1986

STABILITY AND CONTINUITY IN COGNITIVE PERFORMANCE FROM EARLY TO LATE INFANCY

COOPER, RACHEL

http://hdl.handle.net/10026.1/2080

http://dx.doi.org/10.24382/4058
University of Plymouth

STABILITY AND CONTINUITY IN COGNITIVE PERFORMANCE FROM EARLY TO LATE INFANCY

by

RACHEL COOPER

Plymouth Polytechnic

In collaboration with Exeter University

November 1986

Submitted to the Council for National Academic Awards in partial fulfilment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

Stability and continuity in cognitive performance from early to late infancy, by Rachel Cooper

This research is concerned with the stability and continuity of cognitive measures throughout the infancy period.

A review of the literature showed that standardised infant intelligence tests fail to predict subsequent intellectual performance, and one reason for this may be the dominance of motor based items in the tests, and the relative dearth of measures reflecting cognitive ability. Furthermore, it has been suggested that measures from recently developed procedures, which are considered to reflect learning, memory and discriminative ability in infancy may constitute more useful predictive measures of later intellectual functioning. However, the reliabilities and inter-correlations between these measures have not been systematically studied.

At 3 and 5 months of age infants were presented with a variety of tasks, including, visual preferences, visual habituation, concept acquisition, cross-modal integration, operant conditioning and other measures of attention and vocalisation. From these, measures were examined in terms of their across and within sessions reliabilities. Age appropriate tasks were given to the same sample of infants at 9 months of age, and a series of infant intelligence scales and subscales were administered at 18 months. Correlations between the reliable measures across age were examined.

The results indicated that a number of measures at 3 months were reliable, and predictive of performance on tasks at 9 months and 18 months. Specifically, the predictive measures from early infancy tended to reflect aspects of fixation duration, and these were related to measures of attention and manual exploration at 9 months, and verbal ability at 18 months. The predicted measures at 9 months were also related to the 18 month verbal measures, indicating some degree of continuity across these ages. The data are tentatively interpreted in terms of a speed of processing model of intelligence and the implications for models of continuity and the implications for future research are discussed.

DECLARATIONS

While registered for this degree, I have not been a registered candidate for any other CNAA or University award.

This research project was funded by a research studentship grant from the Science and Engineering Research Council during the period from the 1st October 1981 to the 31st December 1984.

None of the material contained herein has been used in any other submission for an academic award.

A course of advanced study has been completed, in partial fulfilment of the requirements for the degree, consisting of guided reading in the area of intellectual development in infancy, under the supervision of Dr Alan Slater; attendence at an advanced course on infant cognitive development (B.Sc. Psychology, Year III special subject) and attendance at four relevant professional conferences.

ACKNOWLEDGEMENTS

My thanks are principally due to Dr Alan Slater for his excellent supervision, patience and support, throughout the completion of this thesis.

I would also like to thank Dr Ian Dennis, particularly for his help with the statistical issues, analyses and with the later part of this thesis.

I am grateful to Dr David Rose, Ms Stephanie Keeble, Dr Morag Donaldson and Dr Mark Terry for their stimulating discussions and guidance on the content of this thesis. Dr Reg Morris and Charlie Attwood gave valuable time in developing the computer programs used throughout the experimental work.

The technicians in the Psychology Department of the Polytechnic, in particular Tony Kirby and Steve Gill, ably provided materials and technical assistance. The Media-Services Department of the Polytechnic library did a marvellous job in providing the stimuli and photographic materials.

I would like to thank Tony Atkin for his help in producing many of the figures and tables presented in this thesis.

I am indebted to the Science and Engineering Research Council for financial support.

Grateful thanks are also due to the mothers and their babies who acted as subjects, and finally, to my parents and friends who have offered so much help and encouragement during the completion of this work.

ABSTRACT

Stability and continuity in cognitive performance from early to late infancy, by Rachel Cooper

This research is concerned with the stability and continuity of cognitive measures throughout the infancy period.

A review of the literature showed that standardised infant intelligence tests fail to predict subsequent intellectual performance, and one reason for this may be the dominance of motor based items in the tests, and the relative dearth of measures reflecting cognitive ability. Furthermore, it has been suggested that measures from recently developed procedures, which are considered to reflect learning, memory and discriminative ability in infancy may constitute more useful predictive measures of later intellectual functioning. However, the reliabilities and inter-correlations between these measures have not been systematically studied.

At 3 and 5 months of age infants were presented with a variety of tasks, including, visual preferences, visual habituation, concept acquisition, cross-modal integration, operant conditioning and other measures of attention and vocalisation. From these, measures were examined in terms of their across and within sessions reliabilities. Age appropriate tasks were given to the same sample of infants at 9 months of age, and a series of infant intelligence scales and subscales were administered at 18 months. Correlations between the reliable measures across age were examined.

The results indicated that a number of measures at 3 months were reliable, and predictive of performance on tasks at 9 months and 18 months. Specifically, the predictive measures from early infancy tended to reflect aspects of fixation duration, and these were related to measures of attention and manual exploration at 9 months, and verbal ability at 18 months. The predicted measures at 9 months were also related to the 18 month verbal measures, indicating some degree of continuity across these ages. The data are tentatively interpreted in terms of a speed of processing model of intelligence and the implications for models of continuity and the implications for future research are discussed.

CONTENTS

CHAPTER 1

PREDICTING SUBSEQUENT DEVELOPMENT FROM INFANT BEHAVIOUR

1.1 Introduction..5

1.2 The history of infant intelligence testing...............7

1.3 A challenge to the failure to predict....................15

1.4 The nature of intelligence in infancy....................22

1.5 The growth of intelligence in infancy....................26

1.6 General discussion.......................................36

CHAPTER 2

THE CHANGING VIEW OF INFANT ABILITIES

2.1 Introduction...37

2.2 New procedures and predictive studies -
A review...40

2.3 Motivational, social and environmental measures.
Predictive studies - A review............................73

2.4 Studies incorporating cognitive and social measures......82

2.5 Conclusion and interpretations of predictive studies.....87

2.6 Conclusions and the aims of the thesis...................95

CHAPTER 3

TASK SELECTION AND PILOT STUDIES

3.1 Introduction...98

3.2 Task selection...99

3.3 Pilot studies..110

3.4 Pilot studies - Visual-based laboratory tasks............113

3.5 Pilot studies -
Operant conditioning and the home based tasks............161

3.6 Pilot studies - Discussion...............................177

CHAPTER 4

DESIGN AND ANALYSES OF THE LONGITUDINAL STUDY

4.1 Introduction..........180

4.2 The longitudinal study..........180

4.3 Subjects..........181

4.4 Analyses..........182

CHAPTER 5

THE TESTING AT 3 MONTHS -
RELIABILITIES AND RELATIONSHIPS BETWEEN MEASURES

5.1 Introduction..........188

5.2 Method..........188

5.3 Procedure..........191

5.4 Measures..........192

5.5 Results and discussions..........193

5.6 Conclusion -
Test-retest and split-half reliabilities..........223

5.7 Across task correlations - 3 month data..........225

CHAPTER 6

THE TESTING AT 5 MONTHS -
RELIABILITIES AND RELATIONSHIPS BETWEEN MEASURES

6.1 Introduction..........234

6.2 Method..........235

6.3 Procedure..........236

6.4 Measures..........238

6.5 Results and discussions..........239

6.6 Results across task correlations 5 months..........266

6.7 Discussion and comparison of 3 and 5 month reliabilities.268

CHAPTER 7

THE TESTING SESSIONS AT 9 MONTHS

7.1 Introduction..273

7.2 The tasks at 9 months...273

7.3 The 9 month testing sessions - Method...........................278

7.4 Results...290

7.5 9 month tasks - Summary and conclusions..........................302

CHAPTER 8

THE 18 MONTH STUDY

8.1 Introduction..304

8.2 The choice of scales at 18 months...............................304

8.3 Design..310

8.4 Measures..311

8.5 Results...312

8.6 18 month data - Summary and conclusions..........................317

CHAPTER 9

PREDICTION OF PERFORMANCE ACROSS AGES -
LONGITUDINAL ANALYSES OF 3, 5,9 AND 18 MONTH DATA

9.1 Introduction..319

9.2 Correlations between 3 and 5 month measures.....................323

9.3 Correlations between 3 and 9 month measures.....................323

9.4 Correlations between 3 and 18 month measures....................345

9.5 Correlations between 5 month measures and 9 month measures..352

9.6 Discussion - 5 month correlations with 3, 9 and 18 months.............353

9.7 Correlations between 9 and 18 month measures....................354

9.8 Conclusions from the longitudinal analyses......................360

CHAPTER 10

DISCUSSION AND CONCLUSIONS

10.1 Introduction....................................363

10.2 Review of thesis....................................364

10.3 Continuity versus discontinuity hypotheses..............368

10.4 The reliability and predictive ability of measures at each testing age....................................370

10.5 The nature of the continuity..............................385

10.6 Models of adult intelligence - Tentative links..........388

10.7 Final conclusions and implications for future research..392

REFERENCES...396

APPENDIX 1

CHAPTER 1

PREDICTING SUBSEQUENT DEVELOPMENT FROM INFANT BEHAVIOUR

1.1 INTRODUCTION

An important goal in developmental research has been to determine whether there are measures in infancy which reliably predict a child's later intellectual ability. It has been well documented that current standardised infant developmental scales, based essentially on motor performance, fail in this respect and there has increasingly been a search for alternative measures. Throughout the mid 1960's methodologies for the study of attention and memory in infancy were developed. These measures have been considered as possible alternatives, as potential predictors, to the conventional tests. The first two chapters of this thesis put this research into context.

Chapter 1 is divided into 6 sections. In the second is presented a brief history of infant testing, and a discussion of the reliability and predictive validity of the standardised infant tests and scales. Infant tests have repeatedly been found not to be predictive of later intelligence, and a currently popular explanation for this is that test items on the scales do not reflect intellectual ability, and that other measures of behaviour in infancy may be more suitable in this respect. This argument is developed in the third section of Chapter 1.

In sections four and five, theoretical issues relating to the question of prediction are considered. These include theories and issues relating to: a) the nature of intelligence in infancy; and b) the nature of intellectual growth throughout infancy, whether this is characterised by quantitative or qualitative change, and the implications of each for the question of prediction.

Chapter 2 describes procedures developed recently which are thought to measure aspects of specifically cognitive functioning in infancy, for example, memory, discrimination and learning ability. It is measures from these tasks taken in infancy which many consider to be potential predictors of cognitive or intellectual functioning in childhood. Also, literature examining the predictive validity of these measures is reviewed. Finally, related psychometric issues are discussed and an overview of the aims of the thesis is presented.

1.2 THE HISTORY OF INFANT INTELLIGENCE TESTING

1.2.1 The Development of Infant Testing

Infant intelligence testing has its roots in the late nineteenth and early twentieth centuries. Initial interest stemmed from a number of influences throughout Europe. For example, in France the need for standardising infant behaviours arose from the suggestion that mentally retarded adults behaved in a similar manner to young children (Binet and Simon, 1905), and in England the work of Darwin and Galton generated interest in the abilities of infants and young children. Darwin (1872) believed that the facial expressions of young children were biological in nature and he studied the emotional development of young children which was characterised by individual differences in rates of development (Goodenough, 1949). Further, in Germany, psychologists, particularly J. Cattell, studying sensation and perception emphasised the importance of the consistency of individual differences.

As a result of such influences, childhood abilities were meticulously studied throughout the early part of this century and the first infant tests were developed. Brooks-Gunn and Weinraub (1983) present an excellent review of the history of infant intelligence testing.

The first real attempt to assess mental development in infancy was made by Kuhlman (1922) who attempted a "downward" extension of the 1916 Stanford-Binet, since the latter did not include items for pre-verbal children. Kuhlman used measures of coordination, speech, imitation and recognition for infant testing. This scale, however, was standardised on few infants and was never adopted widely for general use.

A more popular "downward" extension of the later (1937) Stanford-Binet scale was made by Psyche Cattell (1940). Cattell developed her scale for infants aged between 2 and 30 months and many of the items were drawn from other scales (particularly from Gesell, see below) although she did limit the scale to items easily administered and eliminated many items which were dependent on home learning and those which required gross muscular behaviour.

One of the most important and influential early scale developers was Arnold Gesell who divided infant behaviour into a number of dimensions or "major fields of functional organisation" (Gesell, 1954, p.338). Four such fields were defined:

1. Motor items including balance, movement and coordinated motor activity;

2. Adaptive behaviours defined as prehension and grasping of objects;

3. Language, assessed in infancy by vocal and visual communication (facial gestures and vocalisations);

4. Personal-social items, including reaction to people and environmental events.

The Gesell schedules, which are still used, give behavioural norms for each of these fields for ages at 4 week intervals from birth until 56 weeks, then at 3 month intervals to 24 months and at 6 month intervals up to 42 months. The scales were standardised on a group of infants from middle class families, which ranged in number between 26 and 49 in the first year and between 10 and 18 for subsequent ages. The behavioural norms were determined by the general impression of the examiner concerning the age at which a particular response seemed most characteristic of the infants.

Virtually all scales and tests developed for the infancy period have drawn and adapted items from Gesell's schedules. Other scales include the Griffiths Scale of Mental Development (Griffiths, 1954) which incorporates subscales resembling those of Gesell, and the Bayley Scales of Infant Development (Bayley, 1933, 1969) which consists of separate scales for mental and motor tasks.

1.2.2 Reliability of Infant Tests

Many of the tests mentioned above were originally developed in the early 1930's. Throughout the late 1930's awareness grew of the importance of the reliability of the tests, and as a result virtually all the tests were revised and restandardised. If an infant, for whatever reason, does not perform consistently on a specific measure or test when tested twice over a short period of time, then the test or measure concerned may not be providing any useful information regarding individual or overall development in infancy. Two types of stability were considered:

1. Internal or split-half reliability, relating to the stability of a measure within a testing session;

2. Across session or test-retest reliability, relating to the stability of measures across testing sessions.

Split-half reliabilities for the restandardised tests, where reported, are typically high. Bayley (1970) for example, reports corrected split-half reliability coefficients to average .82 for her first year mental scale. These correlations were lower in the first 3 months (ranging from .51 to .74) but increased between 4 and 36 months (with correlations ranging from .75 to .95).

Test-retest reliabilities are rarely reported within age, since most of the scales were developed by testing infants at one, two or three monthly intervals, rather than say, twice in one week. These short-term test-retest reliabilities are generally not quite as encouraging as the split-half reliabilities, and appear to be affected by the length of the delay between testing sessions. In general, the further apart the two testing sessions the lower the correlations. For purpose of illustration, Table 1:1 shows the correlations between scores taken from the California First Year Mental Scale (Bayley, 1933) from 1 to 15 months and the California preschool scale (Jaffa, 1934) given at 3 monthly intervals to 3 years. Items from both of these scales were used extensively in the development of the Bayley Scales of Infant Development (1969).

TABLE 1:1

Correlation coefficients between age-level standard scores of the California first year mental scale (from 1 to 15 months) and the California preschool scale (at three month intervals up to 3 years). (Both scales were used for the later Bayley - 1969 scales.)

The table is taken from data presented by Bayley (1949).

Average Age (Months)	4,5 & 6	7,8 & 9	10,11 & 12	13,14 & 15	18,21 & 24	27,30 & 36
1,2 &3	.57	.42	.28	.10	-.04	.09
4,5 &6		.72	.52	.50	.23	.10
7,8 &9			.81	.67	.39	.22
10,11 &12				.81	.60	.45
13,14 &15					.70	.54
18,21 &24						.80

It can be seen from the table that the highest correlations emerge between adjacent testing ages (those correlations lying on or close to the major diagonal), and as the period of time between testing sessions increases the magnitude of the correlations decreases. Reviews of these and other similar data have been presented by Thomas (1970) and McCall (1979).

Since the majority of the available tests yield good split-half and moderate test-retest reliabilities, at least over fairly short time periods, it was not long before the question of predictive validity was raised.

1.2.3 Predictive Validity of Infant Tests:

"Diagnosis involves Prognosis" (Gesell and Armatruda, 1941, p.349)

Most of the early infant scales were developed in an attempt to establish normative scales of behavioural development. However, although these early tests were not specifically developed for predictive purposes, their predictive validity became of increasing interest (Brooks-Gunn and Weinraub, 1983). In particular the relationship between these standardised tests in infancy and the subjects' later performance on standardised childhood intelligence tests was examined, and within the normal range the resulting correlations were invariably close to zero. One review of a number of studies which examined the predictive validity of these tests has been produced by Stott and Ball (1965). They concluded:

"Although the tests have been found to be valuable

> aids in the overall appraisal of health and developmental status of babies, ... (the) lack of predictive value has raised serious doubts of their validity as tests of intelligence." (p.24)

Bayley (1955) has also acknowledged that the standardised infant tests do not predict subsequent development:

> "These findings (ie. the low correlations) give little hope to our being able to measure a stable and intellectual factor in the very young..." (p.807)

Clarke (1978) has claimed that, overall, the studies reported, which examine the relationship between performance on infant tasks and childhood levels of intelligence, comply with 2 behavioural laws:-

1. The earlier the measure taken, the lower the long-term reliability and
2. The longer the predictive period the lower the reliability.

Indeed, McCall (1979) reviewed the predictive validities from a number of studies and reported median correlations of .06 between infant tests administered between 1 and 6 months and childhood test scores taken between 8 and 18 years, and of .49 between scores obtained at 19 to 30 months and childhood tests from 8 to 18 years. Thorndike (1940) having collated a number of both short and long-term reliabilities from infants ranging in age from birth to about 5 years, concluded that below 4 months of age there is little chance of finding correlations between any tests given more than one month apart. Above the age of 4 months, correlations do begin to increase, but it is not until the child reaches 4 years of age that correlations between intelligence test scores reach approximately .7 over a one year period. In conclusion, it appeared that prediction of intellectual ability, at least from early infancy, within the normal range was not possible, although the ability to predict does improve in late infancy and early childhood.

The failure of the infant tests and scales (at least for the early infancy period), to predict later intelligence scores gave support to a number of discontinuity hypotheses. For instance: that the concept of intelligence is not applicable to early infancy; that intelligence changes qualitatively from infancy to childhood; and that intellectual growth is characterised by variability rather than stability. These issues are considered in later sections of this chapter, where theories of development are considered in terms of their potential implications for the ability to predict intelligence scores.

The following section is concerned with the currently popular view that infant intelligence tests have failed to predict due to the inappropriateness of the test items they employ.

1.3 A CHALLENGE TO THE FAILURE TO PREDICT

1.3.1 Introduction

While some researchers seem to be satisfied with the failure to predict, others have persisted in the search for predictive measures. Indeed the ability to predict intelligence from infancy has important practical, as well as theoretical implications; at present we are not able to detect infants who will experience minor learning and intellectual deficits. The ability to detect such infants in the early months may lead to intervention programs, which may help to alleviate such difficulties. At the very least, the ability to detect such individuals in infancy allows caregivers to prepare for the problems. Such issues, coupled perhaps with a stubborness on the part of psychologists to accept a null hypothesis, has led to a continuing search for predictive measures.

As a result, the conclusion reached from the evidence cited above, which suggested that prediction of intellectual ability is not possible from early infancy, has been challenged. This challenge has not emerged from the quality of the data, but rather from the way in which mental intelligence in infancy has been conceived. Specifically, it has been suggested that the standardised intelligence tests are dominated by motor based items, and do not contain the types of measure which are more likely to be reflectors of intellectual ability (for example, measures of memory, learning and discriminative ability). Furthermore, due to rapid development in technology and methodology over the past few years, we now have the means to measure such abilities in infancy, and these measures may constitute better predictors of later intellectual behaviour. This argument is qualified below.

1.3.2 The Content of Standardised Infant Intelligence Tests

Zelazo (1979) suggested that items on traditional infant tests fall into 3 major categories; 1) gross and fine motor performance, 2) imitative behaviours, which require fine motor ability, and 3) comprehensive and productive language, which, argues Zelazo, also requires a degree of motor ability, although this latter category of behaviour would seem less likely to be dominated by motor ability.

The general dominance of motor orientated tasks in standard infant intelligence tests is also apparent when the individual measures of the scales are examined. The Bayley scale of infant development is a good example, since it is made up of scales designed to reflect specifically mental and motor ability. The motor scale, as one might expect, is based on measures of gross and fine motor ability. Gross motor abilities include measures of when infants are able to turn over in their crib, when they can sit alone for 30 seconds or more (5 to 8 months), throw a ball (9 to 18 months) and in later infancy ability to stand, walk and jump are examined. Measures of finer motor skills include holding a red ring (up to 3 months) and picking up a small pellet with thumb and forefinger at 7 to 12 months.

The mental scale includes a number of measures comparable to the fine motor skills measured in the motor scale. For example, "fingers holes in peg board" and "successively placing pegs in holes" at 6 to 12 and 13 to 20 months, respectively, and items concerned with the manipulation of a small pellet are in fact the same in both the motor and the mental scales. For early infancy, it is extremely difficult, if not impossible, to distinguish types of behaviours being measured by the Bayley scales that are either wholly motoric or wholly mentalistic in nature, the two essentially being interdependent. However, the nature of test items does become more apparent as the testing age of the infant progresses. At least by the time the infant has reached 18 months of age it is possible to differentiate test items with respect to whether they reflect language-based, or other ability (for further discussion of this see Chapter 8). Since during late infancy, the predictive correlations, from infant intelligence tests to childhood IQ scores, do tend to increase and this may be a reflection of the changing nature with age of the infant test items.

In conclusion it seems that in general items on infant tests, at least those for the early infancy period, are largely dominated by motor abilities. The relationship between mental and motor abilities is considered next.

1.3.3 The Relationship between Mental and Motor Abilities

Historically it has been assumed that motor and mental development are closely related in early childhood. Indeed, Piaget viewed the infant as an active seeker of knowledge and regarded sensorimotor interactions as the foundation of later cognition. This reinforced the already popular belief of an intimate relationship between motor behaviour and cognition; cognitions being a matter of sensori-motor actions performed by the subject. However, questions relating to the specific nature of the relationship between mental and motor development arise. For instance: is mental development a consequence of motor activity; are the two equivalent; and are motor behaviours effective means of measuring infant cognition?

The view that motor ability and cognition are closely related in infancy has been challenged by a number of researchers; Decarie (1969) and Kopp and Shapperman (1973) for instance have reported cases where infants with thalidomide poisoning have demonstrated normal cognitive and intellectual development. Also it has been reported that approximately one in three of those children with symptoms of cerebral palsy in infancy exhibit normal intellectual ability in adulthood (Crothers and Paine, 1957; Holman and Freedhiem, 1959). Segalowitz (1980) reviews a case study reported by Jordan (1972) of a woman with the head of a normal 40 year old and a body comparable to that of a neonate, who was considered intelligent and able to lead discussions.

Kopp (1975) suggests that since learning can occur when motoric development is impaired, then motor behaviours may not be important in terms of language development and other intellectual skills. She goes on to suggest that we should perhaps consider attentional, motivational and object-person interactive factors as possible precursors of later intellectual functioning. Stott and Ball (1965) also concluded that infant intelligence may more usefully be defined in terms of cognitive processes and that careful study of alternative behaviours reflecting memory, conceptal and reasoning skills may provide greater indications of the potential for intellectual growth.

The implication apparent here is that the infant behaviours which are measured in the standardised infant tests are unlikely to predict subsequent intelligence scores in later childhood because, at least in early infancy, the test items are essentially motoric in nature, while childhood intelligence tests are concerned with cognitive or mental abilities. As Zelazo (1979) has pointed out, why should we expect a relationship between the age at which a child can sit, stand or grasp an object, etc., and his intelligence score in childhood? There is little, if any, correlation between motor ability and intelligence in adults and we should therefore not expect measures of the former in infancy to predict the latter in childhood.

1.3.4 Are there cognitive measures that will predict?

Clearly the sorts of behaviours that have been considered as potential predictors from infancy have tended to be limited by existing conceptions of the infant's behavioural repertoire, and by technology, and by what could easily be measured. In view of this it is not suprising that tests developed in the early part of this century are dominated by motor responses since this type of response at the time was, with a few exceptions (see references to Marsdon, 1903 and Valentine, 1913-1914 in the following chapter), considered the only method of assessing ability in the infancy period. However Butterworth (1983) appropriately points out that "An intensive period of research over the last fifteen years has had the effect of changing the stereotype of the young baby from one of "incompetence" to one of "competence"" (p.25).

In conclusion, it is apparent that although performance of young infants on infant scales has been found not to predict their later intelligence level, we should not yet completely abandon the search for predictive measures. It has been pointed out that alternative measures are available which may be more useful for predictive purposes. Indeed, a number of researchers have reported significant correlations between these new measures from infancy and later childhood intelligence scores (these studies are reviewed and discussed in Chapter 2.) Hence, although it would appear that infant scales in general fail to predict later intelligence, it is possible that an alternative set of measures might.

1.4 THE NATURE OF INTELLIGENCE IN INFANCY.

1.4.1 Intelligence as a Unitary or Multi-Dimensional Construct

There has been considerable debate in the literature relating to adult intelligence regarding the structure of intelligence, and this is also relevant to conceptions of the nature of intelligence in infancy. For instance, one can ask whether it is characterised by a single unitary construct 'g', or, as a number of authors have argued, whether intelligence in infancy subdivides into a number of at least semi-independent abilities (Lewis and McGurk, 1972). The suggestion that adult intelligence is subdivided into a number of abilities is by no means new; Guilford (1967) has suggested that at least 120 unique intellectual abilities exist and also that to reduce this number by forming, for instance, just two composites (such as verbal ability and spatial ability, the two major group abilities suggested by Vernon 1969), does a great injustice to the nature of intelligence. However, other researchers point out that when factor analytical procedures are applied to data from adult intelligence tests, although at least some semi-independent factors do emerge, they are usually subsumed by a general all-embracing cognitive factor which Charles Spearman referred to as 'g'. Indeed, some authors have suggested that the failure to find prediction from infancy, is due to there being no measurable 'g' in this period. Garrett (1946) for instance, suggested that the concept of intelligence is not applicable to infancy, and only emerges as the child begins to invest words and symbols with meaning, ie. when a child is able to represent ideas. This implies of course, that no measures from the infancy period are ever likely to be predictive. Alternatively, it is possible that the failure to predict intelligence from infancy may be due to the inappropriateness of measures for reflecting or measuring 'g'.

If infant intelligence is considered to be made up of a number of separate abilities (which are at least to some extent independent of 'g'), it is possible that some, but not all, items of the infant tests predict aspects of later intelligence, ie. individual items or groups of items from the scales may predict later intelligence scores with more accuracy than would a single composite measure. Stott and Ball (1965) factor analysed items from five scales: the Cattell at 3 and 6 months; the Bayley at 6 and 12 months; the Gesell at 6 and 12 months the Merrill-Palmer at 24, 30, 36, 48 and 54 months and the Stanford-Binet at 36, 48 and 60 months. One of the conclusions reached from this massive study was that in infancy the scales were more "diversified" and covered a wider range of abilities than the tests with young children. The infancy tests all contained factors characterised by manual abilities as well as cognitive convergent and divergent production skills, but with relatively little emphasis on memory. The tests from 12 to 24 months, however, contain an increasing number of memory based items with only the Merrill-Palmer containing items measuring manual ability at these ages. The Stanford-Binet at 60 months is particularly narrow in factor content, containing two memory and two cognitive factors only. The implication here is that the measures of manual ability in the early months may be different in kind from the types of measures assessed in later scales, hence reducing the predictive correlations to approximately zero.

Further, a number of researchers have performed item and or factor analyses on various infant intelligence tests and have reported that individual items, groups of items, or factors from the tests (particularly those pertaining to infant vocalisations and verbal ability), produced higher predictive validity than the overall test scores (Richards and Nelson, 1939; Nelson and Richards, 1938; Hunt and Bayley, 1971). Richards and Nelson (1939) also reported that a number of what they describe as measures of alertness, (for example 'splashes in bath', 'regards pellett' and 'quietens to music' from the Gesell schedules) correlated slightly higher with later childhood intelligence than with the global infant test measures. For example, 'splashes in bath' from the 6 month Gesell schedules correlated .47 with the overall six month test score but alone produced an r of .58 with scores of the Stanford-Binet taken at 3 years.

A recent study has been reported by Roe, McClure and Roe (1983), in which they administered the Gesell scales to a sample of middle class white English infants at 3, 5, 7, 9 and 15 months and correlated these scores with scores from the same infants on intelligence scales and subscales at 5 and 12 years. They found that performance on the Gesell scales (particularly from 5 months) correlated more highly with Illinois Test of Psycholinguistic Abilities (ITPA) visual-motor scale at 5 years (with r's ranging from .41 to .62), and the performance subscale of the Wechsler Intelligence Scale for Children - Revised (WISC-R) at 12 years (r's ranging from .43 to .66), than with the verbal subscale from the Stanford-Binet at 3 years (r's ranging from -.17 to .25), the Peabody Picture Vocabulary Test (PPVT) at 5 and 12 years (r's ranging from -.26 to .29 and from -.15 to .54, for the two ages, respectively) or the ITPA auditory-vocal scale at 12 years (r's ranging between -.44 to .28). These findings led Roe, McClure and Roe to suggest that the failure in the past to find prediction of later intelligence from infant tests may have been due to a tendency to regard "intelligence" as a unidimensional quality.

In summary, one implication apparent from these findings is that globally, infant tests are not measuring the same ability (or abilities) as do the childhood tests, but there is some indication that some of the test items in the infant tests may be tapping some of the abilities examined in childhood tests. As discussed in section 1.3, infant scales are dominated by motor-based items, which in them selves may be predictive of nothing, and therefore have the effect of obscuring the few items which may reflect other aspects of ability in infancy and constitute predictive measures.

1.5 THE GROWTH OF INTELLIGENCE IN INFANCY

1.5.1 Qualitative and Quantitative Changes in Intelligence and Implications for Prediction

The popular view of intelligence at the time the first infant developmental scales were produced was one which assumed a predetermined or finite potential with which the infant was endowed from birth (Brooks and Weinraub 1976). This view of intelligence originates from Darwin's theories of biological evolution and was reinforced by the work of Galton, Spearman and more recently Burt. Indeed the notion of a fixed intelligence is to some extent implicit in attempts to predict intelligence from measures taken in infancy; intelligence thereby representing a predetermined characteristic of the infant which develops through steady quantitative increments from birth.

Whether or not an ultimate level of intelligence for any given individual is predetermined is, however, of less concern for the question of prediction than the way in which the level is reached. Intellectual growth may be characterised by continuous steady quantitative increments over time or may reflect a pattern of instability with much fluctuation over time. Neither of these possibilities precludes the notion of a fixed level of intelligence, although they do have different consequences for prediction of intelligence. For instance, if a high degree of concordance were to be found between intelligence test scores between all ages then intellectual growth would seem to constitute a continuous and consistent developmental process. However the failure to find such continuity may indicate that either intelligence is a variable attribute or that intellectual growth is characterised by periods of both stability and fluctuation at various points in time.

There is an alternative theoretical view to that which suggests a quantitatively increasing intelligence, which is that intelligence changes qualitatively, over age. Indeed as Wohlwill (1973) has pointed out, 'development' itself implies change. Binet and Simon (1905) were perhaps among the first to suggest that childhood intelligence differed from adult intelligence not only in degree and quantity, but also in form. Others who have also acknowledged the possibility of cognitive growth being characterised by qualitative change include Gesell (1940), who held the belief that the child's mind does not grow merely by linear increments but is characterised by transformations over age.

If intelligence is considered to be characterised by qualitative changes, then intellectual growth can also still be regarded as either stable or variable in nature, and hence a qualitatively changing intelligence does not necessarily preclude the possibility of prediction. For instance, if intellectual development is considered to change qualitatively as well as being stable, then one might find prediction from one stage to the next. If on the other hand, intelligence is considered to change qualitatively and be variable, then prediction is unlikely to be found. Piaget assumed that intellectual growth is characterised by qualitative change, and Piagetian theory, and its implications for prediction is discussed next.

Piagetian Theory

Piagetian theory (Piaget, 1954) is based on the changing nature of intelligence during development. The first two years or so of life for Piaget are characterised by 'sensori-motor' development. Piaget in his book 'The Origins of Intelligence', traced the course of development through six sequential stages, for each of six specific content areas. As actions are performed within and upon the environment and as the infant organises and adapts these actions, he creates 'invariant functional laws' which eventually become 'schemata' and which ultimately determine future interactions with the environment. Each successive stage is characterised by specific schemata organisation. The six content areas covered by Piaget are; circular reactions, imitation, the understanding of time, space and causality, and the permanence of objects. For the purpose of illustration the development of just one of these, the permanence of objects will be described.

The first two stages of Piaget's sequence (0 to 4 months) are dominated by, initially, the use of reflexes and the onset of adaptation and primary circular reactions. These two stages in the development of the permanent object are characterised by a lack of search for a hidden object; for instance when a moving object disappears behind a screen infants tend to continue to fixate at the point at which they last saw the object. Stage 3 (4 to 8 months) is characterised by 'secondary circular reactions' - the development of 'movements centred on a result produced in the external environment and the sole aim of this action is to maintain this result.' (Piaget, 1953, p157). At this stage infants do begin to adjust to a disappearing object. When a moving object disappears, for example, infants tend to anticipate its reappearance by visually following the direction of travel. Infants at stage 3 will also recover a partially covered object. Stage 4 occurs at 9 to 12 months, and Piaget described this as the stage at which infants show coordination of secondary schemata and apply it to new situations. In object permanence, infants will now search for a completely hidden object, but should the object be seen to be hidden in a new hiding position, infants will tend to continue to search in the original hiding position (the "Stage IV" or "AB" error). Stage 5 (occurring at 12 to 18 months) sees the onset of "tertiary circular reaction" and the "discovery of new means through active experimentation" (Piaget, 1966). At this stage infants tend to have overcome the error made at stage 4, although difficulties may still arise if the complete object is not seen to be hidden at the new hiding place, ie. the infants have difficulties with invisible displacements of objects. The final stage 6, sees the onset of the intervention of new means through mental combination. At this stage all errors in the search task have been overcome.

For Piaget development subsequent to the sensori-motor period is defined as being operational in nature and qualitatively different from sensori-motor development.

Although Piaget was not directly concerned with the prediction of intellectual abilities (Uzgiris, 1983) his work clearly has important implications for this area of enquiry.

Piaget, for instance, viewed progression through his stages as cumulative, each new development being dependent on previous attainments and each new stage emerging from the processes involved in the previous stage. Hence, although behaviours characterising intellectual performance are different at different points in time, Piaget seems to suggest that a precocious child at one stage in development should also show precocity at other stages. In other words, although the nature of intelligence changes over age, the relative performance of an individual child within a group may not. This implies that prediction may be possible even though the nature of intelligence undergoes qualitative changes as development continues.

A number of developmental scales based on Piagetian theory have been developed: Uzgiris and Hunt, 1975; Escalona and Corman (1969), being the most popular and most commonly used. Although these scales provide a useful means of examining whether development through stages is stable, empirical evidence is sparse. In fact many of the studies reported have been concerned with determining the stability of individual performance on different types of task within a given stage, rather than assessing stability or prediction across stages. The Uzgiris-Hunt scales (1975) constitute a particularly useful tool for this purpose since they include separate sets of scales for each of 6 content areas, namely: the development of visual pursuit and the permanence of objects; the development of means for obtaining desired environmental events; the development of vocal and gestural imitation; the development of operational causality; the construction of object relations in space; and the development of schemes for relating to objects. However, difficulties arise when one attempts to use these scales to determine concordance of performance across scales at a particular point in time, since defining the stage reached by a given individual appears to be problematic. Corman and Escalona (1969) for example, reported that infants often would pass items spanning two or more scales and Piaget himself has acknowledged that some of the behaviours characterising a particular stage disappear very gradually and that behaviours of a subsequent stage can occur on occasions when infants are still predominantly in the previous stage. One conclusion to be reached from this is that it seems unlikely that attempts to determine the stage which an individual infant has reached at a particular time would be of use for assessing either within stage consistency of performance, or long term predictive validity. Indeed those few studies which have attempted to assess across-stage stability have reported correlations predominantly close to zero (Kopp, Sigman and Parmelee, 1974; Bates et al., 1979). Literature relating to the development of Piagetian stages has been reviewed by Siegler and Richards (1982). They concluded that it is possible to teach children various aspects of concept understanding at an earlier age than Piagetian theory suggests, and also that tasks which are formally similar (and therefore should fit into a particular stage) are often performed by children at different ages. This seems to suggest that performance on similar tasks within a Piagetian stage may not be consistent, and performance of any individual may be affected by previous experience on that particular (or a related) task.

1.5.2 Models of Continuity

In the preceeding sections it is apparent that conceptions of the way in which intellectual development occurs have different implications for the ability to predict intellectual development. Currently we do not know how intelligence manifests itself and develops throughout the infancy period. Theory in the past has by and large been dependant on data from the available standardised intelligence tests, and the most prominent theories have emphasised discontinuity. However, in view of the fact that there are now alternative methodologies which provide measures which appear to reflect intellectual behaviour more closely (see Chapter 2), as yet we are not able to reject the notion of prediction.

There are a number of ways in which continuity of development, from one developmental period to another, may be caused. For instance, supposing two identical measures of behaviour correlate with each other over time, then it would seem reasonable to suggest that the continuity was due to the two identical measures reflecting the same behaviour and underlying process on each occasion. This type of continuity is referred to as continuity of identical behaviour (Bornstein and Sigman, 1986), and is similar to homotypic continuity as defined by Kagan (1971). Kagan (1971), however, does point out that often a similar or identical behaviour may have completely different causes at two different ages. For instance, a child at 3 months may cry due to hunger, while at 9 months, the same child may cry due to separation anxiety. Although the observed behaviors are identical the underlying causes are different. Clearly, where continuity of identical behaviour is concerned, the provision of a theoretical base relating to the processes causing the continuity is important.

A second type of continuity infers a relationship between two measures which are manifestly different, but which can be theoretically related. Kagan refers to this as heterotypic continuity, and this is similar to the continiuty of underlying process as defined by Bornstein and Sigman (1986). For example, the ability of infants to acquire concepts as determined through the use of visual habituation paradigms (which are described in Chapter 2), have been considered to reflect measures of intellectual functioning in infancy. These same visual fixation measures are hardly likely to be useful indices of intelligence when taken from a population of older children, although measures of verbal proficiency might also reflect infants' ability to acquire concepts. Here, two different behaviours which may be related, can be theoretically linked.

A third continuity model implies a relationship between two completely different measures at two different points in time when there is no theoretical link which obviously links the measures. Bornstein and Sigman (1986) suggest that this reflects continuity of developmental status. Essentially, this implies that a correlation may be found between two completely different behaviours which follow similar developmental trajectories, linked by the maturational status of the child, rather than being linked by a common underlying process.

