurement occasions and constructs as well as mean performance scores were reported. As variable-centred approaches analyze group stabilities at aggregate levels, they do not allow for a differential analysis of individual stability or lability of memory performance over time (Asendorpf, 1990; Ghiselli, 1956, 1960). The variable-centered approach focuses on normative stability. Since mere (stability) correlations calculate mean group stabilities at aggregate levels, they do not allow a differential analysis of individual stability or lability of memory performance over time (Asendorpf, 1990). Thereby, individual stability reflects the relative position of an individual within a reference group across time. Following this view, stability is independent of intra-individual change, but rather reflects the absence of inter-individual differences in intra-individual change.

In the variable-centered approach, the focus of interest is on single or combinations of variables, and on correlations between these variables. In contrast, the person-centered approach focuses on the person and therefore on inter-individual differences of intra-individual change. There are exploratory and confirmatory person-centered analysis approaches, i.e. longitudinal cluster analysis and longitudinal configural frequency analysis. Generally, these approaches allocate the subjects under study into different developmental groups. In a next step of analysis, intra-individual change is explainable by important correlating variables.

Using two waves of data (12- and 18-month-old infants) of a longitudinal study, Kolling et al. (in press) analyzed the development of deferred imitation with a person-centred analysis approach. The authors identified subgroups revealing differential growth and stability. However, these subgroups did not differ with respect to cognitive, language and motor development. This study nevertheless demonstrated empirically that the analysis of deferred imitation with a person-centered approach is both fruitful and necessary to understand inter-individual differences in intra-individual change.

Taken together, the evidence reviewed above clearly demonstrates that language and self-development are both theoretically and empirically important correlates for declarative memory development. However, it is far from clear how these interactions lead to inter-individual differences of intra-individual change in infancy.

The present multivariate, longitudinal study investigates co-variations of deferred imitation performance, cognitive, language, social, emotional, motor and self-development through the second year. A multivariate, longitudinal design with three-waves (12-, 18 and 24-month-old infants) was applied. For the analyses of the data, a person-centered approach is used to focus increasingly on inter-individual differences of intra-individual change.

To sum up, the aims of the present longitudinal study are threefold. Firstly, the study focused on the extension of the longitudinal knowledge base of deferred imitation performance with respect to mean growth and stability correlations (variable-centred analysis) through the second year. Secondly, inter-individual differences in intra-individual change are analyzed using group-based individual growth curves (i.e. cluster analysis) of deferred imitation (person-centered analysis). Thirdly, these developmental groups are compared with respect to cognitive, language, social, emotional, motor and self-development (multivariate analysis).

Method

Participants

The sample of the longitudinal study consisted of *N* = 92 infants (*N* = 48 boys) who were recruited via radio announcement and advertisements in childcare centres and local pediatricians and by word of mouth. Four subjects did not continue the study to the second testing because of relocation and one infant was excluded from data analyses because of fuzziness. This means that the statistical analyses were done for *N* = 87 infants. The criteria for admission into the study were no known physical, sensory or mental handicaps, normal length of gestation (over 37 weeks) and normal birth weight (2500–4500 grams). The mean age of the subjects at the first measurement occasion was *M* = 362 days (*SD* = 8.7) and *M* = 551 days (*SD* = 7.9) at the second testing. At the third measurement occasion children were *M* = 731 (*SD* = 10.6) days old. Mean birth weight was *M* = 3393 grams (*SD* = 507). Parents reported an *APGAR* index of *M* = 9.78/9.94 with a minimum score of 7.

Furthermore, the developmental test for 6-month to 6-year-olds (Petermann & Stein, 2000) which aims at the assessment of several psychological aspects (e.g., motor development, language development, cognitive development) demonstrated that all subjects were within normal range in all aspects tested.

Testing environment

All infants were tested individually in the laboratory. Testing took place in a small room that was unfurnished except for the experimental apparatus. During the testing, the infant was seated on his or her parent’s lap at a small rectangular table just opposite the experimenter. Behind and to the left of the experimenter was a video camera that was focused on the child's head and torso and most of the tabletop. A second camera behind the infant on the right recorded the experimenter. The recording apparatus was housed in an adjacent viewing room to reduce auditory distractions. The testing was electronically timed by a computer that mixed elapsed time in 0.10 sec increments directly onto the videotaped records.

Material and Apparatus

Construction principles of tests. The construction and pilot testing of the deferred imitation tests applied in the present longitudinal study took into account most important test development requirements. Generally, structural and differential continuity, the relationship between item difficulty and differential validity as well as the longitudinal shrinkage of variance due to ceiling effects are very important pre-considerations. Structural continuity (Bates & Nowosad, 2005) concerns the degree of constancy in the operational definition of a trait over time, hence developmental measurement equivalence (Hartmann, 2005). Differential continuity (Bates & Novosad, 2005) reflects stability insofar as a person scoring high on a trait remains high in further measurements and vice versa. It is well known that a test composed of 50% difficulty yields the highest potential differential validity. Furthermore, the item difficulties have to be uniformly distributed over the whole item difficulty distribution, ideally from P = .05 to P = .95 (Gulliksen, 1950). In addition, the problem of the shrinkage of the variance had to be taken into account, namely floor and ceiling effects. It can be shown statistically that because of the shrinkage of variance of floor as well as ceiling effect, a lower reliability and lower longitudinal stability correlations (test-criterion correlations) results.

To obtain sound tests for deferred imitation, items for both 12-month-olds and 18-month-olds were pilot tested in several independent studies. Furthermore, control groups (12- and 18- month-olds) were assessed to obtain mean baseline performance and mean test performance. The deferred imitation items finally adopted were chosen among the potential items in the pre-tests using the criteria that (1) they yield good inter-scorer reliability, (2) they involve uniformly distributed item difficulties with a total mean item difficulty of about 50 percent, (3) the items reflect several facets of deferred imitation (number of steps, causal item constraints, goal-relevant *vs*. goal-irrelevant steps), and (4) they comprise formerly unknown actions that infants in the different age groups are able to perform in terms of motor skills. The tests finally applied consisted of these items adjusted for 12-, 18- and 24-month-olds, respectively.

Frankfurt Imitation Test for 12-month-olds (FIT-12) (Goertz, Knopf, Kolling, Frahsek, & Kressley-Mba, 2006). At the first measurement occasion, the Frankfurt Imitation Test for 12-months-olds (FIT-12) was administered. FIT-12 consists of five different object-based actions, two of which are two-step actions and three are one-step actions. The optimal score that could be reached in FIT-12 is 7. The objects were commercially available toys and were specifically adapted for this age-group. Most of the objects and target actions were used in previous research (Knopf, Kraus, & Kressley-Mba, 2006; Kressley-Mba, & Knopf, 2006).

The five actions were presented successively by a model to the infants three times each within a 30-seconds interval. Following a delay of 30 minutes, infants were given the props successively for 30 seconds each. Infants' playing behavior was video-taped. The three scores of FIT-12 reached an inter-rater reliability of r = 89.0-92.7% and a Cohen's kappa of κ = .78- .83. Discrepancies in scoring were resolved by consensus to 100% agreement.

Frankfurt Imitation Test for 18-month-olds (FIT-18) (Goertz, Kolling, Frahsek, & Knopf, under review). The deferred imitation test for 18-months olds (FIT-18), which was administered at the second wave, consists of six actions (one three-step action, four two-step actions, one one-step action). All the action-related objects were commercially available and specifically adapted for this age group. Objects and target actions were used in previous work (Kolling, Goertz, Frahsek, & Knopf, in press). The optimal score that could be reached in FIT-18 is 12.

Again, the six actions were shown successively to each child three times each within a given time interval of about 45 seconds. The retention time was 30 minutes. Then, the children were given the six props successively in the proper order for 30 to 45 seconds each. Playing behavior was videotaped. The four scorers of the FIT-18 reached an inter-rater reliability of r =. 95.7 - 96.4 % and a Cohen's kappa of κ = .91 – .92. Discrepancies in scoring were resolved by consensus to 100% agreement.

Frankfurt Imitation Test for 24-month-olds (FIT-24) (Goertz, Kolling, Frahsek, & Knopf, under review). At the third measurement occasion, the deferred imitation test for 24-months olds (FIT-24) was administered which consisted of eight actions (one six-step action, one five-step action, and six three-step actions). All the objects were commercially available and specially adapted for this age-group. Objects and target actions were used in previous work (Goertz, Kolling, Frahsek & Knopf, subm.). The optimal score that could be reached in FIT-24 is 28 [1].

The actions were presented successively to the child three times each within a given time interval. The retention time was 30 minutes. Following the delay, infants were given the action props successively for a given time interval (30 - 60 seconds) and their play was videotaped. The three scorers of the FIT-24 reached an inter-rater reliability of *r* = 93.9-96.3 % and a Cohen's kappa of *κ* = .88 – .93. Discrepancies in scoring were resolved by consensus to 100% agreement.

Data scoring. All experimental sessions were videotaped by two cameras, one taping the infant and the second the experimenter. Two naïve and independent observers scored target action completion using operational definitions. They were uninformed about the hypotheses of the study and how exactly the experimenter demonstrated the target actions. Target actions were scored as “Yes” or “No” responses.

Distribution and scaling of Frankfurt Imitation Tests. Data for skewness (*FIT - 12* = -0.5, *FIT-18* = 0.2, *FIT-24* = -0.65), and kurtosis (*FIT-12* = -.53 , *FIT-18* = -.34, *FIT-24* = .27), as well as Kolmogorov-Smirnoff normal distribution analysis, showed that the sum scores of the deferred imitation tests can be assumed to be normally distributed. To compare the findings of the different waves, a z-transformation of the performance scores of FIT-12, FIT -18 and FIT -24 was done, so that the test scores were comparable according to mean and standard deviations.

Developmental Test for 6-month to 6-year-olds (ET6-6) (Petermann & Stein, 2000). The developmental test for 6-months to 6-years olds is comprised of six developmental factors, namely motor development, cognitive development, language, social, emotional development and body self-awareness.