Each of these models would have different implications for longitudinal predictive studies. Homotypic continuity predicts that the same measure is related at different points in time. Continuity of underlying process, or heterotypic continuity, predicts that different measures which are conceptually linked should correlate over time, while continuity of developmental status predicts that a greater variety of measures will correlate over time. These models of continuity are discussed further in Chapter 10.

1.5.3 Conclusion

The way in which intelligence develops has a number of implications for whether or not prediction is possible. Primarily, there is the issue of whether intelligence develops in a stable or variable manner. If it is stable, then prediction may be possible, whilst a developing intelligence characterised by variability would be less likely to show prediction. Intellectual development may be characterised by quantitative increment or qualitative transformations over age. Should either of these be characterised by variability, then we may not expect to find prediction, and equally should either be characterised by continuity then prediction may be possible.

1.6 GENERAL DISCUSSION

The initial part of this Chapter focused on the history of infant intelligence testing, and discussed how standardised tests have failed to be predictive of later intellectual development. It was suggested that the most popular recent explanation of this was in terms of the nature of the actual measures used in the tests, which tended to reflect aspects of motor ability rather than intellectual ability. Methodologies have been developed which appear to reflect abilities which can be considered more intellectual in nature, and these measures may be better predictors.

In the second part of the Chapter, a number of theoretical issues were discussed, relating to the nature, and the development, of intelligence in infancy. The implications of these theories for the question of prediction were discussed, and it was apparent that much theorising had been influenced and restricted by the failure of standard tests to predict. If these comparatively new methods of assessing infant ability prove to be more useful for prediction (as the review in Chapter 2 suggests), then it may be possible to further our understanding of intelligence and intellectual growth in infancy.

CHAPTER 2

THE CHANGING VIEW OF INFANT ABILITIES.

2.1 INTRODUCTION

In Chapter 1, it was pointed out that traditional scales of infant development fail to predict childhood intelligence scores. It was suggested that one of the many possible reasons for this was due to the tests themselves being based on essentially perceptual and motor-orientated behaviours rather than cognitive or intellectual behaviours. Tests of infant ability are limited by the conception of infants' behavioural repertoire at any given time, and this chapter focuses on the the ways in which our view of abilities in infancy has changed, over the last twenty years or so, and the implications of these changes for predictive research.

Methodologies and procedures recently developed have been used to measure behaviours which are considered to reflect cognitive or mental functioning in infancy. Furthermore, since it is has become apparent that infants are capable of relatively complex tasks involving perceptual discriminative and memorial skills, it is these abilities which are now considered by many researchers to reflect infant 'intelligence'.

As well as support from theoretical viewpoints related to the processes which govern infants' performance on these tasks, a number of studies provide concurrent validation for the suggestion that these measures may constitute better predictors of later intelligence than traditional infant scales. The following three types of studies can be distinguished:

1. Those which report differences in performance between different populations of infants of the same age, where one population is "at risk" for subsequent cognitive impairment. Pre-term infants for example tend to display a greater than average number of intellectual, behavioural and physical disorders (Sameroff and Chandler, 1975);

2. Those studies which report changes in performance over age, since if older infants perform in a different manner on a task compared to younger infants, then one can often assume that the older infants are displaying a more mature pattern of behaviour. Individual differences in performance at a given age may reflect individual differences in cognitive development, such that younger infants performing most like their older counterparts should be the more cognitively advanced;

3. Those studies which examine the relationship between individual infants' performance on a task and measures from subsequent childhood tests.

A number of researchers have assessed the usefulness of measures, taken from these recently developed techniques and procedures, as predictors of later childhood intelligence. The findings of these studies, however, are not clear-cut, and produce correlations ranging between .0 and .7 from the infancy measure to childhood scores. A correlation coefficient in the region of .7 accounts for only approximately half of the common variance between scores. Furthermore, explanations for these inconsistencies are currently difficult to ascertain for a variety of reasons. For example, there is: rarely any consistency in the dependent measures taken from studies adopting similar procedures; a lack of concern relating to the psychometric consistency of measures; and a lack of consistency in the age of subjects at testing sessions, or the type of test used in childhood. These issues are discussed further following the literature review.

In this chapter a number of recently developed procedures, most of which have been considered potentially predictive of intellectual status, are discussed, along with a review of studies which assess concurrent and / or predictive validities of various measures derived from the tasks. Tasks involving looking responses from the infant are reviewed first. These include, visual preferences, novelty preferences, visual habituation and concept acquisition. Then, other types of tasks considered to produce potentially predictive measures are considered, including auditory habituation and vocal measures. This is followed by a review of predictive studies which employ motivational, social and environmental measures, and then those predictive studies which employ combinations of these measures are discussed. Finally, issues are considered relating to why the predictive studies are on the whole inconclusive, and the aims of the current thesis are then presented.

2.2 NEW PROCEDURES AND PREDICTIVE STUDIES - A REVIEW

As mentioned in the previous section, many of the procedures developed to explore cognitive development in infancy, have used looking behaviours as the infants' response. The first of these to be developed was the visual preference testing technique, and it is this which is discussed first.

2.2.1 Visual Preferences

The development of the visual preference testing technique has been credited to Fantz (1958), although use of the technique infact began in the early 1900's (Marsden, 1903; Valentine, 1913-1914). The technique involves simultaneous presentation of two visual stimuli equi-distant from some central point, and merely looking at the infant's eyes to see which of the two stimuli the infant fixates most. This is usually determined by either presenting the stimuli for a fixed period of time and measuring the proportion of time spent fixating each, or alternatively allowing infants to accumulate a fixed amount of fixation time to either one or a combination of both stimuli. If infants consistently look longer at one of two paired stimuli, this indicates that they can at least differentiate between them.

Even newborn infants will show visual preferences for some stimuli over others and one application of this technique has been to measure visual acuity in both newborn and older infants (Fantz, 1965; Fantz, Ordey and Udelf, 1962). Other aspects of perceptual development have also been examined, including form, contrast and depth perception (see review by Fantz, Fagan and Miranda, 1975).

Furthermore, Fantz and Fagan (1975) reported both age-related changes in preferences for particular types of stimuli and differences in the age at which these preferences occur for different populations (e.g. normal versus disadvantaged populations). It is these findings which lead to the suggestion that preference testing may provide potentially predictive measures. Fantz (1958) and Fantz and Nevis (1967) for example, found consistent preference for a bullseye over a set of stripes in infants older than two months of age, but an absence of this preference in younger infants, a finding which suggests that the more 'mature' infants will prefer bullseyes earlier than those who are less mature. Also Fantz Fagan and Miranda (1975) reviewed studies which report age changes in preferences in relation to size and number of elements, such that as infants reach 3 months of age they show an increasing preference for greater number of elements regardless of their size, where prior to this a preference for greater size of elements, regardless of number is apparent. For instance, a number of infants were shown a variety of pairs of stimuli varying in terms of the number (8, 32 and 128) and size (one, half and quarter inch squares) of elements. At 6 weeks of age infants showed a preference for a stimulus made up of

8 one inch squares, with a pattern of 32 half inch squares a close second. At approximately 10 weeks the 32 half inch squares were preferred, followed by 128 quarter inch squares, whilst at 15 to 20 weeks the 128 quarter inch squares was preferred over all.

Fantz, Fagan and Miranda (1975) also report studies in which a selection of home reared infants of university faculty were compared with infants reared in an institution. The results indicated preference curves similar in shape for each group, suggestive of a similar basic developmental change, but often the curves were displaced from each other along the scale with the difference favouring the home-reared infants. A number of differences in performance between Down's Syndrome infants and normal infants were also reported by Fantz et al. (1975). For example the initial development of the bullseye preference occurred later for the Down's syndrome than the normal subjects.

Theories relating to the development of visual preferences have generally been bound by stimulus characteristics. Greenberg (1971) put forward a theory relating stimulus complexity to preference, referring to a perceptual-cognitive interaction. Karmel (1969), on the other hand, offered a theory relating changes in preferences to contour density and the effect of neurological maturation of the visual system.

However, findings such as those mentioned above led Fantz and Nevis (1967) to suggest that the development of visual preferences may reflect more than maturation of perceptual processes;

"There is some basis for a tentative conclusion that development of selective visual attention to configurational variables represents an early stage of basic conceptual-cognitive development- a stage which may be not only predictive of later stages in this development but also may facilitate further development by making a visual exploration of the environment a more effective learning process." (p. 105)

To conclude, the visual preference testing technique may provide a reflection of cognitive functioning but, since no predictive studies using this procedure have been reported, further research is necessary in order to clarify this position.

2.2.2 Novelty Preference Testing

Novelty testing procedures are thought to reflect memory formation and have therefore been considered to be potential predictors. Novelty preference testing is an adaptation of the visual preference testing technique and was first investigated by Fantz (1964), but has also been researched extensively by Fagan (Fagan, 1976 (a,b), 1977, 1979, Fagan and Singer, 1983). The method involves presenting infants with one stimulus, or more commonly a pair of identical stimuli, during usually one familiarisation trial, after which a further two stimuli are presented, one being familiar, ie. the same as those shown initially, the other being novel. Stimulus presentations either last for a fixed time period (which may start when the infant first fixates the stimulus), or until infants accumulate a certain duration of fixation time towards the stimuli. The test presentation, ie.

the novel stimulus paired with the familiar one, is usually presented over two trials with the left - right position of each stimulus alternated across trials. This controls for effects of position preferences on the part of the infant, or biases on the part of the observer (since usually observers are located on one side of the infant, and hence detection of fixation to each stimulus position may differ in accuracy).

When given a lengthy familiarisation period, infants typically show a visual preference for the novel stimulus, although a number of researchers have reported familiarity preferences to be the norm especially for very young infants, (this finding is discussed in the following section on Visual Habituation.) The tendency of infants to fixate a novel stimulus, as opposed to one to which they had previously been exposed, demonstrates recognition memory in infancy, as well as the ability to discriminate between stimuli.

Fagan (1970, 1972) reported that different types of stimuli produced variations in the age at which novelty preferences were first exhibited. For instance, multidimensionally varying stimuli can be discriminated by infants as young as two and a half months; stimuli which are made up of the same type and number of elements, but which vary in terms of their pattern or arrangement, fail to produce novelty preferences in infants below the age of four months; while discriminations between photographs of faces is not demonstrable until between five and six months. This sequence of development, of the ability to discriminate

different types of stimuli across age, is given further support by Fagan (1974) who found that the study time of the familiar stimuli needed to elicit novelty preferences from five and a half month old infants followed a similar pattern.
Multi-dimensionally varying stimuli produced a novelty preference after an average of 3.4 seconds of fixation to the familiar stimuli, while the change in pattern arrangement showed similar novelty preferences after 17.1 seconds fixation to the familiar stimuli, and 22.4 seconds were required for photographs of faces, and no novelty preferences emerged for a line drawing of a face even after a 35 second exposure to the familiar stimuli.

These results together suggest that this technique can be considered to reflect some form of cognitive processing in infancy. Furthermore, a number of studies, comparing normal samples with ones which can be considered disadvantaged in some way, provide concurrent validity in support of this suggestion.

Miranda and Fantz (1974) compared novelty preferences of 16 Down's Syndrome infants with 16 normal infants across three ages, 8 to 16 weeks, 17 to 29 weeks, and 30 to 40 weeks. Three types of problem were used. The first included pattern and colour as the independent novel features, such that the familiarised stimulus was paired with, both a different pattern with the same colour, and with one of the same pattern of a different colour. In this comparison then, memory for two stimulus features was tested. The second set were photographic prints of two faces, one of a baby, the other of a woman. The third set of stimuli each consisted of 25 small squares arranged in a circular configuration, and the other a checkerboard. Familiarisation periods lasted for 60 seconds, followed by two 10 second testing periods, over which the left-right position of each stimulus was counterbalanced.

Results of this study indicated that the ability to solve each problem differed across ages and across groups. For the first stimulus set preference for novelty for the normal infants began at 8 to 16 weeks, while for the Down's infants the preference did not show until between 17 and 29 weeks. Preference for novelty for the remaining two stimulus sets began between 17 and 29 weeks for normal infants, but the Down's babies showed a later novelty preference for stimulus set two (at between 30 and 40 weeks) and failed to show any novelty preference for stimulus set three. Although the wide age ranges within testing sessions make it difficult to assess the precise time lag of the Down's subjects, Miranda and Fantz (1974) suggested that the lag was at least 2 months. At the very least, these data do imply a basic developmental process which is delayed in Down's Syndrome babies, a conclusion also reached by Cohen (1981) having also reported significant novelty preferences occurring earlier for normal infants than for infants suffering from Down's Syndrome.

Jacobson, Fein, Jacobson, Schwartz and Dowler (1985) point out that adverse neonatal outcomes have been associated with exposure to environmental toxins, in particular with polychlorinated biphenyls (PCBs). PCBs are a type of synthetic hydrocarbon and are used in many industrial products, such as plastics and copy paper. They also accumulate in air, water, soils and human bodies, since they are difficult to excrete. Some of the neonatal deficits associated with exposure to such toxins appear to be transient, and some infants have continued exposure via breast milk, although as yet we know little of the conditions under which deficits occur, or whether any of the effects appear only later in life.

Jacobson et al. compared performance on novelty preference tests of infants who had been exposed to PCBs and non-exposed control group of infants, at 7 months of age. Infants with mothers who had eaten fish from Lake Michigan, known to be contaminated with PCBs, constituted the 'exposed' group and two measures of exposure were obtained. One was the cord serum PCB level which constituted an intrauterine exposure level; the other was a post-natal measure of exposure, measured from the mother's milk after leaving hospital. Results showed that preference for novelty (measured as the percentage preference for the novel stimuli averaged over 3 novelty preference tasks) decreased significantly with greater levels of PCB exposure although this effect was only found for the intrauterine measure of exposure and was not evident for the post natal measure of exposure. Nevertheless, since the effect of PCB exposure on the novelty preference task was not predicted by smaller birth size nor by the other deficits associated with PCB exposure, the infants performing 'poorly' on the preference task could not be detected by such measures at birth. Hence novelty preference testing may be potentially useful for assessing damage from exposure to toxins.

These studies, then, do suggest that novelty preferences can change over age (although this is affected by stimulus characteristics), and can be used to differentiate between normal and disadvantaged samples. Furthermore, Fagan and McGrath (1981) adopted the novelty preference testing technique and reported significant correlations between novelty preferences from a total of 93 infants at 5 or 7 months, and subsequent childhood intelligence scores. Subjects had been given either pairings of photographs of different faces and / or pairs of abstract patterns at 5 or 7 months, and a series of vocabulary and language subscales (including the Peabody Picture Vocabulary Test (PPVT), the Stanford-Binet picture vocabulary section, and the vocabulary sub-tests of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) and the Wechsler Intelligence Scale for Children - Revised (WISC-R) at 3.8, 6.5 or 7.5 years. Each child in this study was given a recognition memory score constituting the overall mean of their novelty preferences in infancy, and a vocabulary score for the childhood measures defined as the mean of the standard scores obtained on vocabulary tests. Results showed that the novelty preference measure correlated significantly with the vocabulary score of infants at the younger (3.8 years) and the older (6 to 7 years) age groups. The correlations were .37 and .57 respectively.

Rose and Wallace (1985) followed-up a group of pre-term infants, for whom novelty preference data had been collected when the infants were six months of age. Premature infants were defined as those having a gestational age of 37 weeks or less and also a birth weight of less than 2,500 grams. Other measures taken from these infants included the Bayley scales (both mental and psychomotor indexes) at 6, 12 and 24 months, the Stanford-Binet at 34 and 40 months and the Wechsler Intelligence Scale for Children - Revised (WISC-R) including scores for verbal and performance IQ as well as the full-scale, at 6 years. The results indicated that visual novelty scores correlated significantly with all of the cognitive measures beginning with those administered at 24 months, although of the 24 month measures, only the mental developmental index of the Bayley scales correlated significantly with the novelty preference scores. The Stanford-Binet at 34 and 40 months correlated with the 6 month novelty measures with r's of .66 and .45, respectively, while the 6 month novelty scores correlated .56, .52 and .56, with the WISC-R IQ (full scale), verbal IQ and performance IQ scales, respectively, at 6 years. Since none of the correlations with the Bayley psychomotor test were significant, the authors concluded that recognition memory, as measured by the novelty preference testing technique, is specifically related to later measures of cognitive abilities.

Results such as those mentioned above led Fagan (1984) to claim that the ability of infants to detect a similarity between, and to abstract common features from a stimulus set in early infancy, is an intellectual process and mirrors the type of processes required to perform tasks in later infancy and early childhood intelligence tests. For instance, in vocabulary and comprehension tests infants and children are asked to detect and point to particular objects and in order to be able to do this the child must be aware of particular attributes common to different instances of an object, and the features distinguishing a particular object from other objects. This suggests that to perform successfully on these tasks, children need to be able to detect invariants and abstractions of features relating to objects. Novelty testing shows that infants can do this at a very early age.

2.2.3 Habituation Procedures

The visual preference and novelty preference testing techniques were developed more or less concurrently with other similar methods allowing measurement of performance considered to reflect aspects of cognitive functioning. Perhaps the most popular and prolific of these, in terms of the extent to which they have encouraged research, have been habituation procedures.

Habituation procedures involve the repeated presentation of usually a single stimulus, to which, over trials, infants typically show a decrement in fixation time. Following this, a novel stimulus is presented, to which the infant often 'recovers' in terms of showing an increase in fixation time towards this stimulus. These changes in fixation time seem to be apparent in infants from birth. Slater, Morison and Rose (1982) found evidence to suggest that even newborn infants would habituate visual attention to a familiar stimulus and subsequently recover to a novel one.

Most theories of the basic habituation paradigm (for a review see Lewis and Baldini, 1979) accept Sokolov's (1963) basic principle, that habituation is accompanied, or caused, by the formation of an internal representation of the stimulus by the infant, which is developed over trials. When the infant is presented with a novel stimulus it is compared with the current engram and because there is a mismatch, a longer fixation time is exhibited to the novel stimulus. Measures from this task are therefore considered to reflect aspects of cognitive functioning in infancy: the extent and rate of decline in responding to the familiar stimulus have been considered to reflect the speed at which individual infants are able to form the engram or internal representation of the stimulus, and it has rarely been disputed that some form of memory ability is necessary for infants to habituate to a familiar stimulus and subsequently recover to a novel one (Werner and Perlmutter, 1979).

Two major habituation procedures have been developed. In the first, the infant is presented with a fixed number of 'familiarisation' trials, each lasting for a specific length of time, followed by novel trials of the same duration. This procedure is referred to as the fixed trial procedure. Trial periods may begin as soon as the stimulus is presented to the infant, or when the infant is first judged to have fixated the stimulus (the latter of these two possibilities is favourable, since in the former the infants may not fixate the stimulus at all on some trials). The second procedure allows infants to reach a prespecified criterion of fixation decrement relative to the initial fixation times. This is referred to as the infant control procedure and was developed as an alternative to the fixed trial procedure by Horowitz, Padan, Bhana and Self (1972). The most commonly used decrement is for the duration of three consecutive looks, from trial four on, to total half or less of the total fixation time in the first three trials. Following the familiarisation period a novel stimulus is presented. Usually, in both habituation procedures, single stimuli are presented on test (novel) trials although some researchers do use a novel stimulus paired with the familiar in test trials, for a fixed time period, or accumulated fixation duration.

Support for the suggestion that measures of habituation do reflect measures of cognitive functioning comes from studies which show that age changes occur in rates of habituation; older infants give shorter fixations and require less time to habituate (Lewis, Goldberg and Rausch, 1968; Lewis Goldberg and Campbell, 1969; McCall, Hogarty, Hamilton and Vincent, 1973).

Further a number of studies provide concurrent validity for the suggestion that habituation paradigms do in fact tap individual differences in cognitive functioning. Several studies have compared performance on habituation tasks between full-term and pre-term infants. Werner and Siqueland (1978) examined performance of pre-term infants with a mean gestational age of 35.03 weeks, and aged on average 6.5 weeks at the time of testing. They adopted an infant control procedure of habituation, using high-amplitude sucking rate as their dependent measure: first a base-line measure of sucking rate on a blind teat was established for each infant, and then a relative decrease from this sucking rate to a pre-set criterion was required to a repeated visually presented stimulus, followed by subsequent increase to a novel stimulus. A number of significant correlations were reported between the recovery measure (ie. response to novelty) and measures of mortality risk ($r= .62$) and the number of medical complications ($r= .57$) and the gestational age ($r= .50$) of the infants.

It seems then that habituation measures do correlate with aspects of risk and number of "aberrations" in infancy. Further corroborative evidence has been reported by Schexnider, Bell, Shebilske and Quinn (1981) who examined the effect of the number of minor physical anomalies, indicating disorders of fetal development, suffered by individual male infants on measures of habituation. The anomalies they assessed included aberrations of: the hands such as curved fifth finger; head, including soft, pliable, assymetrical or malformed ears, high steeped palate, outsized head circumference; and feet, such as webbed toes or the third toe being longer than the second and / or large toe. Schexnider et al. claim that these anomalies may affect attention in infants, especially since children with high numbers of them are reported often to be hyperactive. Two groups of 12 month old infants were identified: a high anomaly group showing 5 or more physical anomalies; and a low anomaly group consisting of infants with 3 or less. All infants were habituated, by means of fixed trial procedures, to two sets of stimuli, one made up of 3-Dimensional dolls, the other consisting of curved and straight lines. The first of these failed to differentiate between groups, but for the second set of stimuli a difference in the recovery measure between groups was reported, in terms of both the total fixation time and the longest single fixation to stimuli. The high anomaly group showed significantly less recovery than the low anomaly group, and further analysis revealed that the low anomaly group in fact continued to show a decrease in fixation to the novel stimuli.

Further a number of studies have been reported which examine the relationship between measures of habituation in infancy and measures of intelligence in later childhood. Overall these produce a confusing picture of the value of these measures for prediction of later ability.

A series of studies by Miller and co-workers (Miller, Ryan, Short, Reis, McGuire, and Culler, 1977; Miller, Ryan, Aberger, McGuire, Short and Kenny, 1979; Miller, Spiridigliozzi, Ryan, Callan and McLaughlin, 1980) reported relationships between habituation measures taken at 2, 3 and 4 months, and at 27, 39 and 51 months; scores from a group of object concept and operational causality tasks, derived from the Uzgiris and Hunt (1975) Piagetian-based scales at 15 months; and measures of language comprehension and production at 27 months and at 39 months, as well as measures of memory and performance on a paired associates task at 39 months. The sample of infants at the outset of this study numbered 48, although only 24 children completed all of the testing sessions up to 51 months. The habituation tasks at each age were presented in the form of a fixed trial procedure with 8 presentations of the familiar stimuli (Miller and co-workers were only concerned with habituation rates and not with recovery measures since no novel stimuli were presented at the end of the habituation sequences). Two indices of habituation ratio were examined; the first was in terms of first fixation time per trial, the second in terms of total fixation time per trial. Both ratios were calculated in the same way, using the formula

$$((S7 + S8)-(S1 + S2)) / (S1 + S2)$$

where Sx = either the duration of the first fixation in trial x or the total fixation duration on trial x for the two ratios respectively. Miller et al. (1977) reported that the measure of the ratio of the the first fixation duration taken at 2, 3 or 4 months, correlated significantly (r= -.49) with scores from a group of object concept and operational causality tasks, derived from the Uzgiris and Hunt (1975) Piagetian-based scales, at 15 months. Subjects with greater habituation ratios in terms of first fixation duration earned higher scores on the cognitive tasks. However, ratios taken from the total fixation time per trial were not correlated with the 15 month scores. The pattern of intercorrelations reported by Miller et al. (1979) and Miller et al. (1980), show that, overall, neither measure of habituation ratio produced a clear pattern of predictions over this period. However, a number of significant correlations did emerge. For example, total fixation ratio of habituation taken at 2, 3 or 4 months correlated significantly with performance on a memory task and a paired-associate learning task at 39 months (r's of -.40 and -.26 respectively), whereas first fixation duration ratio of habituation also at 2, 3 or 4 months, correlated with language comprehension scores (r= -.39) and memory (r= .42) at 39 months (note that this correlation was the only one to go in the unexpected direction - this is somewhat difficult to interpret and may be spurious). None of the fixation measures at 2, 3 or 4 months predicted any of the cognitive measures at 51 months. The studies reported by Miller and co. workers produces a confusing pattern of results, with no clear indication that any particular measure is a reliable predictor over all age periods.

Lewis and Brooks-Gunn (1981) reported a number of significant relationships between measures from habituation procedures and subsequent novelty scores. Lewis and Brooks-Gunn followed up 2 separate groups of infants, with 22 infants in the first sample, and 57 in the second. All infants had been given a fixed trial habituation task at 3 months. They used a measure of 'duration of fixation' (total fixation time divided by number of fixations) to calculate both habituation and recovery to novelty scores. The habituation measure was calculated using the formula:

(Trial 1 - Trial 6) / Trial 1

and the novelty score by:

(Trial 6 - Trial 7) / Trial 6

Both of these scores were related to the Mental Development Index of the Bayley Scales at 24 months: for the first group (N=22), r's were reported of .61 and .52 for the habituation and recovery measures, respectively, although for the second group (N=57) only the recovery measure was significant (r= .40). The correlations indicate that the most effective habituators, and those infants who showed most recovery, had the higher intelligence scores as children. It is not clear why only the recovery measure was significant for the second group, but this may suggest that the significant relationship between habituation rates and the Bayley found for group 1 is suspect.

Ruddy and Bornstein (1982), obtained measures from 4 month old infants on a habituation task. These infants were also administered the Bayley scales, and tests of vocabulary at 12 months. In the habituation task a fixed trial procedure was employed, during which infants were presented with fifteen, 10 second presentations of a stimulus made up of 5 vertical red bars, followed by 4 trials of faces. The measures of habituation derived were based on the infants' first fixation duration on each trial, and two measures were calculated. The first was a measure of amount of habituation and was calculated as the percentage decline from the average of the first fixations in the first two trials to that of the final two trials; the second was a measure of habituation rate, measured as the first pair of consecutive trials for which the average first look declined to 80 percent or less of the average for the first two trials, without a subsequent rebound to above 90 percent. The longitudinal analyses revealed that the faster habituators and those who habituated most at 4 months, had the higher Bayley scores (r= .49 and .46, respectively), and greater speaking vocabularies (r= .39 and .52) at 12 months.

Fourteen of the infants who had taken part in this study, were administered the Wechsler scales at 4 years, and Bornstein (1985) reports a correlation of .54 between this measure and the habituation rates at 4 months.

In summary then, it seems that habituation procedures may provide measures which are useful for prediction,and this includes habituation rates and recovery measures. However the pattern is rather jumbled and although, as mentioned above, habituation rates are considered to reflect speed of engram formation or something similar, it is not yet clear which measures are likely to constitute the better predictors, nor what the measures are reflecting. For instance, the habituation procedures are similar to novelty preference testing procedures, in that a preference for novelty is sought in both. One question then, relates to whether the habituation procedures offer any different information about processing in infancy than does novelty preference testing techniques.

In the habituation procedure, emphasis is placed on a decline in fixation to a stimulus over trials with subsequent recovery to novelty, whilst the novelty preference technique seeks novelty preferences usually after a brief exposure to the familiar stimulus. Indeed preference for novelty in 5 month old infants (Cornell 1979) and in 6 and 9 month olds (Rose, 1981) has been reported following just 5 and/or 10 seconds familiarisation time. Further, since memorial ability in these tasks is defined in terms of the novelty preferences, and since novelty preferences have been reported after very brief familiarisation periods (particularly when compared to the amount of familiarisation time typically given in visual habituation procedures), the meaning of the decline in responding in the habituation procedures is somewhat unclear. This issue however, has in part been resolved from interest in a number of studies which report familiarity preferences in infants under 2 months of age (Friedman, Nagy and Carpenter, 1970; Friedman and Carpenter, 1971; Weizmann, Cohen and Pratt, 1971; Friedman, 1972; Wetherford and Cohen, 1973). The finding of significant preferences for familiar stimuli led Hunt (1970) to suggest that as infants develop the ability to recognise novelty, they first take pleasure in recognition itself, in terms of recognising familiarity. A study by Rose, Gottfried, Melloy-Carminar and Bridger (1982) found that shifts in preference were affected by both the age at which infants were tested and familiarisation time, and that familiarity preferences were not always characteristic of very young infants, especially if they were allowed lengthy familiarisation periods. Three and a half and six and a half month old infants were used for this study. The younger infants were allowed 5, 10, 15, 20 or 30 seconds of familiarisation, whilst the older group were allowed 5, 10 or 15 seconds fixation to the familiar stimuli. All the stimuli were three-dimensional white plastic shapes. Results indicated a significant preference for the familiar stimulus after 5 seconds exposure for both age groups, but this shifted to a novelty preference after 30 and 15 seconds exposure for the younger and older groups, respectively. Rose et al. concluded that at the start of processing infants prefer familiarity, but as processing gets more advanced, novelty is preferred.

Furthermore, Rose and Slater (1983) examined the possibility that memories after brief exposures have different characteristics than those formed after longer familiarisation periods. In part of one experiment, 3 and 5 month old infants were given both a novelty preference task and an habituation task. In the novelty preference task, infants were allowed 10 and 20 seconds familiarisation time for the older and younger groups respectively, and in the habituation task an infant control procedure was adopted although paired stimuli were presented in the post-criterion trials. In both conditions the familiar stimuli consisted of a simple yellow shape, and the novel stimuli a different yellow shape paired with the original. Results showed that when infants had been fully habituated, a significant preference for the novel shape emerged, although no significant preference emerged for the brief exposure condition.

In a further experiment reported by Rose and Slater (1983) infants were again given both a novelty preference and an habituation task, although in all tasks a 30 second delay was included prior to the test trials. The habituation procedure still produced novelty preferences after this delay, but no significant preferences emerged from the brief exposure condition. These results suggest that the memory trace developed in brief exposures is less robust and more susceptible to decay than that developed in full habituation procedures.

Measures of decline and of novelty preference then, may well reflect measures of some form of intellectual functioning. Individual differences in rates of habituation may reflect processing speeds, although one can never assume that infants who exhibit similar measures of decline are equivalent in terms of the extent to which the stimulus has been processed. Similarly, since novelty preferences of equivalent magnitude can be obtained from novelty preference tests following brief exposure and habituation procedures (Rose and Slater, 1983), one cannot assume that the size of this measure reflects the extent of processing achieved in the familiarisation phase. So although there is abundant evidence to suppose that measures from novelty preference and habituation testing procedures do reflect some aspect of cognitive functioning, as yet it is neither clear which measures should constitute the most appropriate predictors, nor what individual measures are reflecting.

2.2.4 Auditory Habituation

As reviewed above measures from novelty preference and habituation techniques in the visual modality, appear to be potentially reflective of later intellectual ability. The relationship between measures of habituation and recovery in the auditory modality and subsequent intelligence test scores has been investigated by O'Conner, Cohen and Parmelee (1984). Preterm infants were defined as those who have a gestational age of 37 weeks or less and also a birth weight of less than 2,500 grams. At 4 months corrected age, pre- and full-term infants were presented with a sequence of pure tones alternating between 400 and 1000 Hz presented continuously for 2 seconds, each for a period of 12 seconds, and the inter-stimulus interval was also 12 seconds. Infants were presented with this stimulus sequence until they reached a criterion of habituation measured by cardiac de celeration: criterion was met when no apparent decceleration to the stimulus was recorded for 3 successive trials. Presentation of 3 novel tones followed and the change in heart rate (in terms of deceleration) to the first novel tone was taken as the recovery measure (no reason is given as to why response to the first novel tone only was used as the recovery measure). The Stanford-Binet was administered to the subjects when they reached 5 years of age. No difference was found between full- and pre-term infants of equivalent conceptual age on the following measures: heart rate change in response to novelty in infancy and the Stanford-Binet scores in childhood. The predictive analyses were therefore done with the combined sample. These revealed that the magnitude of cardiac change in response to novelty correlated significantly (r= .6) with the 5 year intelligence score, such that those infants who showed the greater cardiac deceleration to novelty were also those infants who scored highest on the intelligence measure. These results suggest that individual differences in perceptual memory within the auditory modality may be predictive of later intellectual performance.

2.2.5 Concept Acquisition

Habituation procedures have been used to investigate concept acquisition in infancy. Infants are presented with a series of concept exemplars, or stimuli which can be categorised on one or more dimensions, and a decrement in fixation over time typically occurs. The test phase involves presentation of a novel stimulus which does not represent a concept exemplar, to which the infant tends to exhibit increased fixation. Both the fixed trial and infant control procedures, described under Visual Habituation can been used in concept acquisition tasks.

Using this technique and also adaptations of the novelty preference testing procedure, infants have been reported to be able to distinguish between male and female faces by 28 weeks of age (Cornell, 1974; Fagan, 1976); distinguish a specific female face from female faces in general, regardless of orientation, by 30 weeks (Cohen and Strauss, 1979); and to abstract a prototypical representation of a face by means of averaging a number of presented feature characteristics at approximately 43 weeks of age (Strauss, 1979).

The theories related to how infants are able to acquire concepts are similar to those relating to habituation and visual memory (see previous sections), and hence measures from this type of task (for example, the number of category exemplars required to acquire a concept reflected by trials to criterion in the habituation procedures, and the extent of novelty preferences in novelty preference and habituation procedures), have been considered to reflect aspects of cognitive functioning in infancy.

Furthermore, there is an indication that age changes occur in performance on this type of task. Cohen and Strauss (1979) report that 30 week old infants respond in a way which is compatible with the suggestion that they are capable of abstracting conceptual categories related to the human face including: (a) a particular orientation of faces; (b) a particular face regardless of orientation; and (c) female faces in general. The younger groups of infants used in this study (aged 18 and 24 weeks) appeared to be sensitive to the first of these, ie. change from side to frontal orientation of faces, but they did not show evidence of having abstracted invariant features of a face or female faces in general.

The importance of the ability of infants to be able to acquire concepts or to abstract relational information for normal cognitive development, has been emphasised and researched by Caron and Caron (1981). They compared the performance of a sample of healthy full-term infants with that of a group of pre-term infants who had a mean gestational age of 33 weeks and all of whom had been in intensive care for complications varying in degree of severity. The tasks used involved showing infants a number of concept exemplars in a fixed trial habituation sequence. The stimulus sets used reflected relational invariance, for example a small figure being consistently above a larger one, or a series of faces all of which were smiling, and so on. Novel stimuli were paired with relationally invariant ones, and represented a deviation from the norm, ie. in the above examples novel stimuli would consist of a small figure below a larger one, and a sad non-smiling face, respectively. The infants were tested at 52, 58, 61 and 64 weeks' conceptual age. Infants were presented with a series of familiar invariant stimuli in an infant control procedure. Infants were considered to have reached criterion when either fixation time on 3 consecutive trials was 50 percent or less than fixation time on the first 3 trials, or when infants looked for 8 seconds or less on any 3 consecutive trials (this latter absolute criterion was included to allow those infants who tend to give very brief fixations at the start of the habituation sequence a chance to reach criterion). Further familiar and novel trials were presented after either criterion was reached. Five of the dependent measures were the following. 1. A measure of configurational discrimination: the total fixation to the two novel configurations relative to that to the final two familiar configurations. 2. A score of component discrimination: this being the total fixation time to the two familiar configurations (presented after the habituation criterion had been reached), relative to the last two familiarisation trials. 3. Total fixation time to criterion. 4. Number of trials to criterion. 5. Total fixation time throughout the complete task. The results indicated that at each age the pre-term infants had a lower configurational score than the full-term infants: this was significant at all ages except at 64 weeks conceptual age. Pre-term infants all recovered to a change in components, and this did not differ from that shown by the full-term infants. However the pre-terms' pattern of responding was less conceptual, with 71 percent of full-term infants responding to a configurational rather than component change, while only 36 percent of pre-term infants responded so. Attention measures failed to reveal any clear differences between groups. Caron and Caron conclude that pre-term dysfunction is not an information processing dysfunction per se, but rather is a deficiency in the content in information processing such that pre-term infants appear to have encoded part of the stimulus information during habituation, but were insenstitive to the invariant relationships linking the elements.

Again then the implication here is that concept acquisition may reflect apects of cognitive functioning, although to date no predictive studies have been reported using this measure.

2.2.6 Vocal Measures

One study in particular investigated the predictive validity of a measure in infancy reflecting vocalisation. Roe (Roe, 1978; Roe and McClure, 1980) reported high correlations between a measure of differential vocal responsiveness (DVR) taken at 3 months from 12 male infants and performance on a series of verbal sub-scales of intelligence tests at 3, 5 and 12 years. The DVR measure used by Roe was the difference between the amount of vocalisation elicited from the infant by the mother and that elicited by a stranger, each over separate 3 minute periods. On the basis of this measure, subjects were divided into high and low scoring groups. The two groups gave a significant difference between the respective means of scores on the Stanford-Binet at 3 years ($t=2.25$ $p<.03$) and also the high scoring DVR group had a significantly higher mean score on the Illinois Test of Psycholinguistic Abilities (ITPA) than those in the low DVR group at 5 years. The ITPA is made up of a number of subscales and each of these was significantly correlated with the DVR scores, with mother's educational level and parents' socio-economic level partialled out. Significant correlations emerged with the following scales: auditory association ($r= .79$); visual association ($r= .80$); auditory reception ($r= .59$); visual reception ($r= .64$); and the grammatical closure subtest ($r= .77$). All of these correlations are positive suggesting that infants with the highest DVR scores also had the highest scores on intelligence subscales. The correlations presented by Roe and McClure (1980) are very impressive particularly with measures of subsequent verbal ability: the verbal scale of the WISC-R at 12 years correlated .85 with the 3 month DVR measure, such that the greater the difference measure the higher the subsequent intelligence measure. Roe and McClure argue that it is the ability of the infants to respond differentially to the mother as compared to the stranger which is of importance, since measures of the sheer amount of vocalisation elicited from the infant were not predictive. However, a point of caution should be made in relation to these results: the sample size used by Roe was small, only 12 infants, and the use of parametric correlations produces a possibility of inflating the rs if any outlying individuals are included in the sample.

2.2.7 Conclusion: Visual Preference, Novelty Preference, Visual Habituation, Auditory Habituation and Vocalisation Measures as Potential Predictors of Later Intelligence

A summary of the predictive studies described in previous sections is presented in Table 2.1.