Data analysis

In a first step of data analysis, the percentage of missing data was analyzed and missing values were substituted. Secondly, in total group analyses both mean memory performance and stability correlations were calculated for the total sample (variable-centered). Thirdly, developmental groups are separated with cluster analytic procedures with pre-transformed data (person-centered). Fourthly, differences between these developmental groups are reported (multivariate analysis).

Results

Longitudinal missing data analysis

From the total longitudinal sample ($N = 87$), infants with a missing value in deferred imitation testing within any of the measurement occasions ($n = 9$) were excluded for further analysis. The analysis presented above was therefore done with $N = 78$ infants with deferred imitation scores for all three measurement occasions. The percentage of missing values of the Developmental Test for 6-month to 6-year-olds (ET6-6, Petermann & Stein, 2000) was 5.86%. To maximize the power of the statistical analyses, missing data of ET6-6 were substituted with individual mean substitution. Out of the sample of $n = 78$ infants, $n = 3$ infants could not be tested with the Developmental Test for 6-month to 6-year-olds (ET6-6, Petermann, & Stein, 2001), therefore group comparisons are based on $n = 75$ infants.

Total group analysis

Mean memory performance ($n = 78$) increased from $M_{t1} = 4.2$ ($SD_{t1} = 1.5$) over $M_{t2} = 6.9$ ($SD_{t2} = 1.9$) to $M_{t3} = 17.0$ ($SD_{t3} = 3.8$). An analysis of variance for repeated measures showed that the linear development trend is significant ($F = 963$, $df = 1$, $p < .05$).

Stability correlations. Figure 1 depicts the stability correlations of deferred imitation performance.

Insert Figure 1 about here

As can be seen from Figure 1, the stability correlation between the first and second measurement occasion is lower than the correlation between second and third measurement occasion. However, as both stability correlations are relatively low, further person-centered analyses were done in order to get more insight in inter-individual differences of intra-individual change.

Exploratory person-centered analysis

Cluster analysis. First, the deferred imitation data were separately z-standardized (per subject and time point) to eliminate mean and variance differences of the deferred imitation tests. For the analysis of individual consistency, several scaling procedures were considered for analysis, i.e. (1) *absolute* difference scores, |D| (Ghiselli, 1956, 1960), (2) individual consistency scores (Asendorpf, 1990), (3) absolute difference score (Ghiselli, 1960), and (4) relative difference scores (Zedeck, 1971). Then cluster analyses with Ward's (1963) method and the Euclidian distance (both squared and un-squared) measures with these scores were computed. The hierarchical cluster analysis used Ward's (1963) method, since its properties include non-overlapping clusters, distance rather than a correlational measure, and preservation of unequal cluster sizes. The decision for an optimal number of clusters was guided by the following criteria (1) the accepted solution has to be meaningful, (2) reasonably equal sample sizes per cluster should result, and (3) there should have been adequate validity of the cluster solution.

Following these criteria the best cluster solution was established with Ward's method (Euclidean distance) using relative difference scores. As deferred imitation data are very little understood with respect to inter-individual differences in intra-individual change, only two cluster solutions were taken into account to avoid the problem of too difficult to interpret cluster solutions. This hierarchical cluster analysis with the D scores as variables resulted in (1) the most optimal cluster solution with respect to equal cluster group sizes (n_{C1} = 45, n_{C1} = 33), and (2) the cluster solution showed adequate validity with respect to the developmental correlates assessed.

In a next step of the analysis, stability correlations and means of deferred imitation performance were calculated for the two cluster groups, which are reported below.

Cluster group analysis

Growth. Figure 2 shows mean growth patterns for both cluster groups and the total group.

Insert Figure 2 about here

The *first group* (n = 45) increases from M_{t1} = 3.4 (SD_{t1} = 1.4) over M_{t2} = 7.7 (SD_{t2} = 1.9) to M_{t3} = 17.2 (SD_{t3} = 3.5). The *second group* (n = 33) increases from M_{t1} = 5.1 (SD_{t1} = 1.0) over M_{t2} = 5.8 (SD_{t2} = 1.3) to M_{t3} = 17.2 (SD_{t3} = 4.3). An analysis of variance for repeated measures revealed a significant linear trend (F = 958, df = 1, p < .05) and a significant interaction (F = 4.0, df = 1, p < .05).

Standardized deferred imitation scores. As mean growth of the deferred imitation performance is dependent on the increasing total sum scores of the FIT tests, z-scores for both groups are depicted in Figure 3.

Insert Figure 3 about here

Stability. Figure 4 depicts the stability correlations for both groups. The *first group* (n = 45) shows high, significant correlations between t_1 and t_2 (r = .66, $p < 0.01$) and between t_2 and t_3 (r = .54, $p < 0.01$). The *second group* (n = 33) shows a high, significant correlations between t_1 and t_2 (r = .53, $p < 0.01$) and a moderate correlation between t_2 and t_3 (r = .29, ns).

Insert Figure 4 about here

Multivariate analysis. Furthermore, the *first* and *second group* were analyzed for mean differences in the developmental characteristics obtained with the Developmental test for 6-month to 6-year-olds (ET6-6; Petermann & Stein, 2000). Table 1 depicts the differences for the factors of the Developmental test for 6-month to 6-year-olds for the two different groups.

Insert Table 1 about here

Table 1 reveals that the second group has significantly higher scores in body self-awareness, language, and motor development at the second measurement occasion (18 months) than the first group.

Discussion

The present multivariate, longitudinal deferred imitation study aimed to extend previous research by setting the focus on inter-individual differences of intra-individual change through the second year. It was shown that deferred imitation performance of the total sample increased from M_{t1} = 4.2 (SD_{t1} = 1.5) over M_{t2} = 6.9 (SD_{t2} = 1.9) to M_{t3} = 17.0 (SD_{t3} = 3.8) target actions. This result replicates earlier findings (Heimann & Meltzoff, 1996; Nielsen & Dissayanake, 2004) and extends them using a test consisting of more items with different steps and structures. As those studies used deferred imitation tests with fewer items, the present study shows more clearly than before that declarative memory performance is increasing tremendously through the second year. Stability correlations for deferred imitation were lower between 12 and 18 months (r = .17) than between 18 and 24 months (r = .37*). The stability indices are almost identical to those reported by Nielsen and Dissanayake (2004) providing a cross-validation of the former results. Therefore, the conclusion that, from an empirical perspective, inter-individual differences in intra-individual change (developmental dynamics) are high in the first two years of life is reasonable.

In the core analysis of the present study, inter-individual differences of intra-individual change were analyzed longitudinally. A cluster analysis with the transformed deferred imitation data (relative z-difference scores), revealed two cluster groups. The analysis of variance indicated a significant time effect and a significant interaction effect demonstrating that both cluster groups are improving with respect to declarative memory performance. Furthermore, the significant interaction effect shows that both groups are developing differentially. Whereas the first group has a lower memory performance at the first measurement occasion than the second cluster group of approx. 1 standard deviation, at the second measurement occasion the first group outperforms the second group by approx. 1 standard deviation. At the third measurement occasion differences between the groups equal out.

With respect to stability, the first group shows high, significant correlations between all three measurement occasions ($r_{t_1t_2} = .66^*$, $r_{t_2t_3} = .54^*$). The second cluster group demonstrates a high, significant correlation between t_1 and t_2 ($r_{t_1t_2} = .53^*$) and a somewhat lower correlation between t_2 and t_3 ($r_{t_2t_3} = .29$).

The multivariate analysis of the developmental correlates, i.e. the factors of the Developmental Test for 6-month to 6-year-olds (ET6-6, Petermann & Stein, 2001), yielded a significant difference for the total score of the developmental test at time 2, such that infants of group 2 showed a higher mean developmental sum score at the age of 18 months. This result is due to the differences in the developmental factors of language, motor and body self-awareness development. As Kolling et al. (in press) analyzed deferred imitation performance with a person-centered approach (two waves), but could not explain differentiating factors between developmental groups, this result demonstrates how important multi-waved, multivariate longitudinal approaches are for the explanation of inter-individual differences in intra-individual change. The significant difference in the factor of body self-awareness (i.e. representation and knowledge of own body and body of others) indicates that self-development is an important factor for the developmental dynamics found in the 18-months period. Even if the present results do not yet show unambiguously exactly how the development of the self interacts with declarative memory, the results of the present study indicate that the developmental theory of episodic memory (Howe & Courage, 1997; Knopf, Mack & Kressley-Mba, 2005) is worth being considered empirically.

The other significant difference between the groups at time two was for language development. Recent research demonstrated that language development plays an important role in the development of deferred imitation (Heimann et al., 2006; Herbert & Hayne, 2000b; Strid, Tjus, Smith, Meltzoff, & Heimann, 2006). Therefore, it is reasonable to assume that in this age period, declarative memory is related to language development. How exactly language relates to declarative memory needs to be investigated in more detail in future longitudinal research.

Furthermore, the two cluster groups differed with respect to motor development. However, although it is certainly arguable that motor components might be subsidiary in the action-based deferred imitation task, the relatively small difference between groups does not contrast to the importance of language and self-development highlighted above.

The analysis of the present study used a hierarchical clustering algorithm (Ward, Euclidian distances using relative difference scores) to extract subgroups. There are, however, some cautions about cluster analysis. Firstly, most cluster algorithms are relatively simple procedures and therefore are more heuristic than, for example, factor analysis. Secondly, different clustering methods can generate different solutions to the same data, a phenomenon in most areas of cluster analysis research. Thirdly, the strategy of cluster analysis is structure-seeking although its operations are structure imposing, i.e. the cluster algorithms always places objects into groups and these groups may be different when different clustering algorithms are used.

The data found in the present study have to be interpreted with these considerations in mind. Nonetheless, the present study demonstrates that inter-individual differences of intra-individual change can be found in deferred imitation and that these differences are explainable by theoretically important factors, namely language as well as self-development.