Results of the predictive studies mentioned so far lead one to conclude that measures in infancy, which appear to reflect cognitive abilities, can predict subsequent intelligence levels in childhood with greater accuracy than traditional infant tests. For instance, Fagan and Singer (1983) having examined data from 27 studies comparing correlations from traditional infant tests with later intelligence, report the following mean correlations: from 3 - 4 months to between 3 and 6 years, .06; from 5 - 7 months to

TABLE 2:1 Cognitive Measures in Early Infancy in Relation to Various Measures of Cognitive Competence in Late Infancy and Childhood. Longitudinal Studies

AUTHORS	N	INFANCY MEASURE	AGE	CHILDHOOD MEASURE	AGE	CORRELATION
Ruddy & Bornstein (1982)	20	Habituation Rate	4m	Bayley MDI	1 yr	-.49
		Habituation Decline	4m	Bayley MDI	1 yr	-.46
		Habituation Rate	4m	Vocabulary Score	1 yr	-.39
		Habituation Decline	4m	Vocabulary Score	1 yr	-.52
Bornstein (1985)	20	Habituation Decline	4m	WPPSI	4 yrs	-.54
Cohen & Parmelee (1983)	100 (pre-term)	Visual Fixation Duration	term	Stanford-Binet	5 yrs	-.29
Fagan & McGrath (1981)	93	Novelty Preference	5-7 months	Vocabulary Score	3.8 yrs	.37
				Vocabulary Score	6-7 yrs	.57
Lewis & Brooks-Gunn (1981)	22	Habituation Decline	3m	Bayley MDI	2 yrs	-.61
		Novelty Response	3m	Bayley MDI	2 yrs	.52
	57	Novelty Response	3m	Bayley MDI	2 yrs	.40
Miller et. al. (1977)	48	Habituation Decline	2-4m	Operational Causality tasks (Uzgiris & Hunt)	1.3 yrs	-.49
Miller et. al. (1979)	29	Habituation Duration	2-4m	Paired Associate Learning	3.3 yrs	-.26
		Habituation Decline	2-4m	Memory tasks	3.3 yrs	-.40
		Habituation Duration	2-4m	Language Comprehension	3.3 yrs	-.39
O'Conner, Cohen & Parmelee (1984)	28	Auditory Response to Novelty (Habituation)	4m	Stanford-Binet	5 yrs	.60
Roe (1978)	12 (male)	Differential Vocal Responsiveness	3m	ITPA subscales:	5 yrs	
				Auditory Association		.79
				Visual Association		.80
				Auditory Reception		.59
				Visual Reception		.64
				Grammatical Closure		.77
Roe & McClure (1980)	12 (male)	Differential Vocal Responsiveness	3m	WISC-R	12 yrs	.85

TABLE 2.2: Motivational, Social and Environmental Measures in Relation to Various Measures of Cognitive Competence in Late Infancy and Childhood. Longitudinal Studies.

AUTHORS	N	INFANCY MEASURE	AGE	CHILDHOOD MEASURE	AGE	CORRELATION
Ruddy & Bornstein (1982)	20	Mother's encouragement to stimuli in the environment	4m	Vocabulary score	1 yr	.55
Bornstein (1985)	20	Mother's encouragement to stimuli in the environment	4m	WPPSI	4 yrs	.51
Bakeman & Brown (1980)	43 (low-income)	Infant responsiveness in feeding	5m	Social competence	3 yrs	.32
				Social participation	3 yrs	.32
Cohen & Beckwith (1979)	50 (pre-term)	Social Behaviour	1m	Receptive Language	2 yrs	.44
		Smiling at Mother	1m	Bayley MDI	2 yrs	.42
		Social Behaviour	1m	Bayley MDI	2 yrs	.33
		Social Behaviour	1m	Gesell Schedules	2 yrs	.41
		Smiling at Mother	1m	Gesell Schedules	2 yrs	.39
		Mutual Gazing	3m	Gesell Schedules	2 yrs	.33
		Smiling at Care-giver	3m	Gesell Schedules	2 yrs	.30
		Mutual Gazing	3m	Bayley MDI	2 yrs	.32
		Mutual Social Behaviour	8m	Gesell Schedules	2 yrs	.40
		Smiling at Observer	8m	Gesell Schedules	2 yrs	.32
Cohen & Parmelee (1983)	100	*Schemas in object exploration	8m	Stanford-Binet	5 yrs	.25
		Infant/care-giver behaviour	8m	Stanford-Binet	5 yrs	.28
		Maternal Education	8m	Stanford-Binet	5 yrs	.46
Crockenburg (1983)	25	Smiling and eye contact	3m	Bayley MDI	1.9 yrs	.61
		Mother's response to distress	3m	Bayley MDI	1.9 yrs	.44
		Mother's Education	3m	Bayley MDI	1.9 yrs	.66
		Mother's routine contact	3m	Bayley MDI	1.9 yrs	-.44
Messer et. al. (1983)	75	*Peripheral Exploration of objects	6m	McCarthy Scales	2.6 yrs	.45 (females)
		*Task Directed Behaviour to objects	6m	McCarthy Scales	2.6 yrs	.36 to .66
		*Goal Directed Behaviour	12m	McCarthy Scales	2.6 yrs	.38 to .48

continued / ......

<u>AUTHOR</u>	<u>N</u>	<u>INFANCY MEASURE</u>	<u>AGE</u>	<u>CHILDHOOD MEASURE</u>	<u>AGE</u>	<u>CORRELATION</u>
Olson, Bates & Bayles (1984)	121	Maternal Stimulation	6m	Bayley MDI + PPVT	2 yrs	.31
		Maternal close contact	6m	Bayley MDI + PPVT	2 yrs	.23
		Baby communication	13m	Bayley MDI + PPVT	2 yrs	.42
		Maternal Teaching	13m	Bayley MDI + PPVT	2 yrs	.49
		'HOME' Scales	6m+13m	Bayley MDI + PPVT	2 yrs	.23 to .38
O'Conner, Cohen & Parmelee (1984)	28	Mother's Education	4m	Stanford-Binet	5 yrs	.42
Siegel (1982)	96	'HOME' Scales	12m	Language Comprehension	4 yrs	.27 to .37
VanDoornink et. al. (1981)	50 (low-income)	'HOME' Scales	12m	Reading + Maths ability	5-9 yrs	.37

* denotes motivational measures. Remaining infancy measures are social and environmental measures.

between 3 and 6 years, .15; and from 8 - 11 months to between 3 and 6 years, .21. The studies using alternative cognitive measures reviewed in the previous sections, however, with infant testing ages ranging from 2 to 6 months, and childhood tasks taken between 2 and 12 years produce a mean correlation of .53. These studies then provide a strong indication that greater prediction is possible from cognitive based measures than from standardised infant tests, in spite of the fact that the studies reported to date do not produce a clear indication of which particular measures are the most potentially predictive. This issue is discussed further later in this chapter.

There is a further body of literature which assesses the predictive validity of measures reflecting social, environmental and motivational factors in infancy. These studies are summarised in Table 2:2, and they are discussed next.

2.3 MOTIVATIONAL, SOCIAL AND ENVIRONMENTAL MEASURES. PREDICTIVE STUDIES - A REVIEW

This section reviews studies which have looked at the predictive validity of social, interactive, environmental, and what can be loosely defined as motivational measures, from infancy to childhood intelligence. Studies dealing with motivational measures are discussed first.

2.3.1 Motivational Measures

Motivational measures, as defined here, reflect the ways in which infants approach particular tasks and objects. Messer, McCarthy, MacQuiston, MacTurk, Yarrow and Vietze (1986) examined infants' exploratory behaviour towards objects on a number of tasks and looked at the relationship of these measures with later intelligence scores. Twelve age-appropriate tasks designed to elicit exploratory behaviours were presented to the infants at 6 and 12 months. The tasks were divided into 3 major types: first, 'effect production tasks', which involved manipulating a piece of equipment in order to produce a sensory effect; second, 'practising sensori-motor skills tasks', for example removing objects from, or putting objects into, a container at 6 and 12 months, respectively; and third, 'problem solving tasks', such as obtaining an object which is out of reach, and removing and replacing objects to and from specific positions in a peg-board. From these tasks measures were also taken of task involvement.

These were graded as follows: (1) the time spent by infants merely looking at objects was considered a period of low-level involvement with the task; (2) peripheral exploration, defined as minimal contact with usually just one modality other than looking; (3) general exploration, when infants performed two simultaneous behaviours in exploring the task; (4) task directed behaviour, which included more positive schemes which appeared to be attempts to master the object and task, but are not the appropriate behaviours for reaching the goal; and (5) goal directed behaviours; which included behaviours leading to success on the task and those which re-set the task to allow the goal to be repeated. The measure for each of these was the amount of time infants spent in each category, and the categories, as listed above were considered to reflect increasing task involvement. These measures were obtained at both 6 and 12 months and the McCarthy scale (McCarthy, 1972) was administered at 30 months. The McCarthy scale is an overall scale of cognitive development and also gives scores for motor and memory sub-scales.

The results showed that infants' success on tasks (success being defined by the number of times infants achieved one of the goals posed by the toys) did not predict later ability as measured by the McCarthy scales. These correlations did not differ from chance expectancy, and those which were significant did not exceed the magnitude of .38. However, measures reflecting task involvement did in general predict later competence more convincingly. At 6 months measures of peripheral exploration were positively related to the McCarthy scales for females (this measure correlated with the combined McCarthy scales, r= .45, and with the memory and motor subscales, r= .43 and .49 respectively). For both sexes, task directed behaviours also correlated significantly with all the McCarthy measures (rs ranging from .36 to .66) whilst measures of looking without manipulation were only related to the memory subscale. Essentially, children at 6 months who engaged in less directive forms of involvement had higher McCarthy scores and these correlations were stronger and more abundant for females. The measures at 12 months, again produced more indication of prediction for females, although the types of measures which were predictive at this age were those reflecting more directive levels of task involvement, with goal directed behaviours on practising sensori-motor skills and problem solving producing r's of .48 and .38, respectively with the McCarthy scales. In discussing the implications of these results Messer et al. refer to Piagetian theory, pointing out that at 6 months infants are likely to be in Piaget's stage 3, during which infants tend to explore objects by utilising their established repertiore of actions including mouthing, manipulating and shaking objects (peripheral and general exploratory behaviours in this study were characterised by these types of behaviour). According to Piaget, at 6 months infants are not able to understand means-ends relationships and therefore may be unlikely to perform the actions leading to task success. At 12 months however, infants do understand means-ends relationships so at this age one might expect (as Messer et al. found) task directed behaviours to be more likely to be predictive.

Kopp and Vaughn (1982) reported a relationship in 76 pre-term infants between "sustained attention" at 8 months and the Bayley Mental Index and Gesell Developmental Quotient at 2 years. The Gesell scale was also administered at 9 months. The infants had a mean gestational age of 32 weeks at the time of birth and a mean birth weight of 1,979.5 grams. At 8 months of age from birth infants were presented with a group of small cubes for a 2 minute period and the occurrence of exploratory behaviours (eg. contact, mouthing, manipulation and so on) was recorded. These measures formed the basis of the "sustained attention measure". This measure failed to correlate significantly with either outcome measure, although did significantly contribute towards prediction for both (as assessed by regression analyses). The other measures contributing towards prediction of the Developmental Quotient and Mental Index were socio-economic status, gestational age, ethnicity, and 9 month Gesell scores.

These motivational or attentional aspects of performance in infancy appear to predict, at least to some degree, later intellectual functioning. It is interesting to note that these measures, in some cases, do not reflect achievement on specific tasks but rather measures of how infants approach specific tasks presented to them. These types of measures are potentially important in the search for prediction, since the way in which infants approach and react in different situations is not only likely to affect their performance, but also the way in which others (ie. caregivers) react to them.

A considerable number of studies have examined the predictive validity of measures relating to social and environmental factors experienced by the infant. A number of these studies are reviewed next.

2.3.2 Social and Environmental Measures

VanDoornink, Caldwell, Wright and Frankenburg (1981) found that measures from the "HOME" (Home Observation for Measurement of the Environment) inventory, designed to assess parent support of development throughout infancy, at 12 months correlated significantly and positively with later school achievement (incorporating current reading and maths cirriculum levels and letter grades) between 5 and 9 years for a sample of 50 low-income children (r= .37). The most predictive sub-scales of the HOME were, emotional and verbal responsivity of the mother, and provision of appropriate play materials. The HOME was not useful for differentiation of middle-income families since there was insufficient variance amongst the scores.

Olson, Bates and Bayles (1984), observed interactive behaviours between mothers and infants at 6, 13 and 24 months of age for two 3 hour periods for a sample of 121 mother-infant pairs. Also the Bayley scale mental index was administered at each age, whilst at 24 months, additional measures including the Peabody Picture Vocabulary test (PPVT) and other measures of language and communicative abilities were assessed (these measures, since they intercorrelated highly, were combined to form a measure called

Mature Child Communication). At each age, measures were factor analysed to determine principle components of behaviour within each age. The HOME inventory was also administered at each age. Only the longitudinal predictive data will be reported here, and these results suggest that there is a relationship between measures of mother-infant behaviours at 6 and 13 months and a composite measure of cognitive ability (ie. Bayley Scale mental index and PPVT combined) at 24 months. For example, measures of maternal stimulation, and maternal close contact, at 6 months correlated positively at .31 and .23, respectively, with the 24 month measure, and at 13 months the strongest predictors of the 24 month measure included baby communication (r= .42), and maternal teaching (r= .49). Further, subscales from the HOME inventory also correlated significantly with the 24 month measure. These included: maternal emotional and verbal responsitivity (r= .23 at 6 months and .26 at 13 months); provision of appropriate play materials (r= .28 at 6 months and .24 at 13 months) and maternal involvement (r= .38 at 6 months and .32 at 13 months) and avoidance of restriction and punishment (r= .26 at 13 months).

Subscales from the HOME inventory in particular appear to constitute useful measures for prediction of later cognitive competence, and this is given further support from a study by Siegel (1982). She reports 12 month HOME scores to correlate significantly with language comprehension scores from the Reynell Language Scales at 4 years. The HOME subscales included were provision of appropriate play materials (r= .37), opportunities for variety and daily stimulation (r= .27).

Bakeman and Brown (1980) reported a study in which they compared interaction between mother and both pre- and full-term infants before and after hospital discharge, and also interaction at 9 and 20 months, with cognitive ability (as measured by the Stanford-Binet) and social competence and participation at 3 years. 21 pre-term infants and 22 full-term infants and their mothers were used. All the mothers were from low-income families. The pre-term infants weighed between 1,000 and 1,950 grams and had spent no longer than 24 hours in intensive care. The sample of pre-term infants had an average gestational age of 32.4 weeks. Results indicated that no measure of mother-infant interaction during feeding predicted subsequent Stanford-Binet scores at 3 years, despite the fact that significant differences on these measures were found between pre- and full-term infants. Higher ratings of infant responsiveness during feeding were positively related to social competence, and participation, at 3 years (rs = .32 in both cases). The extent to which mothers vocally and emotionally responded to their 20 month old infants was positively related to the Stanford-Binet scores at 3 years (r= .42), even though only on the latter measure could pre- and full-term infants be differentiated.

Cohen and Beckwith (1979) also report significant relationships between early interactive measures and later cognitive performance for a sample of 50 pre-term infants. Their criteria for subject inclusion were a gestational age of 37 weeks or less and birth weight of less than 2,500 grams. At 1, 3 and 8 months (controlled such that all infants were tested at equivalent conceptional ages) all infants were assessed on at least a half hour period (at 8 months usually a one and a half hour period) of mother infant interaction, which included feeding and waking periods. Measures from each age were factor analysed to determine major factors of behaviour at each age. Other measures included the Obstetrical Complications Scale (see Littman and Parmelee, 1978) including health and history of the mother, pregnancy events and intrapartum and postpartum events and the Postnatal Complications Scale (see Littman and Parmelee, 1978) which included measures of respiratory difficulties, convulsions, disturbances and so on. At 24 months the Gesell Developmental Schedules were given along with an adapted version of the Casati-Lezine sensori-motor scales (Kopp, Sigman and Parmelee, 1974); receptive language skills were also assessed and the Bayley Mental Scales administered at 25 months. Correlation analyses indicated that competence on the Gesell at age 2 was predicted at each age by mother-infant interaction and in addition the Gesell score at age 2 was also significantly predicted by social behaviour and smiling at mother at 1 month (r's = .41 and .39 for the two measures, respectively), and by mutual gazing at 3 months (r= .33); and mutual social behaviour at 8 months was related to the 2 year Gesell score (r= .40), as was smiling at the observer (r=.32). The 2 year receptive language measure was positively associated with social behaviour at 1 month (r= .44), whilst no behaviour at 3 or 8 months was associated with this 2 year score. The 2 year Bayley mental score was also predicted by smiling at the mother (r= .42), social behaviour (r=.33) at 1 month, and by mutual gazing (r= .32) at 3 months, but not by anything at 8 months.

The reported significant relationship between smiling and eye contact between infants and mothers, and later intellectual ability is further supported by a study carried out by Crockenburg (1983) on 25 mother-infant pairs. From a three and a half hour home observation at 3 months this measure was found to correlate in the region of .61 with scores on the Bayley Mental Development Scale at 21 months. Mothers' responsive attitude to infant distress also correlated with the 21 month score (r= .44) as did mothers' education (r=.66), whilst measures of the extent of routine contact with the infant correlated negatively with the 21 month score (r=-.44).

To summarise, it seems that aspects of the home environment and measures of mother-infant interaction do to some extent predict later intellectual ability.

2.3.3 Conclusion: Motivational, Social and Environmental Measures as potential predictors

It is apparent then that aspects of mother-infant interaction, the home environment, and measures which appear to reflect infant motivation, can also predict later intelligence to a reasonable degree. As can be seen from Tables 2:1 and 2:2 the correlations from the motivational, social and environmental measures are not, overall, especially higher than those from the cognitive based measures. However, the studies discussed here are by no means conclusive, only a few researchers consider how these different measures relate to and or interact with each other, and in few cases have researchers examined the reliabilities of their measures. These issues are discussed later in this chapter.

There are a number of studies which have examined the inter-relationships between the cognitive and motivational, social and environmental measures. Some of these studies have further attempted to establish which type of measure has the greater effect on later intelligence scores by the use of step-wise multiple correlations and path-analyses. It is these studies and their findings which are discussed next.

2.4 STUDIES INCORPORATING COGNITIVE AND SOCIAL MEASURES

2.4.1 Review of Studies Incorporating Cognitive and Social Measures

The study by O'Connor et al. (1984), described in Section 2.2.4 above, in which a relationship was reported between recovery measures in an auditory habituation task and subsequent intelligence scores, also reported that measures of the number of years the mother spent in education correlated significantly with the infants' 5 year Stanford-Binet scores (r= .42). These measures of the mothers' education and the infants' novelty scores, were entered into a step-wise multiple regression to assess their relative contribution to prediction of the 5 year intelligence scores. When the novelty response measure was entered first it produced an R-squared of .36, which increased marginally, although not significantly, to .43 as a result of adding the mothers' education. However, when the latter variable was entered first the amount of variance explained by this alone (17 percent) increased considerably to 43 percent when the novelty preference measure was added. The measures of mother's education and the novelty response predict better in unison than either measure independently, which suggests that childhood performance on intelligence tests is affected (as one might expect) by both of these measures.

Cohen and Parmelee (1983) conducted a large scale longitudinal study in which 100 pre-term infants were assessed on various measures from birth until they reached 5 years of age. 62 of these children were from English speaking families and the remainder were Spanish. Prematurity was defined as a gestational period of 37 weeks or less and a birth weight of 2500 grams or less. At birth infants were rated on obstetric and postnatal complication scales and a measure was also taken of how long infants fixated a visually presented stimulus. At 4 months visual attention was assessed in terms of stimulus and novelty preferences. Caregiver interaction in the home was measured at 1, 8 and 24 months. The Gesell scales were administered at 4, 9 and 24 months, and at 8 months hand manipulation and the way in which infants explored objects was assessed. The Bayley scales were administered at 25 months, receptive language development at 2 years, and the Stanford-Binet at 5 years. Results for the total sample indicated that term visual fixation correlated with the 5 year scores (r=-.29) such that the infants who fixated the stimulus for the shorter duration had higher IQ scores. The infants who demonstrated the more advanced schemas at 8 months also had higher Stanford-Binet scores at 5 years (r= .25). Other measures to correlate significantly with the 5 year score included, caregiver-infant behaviour at 8 months (r= .28) and years of maternal education (r= .46). Regression analyses indicated that the best single predictor was maternal education, although one third of the variance could be accounted for by four variables: maternal education, term visual attention, Gesell scores at 9 months, and manipulative schemas at 8 months. For the English speaking sample alone, however, over one third of the variance (R-squared =.36) was accounted for by 3 measures: the Gesell scores, term visual attention and manipulative schemas at 8 months. This suggests that the term visual attention measure, which was found not to be related to social class, does contribute towards prediction of later intellectual ability.

Bornstein and Ruddy (1984), obtained measures from 4 month old infants from a habituation task (see section 2.2.3), and the infants were also observed in the home for a half-hour period with their mothers. The measures taken included the extent to which mothers directed infants' attention towards objects. At 12 months the Bayley Scales were administered along with a further 30 minute observation of mother and infant during which measures of the extent to which mothers encouraged their infants to attend to the environment, the amount of time spent talking to the child, and the size of the child's speaking vocabulary were taken. The results indicated that all measures at 4 months (that is measures of habituation and of mother-infant interaction) correlated significantly with speaking vocabulary size and the Bayley at 12 months. The 4 month measure of the mother encouraging attention to objects in the environment correlated with the 12 month speaking vocabulary (r= .55). Regression analyses indicated that habituation rate alone accounted for 21 percent of the variance, while encouragement to the environment alone accounted for 31 percent of the variance.

Ruddy and Bornstein attempted to determine whether stimulation from the mother fosters cognitive development or whether a greater amount of vocalisation by individual infants elicits maternal stimulation. Cross-lagged panel analyses were performed and the results indicated that, in fact, maternal stimulation affects infants' vocabulary (maternal stimulation predicted infant vocabulary, r= .55) whilst vocalisation to the mother at 4 months failed to predict maternal involvement at 12 months, r=-.01). These results led the authors to conclude "... that mothers who frequently encourage their young infant's attention foster their cognitive development." (Ruddy and Bornstein, 1984 p.185).

Bornstein (1985) reports a follow-up study in which the sample of infants used in the study reported above were tested again with the Wechsler Scales at 4 years. Measures of habituation and maternal encouragement at 4 months were significantly correlated with the Wechsler, and unrestricted path analyses revealed that habituation measures at 4 months constituted the single strongest influence on the intelligence scores, revealing a beta weight of .58, whilst maternal encouragement produced a weight of .48. The 12 month measures, including the Bayley scales and measures of the extent to which mothers encouraged infants to attend to their environment, were found to have little predictive significance for childhood IQ. This suggests that although measures of mother-infant interaction do influence cognitive performance, these measures have no greater predictive power than measures from habituation tasks.

2.4.2 Conclusion: Studies Incorporating Cognitive and Social Measures

Many of the studies reported indicate that cognitive measures in infancy may reflect some aspect of performance which aids prediction of later intellectual ability, along with environmental and social measures, which also appear potentially predictive. The use of cognitive measures however, is of particular interest for the search for a means of detecting those infants who are likely to suffer subsequent intellectual deficit.

The evidence relating to the ways in which cognitive and social measures influence each other is by no means clear. The sample size in many of the studies is low, which can artifically bias beta weights in regression analyses.

2.5 CONCLUSION AND INTERPRETATIONS OF PREDICTIVE STUDIES

5.2.1 Introduction

In general, the picture created by the predictive studies reviewed is extremely complicated. The studies incorporating social and environmental measures are perhaps the most easily interpretated, largely due to the fact that the measures adopted are in many cases more comparable. For example, specific measures from the HOME inventory seem consistently to constitute potential predictors, and these measures are likely to be easily and consistently assessed by different researchers.

However this is not the case for the measures taken from cognitive predictive studies. Most of the recently reported predictive studies incorporating cognitive measures, constitute post-hoc investigations in that infants who acted as subjects in previous investigations some months or years earlier were traced, and a follow up study of some description was carried out. Fagan and McGrath (1981) for example, found that novelty preference measures taken from a whole host of experiments incorporating different stimulus sets and different age groups of infants predicted to some degree later linguistic scores.

A further important point relating to cognitive studies and which was mentioned earlier, is that many of the studies have adopted a wide variation of measures, and even in cases where comparable measures have been used they have often been calculated in different ways. For example, the amount of fixation decrement to a familiar stimulus has been defined as a ratio from initial trials to later trials in the habituation sequence from a variety of different measures: the first fixation duration of each trial (Miller et al. 1977); mean fixation length per trial (Lewis and Brooks-Gunn, 1981) and total fixation per trial (Bornstein and Ruddy, 1984). This creates difficulties when one tries to determine which measures are potentially the most predictive.

Furthermore, in habituation procedures for example, certain measures are restricted by the particular procedure adopted. Total fixation time is limited in the fixed trial procedure since there is always a maximum amount of time for which infants can fixate the stimulus (this is not the case in the infant control procedure), and also if infants tend to exhibit a large number of fixations by giving a lot of looks away from the stimulus, they may be less likely to get a long fixation duration than infants who exhibit fewer long looks. Also, if an infant is fixating the stimulus at the end of the habituation sequence in a fixed trial procedure, then the measure of number of looks may be spurious. On the other hand, in the infant control procedure total fixation time is dependent on the rate of decline infants show, and infants who exhibit many looks of a similar duration are likely to finish with a large total fixation duration.

Measures of decrement are also differentially affected by the different procedures. If infants exhibit short initial looks at the start of the task, they may have difficulty in reaching the defined criterion in the infant control procedure, although in a fixed trial procedure they may still be able to exhibit a lengthy look before the end of trial one, and hence have a better opportunity show a decrement in later trials.

However, in spite of these difficulties, the predictive correlations resulting from the studies based on cognitive measures are in general encouraging, and the variation in terms of measures, testing conditions, and stimuli in infancy can arguably be considered to add to the robustness of cognitive measures as potential predictors. Alternatively, it may be that had a large number of studies been specifically directed towards the question of prediction, with all infants being tested under the same conditions with the same stimuli, then greater evidence for prediction (i.e. higher correlations between infant and childhood scores) may have emerged. In recognition of this point, Fagan (1984) is currently standardising his measures of novelty preferences for use as tests of infant recognition functioning.

In view of the evidence suggesting that cognitive measures contribute at least as much to our ability to predict as do measures of social and environmental factors, there is a need to create a better understanding of what these cognitive measures are actually reflecting. Clearly further research is crucial. However before discussing which approaches may be the most fruitful, one should perhaps consider in further detail what the current literature has to offer.

2.5.2 The nature of the criterion measures in predictive studies

Typically the measures of later cognitive functioning in the studies discussed above are either scores on standardized infant scales or childhood intelligence tests. Since scores on infant tests tend to correlate highly with performance on childhood tests especially from the age of about 18 months (Caron and Caron, 1981), it seems that intelligence, at least as measured by childhood tests is, to some degree being predicted from infancy.

It is interesting that many authors have used measures of language development as the measures of later intellectual functioning (for example, vocabulary subscales of the Stanford-Binet, WPPSI, and WISC-R, by Fagan and McGrath 1981; speaking vocabularies, by Bornstein and Ruddy, 1984; the ITPA and WISC-R verbal scale by Roe, 1978, Roe and McClure 1980). The use of language-based outcome measures has been fruitful, and the significant ones reported are no lower and often higher, in magnitude than the full-scale outcome measures. The implications and interpretations of this (including those relating to the models of continuity discussed in Chapter 1) are discussed in Chapter 10.

2.5.3 Inter-Relatedness of Measures

It is apparent from the earlier sections of this chapter that a number of different measures in infancy correlate with scores on a relatively small variety of tasks in later childhood (primarily full-scale Childhood Intelligence tests, and sub-scales, and language scales and sub-scales). For example, Roe (1978) found her measure of diferential vocal responsiveness (DVR) to be related to subsequent linguistic ability, and Fagan and McGrath (1981) also report their novelty preference measure to be indicative of later language ability. What is not known is whether the DVR measure is related to the novelty preference measure in infancy. If these two measures, for instance, were correlated with one another, then this might suggest that some form of general discriminative ability in infancy is predictive of subsequent language ability, whilst other types of measures may reflect general intelligence as opposed to specifically language. Information of this nature is crucial for the understanding for the development of intellectual ability throughout the infancy period.

Indeed within the literature relating to environment and mother-infant interaction there has been considerable interest in how these measures inter-relate (see Cohen and Beckwith, 1979; Olson, Bates and Bayles, 1984), but this interest seems currently to be lacking among researchers in the cognitive domain. This is probably in part due to the fact that most of the predictive studies have been done after collection of the initial (infancy) data, and therefore within individual studies, only a limited number of measures from usually one task are available. On the other hand, studies which examine aspects of mother-infant interaction have tended to encompass a much wider range of measures.

2.5.4 The Psychometric Consistency of Measures

One of the most important issues relating to the search for predictive measures concerns the reliability of the measures used. A common principle of tests and measurement is that unless there is evidence of test-retest reliability and internal consistency of measures there is little hope of predictive validity (Brooks and Weinraub, 1983). If a test does not predict itself it will predict nothing else. Furthermore the extent to which a measure is reliable will restrict how predictive a measure can be (McCall, 1981).

As discussed in Chapter 1, the early test developers in the 1930's soon became aware of this, but its importance in relation to more recent searches for prediction from cognitive measures in infancy by and large seems to have been ignored (again, there has been some research on the stability of measures of infant-mother interaction, maternal involvement with infants and infant behaviours (for example, Beckwith, 1978; Lewis and Starr, 1978). There is only a small number of studies concerned with cognitive tasks which have acknowledged the importance of the reliablity of measures, and it is these studies which are discussed next.

Fenson, Sappir and Minner (1974) attempted to determine the within-session stability of the measures of first fixation duration and total fixation time, from two fixed trial habituation procedures administered to one year old infants in the same session. Identical stimuli were used on the two tasks and correlations were calculated for each pair of equivalent trials across the two sessions. The correlations were higher for total fixation duration (r's ranging from .65 to .28, all of which were significant to at least the .05 level) than for first fixation duration (r's ranging from .54 to non significant .18). Fenson et al. also examined the test-retest reliabilities over a 3 week period. Again the measures of total fixation duration were generally higher for total fixation duration than for first fixation duration, although averaged across all stimuli shown on each session both measures had high reliabilities (r= .64 for first fixation duration, and .74 for total fixation duration).

Fenson et al. (1974) also examined the reliability across sessions of the amount of decrement in fixation over trials. Their measure of this was total fixation time in the final two habituation trials subtracted from that exhibited in the initial two trials, divided by the total fixation time in the first two trials. This index was calculated separately for the measures of total fixation and first fixation duration but neither measure produced significant correlations across testing sessions. Hence no evidence was found to suggest that rate of habituation was reliable over a 3 week period.

Miller et al. (1980) reported across-age reliabilities of the decrement ratios to familiar stimuli, including ratios of both first and total fixation durations, from 2, 3 and 4 months to 27 and 51 months (see Miller's studies reviewed above) and found no relationships between these measures over these time periods.

Pecheux and Lecuyer (1983) examined test-retest reliability of habituation measures over a 1 week period in 4 month old infants. They adopted an infant control procedure and infants received either the same or different stimuli on each session. The stimuli used were geometric patterns, bulls-eyes and faces. The number of trials to reach criterion failed over all to emerge as a reliable measure with the correlation from only one of the six stimulus combinations reaching significance, but when total fixation time to criterion was used, then reliability across sessions was found (rs ranging from .48 to .70).

Rose, Slater and Perry (1986) habituated 3 and 5 month old infants to a different stimulus on 3 occasions with a minimum of 24 hours between each testing session. An infant control procedure was again employed and habituation measures were taken from the measures of trials to criterion, first fixation and total fixation duration. A reasonable degree of stability was found for all measures and for both age groups: no one measure emerged as being any more or any less reliable than the others, although reliabilities between sessions 2 and 3 for all measures were higher than for those between sessions 1 and 2, or between 1 and 3.

Overall, these studies indicate that there is some degree of reliability of habituation measures both across and within sessions. The reliabilities seem to lie within measures reflecting quantitative measures of fixation (such as total fixation time and number of fixations) rather than more qualitative measures such as rate of decline to the familiar, or recovery to the novel, stimulus. This is surprising since the qualitative measures have been assumed to be the best predictors. Further research is clearly required to obtain a better understanding of what individual measures reflect.

2.6 CONCLUSIONS AND THE AIMS OF THE THESIS

Having presented the available data pertaining to the search for measures predictive of later intellectual ability, it is apparent that although in general the results are encouraging, in that there are reports of significant correlations between measures in infancy and later intellectual behaviour, the field of research is lacking the psychometric framework required in order to offer any sound practical or theoretical conclusions.

The research reported in this thesis is an attempt to increase our understanding of cognitive or intellectual performance in infancy and to determine whether aspects of mental functioning are stable over a 15 month period.

Two major assumptions were made.

First, if a measure is to be considered potentially predictive it must show test-retest and within-session reliability, unless in the latter case there is sufficient justification for it not to do so. Measures of learning for example, may not be expected to show reliability across sessions since if infants learn that a particular response produces a particular reinforcement on one session this will clearly affect performance on the task in any subsequent session. Similarly, measures of decline in fixation in an habituation procedure may not be expected to show stability within sessions depending on the pattern of habituation displayed.

Second, any relationship (correlation) between measures, either within or across age, implies some underlying ability common to both measures.

The specific aims of the thesis are as follows:

1. To examine the within-session and test-retest reliabilities of a number of measures in infancy which can be thought to reflect cognitive or intellectual behaviour. A series of tasks, including four which were administered in the laboratory, and two in the home, were given twice to the same group of infants at 3, and 5 months of age.

2. To examine the inter-relationships of measures both within and across age. The within-age analyses might reveal something of the nature of intellectual ability in infancy at each age. That is, whether intelligence is made up of general ability (g) with many

infancy measures correlating with one another, or whether different types of ability emerge, for example clusters of related measures. The comparison from 3 to 5 months might determine the extent to which the pattern of results from 3 months is reliable over this 2 month period.

3. To compare performance on the tasks at 3 and 5 months with performance on a different set of age-appropriate tasks at 9 months. This should indicate whether any measures from 3 and or 5 months reflect any aspects of performance on tasks in later infancy.

4. To determine how the measures obtained at 3, 5 and 9 months of age relate to performance on subscales from standard infant tests and language tasks at 18 months.

This program of research was undertaken in the hope that it would aid our understanding of early intellectual functioning, and the implications of the results for practical and theoretical issues are discussed later.

The next chapter is concerned with the choice of tasks and pilot studies for the initial phase of the investigation.

CHAPTER 3

TASK SELECTION AND PILOT STUDIES

3.1 INTRODUCTION

This chapter is concerned with the choice of tasks to be given to the infants at the 3 and 5 month testing ages, and the preliminary pilot studies of these tasks. The choice of tasks for the 9 and 18 month testing ages and the pilot studies are described in later chapters.

This chapter is organised into 3 major sections. Sections 3.2 is concerned with the choice of the tasks, the rationale and very brief details of the tasks themselves. Full details of all of the Pilot studies, the stimuli and procedures used are presented in sections 3.3 to 3.5. The laboratory based tasks requiring a visual response are presented in sections 3.4, and Operant Conditioning and home based tasks in sections 3.5.

3.2 TASK SELECTION

3.2.1 Introduction

One of the conclusions to be reached from Chapter 2 was that a vast number of different types of task have been used as potential predictors, although as yet there are no clear indications that some tasks are, or reasons why some tasks should be, better predictors than others. Issues such as these made the choice of tasks for use in this research difficult. Clearly not all potentially useful tasks could be employed due to a number of constraints, in particular resources and time.

Six tasks in total were selected for this study in the hope that to some degree each could be considered to reflect some different aspect(s) of mental functioning in infancy and therefore also constitute potential predictors. Further, since within testing sessions infants were required to perform a large number of tasks, effort was also made to vary the tasks in terms of the nature of the stimulation and response, in order to obtain as much data from each infant as possible. Nevertheless, many of the tasks required visual responses from the infants to various types of stimuli. The reasons for selecting the tasks are discussed in the following sections. Each study was also included in a set of pilot studies and these studies, the results, and any resultant changes made to the experimental design as a consequence of these pilot studies, are also described in this chapter. The tasks selected are discussed next under the following headings: Visual Preferences; Visual Habituation; Concept Acquisition; Audio-Visual Integration; Operant Conditioning; Differential Vocal Responsiveness; and Attention to a Novel Toy.

3.2.2 Visual Preferences

As outlined in the Chapter 2, the possibility that the visual preference technique provides measures potentially indicative of later intellectual ability, comes from two lines of evidence: first, from reported age differences in the sorts of stimuli for which infants show preferences; and second from studies which compare different subject samples.

Studies relating to these two lines of evidence have been discussed in Chapter 2. In sum, the implication is that the extent to which individual infants show a visual preference for certain stimulus types over others at particular ages may provide some indication of how cognitively advanced they are, and therefore may predict later intellectual ability.

A visual preference task was included in this research. Four pairs of stimuli were used in all, one consisted of a stripe-bullseye comparison, and the other 3 pairs were each made up of a two element and a six element stimulus. Size was held constant within stimulus pairs and varied across stimulus pairs.

This design was adopted as it forced a visual preference of number of elements (large or small) across stimuli of different sizes, and therefore the overall extent of number preference should not be attributable to differences in preference for size.

3.2.3 Visual Habituation

The visual habituation paradigm and its potential for producing possible predictive measures has been discussed in chapter 2, and will only briefly be mentioned here. Essentially, differences in the rates or amounts of habituation and subsequent recovery have been considered to reflect measures of individual processing rates.

The two paradigms which are commonly employed to determine habituation rates and recovery measures of individual infants have also been described in Chapter 2. Historically, the fixed trial procedure was the first of these paradigms to be developed. The infant control procedure was described by Horowitz, Padan, Bhana and Self (1972). Horowitz et al. claimed that the infant control procedure was a more successful technique since data from fewer infants had to be eliminated due to infant fussing.

When visual habituation testing first commenced in the Plymouth laboratory a number of 3 month old infants were administered an habituation task employing the infant control procedure. In many cases however, this was not successful since infants often gave very brief looks at the beginning of the habituation sequence, thus making it difficult for them to reach criterion thereafter. This problem of short initial fixation times has also been reported by Caron and Caron (1981). They attempted to alleviate the problem by allowing infants to reach one of two criteria of habituation. The first of these was the usual one of relative decrease in fixation from the initial fixation time, and the second was that infants were judged to have habituated if they exhibited 8 seconds or less of total fixation time on 3 successive trials from the fourth trial on. However, this method would be of little use when individual differences are of interest since measures of decline in fixation are hardly comparable across infants who have reached different criteria. For the purpose of this research then, the fixed trial procedure was adopted, as indeed it was for many of the predictive studies (including the studies by Miller and co-workers, and Lewis and Brooks-Gunn, 1981) as reviewed in Chapter 2.

Different stimulus sets were used on each of two testing sessions, although the nature of the stimuli (black and white patterns and photographs) remained the same within sessions. Different stimulus sets were used across sessions because if habituation measures are useful for prediction, then performance should be consistent not only across sessions, but also across different stimulus types. All of the infants' fixations towards each stimulus were recorded.

3.2.4 Concept Acquisition

As discussed in Chapter 2 preverbal concept acquisition is typically investigated through the use of habituation procedures. Caron and Caron (1981) have suggested that infants' ability to process relational information is central to other disciplines: memory, perception, psycholinguistics, comprehension and so on. Gibson (1979) claimed that all perceptual development is a hierarchical progression from the detection of isolated features to invariant configurations of features defining individual objects in event contexts. Also, the connection between language development and concept formation is considered to be important; Brown (1973) and Nelson (1977) for instance, have pointed out that initially children tend to name categories and therefore must be able to perceive equivalences between instances on a non-linguistic basis.

In sum, the literature provides a suggestion that the extent to which infants are able to acquire concepts may be indicative of some form of intellectual growth. A concept acquisition task was therefore included in this research and, as for the visual habituation task, a fixed trial procedure was adopted.

Again different stimulus sets on each testing session were used. All the stimuli in this task were made up of black and white patterns. All of the looking responses of the infant were recorded.

3.2.5 Audio-Visual Integration

This task is subsumed within the notion of cross-modal integration. The extent to which infants can co-ordinate information across modalities is important for our understanding of how infants experience their environment.

Spelke (1979), and Spelke and Owlsey (1979), have investigated infants' abilities in this respect, and have shown consistently that 4 month old infants prefer to watch a film which corresponds to concurrent auditory stimulation, as opposed to a film not concordant with sound stimulation. Spelke (1979) suggested that infants were demonstrating this behaviour in one of two possible ways: first, that they were detecting certain familiar objects by looking and listening (for example, infants may have seen and heard a 'patta-cake' sequence and hence have learned the association between sound and sight in this particular event), or second, that infants are able to detect the common temporal structure of the optic and acoustic stimulation, without being dependent on specific prior experience.

In an attempt to distinguish which of these two possibilities was causing the preference for the integrated stimulus, Spelke conducted a series of experiments which involved presenting two similar events simultaneously which differed in tempo, and were played out of synchrony from one another, along with the sound track corresponding to only one of the events. Results indicated that overall infants fixated the event which was integrated with the sound track. This led Spelke (1979) to conclude that infants are able to detect the common temporal structure of visual and auditory events rather than merely having learned the association between sound and a particular event.

Spelke (1979), has further suggested that mature perceivers are the most apt to acquire intermodal knowledge, if they seek and attend to several stimulus modalities at once. Infants of the same age may vary in their capablity of attending to events and objects in more than one modality and hence, individual differences in this respect may reflect differences in cognitive and / or perceptual advancement. A task which assessed the infants' ability to integrate an auditory and visual event was hence included in this research.

Two identical events were presented to the infant on video along with a sound track corresponding to one of these events. Different events were used on each of two testing sessions. On the first testing session the filmed event was a drum being banged, and on the second session a film of a hand-clap was used. Infants' fixations were recorded throughout and the event fixated most was considered the preferred stimulus.

3.2.6 Operant Conditioning

Measures of learning have not, to the author's knowledge, previously been used in predictive studies from infancy. In fact, speed of learning has been reported to be unrelated to intelligence for both slightly retarded and normal human beings (see Estes, 1982).

Estes points out that learning can be associated with acquiring information concerning relationships between situations and actions and their outcomes (as opposed merely to repeating a reinforced response). In other words learning occurs by observation rather than by reinforcement. However for learning to occur at either of these levels, (ie. on the one hand an understanding of relationships between situations and their outcomes, and on the other, a repetition of a previously rewarded response) adaptive behaviour depends to some degree on the acquisition of information relating to the relationship between actions and outcomes. What may differ between these two levels is the level of processing involved, with higher order processes more necessary for the former than for the latter level. Furthermore, since throughout early life infants must acquire an understanding of the relationship between specific environmental events and particular reponses and their consequences (Rovee-Collier and Fagen 1980), then the ability to achieve this may relate to aspects of later intellectual performance.

Rovee-Collier and her co-workers (Rovee-Collier and Fagen, 1980) have developed a simple yet effective technique for studying operant conditioning in infancy. Essentially, infants learn to produce movement of a mobile (the reinforcement) situated above them, by means of foot-kicks.

The procedure they adopt is a conjugate reinforcement schedule and differs from other reinforcement procedures in a number of ways. For instance, in the conjugate reinforcement schedule, the rate or amplitude of a response directly determines the intensity of its consequences; every response produces a reinforcer and reinforcement is continually available without disruption or delay. This procedure has been found to work effectively with infants aged from about 6 weeks, with infants showing rapid and sustained increases in kick rate during reinforcement, and a gradual decrease in responding in the extinction phase (Rovee-Collier and Gekoski, 1979).

Furthermore, Rovee-Collier and Gekoski (1979) suggest that since infants often fixate the mobile, then execute a highly precise response, then fixate the mobile briefly before turning to examine other aspects of their environment, even though the mobile is still moving, then it seems that the reinforcement is not the mobile movement per se, but rather the control which infants have gained over their environment. Also Rovee-Collier, Morrongiello, Aron and Kupersmidt (1978) have demonstrated topographical response differentiation in infants when the 'reinforcing' leg is changed throughout a session. Although it remains debatable as to whether performance does require higher order processes (which according to Estes are more likely to be related to intellectual performance), these lines of evidence do tend to suggest that infants acquire a detailed understanding of the learning situation, and hence the task was included in this study.

A conjugate reinforcement paradigm was adopted for this research, similar in design to that developed by Rovee-Collier and co-workers. A basal measure of kicking rate was taken initially, then infant foot-kicks produced movement of an over-head mobile, and finally, an extinction phase was administered. From this task various measures of learning were taken, the details of which are presented in the Pilot Studies section.

3.2.7 Differential Vocal Responsiveness (DVR)

The DVR measure and the extent to which it has been found to be predictive of later verbal ability at least in a small male sample (Roe, 1978; Roe and McClure, 1981) was discussed in Chapter 2. In view of the high correlations from the DVR measure reported, the task was included in this research program and the procedure adopted was identical to that used by Roe.

3.2.8 Attention to a Novel Toy

This task as employed in this research program has not, to the knowledge of the author, been adopted elsewhere for this type of research. Essentially this task involved presenting a brightly coloured 3-dimensional object to the infant in the home and taking measures of the time spent fixating the object and the number of fixations produced (for details see Method section). The purpose of including this task was to assess measures of fixation towards novel objects in an environment which was familiar to the infant. The fact that this task was administered in the home also acts as a control for the possibility that some infants may be affected more by a change in environment than others, and hence this task may provide a 'cleaner' measure of attention patterns than the laboratory based tasks.

3.3 PILOT STUDIES

3.3.1 Introduction

Each of the tasks selected for use in this research was initially included in a series of pilot studies. There were three major purposes of these preliminary studies. The first was to overcome any 'teething' difficulties likely to arise in the major experimental phase of this work. These included familiarisation of the experimenter with equipment, designs, computing operations and procedures involved with the running of the experiments, as well as establishing rapport with both mothers and infants acting as subjects and gaining awareness of different infant states.

The second purpose of the pilot studies was to ensure that the measures derived from each task produce variation in scores between individuals. For a measure to be potentially predictive, it must produce a range of scores over subjects.

Third and finally, in the case of the four laboratory based tasks which measured aspects of looking behaviour in infants (ie. Visual Preferences, Visual Habituation, Concept Acquisition and Audio-Visual Integration), the pilot studies were also used to determine whether any of the measures taken from the tasks showed test-retest reliability (or, across different stimulus types, equivalent form reliability), and hence constitute potentially predictive measures. Further discussion of the types of reliabilities taken from measures is presented in Chapter 4.

The remaining tasks were not piloted for test-retest reliability. In the case of the Operant Conditioning task, lack of reliability would be difficult to interpret since the effects of learning a particular response in session one is likely to affect performance on session two. The way in which performance is likely to be affected however is somewhat unclear; Rovee-Collier and Fagen (1980) suggest that memory can be assessed by comparing the number of kicks in the extinction phase of the first session with the number recorded in the base-rate phase of session two. If the number of kicks in both of these periods is high then the infant supposedly has a good memory for the kicking response and the reinforcement associated with it. However, what exactly the infant has remembered about the initial learning phase is difficult to establish. For instance, if the infant learns that the foot-kicks produce an interesting response, then infants who show a high kicking rate at the start of the second session may be considered to have good memories. On the other hand, although infants may remember the response, they may also remember that at either the beginning and / or the end of the session, no reinforcement occurred, and if this is the case then one should not assume that a low kicking rate at the start of session 2 fails to reflect memory.

The Differential Vocal Responsiveness task is also difficult to assess in terms of test-retest reliability. A stranger on session one is a little less strange on session 2, especially when other interactions with the 'stranger' have occurred in between times. If no reliable measures emerge from this task, therefore, one would not be justified in eliminating it. The DVR task was therefore practiced on several single occasions during the pilot studies, although reliabilities of the measures were examined in the major longitudinal study. Finally, the Attention to the Novel Toy task was also practised during the pilot studies with reliabilities being established in the major study.

The remainder of this chapter is concerned with details of each pilot study, analyses, and discussion of the tasks, beginning with those laboratory tasks for which reliabilities were examined (sections 3.4), followed by the. Operant Conditioning and home-based tasks (sections 3.5).

3.3.2 Subjects

A total of 32 infants were recruited through advertisments in local health clinics and the local media. All the infants were white, full-term and had no known mental or physical health defects. All came from middle-class families. Sixteen were male and sixteen were female, eighteen were first-born infants with the remainder having between 1 and 3 siblings living at home.

All infants at the time of testing were within one week of their 3 month 'birthday'. In the case of Visual Preferences, Visual Habituation, Concept Acquisition and Audio-Visual Integration, testing continued until a total of 12 infants had completed each task on two occasions separated by a minimum of 24 hours and a maximum of one week. The remaining tasks were piloted with 6 of these infants on one session only.

3.4 PILOT STUDIES: VISUAL-BASED LABORATORY TASKS - Visual Preferences, Visual Habituation, Concept Acquisition and Audio-Visual Integration.

3.4.1 Design

Each infant was brought to the laboratory at Plymouth Polytechnic by the mother on two occasions. The mother was encouraged to choose, where possible, times between feeds and when the infant was typically awake. Transport to and from the laboratory, by taxi, was provided and paid for by the Polytechnic, or petrol money refunded and parking facilities arranged for those mothers preferring to use their own cars.

On each visit the 4 tasks were administered to infants in a fixed order of presentation:

A) Visual Preferences
B) Visual Habituation
C) Concept Acquisition
D) Audio-Visual Integration

A fixed order of testing was adopted because of possible order or fatigue effects influencing performance on individual tasks. If the ordering of tasks was counter-balanced or randomised across sessions for individual infants then failure to find across session reliability may be due to fatigue effects. On the other hand, if all tasks are administered in the same order for all infants on each testing session then presumably effects in the data resulting from ordering factors will be closely approximated for all infants.

In the case of failure to complete a given task, due to fussing, sleepiness, intervention by the mother, or failure to respond, the task was removed and the next one in sequence attempted if the infant returned to a suitable state. Details of the number of infants failing to complete any given task, and the reasons, are presented in the Results sections. In the event of an infant in any session failing to complete any of the tasks, an alternative appointment was arranged with the mother. This occurred on 3 occasions during the pilot studies.

3.4.2 Apparatus

For the Visual Preference, Visual Habituation, Concept Acquisition and Audio-Visual Integration tasks, the infant was seated in a commercially bought infant chair, reclining at an angle of 45 degrees. The chair was secured by clips in a viewing chamber, such that the infant faced a metre square screen at a distance of about 60 cm. The remainder of the chamber consisted of a wooden base, surrounded on each side and above the screen by brown cotton curtains, which also extended behind the infants head. This apparatus is shown in diagrammatic form in Figure 3:1 and a photograph of an infant being tested in this apparatus is presented in Figure 3:2.

Fixation patterns by the infant towards the stimuli were recorded by an E (the author), who observed the infant through a small hole (1 cm. square) situated at the side of the projection screen. Recordings were made via buttons on a hand-held device, which was interfaced to an Apple II computer. For each task the computer indicated with a brief tone when the criterion for a stimulus change was met (see Method section). In addition a video camera was fixed above and behind the viewing screen which allowed recording of all the infants' fixation patterns, via a Beetamax video recorder.

For the Visual Preference, Visual Habituation and Concept Acquisition tasks, a Kodak carousel slide projector was used to back-project the stimuli onto the screen. The projector was situated on a small wooden table behind the screen. Each projected stimulus measured 10 by 13.5 cm. and the mid-point of each stimulus was level with the infant's eyes (see Method section). For the purpose of illustration Figure 3:2 shows an infant in the chamber, being given the visual preference task.

In the Audio-Visual Integration task a small wooden table, supporting two 10 inch black and white Sony monitors, was placed

Figure 3:1. HABITUATION CHAMBER - details and design.

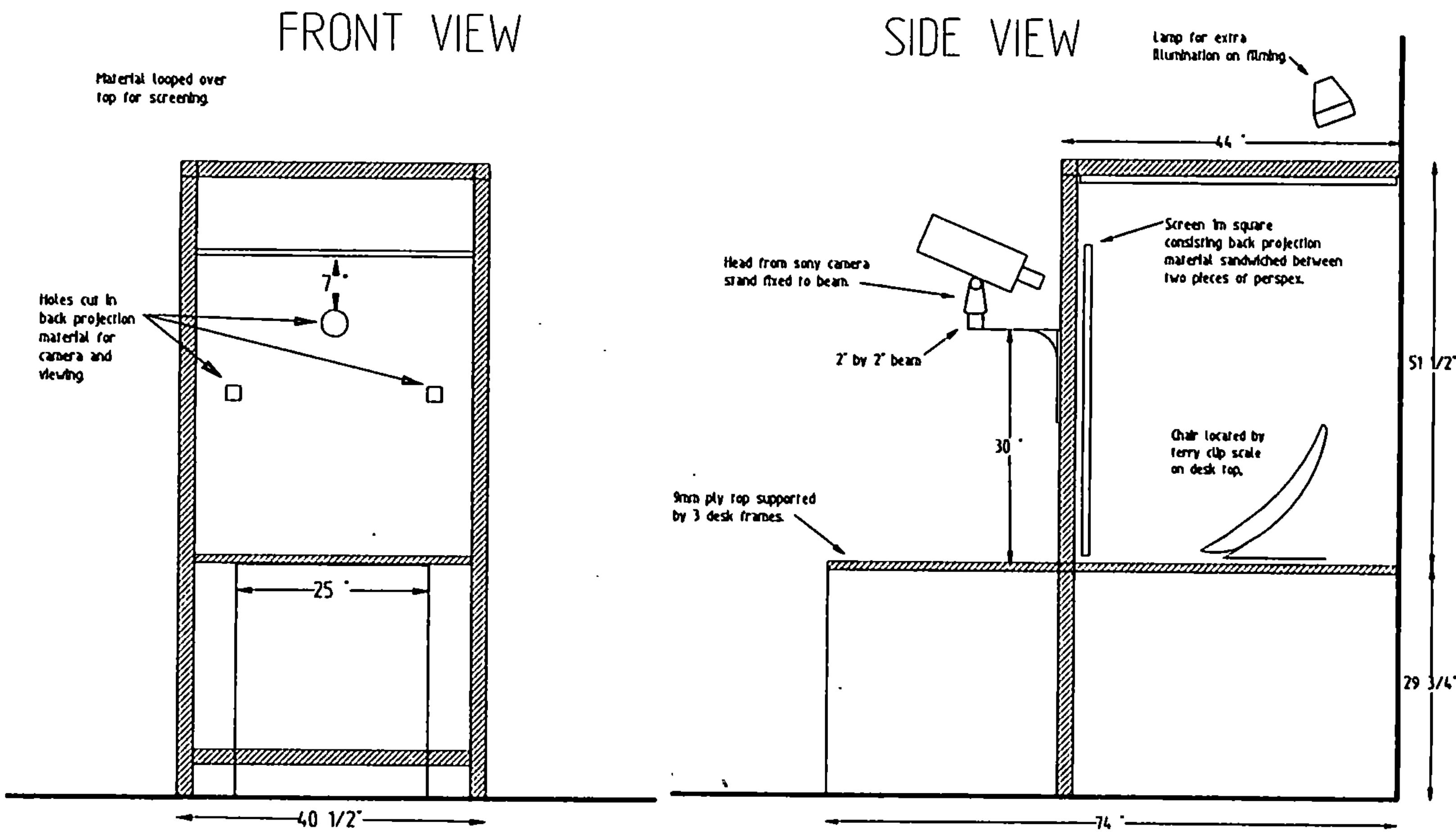
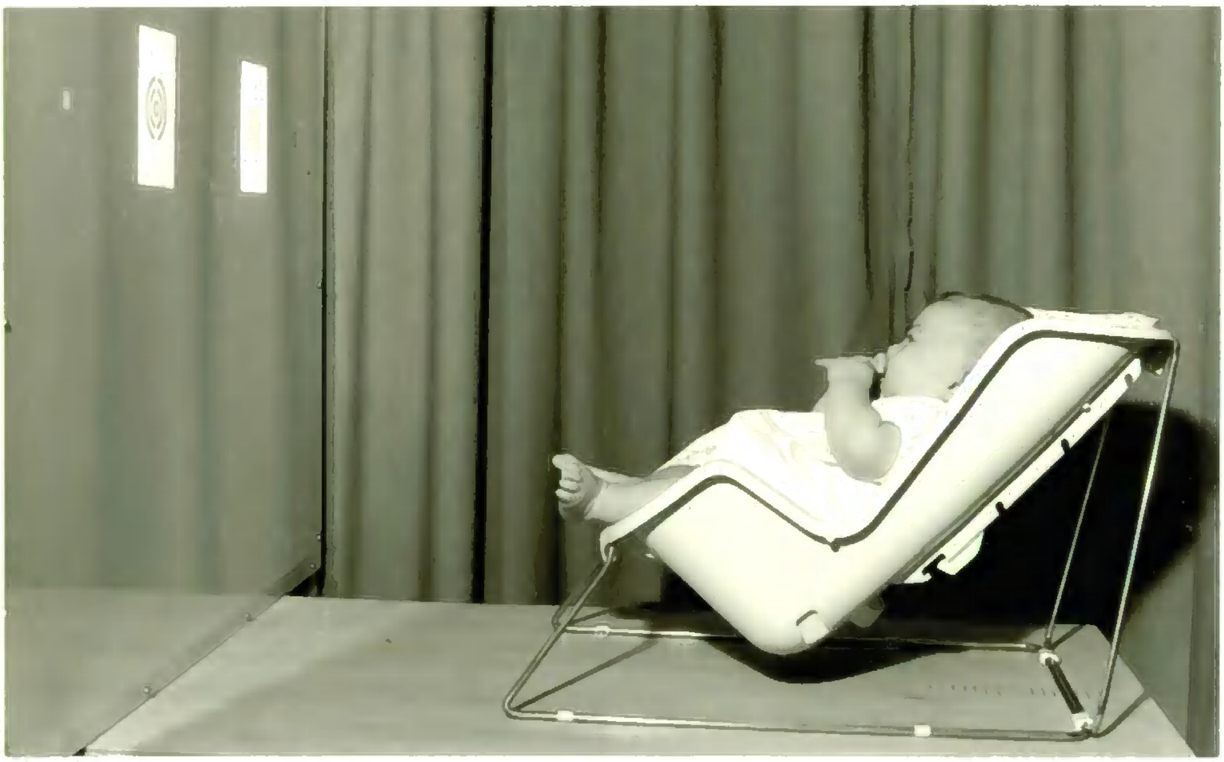


FIGURE 3:2 THE INFANT HABITUATION CHAMBER

The photograph shows an infant being given the Visual Preference task in the habituation testing chamber.



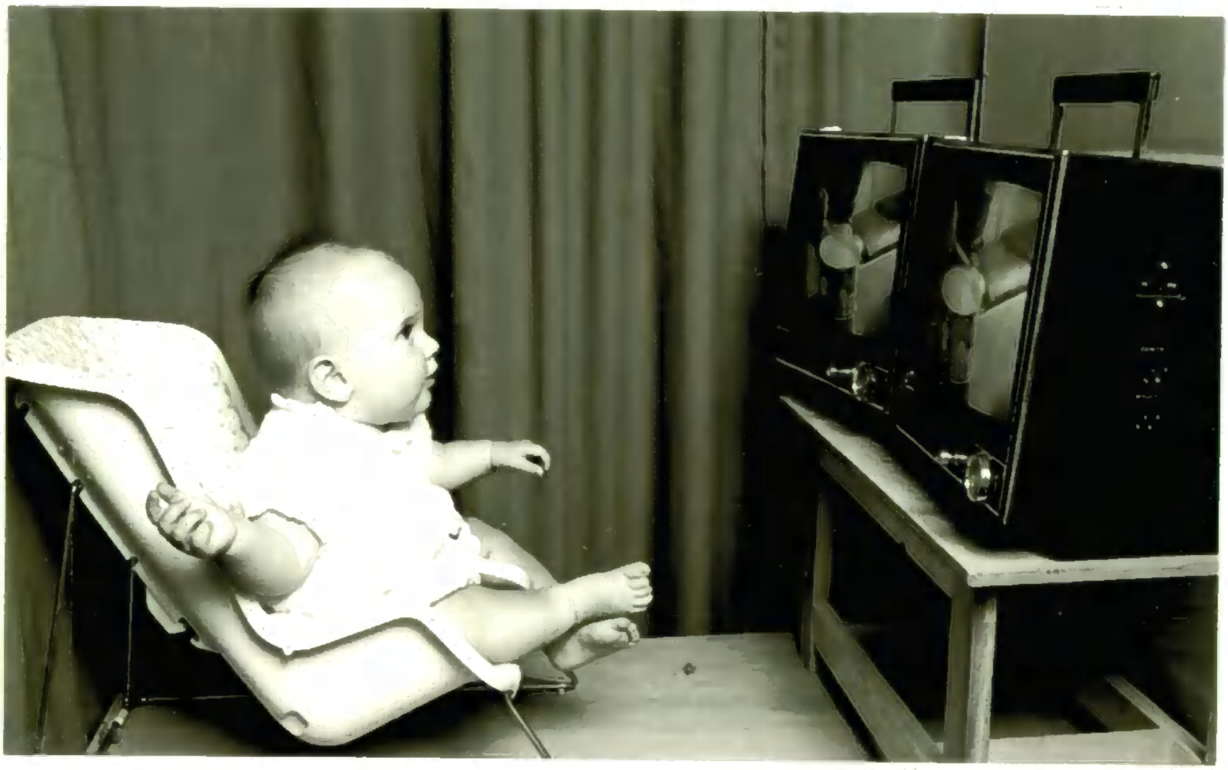
in front of the viewing screen. The monitors were placed side by side and the infant chair was repositioned such that the infant was situated at approximately 70 cm. from the monitors. The videos were presented simultaneously (see Stimuli section) and were played through two portable Panasonic recorders which were situated beneath the chamber. Each monitor had ear-plug leads connected which, via a manually controlled switch fixed behind the screen, allowed the sound track from either video to be presented through a single speaker. The speaker was also placed behind the screen in a central position, so that auditory localising cues were identical for each screen. The intensity of the sound track, as measured from the position of the infant, was approximately 65 dba. For illustration Figure 3:3 shows an infant being given this task.

3.4.3 Procedure

When mother and infant arrived in the laboratory, E initially engaged the mother in informal conversation accompanied with refreshments. Following this, E explained fully to the mother the aims of the research and details of the testing procedures. The mother was asked to remain as quiet as possible throughout testing periods. E also interacted with the infant, until the infant appeared to be in a quiet and alert state, at which point the infant was placed in the infant chair in the testing chamber. The mother remained seated in the laboratory and observed her infant throughout on the Beetamax moniter, but during the testing session the infant could not see his / her mother. Throughout testing,

FIGURE 3:3 THE AUDIO-VISUAL INTEGRATION TASK

The photograph shows an infant being given the Audio-Visual Integration task.



the room was kept at a low level of illumination, by means of a small 'angle-poise' lamp situated behind the infant in the habituation chamber.

Before any testing procedures began, the infant was presented with a neutral grey stimulus on the screen, and when the infant fixated this and seemed settled in his new surroundings, the curtain was drawn and testing began.

Between tasks the infant, if content, was left in the chamber with the curtain open, whilst the E prepared the equipment for the next task. This usually took no more than between 1 and 2 minutes. However, prior to the Audio-Visual Integration task, the infant was always removed from the chamber in order for the table and moniters to be positioned in front of the screen.

If the infant fretted between tasks, he / she was removed from the chamber and testing continued when the infant was again in a quiet and alert state.

3.4.3.1 Visual Preferences

Infants were presented pairs of stimuli by means of back-projection on to a screen. The screen and the testing chamber are described in the Apparatus section and illustrated in Figures 3:1 and 3:2.

Infants were presented with 4 stimulus pairs, and each pair was presented twice over a total of 8 trials. The left - right position of each stimulus within each pair was reversed over the two presentations in order to control for both positional preferences on the part of the infant, and scoring bias on the part of the E, due to her being situated to one side of the infant. The 4 stimulus pairs used are shown in Figure 3:4. The stripes and bullseye were both presented as black elements on a white background, and the projected size of each was 10 x 13.5 cm. They were presented symmetrically to either side of the mid point of the screen and the distance between them was 20 cm (measured from the nearer edges of the two stimuli).

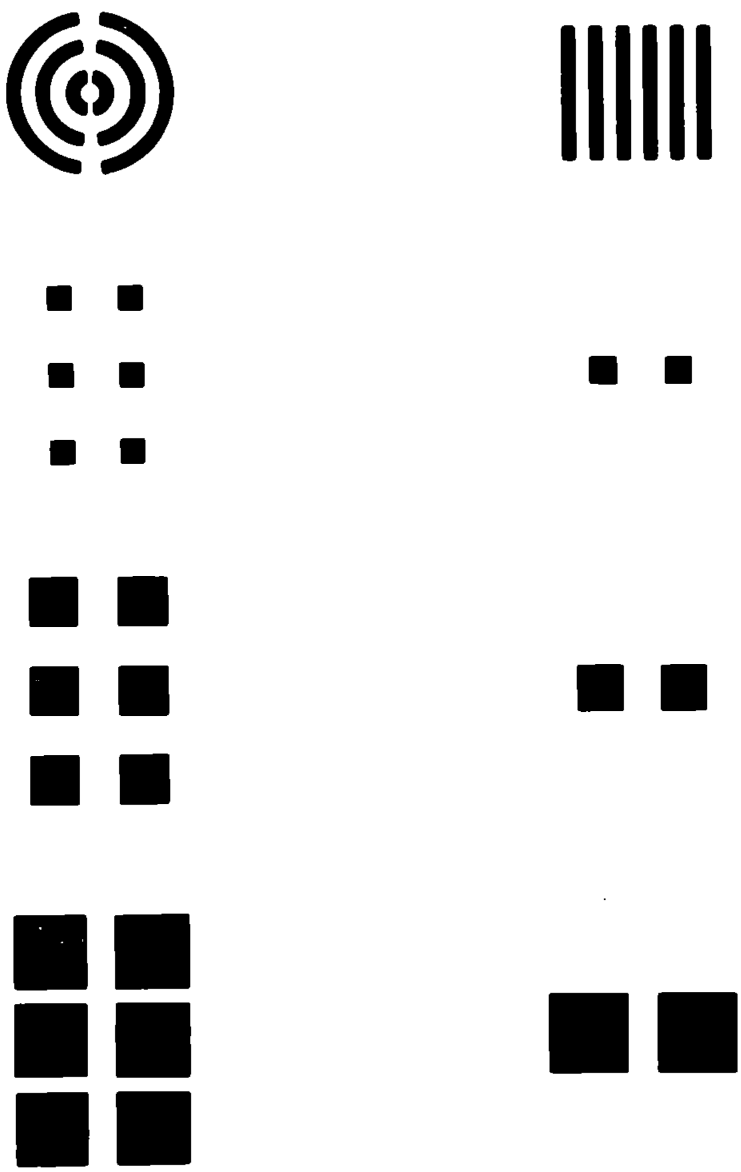
The remaining 3 pairs of stimuli were, 2 small squares paired with 6 small squares, 2 medium squares with 6 medium squares, and 2 large squares paired with 6 large squares. The projected size of these squares was: small 1.5 cm.; medium 2.5 cm.; and large 3.5 cm. When projected they were white squares on a black back-ground (the reverse of that shown in Figure 3:4).

For the second testing session all the slides were rotated 90 degrees so that the horizontal lines became vertical and the pattern of the bulls-eye changed, and the columns of squares presented in session 1, became rows in session 2.

Each stimulus pair was presented until a cumulative total of 10 seconds of fixation time towards the stimuli (either to one or to a combination of both) had been recorded. The next pair of stimuli were then presented, the inter-trial interval consisted of

FIGURE 3:4 STIMULI USED FOR THE VISUAL PREFERENCE TASK

The photograph shows each of the four pairs of stimuli presented in the task.



a slide change, and was approximately 1.5 seconds duration.

The order of presentation of the pairs of slides was selected initially at random, and maintained over all testing sessions. The sequence was as follows:

Right of midline	Left of midline
6 small squares	2 small squares
2 medium squares	6 medium squares
bulls-eye	stripes
6 large squares	2 large squares
6 medium squares	2 medium squares
stripes	bulls-eye
2 large squares	6 large squares
2 small squares	6 small squares

3.4.3.2 Visual Habituation

In this task infants were presented with single stimuli back-projected onto the screen (see apparatus section). Eight trials were presented in all, each for a period of 30 seconds (timed by the computer program) which commenced when the infant was first judged to have fixated the stimulus. The first 6 trials were repeated presentations of the same 'familiar' stimulus whilst trials 7 and 8 consisted of a different or novel stimulus. On session 1 the stimuli consisted of black and white shapes made up of both an internal and external element both of which changed from familiar to novel trials. On session two the stimuli consisted of complex black and white photographs, the familiar one of a building and the novel, one a car. The inter-trial interval consisted of a slide change of approximately 1.5 seconds duration. The projected size of the stimuli was 10 x 13.5 cm. and each subtended an angle of approximately 9.5 degrees from the infants'eyes. The stimuli are shown in Figure 3:5. The infants' fixations throughout the task were recorded.

3.4.3.3 Concept Acquisition

Infants were presented, again by means of back projection, a total of 10 single stimuli. The first 8 of these constituted a series of exemplars of a concept, while stimuli 9 and 10 represented non-concept exemplars. Each stimulus was presented for a 30 second period which commenced when the infant first fixated the stimulus. Inter-trial interval again consisted of a slide change (approximately 1.5 seconds in duration) and infants' fixations throughout the task were recorded.

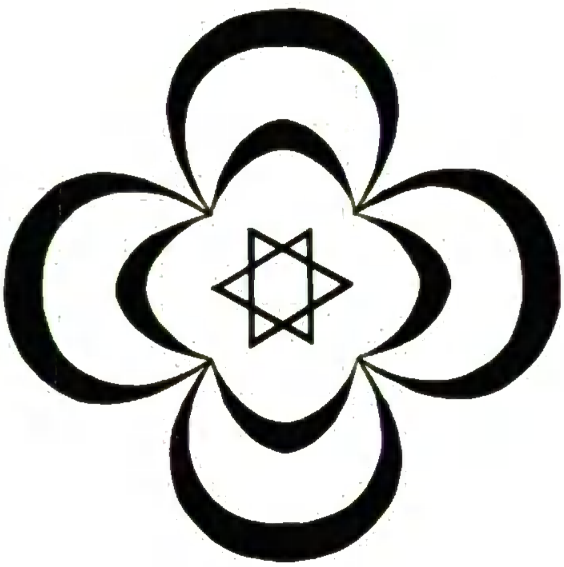
The stimuli used in this task are shown in Figure 3:6. All were presented as black figures on white backgrounds, had an overall projected size of 10 x 13.5 cm. and subtended an angle of 9.5 degrees from the infant's eyes.

The stimulus set employed on the first testing session was made up of a series of 10 slides, 8 of which represented schematic faces, each identical in every respect except for the shape of the eyes and nose of the figure. The remaining 2 slides of this set were made up of the same number and type of elements as the faces but were arranged in a non-facial symmetrical arrangement.

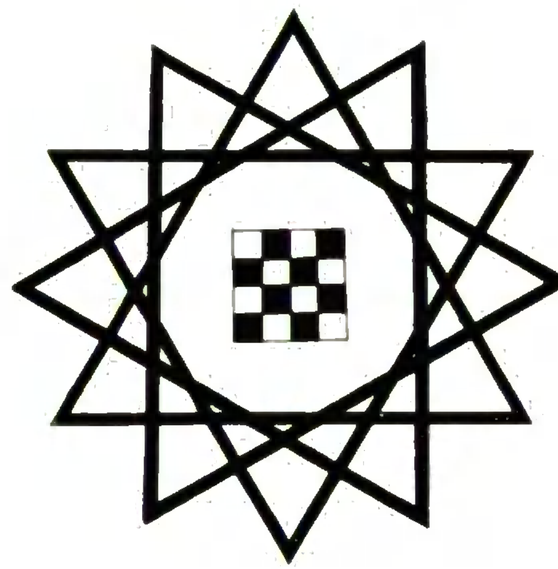
FIGURE 3:5 STIMULI USED FOR THE VISUAL HABITUATION TASK

SESSION 1

familiar



novel



SESSION 2

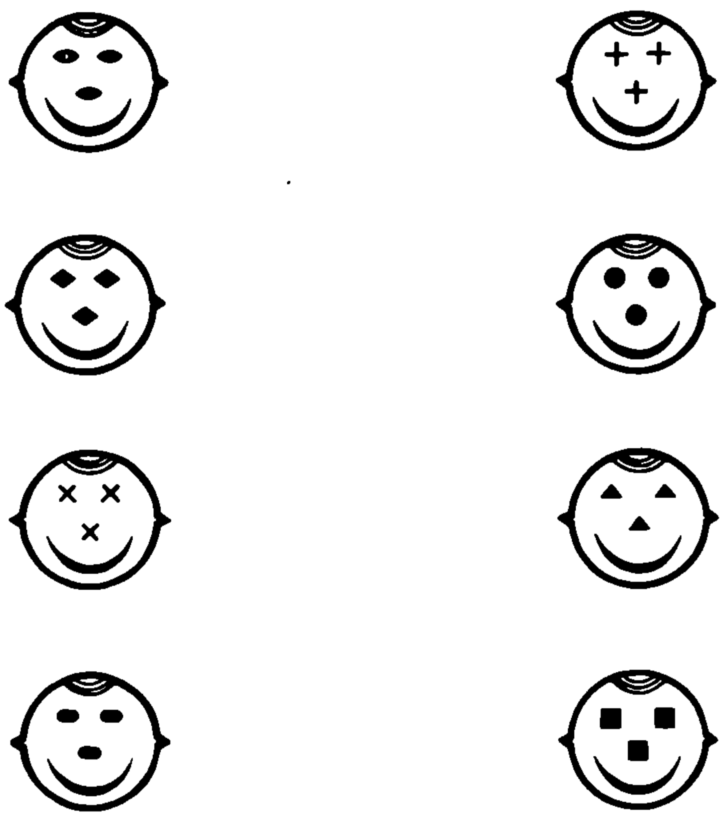
familiar

novel

FIGURE 3:6 STIMULI USED FOR THE CONCEPT ACQUISITION TASK

SESSION 1

familiar concept exemplars



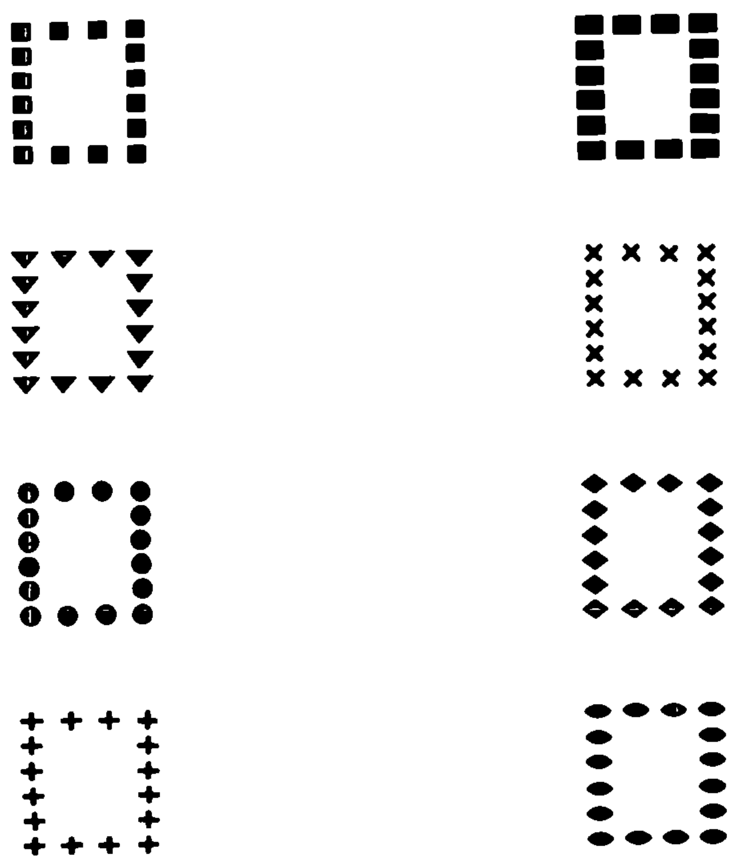
novel non-concept exemplars

continued.....

FIGURE 3:6 (continued)

SESSION 2

familiar concept exemplars



novel non-concept exemplars

Of the set of stimuli used on session 2, eight were made up of 16 identical shapes forming a rectangular arrangement with the remaining two (ie. the novel stimuli) made up of 16 elements arranged in a novel circular pattern.

3.4.3.4 Audio-Visual Integration

In this task infants were presented with two simultaneous video displays.

On the first testing session the films consisted of a hand banging a drum by means of a drum stick, at an approximate rate of 1 beat every 1.5 seconds. On the second testing session the films consisted of a hand clap also at a rate of approximately 1 clap every 1.5 seconds. Two copies were made of each video, and each lasted for about 2 minutes and 10 seconds. When presented to the infant, the two videos were started with a difference of about .7 of a second between them. This had the effect of them being out of synchrony with each other, such that as the drum stick on one came into contact with the drum, the stick on the other was high in the air, and similarly as the hands met for the clap on one video they were far apart on the other.

The duration of the display on each testing session was exactly 2 minutes, and this period was divided into four consecutive 30 second periods. Throughout the presentation, the sound track synchronised with one of the videos could be heard, this was counter-balanced across the two videos using an 'ABBA' design over the four 30 second periods. In the first 30 second period, the sound was synchronised with the video on the left, then with the stimulus on the right for the following two periods, and in the final period the sound was synchronised with the video on the left.

On session 1 all infants viewed the drum beat and on session 2 all viewed the hand clap. Throughout this task infants' fixations to the videos were recorded.