Further multivariate, longitudinal studies of deferred imitation are necessary to understand memory development in toddlers. To increase the understanding of continuities of memory development from infancy to childhood further studies might combine non-verbal and verbal memory in line with important developmental correlates to understand memory development and its correlates. If the number of measurement occasions and the sample size is appropriate, the application of newer statistical techniques (e.g. growth curve modelling) for the analysis of the data will provide a better understanding of correlates of intra-individual change of memory. Furthermore, these future studies should take into account exploratory and confirmatory person-centered data analysis approaches to understand the cognitive processes and structures of infant cognition from an individual perspective.

References

Asendorpf, J. B. (1990). The measurement of individual consistency. *Methodika, 4,* 1-13.

Barr, R., & Hayne, H. (1996). The effect of event structure on imitation in infancy: Practice makes perfect? *Infant Behavior & Development, 19*, 253-257.

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior & Development, 19*, 159-170.

Barr, R., & Hayne, H. (2000). Age-related changes in imitation: Implications for memory development. In C. Rovee-Collier, L.Lipsitt, & H. Hayne (Eds.), *Progress in infancy research* (Vol. I, pp. 21-67). Mahwah, NJ: Erlbaum.

Barr, R., & Hayne, H. (2003). It's not what you know, it's who you know: Older siblings facilitate imitation during infancy. *International Journal of Early Years Education, 11,* 7-21.

Bates, J. E., & Novosad, C. (2005). Measurement of individual difference constructs in child development, or taking aim at moving targets. In D. M. Teti (Ed.), *Blackwell handbook of research methods in developmental science* (pp. 103-122). Oxford: Blackwell Publishing.

Bauer, P. J., Heertsgard, L. A., & Wewerka, S. S. (1995). Effects of experience and reminding on long-term recall in infancy: Remembering not to forget. *Journal of Experimental Child Psychology, 59,* 260-298.

Collie, R., & Hayne, H. (1999). Deferred imitation by 6- and 9-month-old infants: More evidence for declarative memory. *Developmental Psychobiology, 35,* 83-90.

Goertz, C., Knopf, M., Kolling, T., Frahsek, S., & Kressley-Mba, R. A. (2006). Entwicklung und Erprobung eines Messinstruments zur Erfassung des deklarativen Gedächtnisses Einjähriger: Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT-12). [The Frankfurt Imitation Test for 12-month-old infants: Testing a series of items for measuring declarative memory in 1-year-olds.]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38,* 88-96.

Goertz, C., Kolling, T., Frahsek, S., Stanisch, A., & Knopf, M. (2007). Assessing declarative memory in 12-month-old infants: A test-retest reliability study based on the deferred imitation task. *European Journal of Developmental Psychology*, in press. Epub ahead of print retrieved May 28, 2007, from http://www.informaworld.com/smpp/content~content=a778058220~jumptype=rss

Goertz, C., Kolling, T., Frahsek, S. & Knopf, M. (2006). Der Frankfurter Imitations-Test für 18 Monate und 24 Monate alte Kinder (FIT18-24). [The Frankfurt Imitation Test for 18-month-old and 24-month-old infants: Testing a series of items for measuring declarative memory in 1.5-2 year-olds.]. *Manuscript submitted for publication.*

Ghiselli, E. E. (1956). Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology, 40,* 374-377.

Ghiselli, E. E. (1960). The prediction of predictability. *Educational and Psychological Measurement, 20,* 3-8.

Gulliksen, H. (1950). *The theory of mental tests*. New York: Wiley.

Hayne, H., Boniface, J., & Barr, R. (2000). The development of declarative memory in human infants: Age-related changes in deferred imitation. *Behavioral Neuroscience, 114,* 77-83.

Hayne, H., McDonald, S., & Barr, R. (1997). Developmental changes in the specificity of memory over the second year of life. *Infant Behavior & Development, 20,* 233-245.

Hartmann, D.P. (2005). Assessing growth in longitudinal investigations: Selected measurement and design issues. In D.M. Teti (Ed.), *Blackwell handbook of research methods in developmental science* (pp. 319-339). Oxford: Blackwell Publishing.

Heimann, M., & Meltzoff, A. N. (1996). Deferred imitation in 9- and 14-month-old infants: A longitudinal study of a Swedish sample. *British Journal of Developmental Psychology, 14,* 55-64.

Heimann, M., Strid, K., Smith, L., Tjus, T., Ulvund, S. E., & Meltzoff, A. N. (2006). Exploring the relation between memory, gestural communication, and the emergence of language in infancy: A longitudinal study. *Infant and Child Development*, *15*, 233-249.

Herbert, J., & Hayne, H. (2000a). Memory retrieval by 18-30-month-olds: Age-related changes in representational flexibility. *Developmental Psychology, 36,* 473-484.

Herbert, J., & Hayne, H. (2000b). The ontogeny of long-term retention during the second year of life. *Developmental Science, 3*, 50-56.

Jones, E. & Herbert, J. (2006). Exploring memory in infancy: Deferred imitation and the development of declarative memory. *Infant and Child Development, 15,* 195 – 205.