3.4.4 Measures

A large number of measures can be extracted from each of the tasks chosen for this study, and the literature to date offers no clear indication as to which measures will be the most useful for predictive research. Hence, a variety of measures was taken from each task. These included measures reflecting qualitative aspects of performance, for example, the amount of time spent fixating the bulls-eye or greater number of elements in the Visual Preference task, the amount of habituation and extent of recovery shown in the Visual Habituation and Concept Acquisition tasks, and time spent fixating the integrated video in the Audio-Visual Integration task; whilst other measures reflected quantitative measures of performance. These included measures such as the total amount of time spent fixating stimuli, the total number of fixations towards stimuli, mean fixation length and so on. From previous research, one assumes that the shorter fixation durations are exhibited by the "brighter" infants.

It is important to point out that all of these measures are not necessarily independent of each other. For example, mean fixation length is determined by measures of total fixation duration and number of fixations. This is important when interpretation of the data is considered, and this point will be raised in later sections where appropriate.

The measures extracted from each of the four tasks was as follows:

3.4.4.1 Visual Preferences

1. Percentage fixation to greater number of elements.

2. Percentage fixation to the bullseye.

3. Percentage fixation to the stimulus on the left of the infant's midline.

4. Total number of fixations: this was a measure of the total number of individual fixations made by an infant throughout the task.

5. Single longest fixation: the longest individual fixation recorded throughout the task (this measure had a maximum value of 10 seconds).

6. Mean longest fixation: the mean of the longest individual fixation in each of the 8 trials.

7. First fixation duration: this measure was the length of the very first fixation on trial 1. This measure had a maximum value of 10 seconds. (Since the stimulus on each trial was removed after 10 seconds of looking, it was not possible for the infant to look for longer than this.)

8. Total number of looks across from one stimulus to another: the number of occasions that the infant looked across from one of the stimuli to the other without in between directing attention away from the stimuli.

9. Mean completion time: this was determined by summing the time taken to complete the required 10 seconds of fixation time on each trial and dividing this by the number of trials.

3.4.4.2 Visual Habituation

A number of these measures, which are starred *, were calculated separately for familiar and novel trials.

1. Total fixation duration *: this was the total time throughout the task that infants spent fixating the stimuli. (The measure had maximum value of 240 seconds.)

2. Total number of fixations *: this was the total number of

separate fixations made to the stimuli throughout the task.

3. Mean fixation length *: the total fixation duration was divided by the total number of fixations throughout the task.

4. Longest fixation duration *: the longest individual look elicited throughout the task. (Maximum value of 30 seconds.)

5. Mean longest fixation per trial *: the mean of the longest individual fixation in each of the 8 trials.

6. First fixation duration: the length of the very first fixation on trial 1. (Maximum value of 30 seconds.)

7. Mean latency to fixate *: this was a measure of the length of time between the onset of the stimulus and the infant's first fixation towards the stimulus. The time was summed over each trial and divided by the number of trials.

8. The percentage change in fixation over presentations of the familiar stimulus: the amount of fixation time elicited on trial 6 was compared to that on trial 1 and the figure taken as a percentage increase or decrease.

9. The percentage change in fixation from the familiar to the

novel trials: the percentage difference (increase or decrease) from fixation time on trial 6 to the mean fixation time over trials 7 and 8.

3.4.4.3 Concept Acquisition

As for the visual habituation task, a number of these measures which are starred *, were also calculated separately for novel and famliar trials.

1. Total fixation duration *: this was the total time throughout the task that infants spent fixating the stimuli. (The measure had maximum value of 300 seconds.)

2. Total number of fixations *: this was the total number of separate fixations made to the stimuli throughout the task.

3. Mean fixation length *: the total fixation duration was divided by the total number of fixations throughout the task.

4. Longest fixation duration *: the longest individual look elicited throughout the task. (Maximum value of 30 seconds.)

5. Mean longest fixation per trial *: the mean of the longest individual fixation in each of the 10 trials.

6. First fixation duration: the length of the very first fixation on trial 1. (Maximum value of 30 seconds.)

7. Mean latency to fixate *: this was a measure of the length of time between the onset of the stimulus and the infant's first fixation towards the stimulus. The time was summed over each trial and divided by the number of trials.

8. The percentage change in fixation towards the familiar concept exemplars: the amount of fixation time elicited on trial 8 was compared to that on trial 1 and the figure taken as a percentage increase or decrease.

9. The percentage change in fixation time from the concept exemplars to the novel non-concept exemplars: the percentage difference (increase or decrease) from fixation time on trial 6 to the mean fixation time over trials 7 and 8.

3.4.4.4 Audio-Visual Integration

1. Percentage fixation time to the synchronised video.

2. Percentage fixation time to the video on the left of the infant's midline.

3. Total fixation time: the total time throughout the task that the infant fixated the videos (maximum value of 120 seconds).

4. Total number of fixations: the total number of separate fixations throughout the task.

5. Mean fixation length: the total fixation time divided by the total number of fixations.

6. Longest fixation duration: the longest single fixation elicited throughout the whole task (maximum value of 30 seconds).

7. Mean longest fixation duration: the longest single fixation for each 30 second period was summed over the four periods and the total was divided by 4.

8. Total number of looks across from one video to the other: the number of occasions the infant fixated one video followed directly by the other, without the infant fixating elsewhere in between the two fixations.

9. First fixation duration: the length of the first fixation elicited in the task (maximum value of 30 seconds).

3.4.5 Test-Retest Reliabilities

For all tasks, test-retest reliabilities of measures were determined by correlating the performance on the first testing session with performance on the second testing session. All of the correlations were 1-tailed Pearson's. Pearson correlation coefficients were used since this takes into account the variance between measures, as well as the degree of concordance in the ranks of the measures, over the two sessions. 1-tailed tests were used since only positive correlations can be meaningfully interpreted for the purpose of predictive studies. However, the extent to which correlation is significant is not as important as the magnitude of the correlation itself, when the reliability of measures is of interest. With a large sample size, a test-retest correlation in the region of .3 may reach significance, although an r of this magnitude hardly suggests a measure is reliable. This point is raised further in Chapter 4 and later sections where appropriate.

3.4.6 Split-Half Reliabilities

For these four tasks the split-half reliabilities for each measure (where appropriate) were obtained for each testing session. Again 1-tailed Pearson correlations were used. It is necessary here to describe the way in which these measures were calculated for each task.

3.4.6.1 Visual Preferences

As mentioned in the method section each pair of stimuli in this task was presented over two trials (to allow for counter-balancing the position of individual stimuli). The split-half reliability for this task then, was determined by taking the average score of each measure from stimulus pairs 1 to 4, and correlating this with the average of the same measure from trials 5 to 8. For example, for percentage fixation to the bullseye, the measure from the first presentation of this pair was correlated with performance on the second presentation of the pair. For percentage fixation to the greater number of elements, performance over the first 3 presentations of this comparison was averaged and correlated with performance on the second three presentations. For all remaining measures, performance on the first four trials was averaged and correlated with the average of the scores on the second four trials.

3.4.6.2 Visual Habituation

In this task there is an under-lying assumption that infants will respond differently to the novel stimuli than to the previously presented familiar stimuli. Therefore, split-half reliabilities were calculated independently for the familiar and novel stimulus sets.

For the familiar stimuli performance on trials 1 to 3 was combined and correlated with performance on trials 4 to 6. (See Note 1, below) For the novel trials, performance on trial 7 was correlated with that on trial 8. For some measures in this task, for example percentage decrease in fixation to familiar trials and subsequent percentage increase to novel, split-half reliabilities are not measurable.

3.4.6.3 Concept Acquisition

Since this task also adopts an habituation procedure, split-half reliabilities were again calculated independently for the familiar and novel stimulus sets.

For the familiar concept exemplars, performance on trials 1 to 4 was combined and correlated with performance on trials 5 to 8 (see Note 1). For the novel trials, performance on trial 9 was correlated with that on trial 10. Again split-half reliabilities are not possible for measures of percentage decrease in fixation to familiar trials and subsequent percentage increase to novel.

Note 1. It necessary to point out here that a change in responding is also expected over familiar stimulus trials. However, if habituation is a stable characteristic of infant performance, then the decline in fixation might be consistent over trials, and hence we may expect within session reliability to hold over trials in the habituation sequence. It is therefore meaningful to measure split-half reliabilities by comparing performance on the first half of habituation trials with that on the second half of the trials.

3.4.6.4 Audio-Visual Integration

This task was administered in a continuous sequence of four 30 second periods. Both the first minute and second minute periods had a synchronised sound sequence with each of the two videos. In order to calculate the split-half reliabilities, performance during the first minute of the task was correlated with performance on the second.

3.4.7 RESULTS AND DISCUSSION

3.4.7.1 Inter- and Intra-Observer Reliability

Inter-observer reliability was obtained by recruiting 4 female under-graduate students for the purpose of recording infants' fixation patterns throughout tasks, from the Beetamax video recordings. Each student scored the data for all infants on one particular task. Reliabilities were determined by means of Pearson correlations between measures derived from students recordings and the scores as measured by E during the experimental session, for each subject on each trial. This was done on a selection of measures on each task.

Also intra-observer reliabilities were obtained by E (the author) who rescored all of the pilot data from the video recordings. The mean and ranges for all reliabilities, for each task are presented in Table 3:1. All of the r's, for both inter- and intra-observer reliabilities, were reasonably high, although intra-observer reliabilities yielded consistently higher correlations than did inter- reliabilities. This was not due to familiarity of

TABLE 3:1

INTER- AND INTRA-OBSERVER RELIABILITIES

TASK	INTER RELIABILITY AVERAGE r (RANGE)	INTRA RELIABILITY AVERAGE r (RANGE)
VISUAL PREFERENCES	.76 (.36 to .98)	.87 (.65 to .99)
VISUAL HABITUATION	.83 (.39 to .99)	.91 (.79 to .99)
CONCEPT ACQUISITION	.83 (.48 to .98)	.94 (.79 to 1.0)
AUDIO-VISUAL INTEGRATION	.87 (.61 to .98)	.92 (.79 to 1.0)

individual scores, since the rescoring by E was done several months after the initial testing period, and at a time when the longitudinal study was in progress. The higher intra-observer reliabilities were probably due to the greater prior experience of the E compared to students who acted as observers.

3.4.7.2 Sex Differences

Preliminary analyses revealed no major sex differences for any of the tasks, the results presented here were therefore collapsed across sex.

3.4.7.3 Visual Preferences

A total of 3 male and 9 female infants completed this task successfully on two testing occasions. The data from a further 5 infants were eliminated due to equipment failure.

Descriptive statistics for all measures are presented in Table 3:2. It is clear that a good range of individual differences are emerging from the measures in this task.

The data were first analysed to determine whether the expected preferences emerged (ie. for the bullseye over the stripes, and for the greater versus the smaller number of elements). The percentage fixation time to the bullseye was averaged over the two presentations within each session, as was the percentage fixation to the greater number over the 6 presentations. These measures

TABLE 3:2

VISUAL PREFERENCES PILOT STUDY DATA

DESCRIPTIVE DATA N=12

MEASURE	SESSION	MEAN	SD	RANGE
% FIXATION GREATER NUMBER	1	63.29	12.23	49.5 to 87.66
	2	61.82	14.30	30 to 82.2
% FIXATION BULLS-EYE	1	58.75	26.82	0 to 100
	2	60.19	17.17	38 to 86.9
% FIXATION LEFT	1	65.74	18.71	28.6 to 91.7
	2	60.51	21.48	22.5 to 98.1
TOTAL NUMBER FIXATIONS	1	30.17	15.04	14 to 65
	2	37.17	15.3	20 to 69
LONGEST FIXATION DURATION	1	6.10	3.47	1.308 to 10
	2	5.51	3.16	1.108 to 10
MEAN LONGEST FIXATION	1	4.89	3.94	1.375 to 10
	2	3.99	4.11	1.113 to 10
FIRST FIXATION DURATION	1	2.91	4.01	0.918 to 10
	2	3.12	6.31	1.063 to 10
NUMBER FIX ACROSS STIMULI	1	19.80	11.90	4 to 59
	2	24.60	14.70	0 to 62
MEAN COMPLETION TIME	1	20.83	15.33	12.212 to 68.775
	2	23.73	13.42	12.975 to 54.187

were then analysed to determine whether they differed significantly from 50 percent, by means of two tailed t-tests for related samples. On session 1 the bullseye preference was not significant ($t=1.303$ NS) but the preference did reach significance on session 2 ($t=2.52$ $p<.005$). For the greater versus smaller number comparison, the opposite pattern emerged, while a significant preference emerged for the greater number on session 1 ($t=4.968$ $p<.001$), no significant preference emerged for the second testing session ($t=1.151$ NS).

The test-retest and split-half reliabilities of the measures from this task are presented in Table 3:3.

Only the percentage fixation to the bullseye and to the stimulus on the left of the infants' mid-line showed reliability both across and within sessions. The only other measure to show reliability was average completion time, although this was only reliable within sessions which suggests that this measure is characterised by transient state characteristics, as opposed to characterising a stable trait characteristic of infant performance.

The preference for greater number of elements found in session 1 disappeared on session 2, and this may indicate a ceiling effect emerging on session 1 which was lost on session 2 (perhaps due to infants having remembered the greater number stimuli from session 1 particularly well, and hence showing a novelty - ie. the smaller number of stimuli - preference on session 2). However since from

TABLE 3:3

VISUAL PREFERENCES PILOT STUDY DATA

TEST-RETEST (TRT) AND SPLIT-HALF (SH) RELIABILITIES N=12

MEASURE	TRT	SH SESS. 1	SH SESS. 2
% FIXATION GREATER NUMBER	-.02	.42	-.14
% FIXATION BULLS-EYE	.59 *	.43	.61 *
% FIXATION LEFT	.59 *	.82 **	.81 **
TOTAL NUMBER FIXATIONS	-.07	.46	.32
LONGEST SINGLE FIXATION	.36	.26	-.01
MEAN LONGEST FIXATION	.35	-.06	.32
FIRST FIXATION DURATION	.21		
NUMBER LOOKS ACROSS STIMULI	.17	.43	.26
MEAN COMPLETION TIME	.08	.83 **	.73 **

* p < .05 (1-Tailed)
** p < .01 (1-Tailed)

Table 3:2, it can be seen that the percentage preference for greater number declines from 63.29 to 61.82 from session 1 to session 2, there is no indication of a disappearing ceiling effect across the two sessions. The effect was more likely due to the slightly greater range in scores on session 2. Furthermore, since the measure of preference for greater number is also not reliable, this comparison was eliminated from the major longitudinal study at 3 months.

The stripe-bullseye comparison on the other hand, did show significant test-retest reliability, and split-half reliability on session 2, although split-half reliability on session 1 just failed to reach significance. However, since it was clear that all infants did not show a strong bullseye preference, it seemed acceptable to include this comparison in the major study.

Conclusion

The reliabilities from this task are on the whole disappointing, with only the stripe-bullseye comparison producing a clear indication of reliability. Hence, only this comparison was considered potentially useful at 3 months for the major longitudinal study.

3.4.7.4 Visual Habituation

A total of 7 males and 5 females completed this task on two testing sessions within one week of each other. Data from a further 20 infants were not used, primarily due to equipment failure or errors by the experimenter (in terms of presentation of an incorrect slide sequence) (n=14), with the data from the remaining 6 infants being eliminated due to fussing (n=5) and interruption from the mother (n=1).

Descriptive statistics for the measures from this task, for all measures combined, are presented in Table 3:4. Again measures from this task are showing considerable variability across infants.

Test-retest and split-half reliabilities are presented in Table 3:5 (A through C). It can be seen from this table that none of the measures produced significant test-retest reliability, although the split-half reliabilities did show greater evidence of reliability (see Tables 3.5B and 3.5C). This implies that whilst infants did show a greater degree of stability in performance within a session, no stability across sessions emerged. These measures then would not appear to be useful potential predictors of later functioning. These findings of course do not preclude the possibility that different habituation procedures, possibly with different stimuli, might give good test-retest and split-half reliabilities.

Further analyses were conducted to determine whether infants did overall produce the expected pattern of fixation to the stimuli (ie. a decrease in fixation time to the familiar stimuli with a subsequent increase to the novel). The mean fixation durations

TABLE 3:4

VISUAL HABITUATION PILOT STUDY DATA

DESCRIPTIVE DATA N=12

MEASURE	SESSION	MEAN	SD	RANGE
TOTAL FIXATION DURATION	1	135.44	56.56	17.276 to 222.378
	2	136.30	44.42	71.123 to 218.885
TOTAL NUMBER FIXATIONS	1	22.92	6.17	15 to 31
	2	33.00	16.35	11 to 77
MEAN FIXATION DURATION	1	6.37	3.45	1.152 to 13.898
	2	6.37	3.45	1.319 to 19.896
LONGEST FIXATION DURATION	1	14.3	6.0	1.29 to 29.952
	2	12.80	4.2	0.904 to 29.952
MEAN LONGEST FIXATION DURATION	1	12.4	6.2	1.17 to 27.321
	2	13.6	5.7	1.763 to 26.315
FIRST FIXATION DURATION	1	4.2	14.5	0.713 to 29.952
	2	3.2	12.9	0.912 to 29.952
LATENCY TO FIXATE	1	5.67	5.55	1.5 to 17.587
	2	4.43	2.1	1.42 to 16.57
% DECLINE FIXATION TRIAL 1- FIXATION TRIAL 6	1	-9.24	443.09	-1492.85 to 82.36
	2	21.89	52.2	-92.08 to 87.49
% RECOVERY FIXATION TRIAL 6- FIXATION NOVEL	1	0.63	28.63	-21.36 to 64.55
	2	11.48	50.13	-185.0 to 30.28

TABLE 3:5

VISUAL HABITUATION PILOT STUDY DATA

TEST-RETEST (TRT) AND SPLIT-HALF (SH) RELIABILITIES N=12

MEASURE	TRT
3.5A ALL TRIALS COMBINED	
TOTAL FIXATION DURATION	.42
TOTAL NUMBER FIXATIONS	.02
MEAN FIXATION DURATION	.29
SINGLE LONGEST FIXATION DURATION	.39
MEAN LONGEST FIXATION DURATION	.37
FIRST FIXATION DURATION	-.20
MEAN LATENCY TO FIXATE	-.23
% DECREASE (TRIALS 1 TO 8)	-.15
% INCREASE (TRIALS 8 TO NOVEL)	-.16

MEASURE	TRT	SH SESS. 1	SH SESS. 2
3.5B FAMILIAR TRIALS			
TOTAL FIXATION DURATION	.39	.26	.32
TOTAL NUMBER FIXATIONS	-.06	.62 *	.48
MEAN FIXATION DURATION	.19	.41	.46
SINGLE LONGEST FIXATION DURATION	.40	.47	.42
MEAN LONGEST FIXATION DURATION	.35	.41	.36
MEAN LATENCY TO FIXATE	.02	-.14	.06
3.5C NOVEL TRIALS			
TOTAL FIXATION DURATION	.31	.27	.46
TOTAL NUMBER FIXATIONS	-.10	.41	.09
MEAN FIXATION DURATION	.01	.30	.17
SINGLE LONGEST FIXATION DURATION	.38	.41	.21
MEAN LONGEST FIXATION DURATION	-.17	.21	.19
MEAN LATENCY TO FIXATE	-.01	.14	-.21

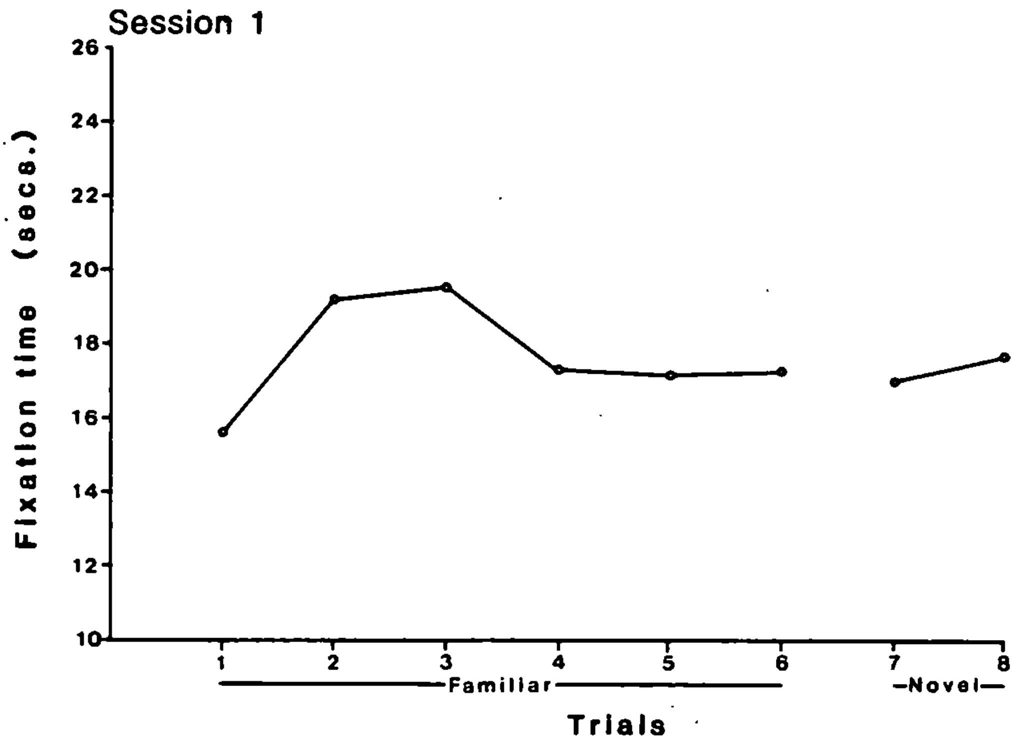
* $p < .05$ (1-tailed)

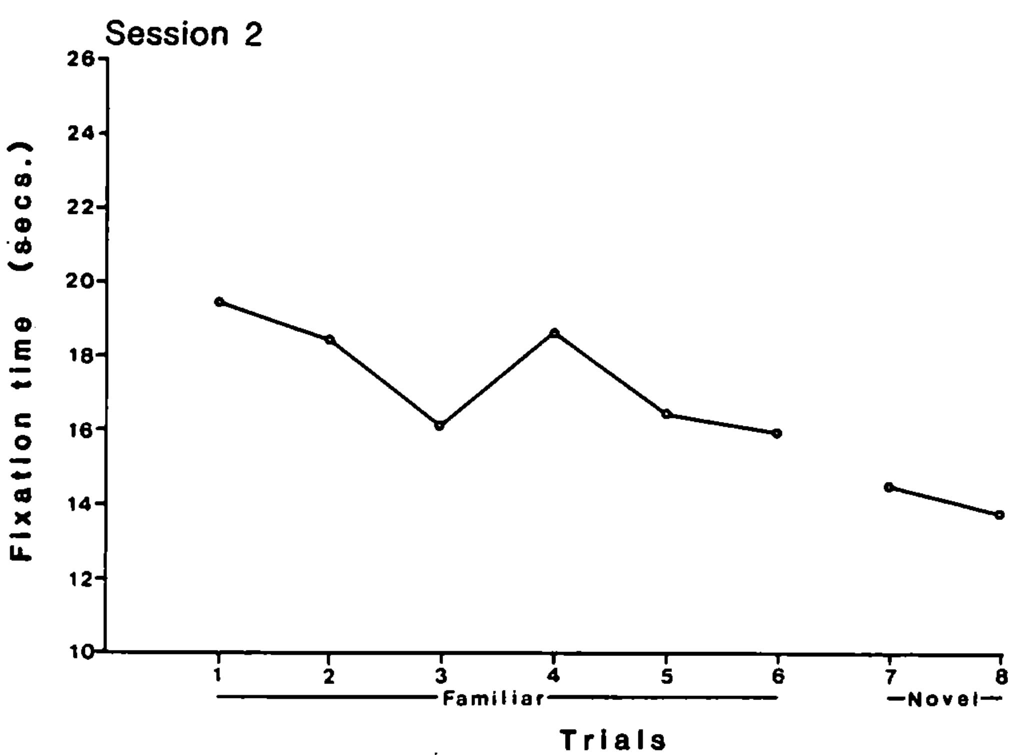
for all 12 infants, on each trial are shown in Figure 3:7. Trend analyses were used to determine whether there was any significant decrease in responding over trials. On neither session did these reach significance ($F= 0.026$ ($df= 1,55$) and 2.124 ($df=1,55$), NS for sessions 1 and 2, respectively). To determine whether there was a significant decline from the initial familiar trials to the later familiar trials, disregarding fixation measures from trials in the middle of the sequence, t-tests for related samples were performed on the difference between the average fixation duration on trials 1 and 2, with the average fixation duration on trials 6 and 7. On neither session was a significant decline found ($t= -0.24$ and 1.25 on the two sessions, respectively). Furthermore, t-tests comparing the average fixation time on trials 5 and 6, with that on trials 7 and 8, revealed that neither was there a significant increase to the novel stimulus ($t= 0.04$ and -0.99 on sessions 1 and 2, respectively).

Conclusion

With the particular design in this laboratory, infants failed to demonstrate the expected decline in fixation to the familiar stimulus, and neither was there a significant increase in fixation to the novel trials. This, coupled with the failure of the measures to show test-retest reliability leads to the conclusion that the habituation task, at least with these stimuli and this equipment, is not a particularly useful task for predictive purposes. This does not necessarily imply that this task under any circumstances is not useful for this type of

Figure 3:7. VISUAL HABITUATION - PILOT STUDY.
The figure shows the average total fixation duration for each trial. n=12





research, but for the purposes of this particular research poject, and the time limits imposed, it was madę priority to pursue the more encouraging tasks.

3.4.7.5 Concept Acquisition

4 males and 8 females completed this task, whilst data from a further 10 infants were eliminated (6 infants due to fussing and 4 due to equipment failure).

Descriptive statistics for the measures in this task, for all trials combined, are presented in Table 3:6. Again reasonable variation is found between individuals for all measures.

Test-retest and split-half reliabilities are presented in Table 3:7 (A through C). These data show considerably more test-retest and split-half reliabilty than was found in the visual habitaution task. Total fixation duration, longest single fixation duration and mean longest fixation duration emerge as moderately reliable measures. Furthermore, in many cases these measures are accompanied by high split-half reliabilities. These are most frequent within familiar trials, but since familiar trials do make up four-fifths of all the trials on this task, high within session reliability across these trials, is a good indication of within session stability.

The measures showing moderate test-retest and split-half reliabilities in this task, reflect quantitative measures of

TABLE 3:6

CONCEPT ACQUISITION PILOT STUDY DATA

DESCRIPTIVE DATA N=12

MEASURE	SESSION	MEAN	SD	RANGE
TOTAL FIXATION DURATION	1	211.78	60.192	119.56 to 295.072
	2	216.74	64.27	115.556 to 296.277
TOTAL NUMBER FIXATIONS	1	27.58	10.99	14 to 44
	2	30.67	15.66	16 to 58
MEAN FIXATION DURATION	1	9.68	6.88	2.85 to 20.67
	2	8.16	5.48	2.84 to 18.52
LONGEST FIXATION DURATION	1	17.40	7.10	0.536 to 29.952
	2	15.3	10.57	0.15 to 29.952
MEAN LONGEST DURATION	1	14.33	6.69	5.837 to 25.883
	2	13.00	6.86	4.911 to 26.856
FIRST FIXATION DURATION	1	5.76	15.9	1.69 to 29.952
	2	12.78	13.94	0.712 to 29.952
LATENCY TO FIXATE	1	7.03	7.99	0.73 to 23.344
	2	3.06	2.62	0.78 to 10.81
% DECLINE FIXATION TRIAL 1- FIXATION TRIAL 8	1	0.65	9.96	-11.59 to 21.6
	2	1.89	10.73	-21.24 to 18.174
% RECOVERY FIXATION TRIAL 8- FIXATION NOVEL	1	-0.98	10.13	-24.89 to 14.7
	2	-2.05	7.77	-17.2 to 11.977

TABLE 3:7

CONCEPT ACQUISITION PILOT STUDY DATA

TEST-RETEST AND SPLIT-HALF RELIABILITIES N=12

MEASURE	TRT
3.7A ALL TRIALS COMBINED	
TOTAL FIXATION DURATION	.52 *
TOTAL NUMBER FIXATIONS	.45
MEAN FIXATION DURATION	.21
SINGLE LONGEST FIXATION DURATION	.54 *
MEAN LONGEST FIXATION DURATION	.53 *
FIRST FIXATION DURATION	.03
MEAN LATENCY TO FIXATE	-.02
% DECLINE (TRIALS 1-8)	-.14
% RECOVERY (TRIALS 8-NOVEL)	.02

MEASURE	TRT	SH SESS. 1	SH SESS. 2
3.7B FAMILIAR CONCEPT EXEMPLAR TRIALS			
TOTAL FIXATION DURATION	.56 *	.63 *	.59 *
TOTAL NUMBER FIXATIONS	.59 *	.51 *	.34
MEAN FIXATION DURATION	.33	.41	.41
SINGLE LONGEST FIXATION DURATION	.52 *	.41	.52 *
MEAN LONGEST FIXATION DURATION	.53 *	.46	.53 *
MEAN LATENCY TO FIXATE	-.03	-.14	.16
3.7C NOVEL NON-CONCEPT EXEMPLAR TRIALS			
TOTAL FIXATION DURATION	.01	.32	.42
TOTAL NUMBER FIXATIONS	-.18	-.01	-.01
MEAN FIXATION DURATION	.01	-.02	.23
SINGLE LONGEST FIXATION DURATION	.38	.41	.21
MEAN LONGEST FIXATION DURATION	.42	.39	.40
MEAN LATENCY TO FIXATE	.39	.45	.27

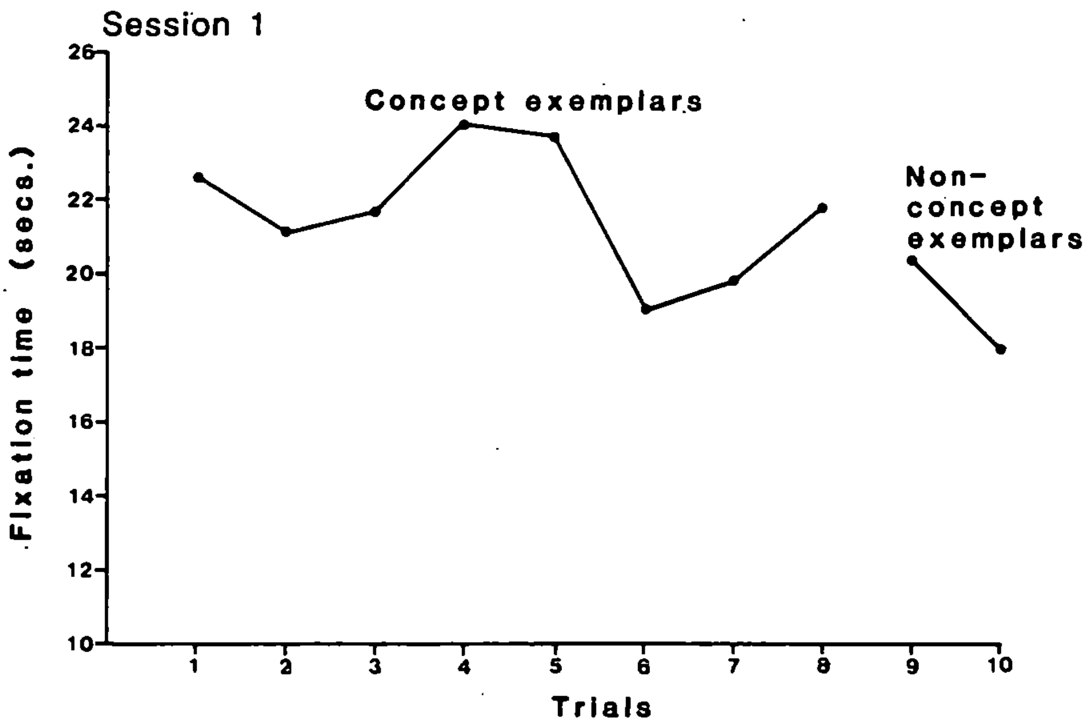
* $p < .05$ (1-tailed)

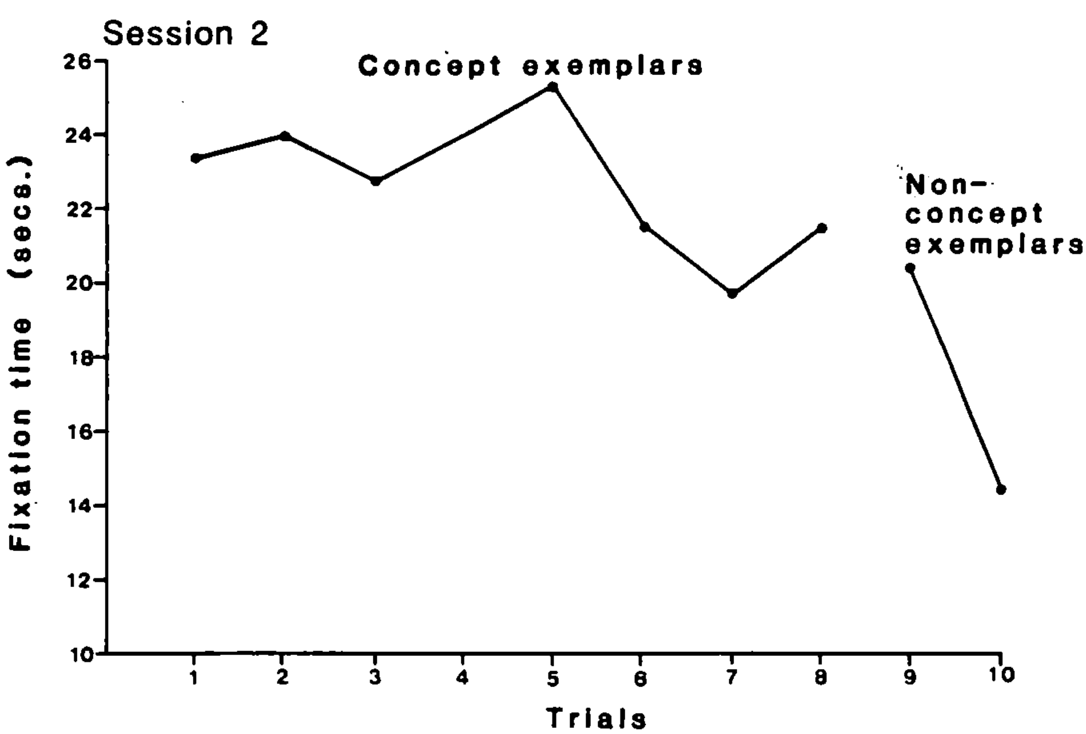
fixation; the qualitative measures of decrease from trial 1 to trial 8 and increase from trial 8 to the novel trials do not emerge as reliable across sessions. The average total fixation times across all infants for each trial on the two sessions, are presented in Figure 3:8. Trend analyses revealed that only on session 2 did infants significantly decrease in fixation to the concept exemplars ($F= .66$ ($df=1,77$) NS and 4.12 $p < .05$ ($df=1,77$), for respective sessions), although t-tests comparing the mean fixation time on trials 1 and 2 with that on trials 7 and 8 did not reveal a significant decline for either session ($t= 0.398$ and 1.29 for sessions 1 and 2, respectively). Furthermore in neither session was there an increase in responding to the novel concept exemplars. T-tests comparing the average fixation time from trials 7 and 8 with that from the two novel trials, revealed an overall decrement in fixation to the novel stimuli in both sessions, although in neither session did these reach significance ($t=-1.29$ and $t=-2.08$, sessions 1 and 2, respectively).

Conclusion

Since the expected pattern of responding was not found, it again cannot be concluded that these qualitative measures per se are not potentially reliable. However, because a number of other measures were found to be reliable from this task it was used in the longitudinal study.

Figure 3:8. CONCEPT ACQUISITION – PILOT STUDY
The figure shows the average total fixation duration for each trial. n = 12





3.4.7.6 Audio-Visual Integration

7 male and 5 female infants completed this task successfully on two occasions, while data from a further 15 infants were lost all due to equipment failure and experimenter error.

Descriptive data from the measures in this task are shown in Table 3:8.

Measures from this task again show good variability amoung infants. T-tests revealed that on neither session did infants overall show a significant preference for the synchronised video ($t = 1.29$ and -0.02 NS in both cases, for respective sessions). Neither were there any side preferences for either video ($t = -0.38$ and 0.77 NS in both cases, sessions 1 and 2, respectively).

The test-retest reliabilities for this task are presented in Table 3:9. The measure of percentage fixation to the synchronised video failed to produce any strong indication of reliability. Percentage fixation to the video on the left of the infants' mid-line was reliable both across and within sessions. The measure of total fixation duration produced reliability across sessions although within sessions the correlations just failed to reach significance. The remaining measures (ie. total number of fixations, mean fixation duration, longest single fixation duration, mean longest fixation duration, total number of looks across videos and first fixation duration) produced a high negative reliability, which is difficult to interpret, and may

TABLE 3:8

AUDIO-VISUAL INTEGRATION PILOT STUDY DATA

DESCRIPTIVE DATA N=12

MEASURE	SESSION	MEAN	SD	RANGE
% FIXATION SYNCHRONISED VIDEO	1	55.38	13.10	37.667 to 80.445
	2	49.92	7.06	30.52 to 64.149
% FIXATION LEFT VIDEO	1	46.39	32.43	2.32 to 99.295
	2	51.96	7.21	2.055 to 89.544
TOTAL FIXATION DURATION	1	110.07	12.69	83.782 to 119.806
	2	104.48	29.20	96.526 to 119.489
TOTAL NUMBER FIXATIONS	1	22.17	12.04	7 to 43
	2	33.00	13.11	11 to 53
MEAN FIXATION DURATION	1	7.47	5.83	1.948 to 19.968
	2	7.27	5.93	2.214 to 10.793
LONGEST FIXATION DURATION	1	15.90	13.77	1.036 to 29.952
	2	14.70	12.80	2.63 to 29.952
MEAN LONGEST FIXATION DURATION	1	12.60	12.71	0.897 to 26.41
	2	10.60	13.92	1.07 to 27.413
NUMBER FIXATIONS ACROSS VIDEOS	1	18.80	11.93	4 to 41
	2	21.47	15.66	2 to 48
FIRST FIXATION DURATION	1	14.7	12.01	1.376 to 28.139
	2	12.74	14.85	2.09 to 29.952

TABLE 3:9

AUDIO-VISUAL INTEGRATION PILOT STUDY DATA

TEST-RETEST (TRT) AND SPLIT-HALF (SH) RELIABILITIES N=12

MEASURE	TRT	SH SESS. 1	SH SESS. 2
% FIXATION SYNCHRONISED VIDEO	.40	.41	.36
% FIXATION LEFT VIDEO	.56 *	.86 **	.51 *
TOTAL FIXATION DURATION	.50 *	.49	.47
TOTAL NUMBER FIXATIONS	-.70	.47	.43
MEAN FIXATION DURATION	-.45	.42	.59 *
LONGEST SINGLE FIXATION DURATION	-.49	.39	.40
MEAN LONGEST FIXATION DURATION	-.37	.39	.09
TOTAL NUMBER LOOKS ACROSS VIDEOS	-.57	.38	.45
FIRST FIXATION DURATION	-.36		

* $p < .05$ (1-Tailed)
** $p < .01$ (1-Tailed)

have occurred by chance.