Knopf, M., Kraus, U., & Kressley-Mba, R.A. (2006). Relational information processing of novel unrelated actions by infants. *Infant Behavior & Development, 29,* 44-53.

Knopf, M., Mack, W., & Kressley-Mba, R. (2005). Wissen und Erinnern. Zur Genese des episodischen Gedächtnisses bei Säuglingen und präverbalen Kindern. [Knowing and remembering: On episodic memory in infants and preverbal children.]. *Psychologische Rundschau, 56,* 113-122.

Kolling, T., Goertz, C., Frahsek, S., & Knopf, M. (in press). Stability of deferred imitation in 12- to 18-month-old infants: A closer look into developmental dynamics. *European Journal of Developmental Psychology.*

Kressley-Mba, R. A., Lurg, S., & Knopf, M. (2005). Testing for deferred imitation of 2- and 3-step action sequences with 6-month-olds. *Infant Behavior & Development, 28,* 82-86

Kressley R. A., & Knopf, M. (2006). A comparison of between- and within-subjects imitation design. *Infant Behavior & Development, 29*, 564-573.

Mandler, J. M., & McDonough, L. (1995). Long-term recall of event sequences in infancy. *Journal of Experimental Child Psychology, 59,* 457-474.

Mandler, J. M. (2004). *A Synopsis of the foundations of mind: Origins of conceptual thought*. New York: Oxford University Press.

McDonough, L., Mandler, J., McKee, R., & Squire L. (1995). The deferred imitation task as a nonverbal measure of declarative memory. *Proceedings of the National Academy of Science, 92,* 7580-7584.

Meltzoff, A. N. (1985). Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. *Child Development, 56,* 62-72.

Meltzoff, A. N. (1988a). Infant imitation after 1-week delay: Long-term memory for novel actions and multiple stimuli. *Developmental Psychology, 24,* 470-476.

Meltzoff, A. N. (1988b). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development, 59,* 217-225.

Meltzoff, A. N. (1988c). Imitation of televised models by infants. *Child Development, 59,* 1221-1229.

Meltzoff, A. N., & Moore, M. K. (1999). Persons and representation: Why infant imitation is important for theories of human development. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (pp. 9-35). Cambridge: Cambridge University Press.

Meltzoff, A. N. (2002). Imitation as a mechanism of social cognition. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 6-25). Oxford: Blackwell.

Meltzoff, A. N. (2005). Imitation and other minds: The “like me” hypothesis. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science* (Vol. 2, pp. 55-93). Cambridge, MA: MIT Press.

Nielsen, M., & Dissanayake, C. (2004). Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year. *Infant Behavior & Development*, *27,* 342-365.

Petermann, F., & Stein, I. A. (2000). *Entwicklungstest ET 6-6. [Developmental test for 6-month to 6-year-olds.]*. Lisse: Swets & Zeitlinger Publishers.

Piaget, J. (1962). *Play, dreams and imitation in childhood.* New York: Norton. (Original work published 1945)

Spiel, C. (1998). Four methodological approaches to the study of stability and change in development. *Methods of Psychological Research Online, 3,* 7-22.

Strid, K., Tjus, T., Smith, L., Meltzoff, A. N., & Heimann, M. (2006). Infant recall memory and communication predicts later cognitive development. *Infant Behavior & Development*, *29*, 545-553.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40,* 385-398.

Zedeck, S. (1971). Identification of moderator variables by discriminant analysis in a multipredictable group validation model. *Journal of Applied Psychology, 55,* 364-371.

Footnotes

1. Target actions, operational definitions and item statistics for *FIT* -12, *FIT* -18 and *FIT* -24 can be obtained on request from the corresponding author.

2. The target action *tin can* had a baseline performance rate of 20.8%; the target action *toy pig with hat* of 8.3 %; the target action *cup and knife* of 29.2 %; the target action *drum* of 20.8 %; and the target action *mouse* of 4.2 %.

3. The target actions *car, goose, frog, octopus* had a baseline performance rate of 0%. The target action *mouse* had a baseline performance rate of 19.2% and the target action *drum* of 7.7 %.

4. Four scorers rated the video tapes. Scorer 1 and scorer 2 reached an inter-rater reliability of r = 96.1% (Cohen's kappa κ = .92), Scorer 2 and Scorer 3 an inter-rater reliability of r = 96.1% (Cohen's kappa κ = .92) and scorer 3 and scorer 4 an inter-rater reliability of r = 95.8% (Cohen's kappa κ = .91).

Table1 Developmental differences for cluster groups

	t1		t2		t3	
Developmental Factors	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Emotional Development	0.58 (0.19)	0.6 (0.19)	0.72 (0.16)	0.75 (0.15)	0.85 (0.13)	0.87 (0.09)
Motor Development	0.63 (0.14)	0.65 (0.11)	**0.79 (0.12)**	**0.85 (0.15)***	0.77 (0.18)	0.77 (0.17)
Cognitive Development	0.79 (0.13)	0.78 (0.12)	0.54 (0.22)	0.57 (0.24)	0.75 (0.19)	0.81 (0.12)
Language Development	0.63 (0.17)	0.64 (0.19)	**0.62 (0.21)**	**0.75 (0.18)***	0.72 (0.15)	0.74 (0.17)
Social Development	0.63 (0.18)	0.67 (0.18)	0.80 (0.16)	0.84 (0.13)	0.81 (0.18)	0.81 (0.15)
Body Self-awareness	0.19 (0.26)	0.2 (0.31)	**0.53 (0.35)**	**0.70 (0.23)***	0.65 (0.19)	0.64 (0.17)
Total Score	0.60 (0.12)	0.61 (0.10)	**0.67 (0.14)**	**0.75 (0.13)***	0.76 (0.11)	0.77 (0.09)