Conclusion

This task was included in the major longitudinal study since a number of the reliabilities are encouraging, and also due to the fact that no data were disregarded due to infant fussing. Moreover, no ceiling effects emerged in terms of preference for the synchronised video, and the task appears suitable for inclusion for the major study.

3.5 PILOT STUDIES - OPERANT CONDITIONING AND THE HOME BASED TASKS

3.5.1 Operant Conditioning Task

This task was administered in the laboratory. It was not given on both testing occasions, but was piloted with 6 infants. The method used for this task will now be described.

3.5.1.1 Apparatus

For the Operant task, the infants were placed face-up in a commercially obtained infant cot. Over the cot at a height of about 110 cm above the infant's head, was placed a wooden frame which supported a pulley-system. Essentially, this consisted of a length of ribbon which ran over the frame, one end of which was

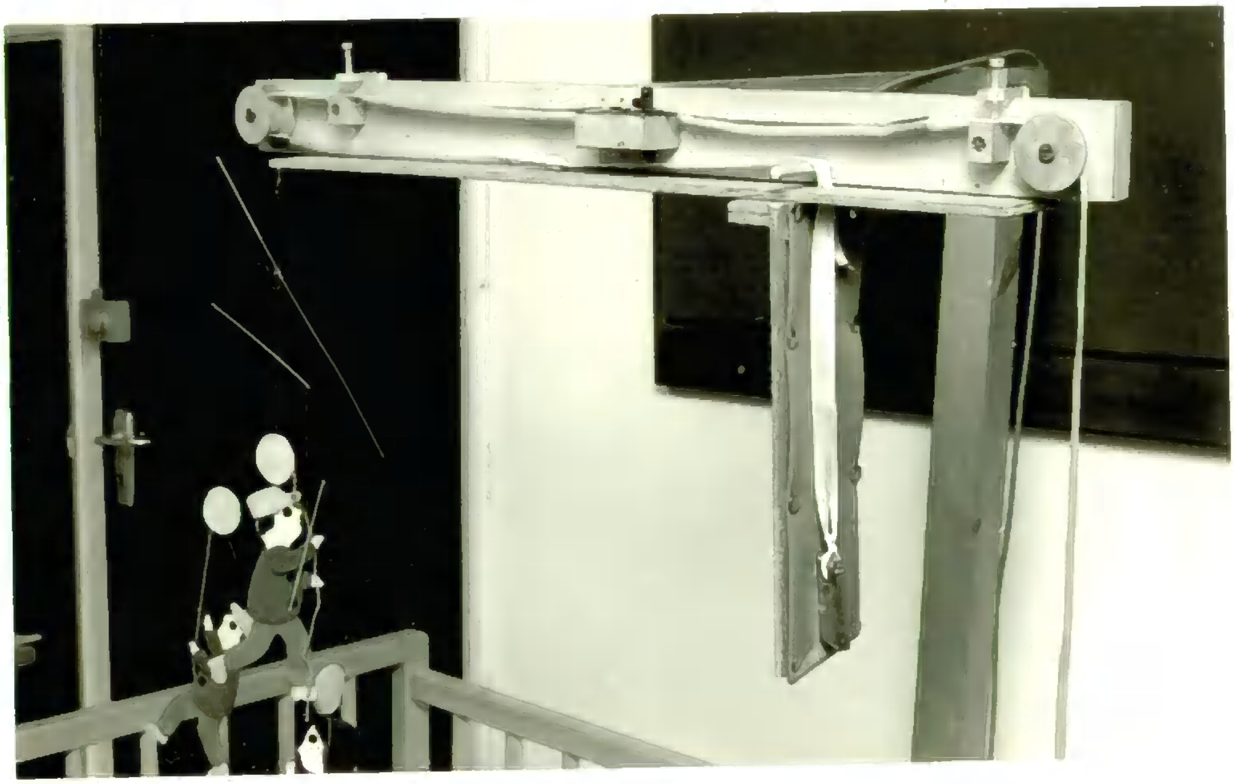
FIGURE 3:9 THE OPERANT CONDITIONING TASK

(i) The photograph shows an infant being given the Operant Conditioning task.

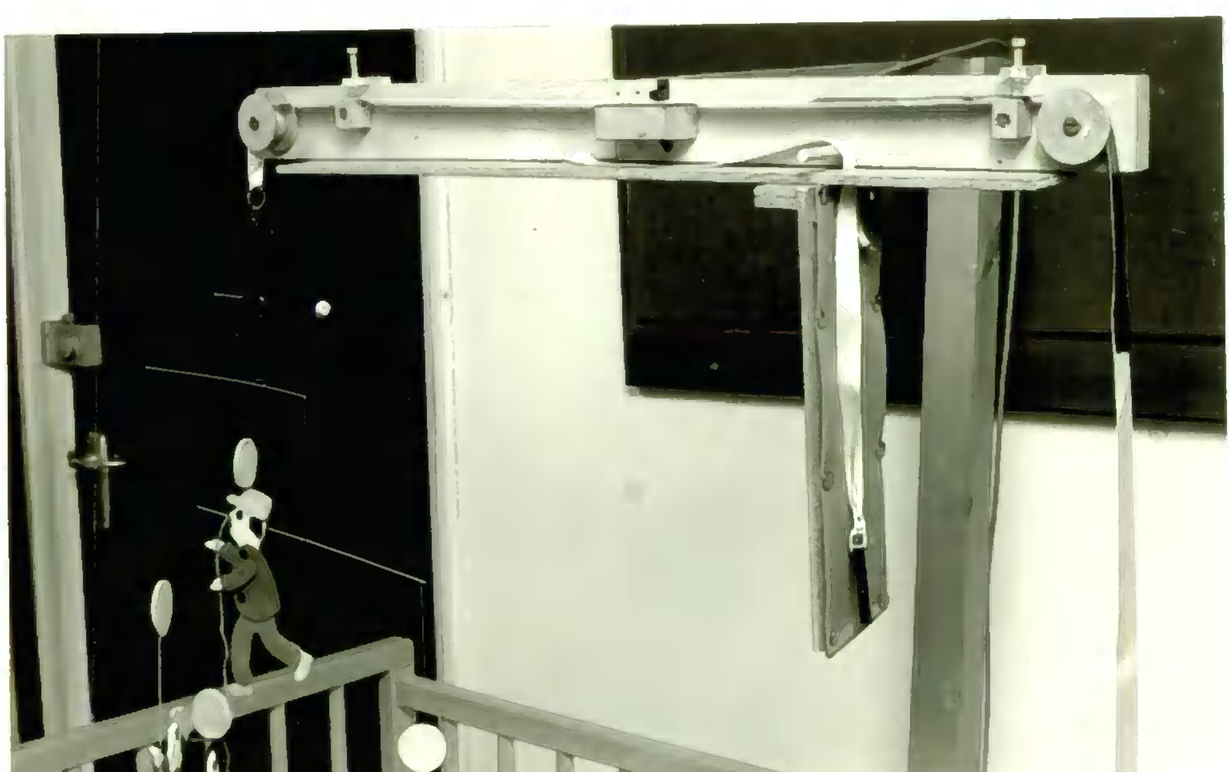
continued....

FIGURE 3:9 (continued)

(ii) The photograph shows the apparatus in the base-rate and extinction phases. The ribbon is attached to the small weight, and not to the mobile.



(iii) The photograph shows the apparatus in the reinforcement phase. The ribbon is attached to the mobile, and not to the small weight.



looped for attachment to the infant's ankle, the other end could either be attached to a mobile (the mobile hanging at a height of 70 cm above the infant's head), or to a small weight, shielded from the infant's view by means of a small ply-wood screen. This apparatus is illustrated in Figure 3:9i. The weight and the mobile produced an equivalent force on the infant's leg. When the ribbon was clipped to the mobile any leg movement by the infant caused commensurate movement of the mobile. When the ribbon was attached to the weight, the weight moved, but no apparent visible event occurred as a result of leg movement. The weight then acted as a control to ensure that infants did not kick merely due to the ribbon attached to their ankle. For illustration of the manipulations made to this equipment see Figures 3:9ii and 3:9iii.

The number of kicks elicited from the infant throughout this task was recorded by means of a small infra-red sensor, which detected and recorded on an electronic event recorder, all occasions when the edge of a piece of black tape, fastened to the ribbon, passed through the sensor. At the start of this task the ribbon was placed in such a position that the edge of the black tape passed through the sensor when a kick upwards of at least 4cm was produced by the infant. This task was also recorded on a J.V.S. reel-to-reel video recorder, the camera of which was situated at the side of the cot allowing the infant's leg movement and general direction of visual attention during the task to be recorded.

A stop-watch was used to determine the time periods throughout this task.

3.5.1.2 Procedure

The Operant Conditioning task was administered on the infant's first visit to the laboratory. The task was administered after the Concept Acquisition task, and prior to the Audio-Visual Integration task.

Infants were placed in a supine position in the infant cot. An initial base-rate of foot-kicking was established in a 2 minute period of non-reinforcement, during which the ribbon was attached to the weight. This was immediately followed by a 5 minute reinforcement phase, during which mobile movement was contingent on foot movement, and finally a further 2 minute period of non-reinforcement was given constituting an extinction phase. The number of foot-kicks produced by the infant during each minute of this task was recorded, along with the total time infants spent fixating the mobile.

3.5.1.3 Measures

A number of measures were taken which were considered to reflect patterns of acquisition of the response (increment in kicking from base-rate to learning phases, and from base-rate to extinction phases), amount of time fixating the mobile, and the general activity level of the infant (total number of kicks). The specific measures were as follows:

3.5.1.4 Results and Discussion

Six infants completed this task when they visited the laboratory. The data from one further infant were not used due to fussing.

Inter-observer reliabilities (established by correlating scores recorded by E with equivalent scores taken by a student from the video recordings), were high. The correlations for the six measures ranged from .73 to .90 (mean =.86).

The descriptive statistics for the 6 measures described above are presented in Table 3:10. The measures appear to produce considerable variability across infants. The mean number of kicks for all infants for each individual minute of testing are plotted in Figure 3:10. This shows that kicking did increase drammatically during the reinforcement phase, as compared with the base-line period and that kicking rate decreased during the extinction phase. T-tests for related samples revealed a significant increase from the average number of kicks recorded in minutes 1 and 2 to the average of kicks recorded in minutes 6 and 7 ($t=3.104 < .05$); and a significant decline in kicking from minutes 6 and 7 to the extinction phase (minutes 8 and 9) ($t=3.269$ $p<.05$).

In conclusion, since this is precisely the pattern of responding

1. Ratio of B to A: the ratio of the number of kicks in the reinforcement period (B) to that in the base-line period (A) (see Note 2).

2. Ratio of C to A: the ratio of kicks in the extinction period (C) to that in the base-rate period (A) (see Note 2).

3. Mins 6 and 7 of B / A: the total number of kicks in the final 2 minutes of the reinforcement period, divided by the number of kicks in the base-line period (see Note 2).

4. Greatest 2 mins in B/A: the greatest total number of kicks elicited over any (not necessarily consecutive) two 1 minute periods in the reinforcement period, divided by the number of kicks in the base-line period (see Note 2).

5. Total fixation time: the total time throughout the task infants fixated the mobile.

6. Total number of kicks: total number of kicks throughout the complete task.

Note 2. For the four measures reflecting aspects of learning ability, it was assumed that the greater the score, the more 'mature' the infant.

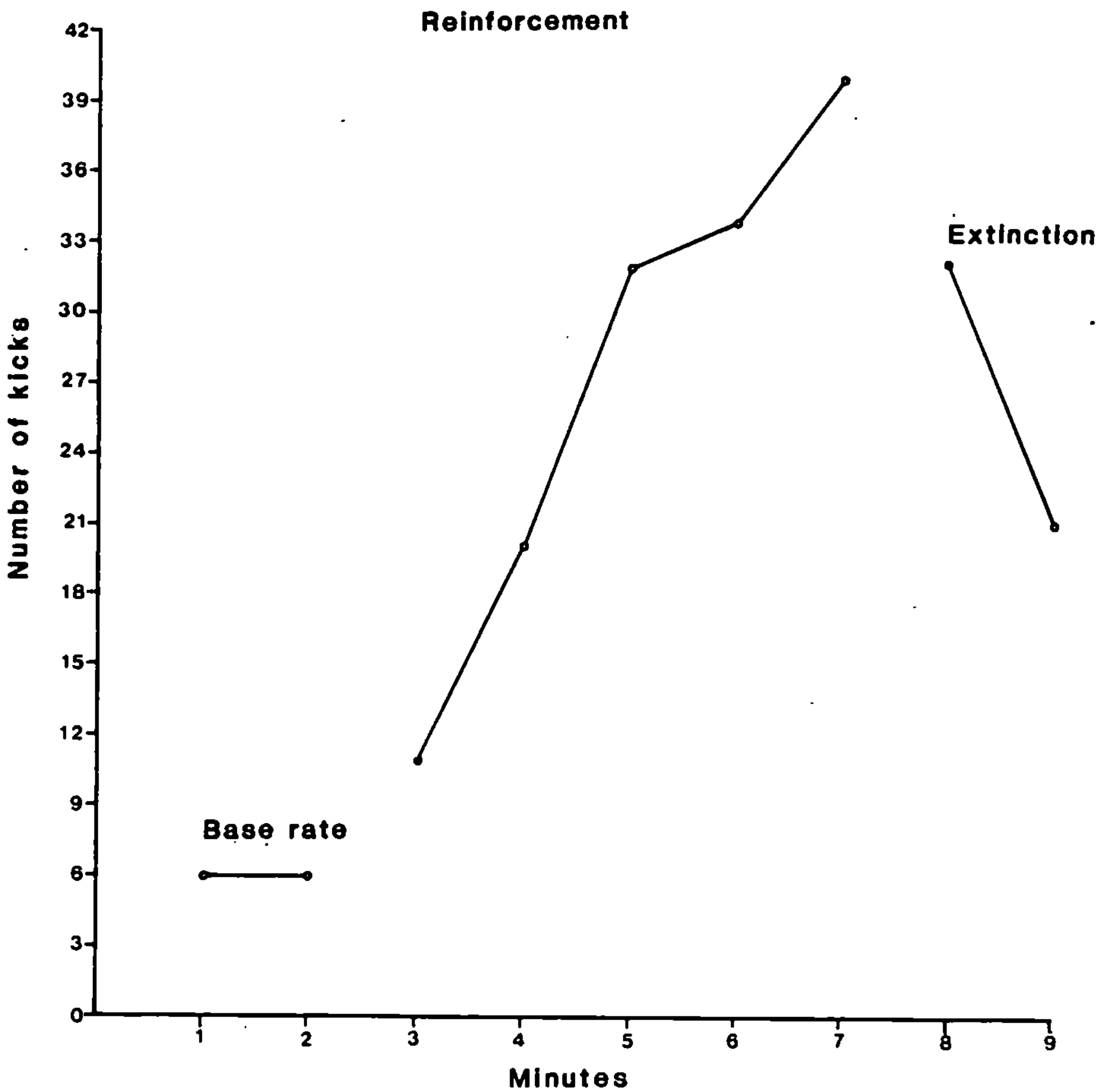
TABLE 3:10

OPERANT CONDITIONING PILOT STUDY DATA

DESCRIPTIVE DATA N=6

MEASURE	MEAN	SD	RANGE
RATIO B TO A	5.47	4.59	1.15 to 12.63
RATIO C TO A	5.49	4.55	1.12 to 13.14
LAST B / A	7.44	6.44	1.41 to 17.57
HIGHEST IN B / A	7.53	6.36	1.41 to 17.57
FIX DURATION	479.1	42.2	415.2 to 514.9
TOTAL NO KICKS	205	101	85 to 320

Figure 3:10. OPERANT CONDITIONING TASK – PILOT STUDY.
The figure shows the average number of kicks recorded in each minute of testing. n=6



one would expect from this task, it was included in the major longitudinal study.

3.5.2 Differential Vocal Responsiveness (DVR)

This task was administered in the infants' home environment. The procedure adopted for this task was identical to that described by Roe (1978).

3.5.2.1 Apparatus and Procedure

The infant was placed in his / her own infant chair or equivalent. The mother was instructed to position herself about 1 metre away from the infant and to talk to her baby for a period of 3 minutes. During this period the mother was instructed to elicit as much vocalisation from the infant as possible, and to stop vocalising if she felt the infant was preparing to vocalise, and not to vocalise concurrently with the infant if possible. During this period the E sat quietly in the same room but out of sight of the infant.

When this 3 minute period was complete, the infant was allowed to rest for a 4 minute period following which the E vocalised with the child for a further 3 minutes in a similar manner to that adopted by the mother. During this period the mother remained out of sight of the infant.

The length of time infants spent vocalising in each condition was recorded by means of a small portable cassette recorder, and the time periods were recorded with a stop-watch.

3.5.2.2 Measures

The measures taken from this task were as follows:

1. DVR measure: the amount of time spent vocalising to the stranger subtracted from that spent vocalising to the mother.

2. Total vocalisation time to the mother: the total vocalisation time throughout the task spent vocalising to the mother.

3. Total vocalisation time to the stranger.

4. Total vocalisation time (to both mother and stranger).

3.5.2.3 Results and Discussion

Six infants completed this task. They were all visited in their homes within one week of their 3 month 'birthday'.

The tape recordings of all sessions were scored by an under-graduate student, and the inter-observer reliabilities for the measures were high (ranging from .79 to .91).

The descriptive statistics for the DVR task are presented in Table 3:11. Again variation between subjects is reasonable for all measures.

In conclusion the good range of scores across subjects for all of these measures justifies the use of this task in the longitudinal study.

3.5.3 Attention to a Novel Toy

This task was also administered in the home to 6 infants following presentation of the DVR task.

3.5.3.1 Apparatus and Procedure

A 3-dimensional coloured object was presented to the infants in this task by the E who manipulated the object by turning it round slowly in order to attract the infant's attention. The infant was then allowed to fixate the object and manipulate it for as long as he / she chose.

The stimulus presented to the infants consisted of a series of different coloured wooden blocks (1 cm thick) glued together to form an object measuring 16 x 13 x 21 cm: this is shown in Figure 3:11.

A stop watch was used to determine the length of time infants fixated and manipulated the object.

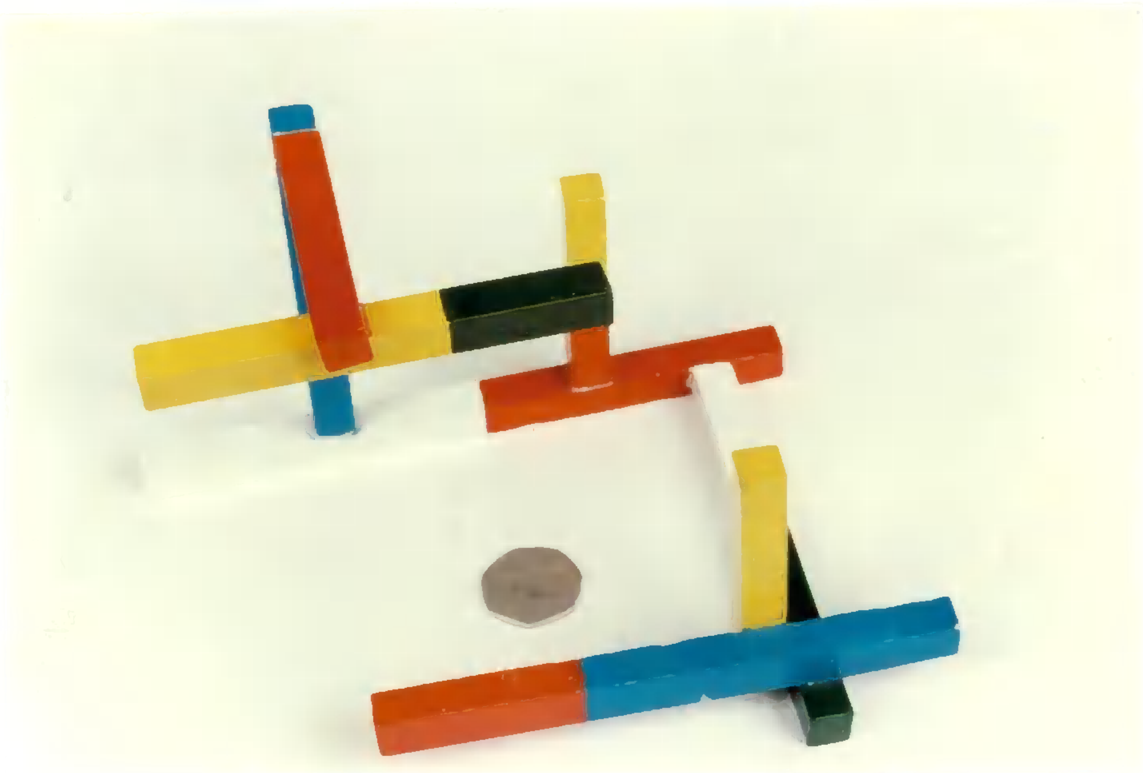
TABLE 3:11

DIFFERENTIAL VOCAL RESPONSIVENESS PILOT STUDY DATA

DESCRIPTIVE DATA N=6

MEASURE	MEAN	SD	RANGE
DVR	8.5	12.4	-3.3 to 27.9
TOTAL VOC MOTHER	19.78	6.7	12.8 to 31.6
TOTAL VOC STRANGER	11.3	7.76	2 to 20.7
TOTAL VOCALISATION	31.08	7.514	22.6 to 39.7

FIGURE 3:11 THE STIMULUS USED FOR THE ATTENTION TO A NOVEL TOY TASK



Total fixation time, total number of fixations and, where appropiate, total manipulation time were recorded.

3.5.3.2 Measures

The measures taken from this task were the following:

1. Total fixation time: the total time spent fixating the object without looking away for longer than a 2 second period. When the infant did look away for longer than 2 seconds, testing stopped.

2. Total number of fixations.

3. Average fixation duration: the total fixation duration divided by the total number of fixations.

3.5.3.3 Results and Discussion

The same six infants completed this task as the DVR task, and both tasks were completed on the same testing session.

Descriptive statistics for the 3 measures are presented in Table 3:11. Again all measures produce good variation between infants.

In conclusion since these measures appear to produce good variation across subjects the task was included in the study.

TABLE 3:12

ATTENTION TO A NOVEL TOY PILOT STUDY DATA

DESCRIPTIVE DATA N=6

MEASURE	MEAN	SD	RANGE
TOTAL FIXATION DURATION	139.17	116.65	12.6 to 321.4
TOTAL NUMBER FIXATIONS	8.43	6.21	1 to 17
MEAN FIXATION DURATION	14.676	4.624	8.13 to 19.55

3.6 PILOT STUDIES - DISCUSSION

The pilot studies allowed the experimenter to become accustomed to the experimental work, both in terms of equipment, and working with infants and their mothers. They also revealed information relating to how the infants could cope with the amount of tasks they were required to do. In fact it was apparent that since no individual infant completed more than three tasks on any experimental session (due to fussing, sleepiness, or as in many cases, experimenter error or equipment failure), the number and or length of tasks needed to be reduced.

This was achieved by eliminating those tasks, or parts of tasks, which were not producing reliable data. These included the Visual Habituation task and the greater versus smaller number of elements comparison in the Visual Preference task. These measures failed to show any indication of even moderate test-retest reliability, and the patterns of responding which they produced were not as one might expect. It is necessary to point out however, that this does not mean that these tasks will not provide potentially predictive measures, and had the experimental conditions and stimuli been altered, then more encouraging results may have emerged. However, in view of the time limitations on this particular research project they were dropped from the study.

The remaining measures and tasks all produced reasonable variation in scores between infants. Of those which were administered on 2 occasions, the test-retest and split-half reliabilities are generally disappointing. Although some of the correlations are significant, few are higher than .5, and hence account for only a quarter of the common variance. The extent to which measures are reliable in turn affects the extent to which measures can predict each other (Cohen and Cohen, 1973). Whether the correlations are significant then, is of little consequence in terms of how predictive they might be. What is important is their magnitude.

However, the infants in these pilot studies varied considerably in terms of state at the time of testing. For instance, many infants cried or fussed at some point in the testing sequence. This in part seemed due to lack of experimenter experience, since toward the end of the pilot study testing period, infants completed more tasks in a quiet and alert state. In view of this the reliabilities can be regarded as somewhat more robust, and had greater care and control been taken over the state in which infants began and completed testing sessions, then reliabilities may have been considerably higher. In view of this it was considered justifiable to continue with at least the initial part of the longitudinal study.

A total of 6 tasks then were included in the main longitudinal study;
Visual Preferences (stripe-bulls-eye comparison only), Concept Acquisition; Audio-Visual Integration, and the Operant

Conditioning tasks were administered in the laboratory; Differential Vocal Responsiveness (DVR), and Attention to a Novel Toy were given in the home. Details of the testing of infants at 3 months of age as the first part of the main study are presented in Chapter 5, along with a discussion of the ways in which measures relate to each other across tasks.

Chapter 4 is concerned with the methods and justification of the statistical analyses used with the data from this study.

CHAPTER 4

DESIGN AND ANALYSES OF THE LONGITUDINAL STUDY

4.1 INTRODUCTION

This Chapter describes the design of the longitudinal study and considers the analyses used, both within each testing age, and the across age longitudinal analyses. A description of the overall design of the longitudinal study is presented first, and this is followed by a description of the analyses used at each age and across age.

4.2 THE LONGITUDINAL STUDY

A group of infants were tested on age-appropriate tasks at 3, 5 and 9 months of age.

Six tasks were administered to the infants at 3 and 5 months of age, and these have been described in detail in Chapter 3 (any variation in procedures at each age are discussed in Chapters 5 and 6 for the 3 and 5 month ages, respectively). Most of the tasks used at these ages have been employed successfully with both 3 and 5 month old infants (see review in Chapter 2).

At 9 months two tasks were administered to the infants in their homes. One was an Object Sorting task, and the other was an Object Search task. Both of these tasks were considered to potentially reflect intellectual ability at this age, the details and justification for the choice of tasks is given in Chapter 7.

When the infants reached 18 months of age, they were visited again and were administered a number of infant intelligence scales and subscales. A strong emphasis was placed on scales and subscales reflecting linguistic abilities, since previous research has indicated that early infancy measures, of the nature used in this study, are generally more predictive of linguistic ability than of complete infant or childhood intelligence scales (see review in Chapter 2). Details of the scales used are given in Chapter 8, and the measures constituted the criterion measures for the longitudinal analyses.

4.3 SUBJECTS

The total number of subjects used in this study was 29. The number was limited by time and financial constraints. In view of the relatively large number of measures taken from infants at each age of testing, this comparatively low number of subjects used does limit the statistical analyses that one can adopt, particularly in terms of the within age and longitudinal analyses. This issue is discussed further in section 4.4.

4.4 ANALYSES

A brief description of the types of analyses adopted at each stage of the research, and justification for their use, is given in the following sections. Test-retest and split-half reliabilities are considered first.

4.4.1 Test-retest and split-half reliability

Any test or measure cannot predict any other test or measure better than it can predict itself. The reliability of a measure sets an upper boundary on how well a test can predict the criteria. If a test is not reliable it will not usefully predict any other measure.

Clearly any reliability estimate is subject to particular sources of error. Generally, test-retest reliability reflects the stability of a measure over a given time period, and is calculated by correlating the scores, on the same measure, from two separate testing occasions, across infants. Where the nature of the stimuli changes slightly (eg. in the Visual Preference, Visual Habituation, Concept Acquisition and Audio-Visual Integration tasks - see Chapter 3), this also constitutes equivalent form reliability. The greatest source of error in the test-retest reliability estimate is caused by the changes which occur within the subjects themselves during the time between the two testing sessions. This is due to maturational factors and to the fact that the test has already been administered on one occasion, which is likely to affect performance on the second testing occasion. The length of time between the two testing occasions is generally left to the researcher's discretion, although particularly with infant subjects the longer the interval between sessions, the greater effect of maturational factors. Within this particular research program, in view of the large number of visits required by any one subject, practical limitations also had to be considered, and as a result of such influences, retest sessions were administered up to one week after the initial testing session. In this research, the minimum period between testing sessions was 24 hours and the maximum was one week.

Split-half reliability is an estimate of the internal consistency of a measure. The split-half methods were based on the proposition that many of the temporary factors which affect the test-retest reliability estimates, could be determined from the scores of a single administration of a test. Essentially, by splitting any test or measure into two comparable halves, and correlating performance on the first half with that on the second half, provides an estimate of the internal consistency on the test or measure. However, even with this procedure, conditions are likely to occur which influence performance on the former or latter part of a test. One way of overcoming some of these influences is to split tests into halves, such that each half contains measurement from both early and later parts of the test.

The Keudar-Richardson formula provides a technique of correlating each possible combination of measures to form two halves, and produces a composite reliability measure from all of these combinations. This procedure was not adopted for this research however, since within some of the tasks used changes in performance were expected (eg. the Visual Habituation and Concept Acquisition tasks).

When a test is divided into two parts the amount of data in the resulting reliability coefficient is half of that in comparable test-retest reliability coefficients, and so a corrective procedure is often used to determine the split-half reliability coefficient of the full length of the test. This is known as the Spearman-Brown formula, and typically has the effect of increasing the split-half reliability coefficient.

For the purposes of this research, these two types of reliability (ie. test-retest and split-half), were established for each measure at each age where appropriate. Specific details of the procedures adopted for calculation of these for individual measures are presented in the experimental chapters (Chapters 3, 5, 6 and 7). For a measure to be regarded potentially predictive it must show test-retest reliability, and preferably within session reliability aswell, although for some of the measures, lack of internal consistency may be due to the nature of the measure itself, for instance, if a change in the amount of responding over trials is expected, as in the Concept Acquisition task.

A further problem relating to the notion of reliability concerns the question of how high a reliability coefficient has to be, in order for it to be considered useful. Unfortunately there are no hard and fast rules relating to this problem, and by and large the decision is left to the individual researcher. A reliability coefficient may be considered useful if it reaches significance, although this does have different implications depending on the subject sample size. For instance, with a subject sample of 100, a correlation coefficient as low as .25 is significant at the .01 level, although a correlation of this magnitude reflects only just over 6 percent common variance between the two measures in the correlation. Since a fairly low number of subjects were employed for the longitudinal research, two-tailed significance levels were used as an indication of the reliability of the measures, although the actual magnitude of the r's produced should also be carefully considered (a reasonable correlation of the magnitude of .7 still only accounts for approximately half of the common variance between the measures). None of the split-half reliabilities were corrected by the Spearman Brown formula since it was considered that as much conservatism as possible was essential for determining reliable measures.

4.4.2 Within-Age Analyses

The measures within each age group which showed reliablity were considered potentially predictive. These were correlated with each other by means of Pearson product moment correlations. These analyses were done in the hope that they would reveal some indication of the types of measures which constituted factors of performance in infancy. These data are presented in chapters 5, 6, 7, and 8, for the 3, 5, 9 and 18 month measures, respectively.

Although multi-variate methods of analyses, such as factor analysis, constitute more appropriate methods of determining factors from the type of data collected in this study, their use was precluded by the low subject numbers in relation to the large number of measures obtained from the data at each age. Although there are no strict rules concerning the subjects to measures ratio, Cattell (1966) has suggested that this should be <u>at least</u> 2.5 to 1, whilst others go as high as 5 to 1. If factor analyses or related procedures were to be adopted in this research, interpretation of the data should be made with great caution; should no significant factors emerge, then this may be in part be due to insufficient data disguising any potential results, and similarly if significant results do emerge then they may be distorted interms of strength and pattern for the same reason.

Where the correlations between measures within tasks are particularly high (for example, the measures of learning in the Operant Conditioning task, see Chapter 5), the measures were combined to produce one composite measure which was forwarded as a potential predictor. Composite measures were formed in the following way. Raw scores, for each measure making up the composite, were transformed into standard scores and these were added. The resulting statistic was forwarded as one measure in subsequent longitudinal analyses. Standard scores were used for this purpose, since they maintain both the hierarchy and distance between measures in the distribution. Alternative methods of creating composites include adding the ranks of scores or normalising and adding scores, although both of these methods may reduce the sensitivity of the measures.

4.4.3 Across-Age Analyses

Relationships between reliable measures from each age with reliable measures from every other age (including the criterion measures taken at 18 months of age), were also done by means of correlation matrices. These data are presented in Chapter 9. For these analyses, aswell as Pearson product moment correlations, Spearmans rho rank order correlation coefficients were also calculated. These were included, since any r's emerging significant from the paramatric correlations, may be caused by just one or two extreme subjects, at either one, or at both ends of the range. Spearmans rho correlation coeficients however, examine the concordance in the rank ordering of scores and one or two extreme subjects at either end of the range may be insufficient to produce significant results.

The next Chapter is concerned with the collection of the data and analyses from the infants at 3 months of age.

CHAPTER 5

THE TESTING AT 3 MONTHS - RELIABILITIES AND RELATIONSHIPS BETWEEN MEASURES

5.1 INTRODUCTION

Six tasks in total were used at the 3 month testing age. These tasks have been described in detail in Chapter 3. Four of these tasks were given in the laboratory and two in the home. This chapter describes the testing of infants at 3 months of age, as the initial phase of the longitudinal study. Test-retest and split-half reliabilities were again examined for a variety of measures from each task, and the relationships between reliable measures across tasks were also determined.

5.2 METHOD

5.2.1 Subjects

A total of 29 infants were recruited through advertisments in local health clinics and the local media. All the infants were white, full-term with no known mental or physical health defects.

All came from middle-class families. 14 were male and 15 were female, 15 were first-born infants with the remainder having between 1 and 4 siblings living at home. Before agreeing to participate in the study the nature of the research and the commitments required on the part of the mother were fully explained.

5.2.2 Design

At the age of 3 months, + / - 1 week, each infant was brought to the laboratory at Plymouth Polytechnic, by the mother on 2 occasions separated by not more than 1 week. On the first visit 4 tasks were administered to the infants in a fixed order of presentation:

A) Visual Preferences

B) Concept Acquisition

C) Operant Conditioning

D) Audio-visual Integration

On the second visit the sequence was repeated with the omission of the operant conditioning task.

Also within 1 week of the infants' 3 month 'birthday' each infant was visited in the home on 2 occasions. 2 tasks were administered in the infants' home environment:

E) Differential Vocal Responsiveness

F) Attention to a Novel Toy.

The visits alternated and always commenced with a laboratory visit. All visits were separated by a minimum of 24 hours and by a maximum of 1 week.

As for the pilot studies, all the tasks were administered in a fixed order which was also maintained over each testing session.

The sequence of task administation was maintained (A through D in the laboratory and E followed by F during the home visits) for all infants throughout testing sessions. In the case of failure to complete a given task, due to excessive crying, excessive sleepiness, intervention by the mother, or failure to respond, the task was removed and the next one attempted if the infant regained a suitable state. The number of infants to complete each task is presented in the Results sections. In the event of an infant in any session failing to complete any of the tasks, an alternative appointment was arranged with the mother, but such that the sequence of testing sessions was not interrupted. This occurred on 2 occasions during the 3 month testing sessions.

5.3 PROCEDURE

5.3.1 Laboratory Tasks

Details of the equipment and procedure for all tasks have been presented in the pilot Studies sections in Chapter 3. The only task to differ in procedure from the pilot studies was the Visual Preference task. The revised procedure of this task is presented next.

Visual Preferences

Infants were presented with one stimulus pair, presented over 2 trials. From trial 1 to trial 2 the projected left / right position of each stimulus was alternated. The stimuli used at this age consisted of the stripe - bullseye comparison used in the pilot studies. The stimuli were presented until a cumulative total of 10 seconds of fixation time, either to one or to both stimuli, had been recorded. The next pair of stimuli was then presented, the inter-trial interval consisting of a slide change of approximately 1.5 seconds duration. On the second testing session stimuli were rotated by 90 degrees, so that the vertical lines became horizontal and the division in the bullseye pattern became vertical (see Figure 3.4).

5.3.2 Home Tasks

Both of the tasks administered in the infants' home environment were given in a similar way as reported in the pilot Studies sections. These tasks at this phase of the research, however, were presented on two testing occasions within one week of each other. This enabled reliabilities for measures from these tasks to be established.

5.4 MEASURES

From each task, a variety of measures was taken and these were the same as for the pilot studies and are listed in the Measures section in Chapter 3. Since the Visual Preference task at 3 months did not include the size - number comparisons, the measure of percentage preference for greater number was not applicable. Remaining measures for this task were calculated on the two stripe - bullseye trials only.

5.4.1 Test-Retest and Split-Half Reliabilities

For the laboratory tasks (with the exception of the Operant Conditioning task, which was again only administered on one occasion, and of the Visual Preference task for which all measures were calculated on the basis of the two trials only) test-retest and split-half reliabilities were calculated in the same way as for the pilot studies (see Chapter 3). For the home tasks, test-retest reliabilties were calculated by correlating corresponding measures across the two sessions. Split-half reliabilities were not taken for the home-based tasks.

Again, reliabilities were determined by means of Pearson correlations, although in order to be more conservative two-tailed tests were administered as opposed to the one-tailed tests used in the Pilot studies (see Chapter 4).

5.5 RESULTS AND DISCUSSIONS

5.5.1 Sex Differences

Preliminary analyses revealed no major sex differences for any of these tasks, the data were therefore collapsed across sex.

5.5.2 Visual Preferences

A total of 29 infants completed this task. Data from 2 infants were eliminated due to equipment failure.

Descriptive statistics from measures on this task are presented in Table 5.1.

The split-half and test-retest reliabilities for the measures from

TABLE 5:1

VISUAL PREFERENCES 3 MONTH DATA

DESCRIPTIVE DATA N=29

MEASURE	SESSION	MEAN	SD	RANGE
% FIXATION BULLSEYE	1	61.43	20.07	23.44 to 100
	2	62.2	14.97	42.47 to 95.81
% FIXATION LEFT	1	49.16	32.43	0 to 100
	2	56.85	26.11	5.18 to 100
TOTAL NUMBER FIXATIONS	1	8.03	5.26	2 to 27
	2	9.52	5.05	2 to 27
LONGEST SINGLE FIXATION	1	6.76	1.96	2.44 to 10
	2	5.49	2.08	2.44 to 10
MEAN LONGEST FIXATION	1	5.26	1.74	2.53 to 10
	2	4.93	2.03	2.5 to 10
FIRST FIXATION DURATION	1	2.85	3.97	1.32 to 10
	2	4.12	5.11	.99 to 10
NUMBER FIX ACROSS STIMULI	1	3.06	2.32	0 to 8
	2	4.86	2.78	0 to 13
MEAN COMPLETION TIME	1	21.76	10.03	6.4 to 63.7
	2	23.97	19.89	11.85 to 94.95

this task are presented in Table 5.2.