Note. Standard deviations shown in parentheses

* = $p < .05$

Figure Captions

Figure 1 Stability correlations of deferred imitation performance

Figure 2 Mean Growth for the cluster groups

Figure 3 Z-standardized deferred imitation scores

Figure 4 Stability correlations of deferred imitation performance for groups

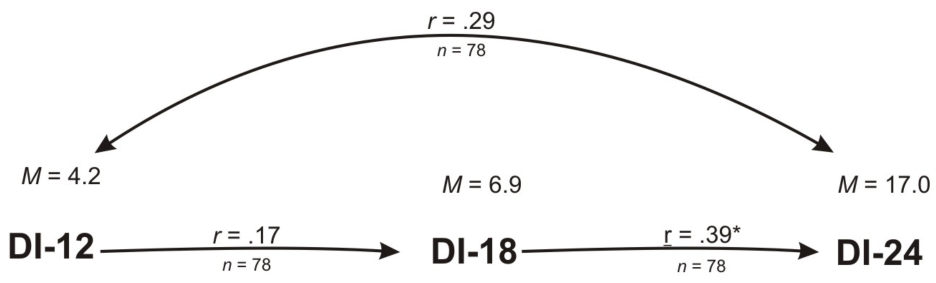


Figure 1

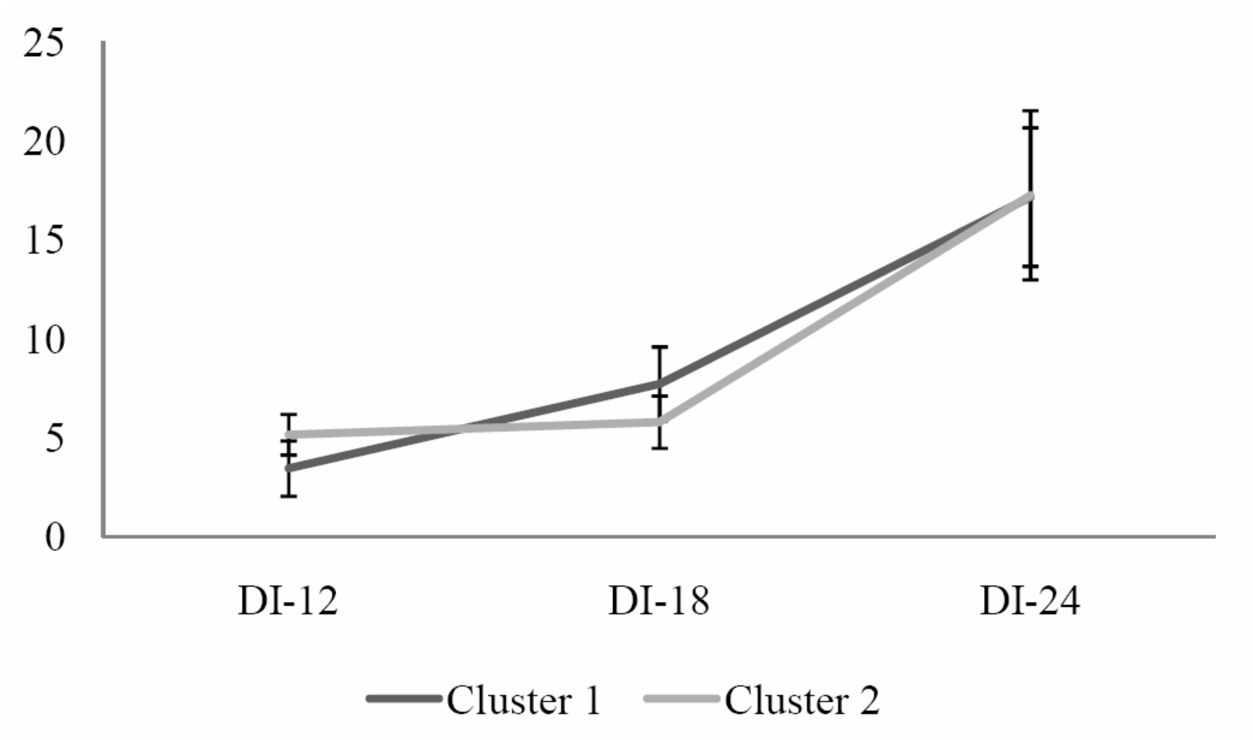


Figure 2

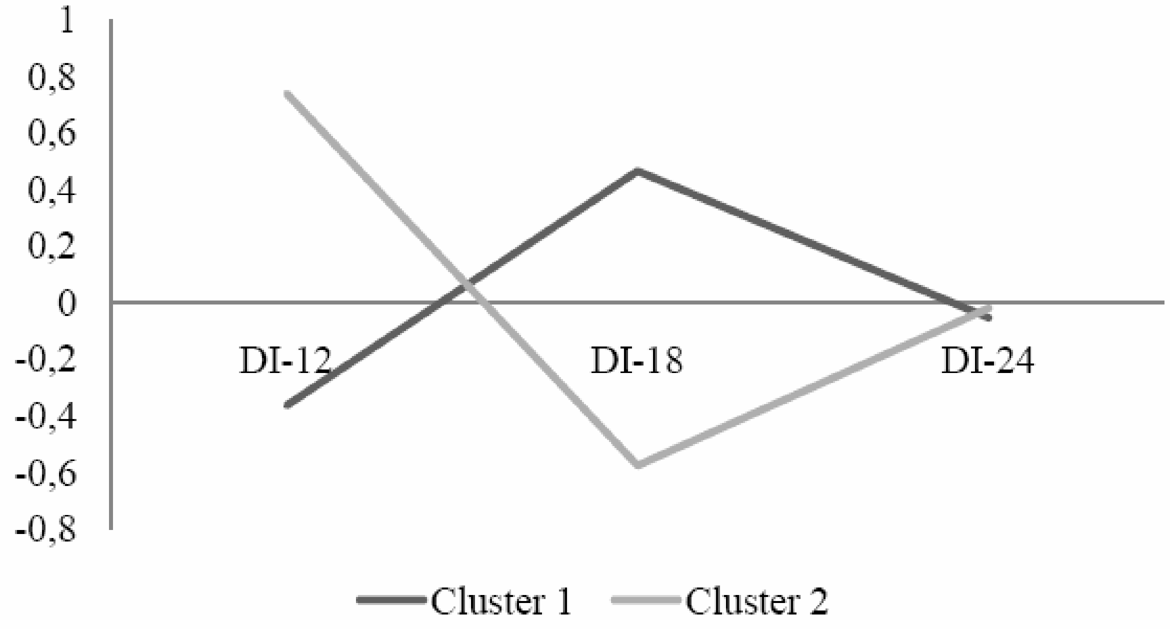


Figure 3

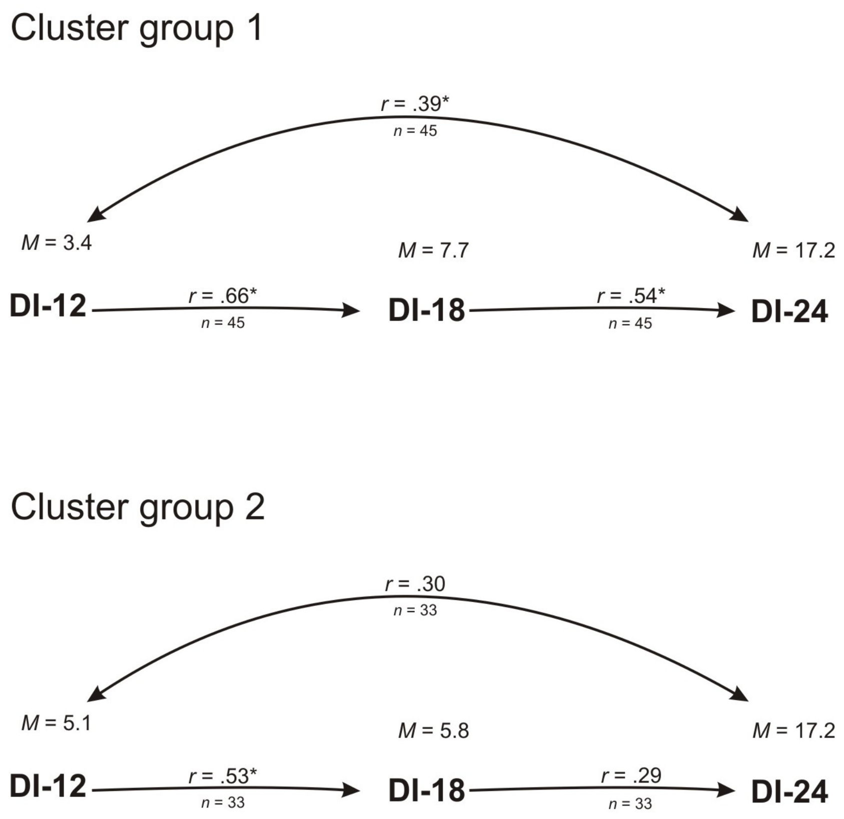

Cluster group 1
r = .39*
n = 45
M = 3.4
M = 7.7
M = 17.2
DI-12
r = .66*
n = 45
DI-18
r = .54*
n = 45
DI-24
Cluster group 2
r = .30
n = 33
M = 5.1
M = 5.8
M = 17.2
DI-12
r = .53*
n = 33
DI-18
r = .29
n = 33
DI-24

Figure 4