The percentage fixation to the bullseye failed to emerge as a consistent measure, either within or across testing sessions, although a significant negative reliability was found for session 1. T-tests for related samples were performed to determine whether infants overall had shown a significant preference for the bullseye over the stripes and this was significant for both testing sessions ($t=3.19$, $p < .001$; $t=4.39$, $p < .005$ for sessions 1 and 2, respectively). Hence, it was thought possible that the failure to find percentage preference for the bullseye a reliable measure might have been due to a 'ceiling effect' obscuring any potential reliability. However, since 8 of the infants on session 1, and 6 of those on session 2 showed greater fixation duration towards the stripes, it is unlikely that a 'ceiling effect' was the cause of the lack of reliability.

A further factor likely to effect reliability of the bullseye preference is position bias. T-tests revealed that on neither session did infants overall show a significant preference for stimuli in a particular position ($t=0.14$ and 1.42, NS for sessions 1 and 2, respectively). However, as can be seen in Table 5.2, the percentage preference to the stimulus on the left of the infants' midline showed a greater degree of reliability than the preference for bullseye, and therefore this may have been partially responsible for the failure of preference for the bullseye to emerge as a reliable measure.

TABLE 5:2

VISUAL PREFERENCES 3 MONTH DATA

TEST-RETEST (TRT) AND SPLIT-HALF (SH) RELIABILITIES N=29

MEASURE	TRT	SH SESS. 1	SH SESS. 2
% FIXATION. BULLSEYE	.11	-.45 *	-.13
% FIXATION LEFT	.44 *	.37 *	.14
TOTAL NUMBER FIXATIONS	.52 **	-.06	.01
LONGEST SINGLE FIXATION	.33	.44 *	.41 *
MEAN LONGEST FIXATION	.15	.18	.48 **
FIRST FIXATION DURATION	.24		
NUMBER LOOKS ACROSS	.30	.13	.24
MEAN COMPLETION TIME	-.29	.55 **	.29

* p < .05 (2-Tailed)
** p < .01 (2-Tailed)

The only other measure from this task to produce a significant correlation with itself across sessions, is total number of fixations, although this was not reliable within sessions. Since it was apparent that this measure was not reliable in the other tasks, it was not considered a potentially useful predictor from this task. For two other tasks, Concept Acquisition and Audio-Visual Integration, this measure gave split-half reliabilities but non-significant test-retest reliabilities (see Tables 5.4B and 5.7).

No other measure from this task produced consistent across- or within-session reliability. Since there is no obvious theoretical justification for it constituting a potentially predictive measure, percentage fixation to the stimulus on the left was not considered worth including in further analysis. It was therefore concluded that the Visual Preference task at 3 months fails overall to produce any measures constituting potential predictors of later psychological functioning. This in part may be due to the task being the first to be administered to infants on each testing session, and / or because the task was comparatively short in duration, requiring a total of only 20 seconds fixation towards the stimuli and therefore not allowing infants to settle into a consistent or 'true' pattern of fixation. In fact a number of infants (4 on session 1, and 2 on session 2) only fixated one of the two locations throughout the whole task.

Conclusion

None of the measures from the Visual Preference task were considered to be potentially predictive at three months.

5.5.3 Concept Acquisition

This task was completed successfully by 27 infants, on two testing sessions. Data from the remaining 2 infants were eliminated due to infant fussing (n=1) and equipment failure (n=1).

Table 5.3 shows the descriptive statistics for the measures from this task.

Table 5.4 shows the test-retest and split-half reliabilities from measures from this task at three months. A number of measures emerged from this task which were reliable both within and across sessions.

Total fixation duration to the familiar concept exemplars, the novel non-concept exemplars, and over both of these combined, produced significant correlations both within and across sessions.

Average fixation duration was also reliable over the 'familiar' trials, and over all trials, although reliabilities are lost for the novel trials. However, since the familiar trials make up

TABLE 5:3

CONCEPT ACQUISITION 3 MONTH DATA

DESCRIPTIVE DATA N=27

MEASURE	SESSION	MEAN	SD	RANGE
TOTAL FIXATION DURATION	1	223.86	70.59	58.47 to 297.98
	2	208.28	55.137	109.62 to 284.8
TOTAL NUMBER FIXATIONS	1	25.15	9.15	17 to 44
	2	36.11	16.76	17 to 99
MEAN FIXATION DURATION	1	10.54	5.81	2.35 to 24.83
	2	7.05	4.07	2.22 to 16.76
LONGEST FIXATION DURATION	1	19.69	7.95	0.98 to 30
	2	16.34	12.41	0.76 to 30
MEAN LONGEST DURATION	1	17.01	7.24	3.80 to 29.69
	2	14.32	5.98	4.88 to 26.78
FIRST FIXATION DURATION	1	11.90	9.34	1.6 to 30
	2	15.04	11.48	1.52 to 30
LATENCY TO FIXATE	1	3.67	2.76	0.49 to 9.6
	2	2.99	1.99	0.53 to 5.51
% DECLINE FIXATION TRIAL 1- FIXATION TRIAL 8	1	-2.46	157.51	-92.21 to 505.90
	2	27.95	75.14	-95.24 to 335.0
% RECOVERY FIXATION TRIAL 8- FIXATION NOVEL	1	5.42	342.64	-49.2 to 1017.77
	2	3.76	49.43	-86.99 to 176.65

TABLE 5:4

CONCEPT ACQUISITION 3 MONTH DATA

TEST-RETEST (TRT) AND SPLIT-HALF (SH) RELIABILITIES N=27

MEASURE	TRT
5.4A ALL TRIALS COMBINED	
TOTAL FIXATION DURATION	.67 **
TOTAL NUMBER FIXATIONS	.02
MEAN FIXATION DURATION	.41 *
SINGLE LONGEST FIXATION	.61 **
MEAN LONGEST FIXATION DUR.	.56 **
FIRST FIXATION DURATION	.36
MEAN LATENCY TO FIXATE	.29
% DECREASE (TRIALS 1-8)	-.04
% INCREASE (TRIALS 8-NOVEL)	-.17

MEASURE	TRT	SH SESS 1	SH SESS 2
5.4B FAMILIAR CONCEPT EXEMPLAR TRIALS			
TOTAL FIX. DURATION	.59 **	.69 **	.74 **
TOTAL NUMBER FIXATIONS	.04	.71 **	.65 **
MEAN FIX. DURATION	.42 *	.63 **	.64 **
SINGLE LONGEST FIXATION.	.44 *	.85 **	.59 **
MEAN LONGEST FIXATION DUR.	.53 **	.52**	.49 **
MEAN LATENCY TO FIX.	.29	.36	-.24
5.4C NOVEL NON-CONCEPT EXEMPLAR TRIALS			
TOTAL FIX. DURATION	.72 **	.61 **	.59 **
TOTAL NUMBER FIXATIONS	-.08	.14	.25
MEAN FIX. DURATION	.20	.30	-.05
SINGLE LONGEST FIXATION DUR.	.52 **	.44 *	.24
MEAN LONGEST FIXATION DUR.	.37	.63 **	.42 *
MEAN LATENCY TO FIX.	.43 *	.38	.42 *

* $p < .05$ (2-tailed)
** $p < .01$ (2-tailed)

four-fifths of the total trials, the measure of mean fixation duration was considered a potential predictor.

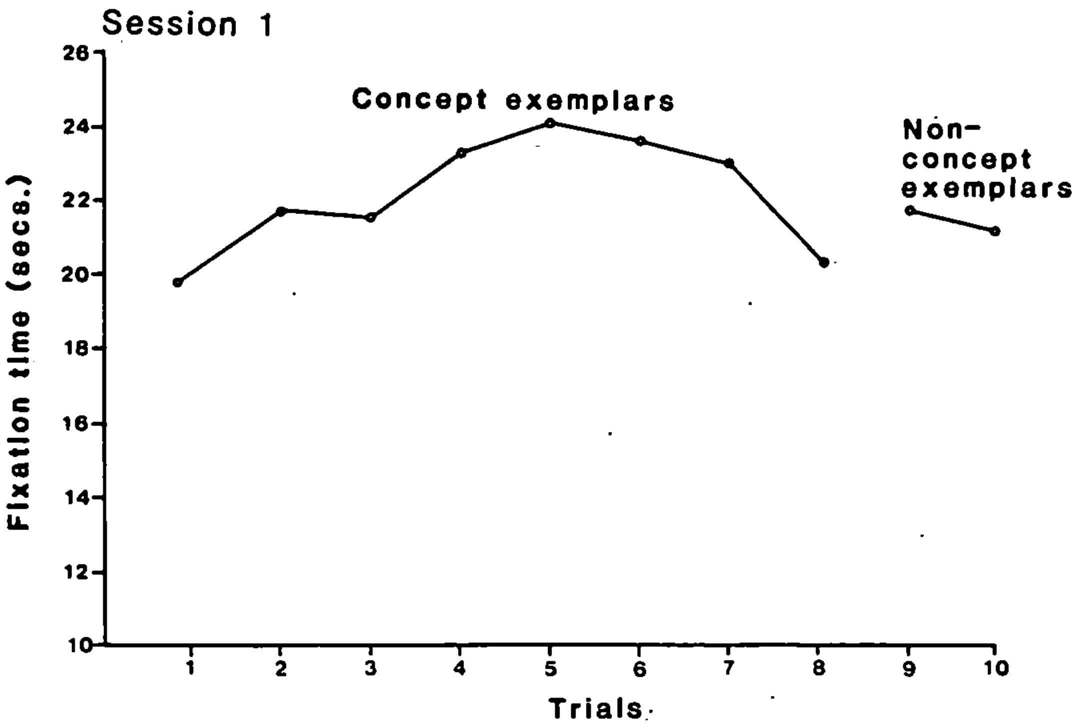
The measure of single longest fixation was also reliable both across and within sessions. Average longest fixation duration is also reliable across and within sessions, although again reliability is lost across sessions for the novel trials.

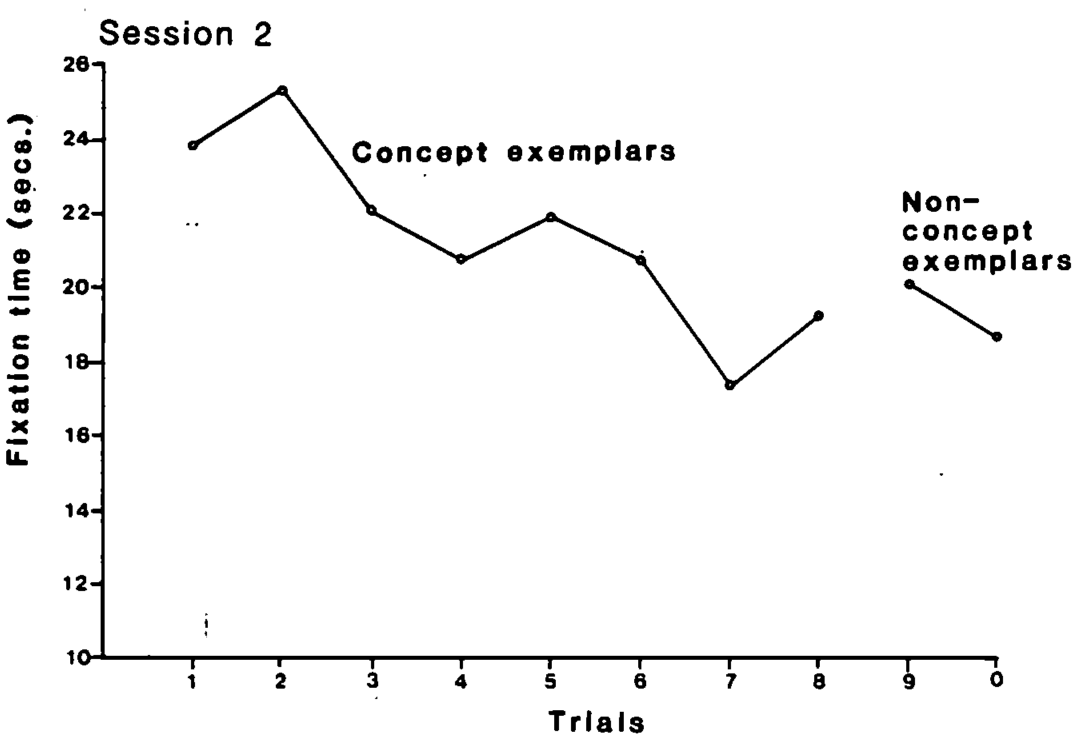
Total number of fixations failed to emerge as reliable across testing sessions but, at least over familiar trials, was characterised by within session stability. This suggests that the measure was perhaps affected by state characteristics of the infant from one testing session to the next, or by characteristics of the different stimuli on each session. Hence this measure was not considered likely to be a useful predictor of later cognitive functioning.

First fixation duration produced no across session reliability (split-half reliability is not appropriate for this measure) and this could not therefore be considered potentially predictive.

The remaining measures, reflecting qualitative aspects of infants' fixation patterns; namely, the percentage decrease in fixation to concept exemplars, and the percentage increase in fixation to the novel non-concept exemplars, failed to show any indication of reliability over sessions. The mean of the total fixation durations over all infants for each trial on each session, were calculated. These are presented in Figure 5:1. Trend analyses

Figure 5:1. CONCEPT ACQUISITION - 3 MONTHS
The figure shows the average total fixation duration for each trial. n = 12





were performed on the average total fixation times on 'familiar' trials for all infants on each session. For session 1 the F value was not significant whilst for session 2 the F value was significant ($F=32.21$ $p < .01$ ($df=1,182$)). To examine whether any decrement in fixation occurred from the first to the final familiar trials, regardless of fixation patterns in intervening trials, t-tests were performed on the mean fixation time on trials one and two with the mean fixation time across trials 7 and 8, and again no significant decrement was found on session 1 but the t value was significant for session 2 ($t=5.97$ $p < .01$). T-tests were also used to determine whether there was a significant increase in fixation to the novel non-concept exemplars. The mean fixation duration from trials 7 and 8 was compared with that from trials 9 and 10, neither of these reached significance ($t=0.18$ and 1.66 for the two sessions, respectively).

Hence, in neither session did infants demonstrate the expected pattern of responding to stimuli in this task. There was no evidence that infants consistently (a) declined in looking over the 'familiar' concept exemplars, or (b) increased looking during the presentation of the 'novel' non-concept exemplars.

The conclusion which was reached regarding these measures from this particular study, was that percentage decrease in responding to the familiar concept exemplars and subsequent increase in responding to the novel non-concept exemplars do not constitute potential predictive measures.

All of the measures which showed test-retest and split-half reliability, ie. total fixation duration, average fixation duration, single longest fixation and average longest fixation duration, reflect quantitative aspects of total fixation time or measures of fixation length, and not qualitative measures of decline to the concept exemplars and recovery to the non-concept exemplars. Clearly, these reliable measures are likely to be highly correlated because, for example, for infants to produce a large mean longest fixation duration, then total fixation duration is also likely to be reasonably high. The inter-correlations between the reliable measures for each session are presented in Table 5:5. All of these correlations were reasonably high, suggesting that infants who fixated stimuli for long periods of time also exhibited longer average fixation times, and conversely infants who gave many short fixations with many looks away from the stimulus, did not achieve such a high total fixation duration as those infants who fixated the stimulus for long periods of time.

Conclusion

The Concept Acquisition task at 3 months produced measures which are reliable both within and between sessions (at least up to a period of one week), and these measures can therefore be considered potential predictors of later ability. These are, total fixation duration, average fixation duration, single longest fixation and average longest fixation duration.

TABLE 5:5

CONCEPT ACQUISITION 3 MONTH DATA

CROSS CORRELATIONS OF RELIABLE MEASURES N=27

	SESSION 1			
MEASURE	TOTAL FIX. DURATION	AV. FIX DURATION	SINGLE LONG. FIXATION	AV. LONG FIXATION
TOTAL FIXATION DURATION	..	.790 **	.769 **	.894 **
MEAN FIXATION DURATION	.697 **	..	.648 **	.912 **
SINGLE LONGEST FIXATION DURATION	.701 **	.569 **	..	.821 **
MEAN LONGEST DURATION	.852 **	.949 **	.720 **	..

SESSION 2

** $p < .01$ (2-tailed)

Correlations between measures from session 1 are presented above and to the right of the major diagonal, and for session 2 below and to left of the major diagonal.

5.5.4 Audio-Visual Integration

21 infants in total completed this task successfully on the two testing occasions. A further 6 infants failed to complete the task due to equipment failure, and data from 2 infants were eliminated due to fussing.

Descriptive statistics from all of the measures from this task are presented in Table 5:6.

T-tests revealed that for session 1 no significant preference emerged for the synchronised video (t=1.60 NS), although the t value was significant on session 2 (t=3.29 p <.01). Also, on neither session was there a preference for one of the video positions (ie. left or right of the infants' mid-line) (t=0.29 and 0.86, NS in both cases for the two sessions, respectively).

The test-retest and within session (split-half) reliabilities of the measures taken from this task are presented in Table 5:7. Percentage fixation to the synchronised video emerged as a reliable measure both across sessions and within session 1, although within-session reliability was lost on session 2. Indeed a similar pattern of reliabilities emerged for total fixation duration. These findings are rather unusual since the two measures appear not to be related to each other (producing r's of -.07 and -.22 for sessions 1 and 2, respectively). Both of these measures, however, were considered potential predictors, in view of the high correlations they produced across sessions and within

TABLE 5.6

AUDIO-VISUAL INTEGRATION 3 MONTH DATA

DESCRIPTIVE DATA N=21

MEASURE	SESSION	MEAN	SD	RANGE
% FIXATION	1	55.30	14.69	37.25 to 89.66
SYNCHRONISED VIDEO	2	59.44	13.0	43.56 to 91.51
% FIXATION	1	52.2	31.1	0.07 to 99.1
LEFT VIDEO	2	54.5	23.8	3.1 to 96.1
TOTAL FIXATION	1	113.61	9.43	85.75 to 119.79
DURATION	2	113.01	7.30	89.45 to 119.42
TOTAL NUMBER	1	27.71	14.24	8 to 63
FIXATIONS	2	30.19	14.80	12 to 50
MEAN FIXATION	1	6.13	4.15	2.65 to 14.9
DURATION	2	4.76	2.39	1.66 to 9.90
LONGEST FIXATION	1	9.04	15.77	1.99 to 30
DURATION	2	16.01	12.86	5.71 to 30
MEAN LONGEST	1	14.58	6.33	6.36 to 27.06
FIXATION DURATION	2	13.55	5.79	4.79 to 24.82
NUMBER FIXATIONS	1	5.09	3.43	1 to 14
ACROSS VIDEOS	2	5.84	3.48	2 to 16
FIRST FIXATION	1	7.12	8.12	2.05 to 30
DURATION	2	6.51	4.30	2.48 to 18.21

TABLE 5:7

AUDIO-VISUAL INTEGRATION 3 MONTH DATA

TEST-RETEST (TRT) AND SPLIT-HALF (SH) RELIABILITIES N=21

MEASURE	TRT	SH SESS. 1	SH SESS. 2
% FIXATION SYNCH. VIDEO	.77 **	.60 **	.19
% FIXATION LEFT	.62 **	.69 **	.69 **
TOTAL FIXATION DURATION	.84 **	.82 **	.18
TOTAL NUMBER FIXATIONS	.21	.71 **	.69 **
MEAN FIXATION DURATION	.50 *	.72 **	.62 **
LONGEST SINGLE FIXATION	.29	.65 **	.64 **
MEAN LONGEST FIXATION	.41	.68 **	.63 **
TOT. NO. LOOKS ACROSS	.26	.68 **	.67 **
FIRST FIX. DURATION	.47 *		

* $p < .05$ (2-Tailed)
** $p < .01$ (2-Tailed)

session 1.

As for the Concept Acquisition task, mean fixation duration was reliable both within and between sessions from the Audio-Visual Integration task.

The only other measure to produce reliability both across and within sessions in this task, was that of percentage fixation to the video on the left of the infant's midline. First fixation duration produced a significant correlation over the two sessions.

A number of other measures; namely, total number of fixations, number of fixations across from one video to the other, average longest fixation duration and single longest fixation duration appear to be affected by the state of the infant at the time of testing, since they produced significant split-half correlations, but showed no evidence of across session reliability.

The correlations between the reliable measures from the Audio-Visual Integration task, are presented in Table 5:8. For this task, none of the reliable measures correlated significantly with each other in either session. The highest correlations, as one would expect, fell between total fixation duration and average fixation duration, although these do not correlate as highly as the equivalent measures did in the Concept Acquisition task. The significant negative correlation between percentage fixation to the synchronised video, and percentage fixation to the video on the left on session 1, is odd, and since for session 2 the

TABLE 5:8

AUDIO-VISUAL INTEGRATION 3 MONTH DATA

CROSS CORRELATIONS OF RELIABLE MEASURES N=21

	SESSION 1				
MEASURE	% FIX. SYNCH.	% FIX. LEFT	TOTAL FIX. DURATION	MEAN FIX. DURATION	FIRST FIX. DURATION
% FIXATION SYNCHRONISED	..	-.516 *	-.078	-.224	-.073
% FIXATION LEFT	.001	..	.097	.227	.028
TOTAL FIXATION DURATION	-.025	-.059	..	.391	.196
MEAN FIXATION DURATION	.018	.083	.387	..	.261
FIRST FIXATION DURATION	-.109	-.348	.049	-.086	..
	SESSION 2				

* $p < .05$ (2-tailed)

Correlations between measures from session 1 are presented above and to the right of the major diagonal, and for session 2 below and to left of the major diagonal

equivalent correlation was zero, perhaps occurred by chance.

Conclusion

In conclusion this task produced a number of measures which can be considered potential predictors of later intellectual functioning. These include, percentage fixation to the synchronised video, total fixation duration, mean fixation duration and first fixation duration. Percentage fixation to the video on the left of the infant's midline was also included in further analyses due to the high reliabilities both across and within sessions, although there is little reason to suspect that this measure should be predictive of later functioning.

In general it appears from Table 5.8 that the potentially predictive measures on this task are reflecting unrelated aspects of performance.

5.5.5 Operant Conditioning

24 infants completed this task successfully. The data from a further 5 infants were not used due to infant fussing.

Descriptive statistics for the 6 measures from this task are presented in Table 5:9.

The mean number of kicks over all infants for each minute of testing are plotted in Figure 5:2. It can be seen that kicking

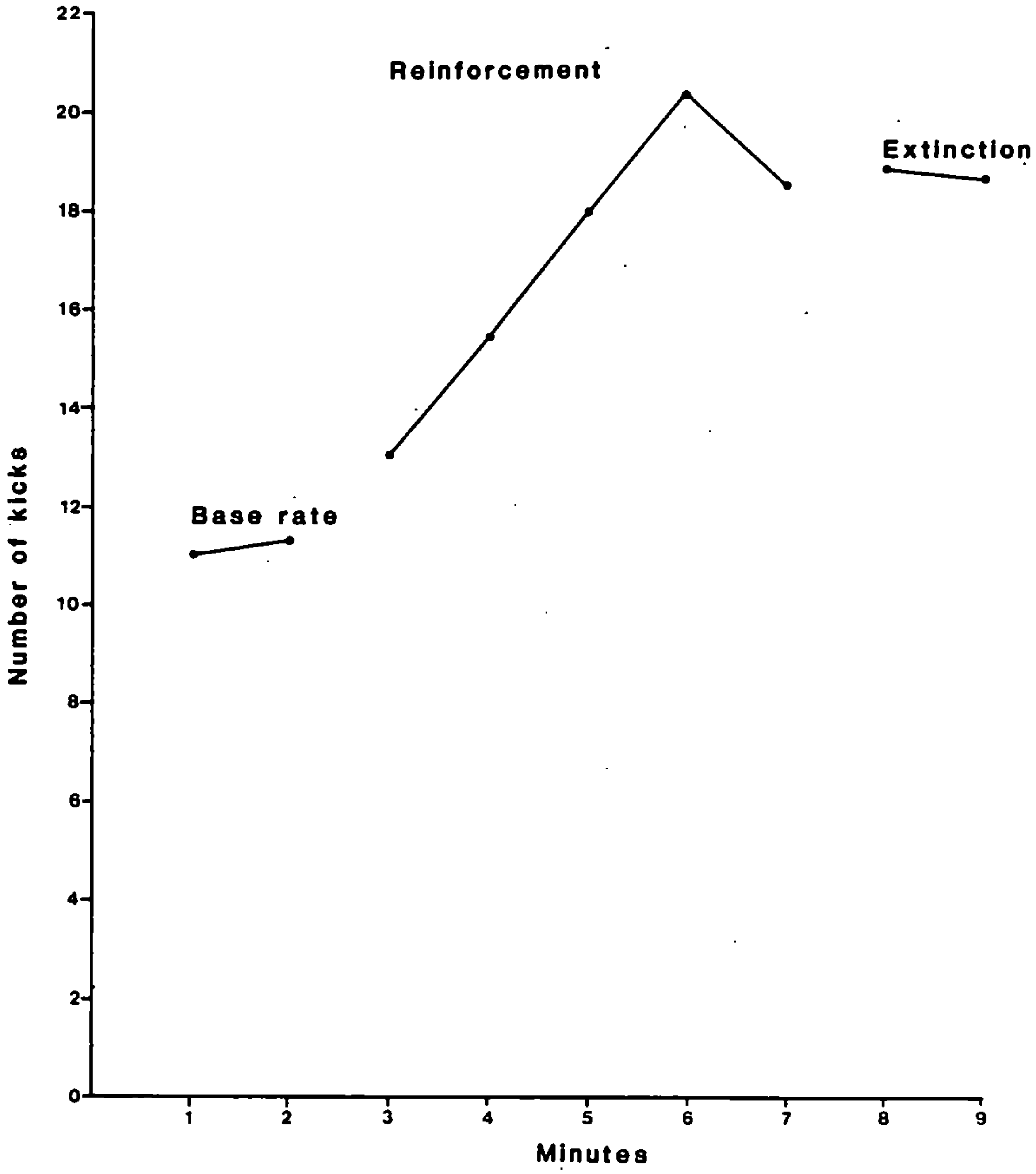
TABLE 5:9

OPERANT CONDITIONING 3 MONTH DATA

DESCRIPTIVE DATA N=24

MEASURE	MEAN	SD	RANGE
RATIO B TO A	2.16	1.58	0.57 to 6.10
RATIO C TO A	2.36	2.2	0.67 to 8.5
LAST B / A	2.49	2.06	0.70 to 9.5
HIGHEST IN B / A	3.06	2.57	0.90 to 9.5
FIX DURATION	489.5	51.7	340.8 to 580.9
TOTAL NO KICKS	145.2	93.6	40 to 430

Figure 5:2. OPERANT CONDITIONING TASK - 3 MONTHS.
The figure shows the average number of kicks recorded in each minute of testing. n=24



did increase during the reinforcement phase as compared with the base-line period. A t-test for related samples revealed a significant increase from the average number of kicks recorded in minutes 1 and 2 to the average number of kicks recorded in minutes 6 and 7 ($t= 3.56$ $p < .01$). There was however no significant decline from minutes 6 and 7 to the extinction phase (minutes 8 and 9, $t= -0.7$ NS).

The inter-correlations of measures from this task are presented in Table 5:10. It is clear that measures reflecting learning rates (the first four measures on the table) are all very highly inter-correlated (producing r's ranging from .7 to .91). This is not suprising, however, since all of these measures are related and many measures overlap. In view of the high correlations between the learning measures these measures were combined to form one composite measure, for the longitudinal analyses (see Chapter 9). The composite measure was derived by converting each raw score for each learning measure to a z-score, and then adding the z-scores for each infant.

The measures reflecting aspects of learning rate, appear to be independent of the two remaining measures, namely, total fixation duration and total number of kicks. These latter two measures are correlated moderately positively with each other, which suggests, as one might also expect, that those infants who gave many kicks also fixated the mobile for a longer time.

TABLE 5:10

OPERANT CONDITIONING 3 MONTH DATA

CROSS CORRELATIONS OF MEASURES N=24

MEASURE	RATIO B TO A	RATIO C TO A	END MINS B / A	GREATEST B / A	TOT FIX DUR	TOT NUMBER KICKS
RATIO B TO A	..					
RATIO C TO A	.70 **	..				
END MINS B / A	.80 **	.78 **	..			
GREATEST IN B / A	.91 **	.88 **	.86 **	..		
TOT FIX DUR	-.09	-.07	.19	-.14	..	
TOT NO KICKS	-.33	-.04	-.08	-.14	.51 *	..

* p < .05 (2-tailed)
** p < .01 (2-tailed)

Conclusion

Although the expected decline in the extinction phase was not found, it was clear that learning had occurred in this task, and therefore the measures from this task were included in further analyses.

5.5.6 Differential Vocal Responsiveness (DVR)

27 infants completed the DVR task in the home on two occasions. Two infants were excluded due to possible visiting times being incompatible with the mother's arrangements.

Descriptive statistics for the measures on this task are presented in Table 5:11.

Table 5:12 shows the test-retest reliabilities from measures taken from this task at 3 months. All of the measures taken from this task were reliable with the exception of total vocalisation time to the stranger. One possible reason for this was that the stranger had become less strange by the second testing session, but a t-test revealed that there was no difference in the extent to which infants vocalised to the stranger across sessions ($t= 1.056$ NS).

The inter-correlations of these measures within sessions are

TABLE 5:11

DIFFERENTIAL VOCAL RESPONSIVENESS 3 MONTH DATA

DESCRIPTIVE DATA N=27

MEASURE	SESSION	MEAN	SD	RANGE
DVR	1	5.8	15.5	-23.1 to 43.8
	2	5.7	13.2	-10.7 to 48.1
TOTAL VOC MOTHER	1	18.1	15.3	0.9 to 53.4
	2	15.8	14.7	0 to 56.3
TOTAL VOC STRANGER	1	12.26	9.71	0 to 35.4
	2	10.09	8.76	0 to 36.4
TOTAL VOCALISATION	1	30.3	20.4	3.4 to 73.5
	2	25.9	20.3	0 to 76.0

TABLE 5.12

DIFFERENTIAL VOCAL RESPONSIVENESS 3 MONTH DATA

TEST-RETEST (TRT) RELIABILITIES N=27

MEASURE	TRT
DVR	.42 *
TOT. VOC. MOTHER	.47 *
TOT. VOC. STRANGER	.33
TOTAL VOCALISATION	.44 *

* $p < .05$ (2-Tailed)

presented in Table 5:13. All the measures correlate significantly with each other, with the highest correlations emerging between measures of total vocalisation duration and vocalisation duration to the mother alone. All the measures were considered seperate potential predictors.

Conclusion

The measures of total vocalisation, and vocalisation to the mother alone, as well as the DVR measure, produced significant correlations over the two sessions, and therefore, must all be considered potentially predictive.

5.5.7 Attention to a Novel Toy

The same 2 infants who failed to complete the DVR task also failed to complete the Attention to a Novel Toy task at 3 months, thus leaving a sample size of 27.

Descriptive statistics for this task are presented in Table 5:14.

Table 5:15 shows the test-retest reliabilities of measures from this task. The measures of total fixation duration and total number of fixations produced high correlations over the two sessions (.7 and .71 respectively), although mean fixation duration failed to produce a significant correlation with itself. Since the correlation between the measures of total fixation

TABLE 5:13

DIFFERENTIAL VOCAL RESPONSIVENESS 3 MONTH DATA

CROSS CORRELATIONS OF RELIABLE MEASURES N=27

		SESSION 1	
MEASURE	DVR	TOT. VOC DUR. MOTHER	TOT VOC DURATION
DVR	..	.80 **	.44 *
TOTAL VOCALISATION DURATION MOTHER	.81 **	..	.89 **
TOTAL VOCALISATION DURATION	.52 **	.92 **	..
	SESSION 2		

* p < .05 (2-tailed)
** p < .01 (2-talied)

Correlations between measures from session 1 are presented above and to the right of the major diagonal, and for session 2 below and to left of the major diagonal

TABLE 5:14

ATTENTION TO A NOVEL TOY 3 MONTH DATA

DESCRIPTIVE DATA N=27

MEASURE	SESSION	MEAN	SD	RANGE
TOTAL FIXATION DURATION	1	126.9	67.1	30.6 to 304.6
	2	131.1	63.7	24.3 to 268
TOTAL NUMBER FIXATIONS	1	7.7	8.25	1 to 32
	2	6.41	4.92	1 to 17
MEAN FIXATION DURATION	1	37.7	38.5	2.6 to 163
	2	36	31.7	3.7 to 119.9

TABLE 5:15

ATTENTION TO A NOVEL TOY 3 MONTH DATA

TEST-RETEST (TRT) RELIABILITIES N=27

MEASURE	TRT
TOTAL FIX. DURATION	.70 **
TOTAL NO. FIXATIONS	.71 **
MEAN FIX. DURATION	.20

** $p < .01$ (2-tailed)

duration and total number of fixations was low for each session (see Table 5:16), and these measures therefore appear not to be related to each other, it is necessary to include them separately as potential predictors.

Conclusion

The two measures of total fixation duration and total number of fixations are forwarded as potentially reliable measures at 3 months.

5.6 CONCLUSION - TEST-RETEST AND SPLIT-HALF RELIABILITIES

From the tasks at three months a number of reliable measures emerged. Those measures which were reliable both across and within sessions are potential predictors although those measures which failed to show within session reliability, but did produce across session reliability can also be considered potential predictors. Hence, those measures which appeared reliable across sessions but which failed to show reliability on one of the individual testing sessions (for example, percentage fixation to the synchronised video and total fixation duration on the audio-visual integration task) are considered potential predictors.

Many of the reliable measures reflect aspects of fixation duration towards different types of stimuli. It is interesting that the

TABLE 5:16

ATTENTION TO A NOVEL TOY 3 MONTH DATA

CROSS CORRELATIONS OF RELIABLE MEASURES

SESSION 1

MEASURE	TOTAL FIXATION DURATION	TOTAL NUMBER FIXATIONS
TOTAL FIXATION DURATION	..	-.02
TOTAL NUMBER FIXATIONS	.10	..

SESSION 2

Correlations between measures from session 1 are presented above and to the right of the major diagonal, and for session 2 below and to left of the major diagonal

measure of total number of fixations was reliable only on the Visual Preference and the Attention to the Novel Toy tasks. This measure had failed to show reliability in the Concept Acquisition and Audio-Visual Integration tasks. The same stimulus was used for each session on the Visual Preference and Attention to a Novel toy tasks, where the stimuli changed across sessions on the other tasks and it would seem that characteristics of individual stimuli do affect the number of looks given by infants.

Further discussion of the measures and their reliability is given in Chapters 6, 9 and 10.

The measures from the 3 month data considered potentially predictive, along with all of the measures from the operant conditioning task, are presented in table 5:17.

5.7 ACROSS TASK CORRELATIONS: 3 MONTH DATA

The next stage of the analyses was to determine the ways in which these apparently reliable measures from the tasks at 3 months are related to each other across tasks. These analyses were carried out to provide information relating to whether reliable measures from the separate tasks are related to any other reliable measures from other tasks.

For these analyses, for each infant the mean score, over the two testing sessions was calculated for each reliable measure, and

TABLE 5.17

MEASURES CONSIDERED POTENTIAL PREDICTORS AT 3 MONTHS

CONCEPT ACQUISITION

1. Total fixation duration
2. Average fixation duration
3. Single longest fixation
4. Average longest fixation

AUDIO-VISUAL INTEGRATION

5. Percentage fixation to the synchronised video
6. Percentage fixation to the video on the left.
7. Total fixation duration
8. Average fixation duration
9. First fixation duration

OPERANT CONDITIONING

10. Ratio of B to A
11. Ratio of C to A
12. Mins. 6 and 7 of B, divided by A
13. Greatest 2 mins. in B, divided by A
14. Total fixation time
15. Total number of kicks

DIFFERENTIAL VOCAL RESPONSIVENESS

16. DVR measure
17. Total vocalisation to the mother
18. Total vocalisation duration

ATTENTION TO A NOVEL TOY

19. Total fixation duration
20. Total number of fixations

TABLE 5.18

INTER-CORRELATIONS BETWEEN RELIABLE MEASURES AT 3 MONTHS

For key to individual measures see Table 5.17

MEASURE NUMBER	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-.18	.31	.49*	.48*	.25	.42	.22	.35	.38	-.22	.22	.19	.00	-.14	-.23	-.09
2	-.11	.11	.42	.77**	.30	.57**	.15	.38	.44*	-.35	.14	.05	.02	.00	-.32	-.16
3	-.09	.22	.48*	.52*	.28	.52*	.18	.25	.24	-.29	.06	.23	.02	-.14	-.31	-.08
4	-.17	.19	.50*	.66**	.24	.51*	.22	.40	.43*	-.26	.14	.08	-.06	-.14	-.35	-.14
5						-.33	-.38	-.41	-.39	-.14	-.09	-.27	-.20	-.10	-.02	-.12
6						.49	.28	.38	.48	.39	.45	-.06	-.28	-.34	.28	-.16
7						.13	-.08	-.17	-.07	.05	.35	.26	.14	.02	.12	-.07
8						.58*	.05	.31	.42	.00	.06	.24	.27	.23	-.12	-.36
9						-.12	.04	.00	-.05	.05	-.01	-.05	.20	.32	-.21	-.26
10												-.12	-.05	.00	.00	-.23
11												-.11	-.08	-.05	-.09	-.15
12												-.15	-.05	.02	.05	-.26
13												-.13	-.07	-.02	-.02	-.29
14												-.30	-.31	-.22	.26	.10
15												.14	.02	-.06	.14	.10
16															.04	.00
17															.01	.11
18															-.02	.17

* $p < .05$ (2-tailed)

** $p < .01$ (2-tailed)

Pearson correlations carried out. Measures from the operant task were included from the single session in which they were obtained.

The cross correlations for all reliable measures with all other reliable measures are presented in Table 5.18, and are discussed in the following sections.

5.7.1 Concept Acquisition and Audio-Visual Integration

A total of 20 infants completed both of these tasks and the correlations between reliable measures from these tasks are presented in Table 5.18 and are in the area surrounded by a red ring.

It can be seen from this table that the measures of percentage fixation to the synchronised video (#5) and percentage fixation to the video on the left of the infant's mid-line (#6) from the audio-visual integration task failed to correlate significantly with any of the measures from the Concept Acquisition task.

On the other hand, measures of total fixation duration (#1), average fixation duration (#2), single longest fixation duration (#3) and mean longest fixation duration (#4) from the Concept Acquisition task, produced a series of high positive correlations with total fixation duration (#7) and average fixation duration (#8) from the Audio-Visual Integration task. The highest correlations emerged between mean fixation duration from both tasks producing a correlation of .77. A measure of visual attention, then, reflecting some aspect of the amount of time infants will fixate stimuli, and the average length of their fixations towards stimuli, appears to be reliable not only within and between testing sessions for the same task, but also across two tasks, which involve presentation of entirely different stimuli types.

First fixation duration (#9) from the Audio-Visual Integration task failed to correlate with any of the measures from the Concept Acquisition task. As can be seen in Table 5:8, first fixation duration also failed to correlate with any other measure within the Audio-Visual Integration task, and it is possible that the measure was affected by the fact that two stimuli were presented in this task, and a shift in attention from one stimulus to the other may have occurred as a result of peripheral distraction. This measure may therefore reflect some aspect of performance in infancy, which is independent of other quantitative fixation measures.

Conclusion

The correlations between measures on these tasks suggest that there is some underlying stable factor relating to aspects of fixation patterns in terms of the total time spent looking at the stimuli, and the average length of fixations toward stimuli, and that this is not related to qualitative fixation patterns as reflected in for example, percentage fixation to the integrated video.

5.7.2 Concept Acquisition and Operant Conditioning

22 infants completed both of these tasks. The correlations between the six measures derived from the Operant Conditioning task and the reliable measures from the Concept Acquisition task are presented in Table 5.18 within the area highlighted by a blue ring.

Two of the measures from the Operant Conditioning task - (i) the ratio of kicks in period B to that in period A and (#10), (ii) the greatest number of kicks in any 2 minute period of session B divided by the number of kicks in A (#13),- both produced significant correlations with two measures from the Concept Acquisition task; mean fixation duration, and mean longest fixation (the r's just miss significance for total fixation duration). The two measures from the learning task are a reflection of the amount of increase in kicking rate as a result of reinforcement, whilst the Concept Acquisition measures are reflecting mean length of fixations.

Conclusion

There is some indication here that measures from the Concept Acquisition task, mean fixation duration and mean longest fixation, are related to some aspect of learning. However, this result is far from conclusive, and was not consistent for all of the learning measures.

5.7.3 Audio-Visual Integration and Operant Conditioning

16 infants completed these tasks.

Table 5.18 shows the correlations between the measures for the 16 infants who successfully completed both of these tasks. These correlations are surrounded by a yellow ring.

The Operant Conditioning measure of the ratio of kicks in period B to that in period A (#10) correlated significantly with mean fixation duration from the Audio-Visual Integration task (#8). This finding supports those which emerged between measures from the Operant Conditioning and Concept Acquisition tasks (see above). However, since the measure of mean fixation duration from the Concept Acquisition and Audio-Visual Integration tasks correlated highly with each other, then should either of these measures produce a significant correlation with operant measures by chance, then it is likely that it would also for the other measure.

No other measures between the Audio-Visial Integration and Operant Conditioning tasks correlated significantly.

Conclusion

The inter-correlations between measures from these two tasks again provide a speculative suggestion that fixation measures may be related to some aspect of learning performance.

5.7.4 Remaining Task Comparisons

The remaining across task comparisons include: Concept Acquisition and DVR, completed by 25 infants (highlighted by green); Concept Acquisition and Attention to a Novel Toy, completed by 25 infants (highlighted by orange); Audio-Visual Integration and DVR, completed by 19 infants (highlighted by brown); Audio-Visual Integration and Attention to a Novel Toy, completed by 19 infants (highlighted by pink); Operant Conditioning and DVR, completed by 24 infants (highlighted purple); Operant Conditioning and Attention to a Novel Toy, completed by 24 infants (highlighted by grey); and DVR and Attention to a Novel Toy, completed by 27 infants (highlighted by black). All correlations were close to zero and none reached significance.

5.7.5 Discussion - 3 Month Correlations Across Tasks

As briefly mentioned earlier in this chapter the correlations of measures from the Concept Acquisition and Audio-Visual Integration tasks lead to the suggestion that certain aspects of fixation (particularly measures of total fixation time and mean fixation time) are reliable not only within and across sessions, but also across tasks, and moreover, across tasks which differ in terms of the nature and amount of stimulation. The inclusion of measures from the Operant Conditioning task led to the suggestion that learning measures might also be related to the measures of fixation, although at this stage the analyses provide no sound evidence for this. The pattern of correlations that has emerged here may be due in part to chance elements, and it is not until relationships between these 3 month measures and measures from tasks at later ages are examined, that a clearer picture can be extracted from these data.

Measures from the two home-based tasks failed to correlate with any of the laboratory tasks or with measures from each other. However since they are reliable across sessions they should be best considered as independent measures, which may or may not correlate with later performance.

In summary the pattern of across task correlations at 3 months leads to the suggestion of a common factor relating to quantitative aspects of fixation and which may also be related to learning ability, as well as a number of independent reliable measures from the tasks administered in the home.

All of these tasks were administered to these same infants at 5 months of age, and it is data relating to the 5 month testing sessions which are presented and discussed in the following chapter.

CHAPTER 6

THE TESTING AT 5 MONTHS - RELIABILITIES AND RELATIONSHIPS BETWEEN MEASURES

6.1 INTRODUCTION

Six tasks were administered to the infants at 5 months of age, and these were the same tasks as those used at the 3 month testing age. This chapter presents the test-retest and split-half reliabilities of all measures at 5 months. Additionally the inter-relationships between reliable measures across different tasks are examined, and discussed in relation to the pattern of reliabilities which emerged at 3 months. Relationships between measures across age (from 3 to 5 months) are presented in Chapter 9.

Some of the 5 month testing procedures differed slightly from those used at 3 months and these will be discussed in the Method sections in this chapter where appropriate.

6.2 METHOD

6.2.1 Subjects

25 of the 29 mothers whose infants had participated in the 3 month session returned for testing at 5 months. Of the remaining four mothers who had participated in the earlier part of the study, three had moved out of the area, and one mother decided not to continue with the program due to illness of her infant. Three of these infants were male and one was female.

6.2.2 Design

At the age of 5 months + / - one week, each infant was brought to the laboratory by the mother, and visited in the home, on two occasions seperated by not less than 24 hours and not more than one week. Visits alternated and always commenced with a laboratory visit. On each visit the tasks were presented in a fixed order, and this was identical to that used at 3 months (see Chapter 5, Design section).

6.3 PROCEDURE

6.3.1 Laboratory Tasks

Full details of the equipment, stimuli and procedures of each task have been presented in the Pilot Studies sections in Chapter 3. All testing procedures were identical to those used for the 3 month testing with the exception of the Visual Preferences and Concept Acquisition tasks, and revised procedures of these tasks are discussed next.

Visual Preferences

At 5 months, as well as the stripe - bullseye comparison, the size versus number of element comparisons, described in the Pilot Studies sections (Chapter 3), were included in this task. The size - number comparisons were not used in the 3 month testing of the major study, partly because the preference measure was not reliable in the Pilot studies, and partly because the infants, within a session, were being given too much to do. The size - number comparison was re-introduced into the 5 month testing sessions for two reasons: first, the preference for greater over smaller number of elements, which develops at approximately 3 months, is still clear for 5 month old infants, and therefore the extent of this preference at this age may tell us something about the intellectual growth of an individual: and second, since infants at 5 months of age seem to be awake for longer periods of time than 3 month infants, lengthening the Visual Preference task is perhaps less likely to affect performance on subsequent tasks at 5 months, than at 3 months.

The procedure for the Visual Preference task at 5 months was identical to that described in the Pilot studies (see Chapter 3).

Concept Acquisition

In the infant control habituation procedure, older infants typically reach the criterion of habituation faster than do younger ones (Werner and Perlmutter, 1979). Rose and Slater (1983) suggest that 3 month old infants habituate in approximately twice the length of time that 5 month old infants require. In view of this, trial length for the 5 month testing sessions was reduced to half of that used with 3 month olds, from 30 to 15 seconds. The number of trials (ie. 8 familiar followed by two novel trials) was kept the same. In all other respects the Concept Acquisition task was identical to that reported in the Pilot studies sections.

6.3.2 Home Tasks

The Differential Vocal Responsiveness task was given in a similar way to that reported in the Pilot studies (Chapter3), and the task was administered on two testing occasions as it was for the 3 month testing sessions (see Chapter 5).

The Attention to the Novel Toy task was also administered in much the same way as in previous testing sessions, although infants were allowed to hold the toy, if they reached for it, and an extra measure of manipulation time was included (see Measures section below). This task was also administered on two occasions as it had been in the 3 month session (see Chapter 5).

6.4 MEASURES

The measures taken from each task were the same as those taken from the tasks in the Pilot Studies and from the 3 month testing. The reduced trial periods in the Concept Acquisition task allowed a maximum total fixation duration of 150 seconds.

The Attention to a Novel Toy task had an extra measure of total manipulation time added. This was a measure of the time (in seconds) infants spent manipulating the object, if they chose to do so.

6.4.1 Test-Retest and Split-Half Reliabilities

These were calculated in the same way as in the Pilot Studies and 3 Month Testing sessions.

6.5 RESULTS AND DISCUSSIONS

6.5.1 Sex Differences

Preliminary analyses revealed no major consistent sex differences, the data were therefore collapsed across sex.

6.5.2 Visual Preferences

16 infants completed this task successfully. Of the remaining 9 infants, one was unable to return to the laboratory for the second session within the required time period, the data of a further 6 infants were eliminated due to infant fussing, and the remaining 2 due to equipment failure.

Descriptive statistics from this task are shown in Table 6:1.

The test-retest and split-half reliabilities for the Visual Preference task at 5 months are presented in Table 6:2.

While there is some indication that a number of measures from this task are reliable within sessions, there were no significant test-retest reliabilities to corroborate their value as potentially predictive measures.

TABLE 6:1

VISUAL PREFERENCES 5 MONTH DATA

DESCRIPTIVE DATA N=16

MEASURE	SESSION	MEAN	SD	RANGE
% FIXATION GREATER NUMBER	1	54.62	13.44	28.22 to 81.75
	2	54.03	8.7	38.51 to 67.47
% FIXATION BULLSEYE	1	61.5	20.4	27.5 to 91.9
	2	55.6	20.0	26.47 to 96.91
% FIXATION LEFT	1	49.5	17.9	23.8 to 78.3
	2	48	12.3	24.1 to 69.8
TOTAL NUMBER FIXATIONS	1	42.7	10.5	24 to 60
	2	47.1	14	23 to 68
LONGEST SINGLE FIXATION	1	7.09	1.34	4.76 to 10
	2	5.97	2.11	3.54 to 10
MEAN LONGEST FIXATION	1	4.65	0.88	3.31 to 6.4
	2	4.01	1.05	2.69 to 5.69
FIRST FIXATION DURATION	1	2.284	1.29	0.46 to 4.69
	2	2.631	1.439	0.62 to 5.36
NUMBER FIXATIONS ACROSS STIMULI	1	23	7.53	7 to 36
	2	30.7	12.6	13 to 52
MEAN COMPLETION TIME	1	21.34	8.13	12.22 to 47.77
	2	16.54	3.07	13.4 to 24.36

TABLE 6:2

VISUAL PREFERENCES 5 MONTH DATA

TEST-RETEST (TRT) AND SPLIT-HALF (SH) RELIABILITIES N=16

MEASURE	TRT	SH SESS. 1	SH SESS. 2
% FIX GREATER NUMBER	-.40	.21	.17
% FIX BULLSEYE	-.54 *	.16	.15
% FIX LEFT	.46 *	.55 *	.12
TOTAL NUMBER FIXATIONS	-.34	.79 **	.92 **
LONGEST SINGLE FIXATION	-.09	.33	.87 **
MEAN LONGEST FIXATION	-.17	.47	.78 **
NO. LOOKS ACROSS	-.16	.73 **	.86 **
MEAN COMPLETION TIME	.01	.33	.03

* $p < .05$ (2-Tailed)
** $p < .01$ (2-Tailed)

A significant negative correlation occurred across sessions for the measure of percentage fixation to the bulls-eye, which is difficult to explain. The test-retest correlation for the percentage fixation to the greater number of elements was also negative and just misses significance ($r=-.4$). One possible expalnation for this is that infants show a strong preference for one type of stimuli of each pair, ie. either the bullseye or the stripes and either the greater or lesser number of elements, on the initial testing session but due to memory factors, for example, show a strong preference for the alternative stimulus on the second testing session. However, related t-tests revealed a significant preference for the bullseye on session 1 ($t=2.25$ $p < .05$) which was lost on session 2 ($t=1.12$ NS), while for greater number of elements, no significant preference emerged on session 1, although significance was reached on session 2 ($t=$ 1.38 NS and 1.85 $p < .05$, for the two sessions, respectively). If memory for stimuli from one session to the next had caused the negative test-retest reliability correlations, then we may have expected significant preferences on both sessions, but in opposite directions across sessions. This pattern of results then fails to support the suggestion that memory for a preferred stimulus on session 1 caused the negative test-retest correlations for the bullseye and greater number preference measures. The negative correlations may have occurred by chance.

On neither session was there a significant preference for the left or right positions.

Conclusion

The Visual Preference task administered at 5 months failed to produce any measures which can be considered potential predictors of later intellectual functioning.

6.5.3 Concept Acquisition

A total of only 12 infants completed this task on two occasions at the 5 month testing age. One infant was unable to return to the laboratory within the specified period for the second testing session, the data from 3 infants were eliminated due to equipment failure, and the data from the remaining nine infants were eliminated due to fussing.

Descriptive statistics for the measures from this task are presented in Table 6:3.

The test-retest and split-half reliabilities are presented in Table 6:4.

All of the measures which produced test-retest reliability at 3 months failed to do so at 5 months. In fact not one measure from the Concept Acquisition task at 5 months showed any evidence of being reliable across sessions.

TABLE 6:3

CONCEPT ACQUISITION 5 MONTH DATA

DESCRIPTIVE DATA N=12

MEASURE	SESSION	MEAN	SD	RANGE
TOTAL FIXATION DURATION	1	63.3	25.6	25.5 to 107.7
	2	54.9	19.9	23.8 to 97.4
TOTAL NUMBER FIXATIONS	1	23.8	7.35	13 to 40
	2	25.42	6.84	17 to 37
MEAN FIXATION DURATION	1	2.90	1.87	1.2 to 8.29
	2	2.17	0.58	1.4 to 3.16
LONGEST FIXATION DURATION	1	8.44	3.97	2.82 to 15
	2	6.97	2.11	3.11 to 9.84
MEAN LONGEST DURATION	1	4.58	2.32	1.55 to 8.33
	2	3.51	1.25	1.69 to 6.14
FIRST FIXATION DURATION	1	4.24	2.37	1.71 to 8.61
	2	4.32	2.55	2.06 to 10.76
MEAN LATENCY TO FIXATE	1	3.44	1.62	0.90 to 6.64
	2	3.27	1.99	0.53 to 5.51
% DECLINE FIXATION TRIAL 1- FIXATION TRIAL 8	1	4.48	216	-76 to 701
	2	-54.47	49.5	-85 to 83.3
% RECOVERY FIXATION TRIAL 8- FIXATION NOVEL	1	-12.3	49.7	-44.1 to 107.1
	2	0.935	42.5	-52.8 to 74

TABLE 6:4

CONCEPT ACQUISITION 5 MONTH DATA

TEST-RETEST (TRT) AND SPLIT-HALF (SH) RELIABILITIES N=12

MEASURE	TRT
6.4A OVER ALL TRIALS	
TOTAL FIX. DURATION	.22
TOTAL NO. FIXATIONS	.23
MEAN FIX. DURATION	-.20
SINGLE LONG. FIXATION	.18
MEAN LONG. FIX. DUR.	-.19
FIRST FIX. DURATION	.38
MEAN LATENCY TO FIX.	.54
% DECREASE (TRIALS 1-8)	-.19
% INCREASE (TRIALS 8-NOVEL)	-.45

MEASURE	TRT	SH SESS. 1	SH SESS. 2
6.4B FAMILIAR CONCEPT EXEMPLAR TRIALS			
TOTAL FIX. DURATION	.23	.69 *	.53
TOTAL NUMBER FIXATIONS	.27	.39	.63 *
MEAN FIX. DURATION	-.22	.89 **	.32
SINGLE LONG. FIXATION	.18	.76 **	.34
MEAN LONG. FIX. DUR.	-.13	.19	.57
MEAN LATENCY TO FIX.	.52	.42	.12
6.4C NOVEL NON-CONCEPT EXEMPLAR TRIALS			
TOTAL FIX. DURATION	.07	-.10	.68 *
TOTAL NUMBER FIXATIONS	.06	.00	.51
MEAN FIX. DURATION	-.27	-.23	.40
SINGLE LONG. FIX. DUR.	.52	-.24	-.42
MEAN LONG. FIX. DUR.	-.05	-.24	.42
MEAN LATENCY TO FIX.	.13	.76 **	.42

* $p < .05$ (2-tailed)
** $p < .01$ (2-tailed)

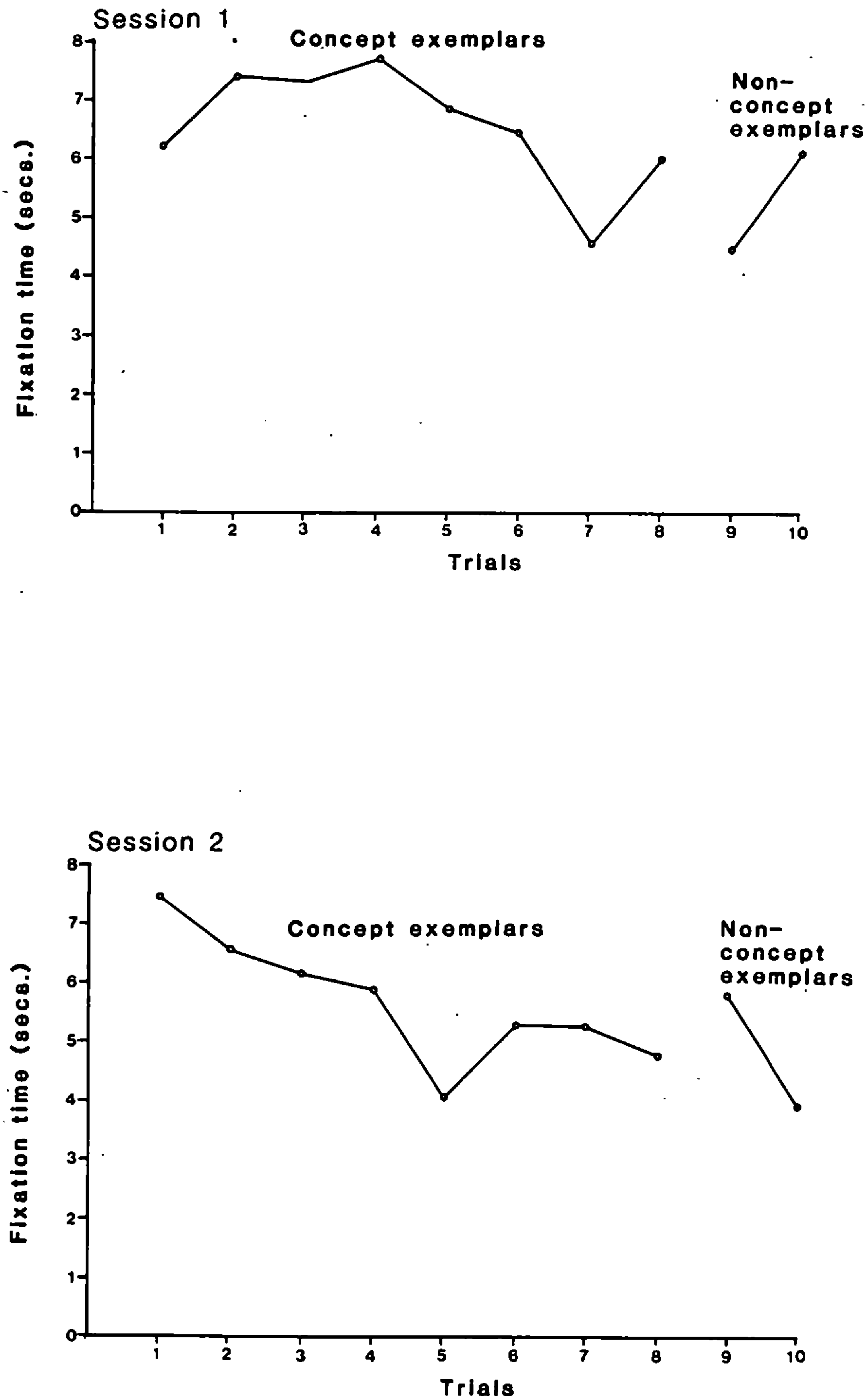
Since a number of significant correlations for the within session comparisons were significant, it must be concluded that performance on this task at 5 months is characterised by state fluctuations from day to day, rather than any stable performance across sessions.

The average fixation durations, over all infants, for each seperate stimulus presentation were calculated, and these are plotted in Figure 6:1. Trend analyses revealed that no significant decrement in fixation across the familiar, concept exemplar trials, occurred in session 1 ($F=2.65$ NS ($df=1,77$)), although this was significant for session 2 ($F=9.67$ $p> .01$ ($df=1,77$)). However, on neither session did t-tests (from the average fixation duration on trials 7 and 8, with the average fixation duration on trials 9 and 10) reveal a significant increase in fixation towards the novel non-concept exemplars. The Concept Acquisition task again failed to produce the expected pattern of responding, as indeed it did on both testing occasions at 3 months.

Conclusion

None of the measures from this task at 5 months can be considered potential predictors.

Figure 6:1. CONCEPT ACQUISITION - 5 MONTHS
The figure shows the average total fixation duration for each trial. n=12



6.5.4 Audio-Visual Integration

14 infants completed this task on the two testing sessions. One infant was unable to return for the second session within the required time period, and data from a further 9 infants were discarded due to fussing, and from one infant due to equipment failure.

Descriptive statistics for the measures from this task are presented in Table 6:5.

Table 6:6 shows the test-retest and split-half reliabilities of measures from this task. The data are characterised by a number of within session reliabilities but with no evidence of reliability across sessions.

T-tests revealed that there was no significant preference for the synchronised video on session 1, although this did reach significance on session 2 ($t=1.73$ NS and 2.43 $p < .05$, for session 1 and session 2, respectively).

Conclusion

None of the measures from this task at 5 months can be considered potential predictors.

TABLE 6:5

AUDIO-VISUAL INTEGRATION 5 MONTH DATA

DESCRIPTIVE DATA N=14

MEASURE	SESSION	MEAN	SD	RANGE
% FIXATION SYNCHRONISED VIDEO	1	55.7	15.1	26 to 80.3
	2	55.72	15.4	23.11 to 80.63
% FIXATION LEFT VIDEO	1	44.84	16.7	24.4 to 76.6
	2	60.9	18.4	18.6 to 87.1
TOTAL FIXATION DURATION	1	89.5	21.7	60 to 115.9
	2	79.8	21.6	50.8 to 116.6
TOTAL NUMBER FIXATIONS	1	37.36	8.43	25 to 51
	2	33.5	11.92	20 to 68
MEAN FIXATION DURATION	1	2.55	1	1.3 to 4.85
	2	2.52	0.69	1.34 to 3.36
LONGEST FIXATION DURATION	1	10.91	4.44	5.17 to 21.94
	2	11.47	4.07	7.49 to 21.02
MEAN LONGEST FIXATION DURATION	1	7.73	2.82	3.5 to 12.43
	2	7.07	2.49	4.88 to 14.67
NUMBER FIXATIONS ACROSS VIDEOS	1	26.36	9.87	8 to 40
	2	22.1	11.1	10 to 55
FIRST FIXATION DURATION	1	2.4	2.24	0.23 to 8.47
	2	2.31	2.21	0.47 to 7.40

TABLE 6:6

AUDIO-VISUAL INTEGRATION 5 MONTH DATA

TEST-RETEST (TRT) AND SPLIT-HALF (SH) RELIABILITIES N=14

MEASURE	TRT	SH SESS. 1	SH SESS. 2
% FIXATION SYNCH. VIDEO	.00	.10	.53
% FIXATION LEFT	.16	.01	.67 **
TOTAL FIXATION DURATION	.34	.20	.59 *
TOTAL NUMBER FIXATIONS	.34	.63 *	.63 *
MEAN FIXATION DURATION	.14	.49	.49
LONGEST SINGLE FIXATION	-.09	.55 *	.28
MEAN LONGEST FIXATION	.05	.53	.29
TOT. NO. LOOKS ACROSS	.53	.16	.46
FIRST FIX. DURATION	.37		

* $p < .05$ (2-Tailed)
** $p < .01$ (2-Tailed)

6.5.5 Operant Conditioning

17 infants completed this task at 5 months. 3 infants failed to complete the task due to fussing, and the data from 5 infants were disgarded due to equipment failure.

Descriptive statistics from the measures from this task are presented in Table 6:7.

The mean number of kicks over all infants for each minute of testing are plotted in Figure 6:2. T-tests for related samples revealed no significant change in responding either from the base-rate phase to the last two minutes of the reinforcement phase, or from the end of the reinforcement phase to the extinction period. The expected pattern of responding had not occurred in this task at 5 months.

The correlations between each of the 6 measures from this task are presented in Table 6:8. All the measures reflecting learning performance are very highly correlated with one another, whilst the 2 remaining measures (number of kicks to a lesser extent than time spent fixating the mobile), appear not to be directly related to the learning measures.

Conclusion

The six measures from this task were carried forward as potential

TABLE 6:7

OPERANT CONDITIONING 5 MONTH DATA

DESCRIPTIVE DATA N=17

MEASURE	MEAN	SD	RANGE
RATIO B TO A	1.53	1.64	0 to 7.28
RATIO C TO A	1.85	2.35	0.34 to 10.6
LAST B / A	1.40	1	0.42 to 3.73
HIGHEST IN B / A	2.22	2.54	0.71 to 11.6
FIX DURATION	458.9	67.9	294.9 to 540
TOTAL NO KICKS	124.7	66.8	49 to 321

Figure 6:2. OPERANT CONDITIONING TASK - 5 MONTHS
The figure shows the average number of kicks recorded in each minute of testing. n = 17

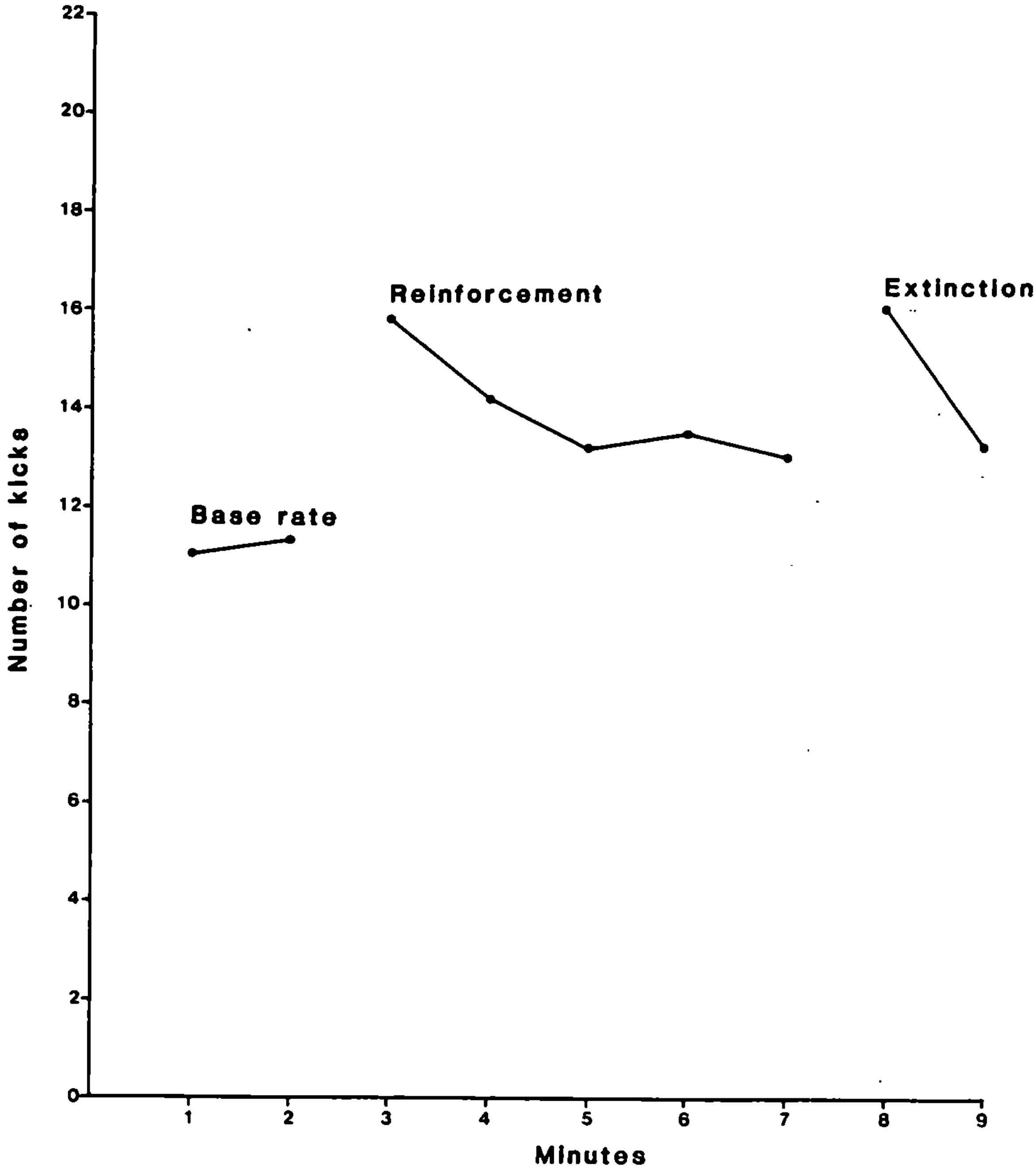


TABLE 6:8

OPERANT CONDITIONING 5 MONTH DATA

CROSS CORRELATIONS OF MEASURES N=17

MEASURE	RATIO B TO A	RATIO C TO A	END MINS B / A	GREATEST B / A	TOT FIX DUR	TOT NUMBER KICKS
RATIO B TO A	..					
RATIO C TO A	.92 **	..				
END MINS B / A	.83 **	.63 **	..			
GREATEST IN B / A	.95 **	.94 **	.74 **	..		
TOT FIX DUR	-.18	-.18	-.04	-.12	..	
TOT NO KICKS	.41	.29	.32	.31	.26	..

** $p < .01$ (2-tailed)

predictors, but in view of the absence of clear learning curves the likelihood of them being predictors must be considered to be low. All the measures were included separately in the analyses (in spite of some of them being strongly related with one another) because of the small number of measures carried forward as potential predictors from the 5 month data.

6.5.6 DVR

All 25 infants completed the DVR task successfully on two sessions.

Descriptive statistics for measures on this task are presented in Table 6:9.

The test-retest reliabilities from this task are presented in table 6:10.

The measures of total vocalisation to the mother and the stranger as well as the total vocalisation duration all produced significant across session reliabilities at 5 months, as they did at 3 months. For the DVR measure test-retest reliability was lost at 5 months.

The failure to find the DVR measure reliable from the DVR task at 5 months, may be due to the possibility that at 3 months (the age at which the infants in Roe's sample were tested), the difference measure reflects the ability of individual infants to

TABLE 6:9

DIFFERENTIAL VOCAL RESPONSIVENESS 5 MONTH DATA

DESCRIPTIVE DATA N=25

MEASURE	SESSION	MEAN	SD	RANGE
DVR	1	3.73	7.29	-6.6 to 23.76
	2	4.24	6.47	-6.6 to 16.6
TOTAL VOC MOTHER	1	12.4	8.13	0 to 28.3
	2	12.47	7.58	0 to 30.2
TOTAL VOC STRANGER	1	8.67	7.55	0 to 22.4
	2	8.24	9.26	0 to 36.8
TOTAL VOCALISATION	1	21.1	13.9	0 to 49.5
	2	20.07	15.6	0 to 67

TABLE 6:10

DIFFERENTIAL VOCAL RESPONSIVENESS 5 MONTH DATA

TEST-RETEST (TRT) RELIABILITIES N=25

MEASURE	TRT
DVR	.39
TOT. VOC. MOTHER	.62 **
TOT. VOC. STRANGER	.58 **
TOTAL VOCALISATION	.64 **

* $p < .05$ (2-tailed)
** $p < .01$ (2-tailed)

discriminate, between the mother and a stranger. By the time the infant reaches 5 months, however, a ceiling effect may have been reached for this discriminative ability, although at this age the sheer amount and nature of vocalisation itself may be of importance for predictive purposes.

Table 6:11 shows the correlations between the reliable measures for both sessions. All measures correlated highly with each other, and this is not suprising since they are all measures of vocalisation time. However, since so few measures at 5 months are forwarded into later analyses as potential predictors, the measures from this task were not combined.

Conclusion

The measures of total vocalisation and the amount of vocalisation to both the mother and the stranger, can be regarded as potential predictors of later intellectual functioning, but not the DVR measure at 5 months.

6.5.7 Attention to a Novel Toy

All 25 infants completed this task in the home at 5 months.

Descriptive statistics for these measures are presented in Table 6:12.

TABLE 6:11

DIFFERENTIAL VOCAL RESPONSIVENESS 5 MONTH DATA

CROSS CORRELATIONS OF RELIABLE MEASURES

SESSION 1

MEASURE	TOT. VOC DUR MOTHER	TOT VOC DUR STRANGER	TOT VOC DURATION
TOT VOC DUR MOTHER	..	.57 **	.89 **
TOT VOC DUR STRANGER	.72 **	..	.88 **
TOT VOC DURATION	.91 **	.94 **	..

SESSION 2

** p < .01 (2-tailed)

Correlations between measures from session 1 are presented above and to the right of the major diagonal, and for session 2 below and to left of the major diagonal

TABLE 6:12

ATTENTION TO A NOVEL TOY 5 MONTH DATA

DESCRIPTIVE DATA N=27

MEASURE	SESSION	MEAN	SD	RANGE
TOTAL FIXATION DURATION	1	26	23.9	4.9 to 100.3
	2	21.24	7.69	6.3 to 39.7
TOTAL NUMBER FIXATIONS	1	4.8	3.65	1 to 14
	2	3.6	3.11	1 to 15
MEAN FIXATION DURATION	1	8.26	6.57	0.98 to 23.77
	2	8.99	6.11	1.09 to 28.3
TOTAL MANIPUALTION DURATION	1	127.3	95.6	15.6 to 372.8
	2	120.9	65.7	33.4 to 290.1

As can be seen in table 6:13, which shows test-retest reliabilities from this task, the only measure from this task at 5 months to show across session reliability was that of total manipulation time.

The previously reliable measures at 3 months of total number of fixations and total fixation time showed no reliability at 5 months.

Conclusion

From this task at 5 months then, only the measure of total manipulation duration is considered a possible predictor.

6.5.8 Discussion - 5 Month Data Test-Retest and Split-Half Reliabilities

From the laboratory tasks at least, the 5 month data are characterised by an abundance of within session, or split-half reliabilities, but not by test-retest reliabilities.

The home based tasks on the other hand, did provide a number of reliable measures. A list of these, along with those from the operant conditioning task which was presented on only one occasion, is presented in Table 6:14.

The next phase of the 5 month analyses involved correlating these

TABLE 6:13

ATTENTION TO A NOVEL TOY 5 MONTH DATA

TEST-RETEST (TRT) RELIABILITIES N=25

MEASURE	TRT
TOTAL FIX. DURATION	.08
TOTAL NUMBER FIXATIONS	.31
MEAN FIXATION DUR.	.26
TOTAL MANIPULATION DURATION	.59 **

** $p < .01$ (2-tailed)

TABLE 6:14

MEASURES CONSIDERED POTENTIAL PREDICTORS AT 5 MONTHS

OPERANT CONDITIONING

1. Ratio of B to A
2. Ratio of C to A
3. Mins. 6 and 7 of B, divided by A
4. Greatest 2 mins. in B, divided by A
5. Total fixation time
6. Total number of kicks

DIFFERENTIAL VOCAL RESPONSIVENESS

7. Total vocalisation to the mother
8. Total vocalisation to the stranger
9. Total vocalisation duration

ATTENTION TO A NOVEL TOY

10. Total manipulation duration

TABLE 6:15

CORRELATIONS BETWEEN MEASURES ACROSS TASKS AT 5 MONTHS

The key to measure numbers is presented in Table 6:14

MEASURE

	7	8	9	10
1	.17	.07	.13	-.04
2	.03	.05	.05	.0
3	.27	-.11	.07	-.01
4	.12	.08	.11	.13
5	-.27	-.52 *	-.44	.4
6	.29	.16	.24	-.3
7				-.4 *
8				-.26
9				-.36

* $p < .05$ (2-tailed)

reliable measures, across tasks. As for the three month data, where the infants were tested on two occasions the mean scores for each infant over the two testing sessions were used.

These results of these analyses are presented in Table 6:15 and are discussed in the following sections.

6.6 RESULTS ACROSS TASK CORRELATIONS 5 MONTHS

6.6.1 Operant and DVR

17 infants completed these tasks at 5 months.

Table 6:15 shows the correlations of measures from these two tasks and are surrounded by a red ring. One significant negative correlation emerged between measures of total fixation time to the mobile and total vocalisation time to the stranger. This result may be spurious, since no other correlations emerged in support of it.

6.6.2 Operant and Attention to a Novel Toy

17 infants completed these two tasks at 5 months.

As can be seen in Table 6:15 (section highlighted by a blue ring) no significant correlations emerged between measures on these tasks at 5 months. This is similar to the pattern of results between these two tasks at 3 months.

6.6.3 DVR and Attention to a Novel Toy

25 infants completed both of these tasks, and the correlations between reliable measures across these tasks are presented in Table 6:15 (highlighted by a green ring).

The correlations between all of these measures are negative, with just one reaching significance, that between total vocalisation duration in the DVR task and the duration of object manipulation in the Attention to the Novel Toy task.

6.6.4 Conclusion

At 5 months the only potentially predictive measures came from the home based tasks and there was little clear evidence to suggest either that these are strongly related to each other or that they are related to the measures from the operant task. The 5 month data then do not provide any suggestion of any common underlying ability, but do suggest the presence of a small number of reliable measures which remain as potential predictors.

6.7 DISCUSSION AND COMPARISON OF 3 AND 5 MONTH RELIABILITIES.

The 3 month data have revealed measures showing both within- and between-sessions reliabilities, some of which also correlated with each other across tasks. Other measures appeared to remain independent potential predictors. This pattern of results was not apparent at 5 months where, at least for the laboratory based tasks, performance seems to be characterised by state fluctuations from day to day, rather than by any stable underlying ability. To some extent, this may be attributable to the tasks themselves, and this is discussed next.

One possibility is that the measures at 5 months may be producing less variance than at 3 months and hence any reliability at 5 months may be obscured. However descriptive statistics have shown that the 5 month measures do not produce standard deviations sufficiently close to zero as to distort potential reliabilities, and, in general, neither do they appear to differ greatly from the equivalent 3 month measures.

5 month olds, and not 3 month olds, appear to be motorically motivated and less content to sit and merely look at objects. This alone may make the tasks less appropriate for the purpose of obtaining individual measures of stability at 5 months. Where possible, performance on tasks across the two ages were compared to determine differences in task behaviour. Compared to the 3

month olds, the 5 month olds differed in performance in 3 ways:

1. Manual exploration dominated performance to a greater extent than looking (eg. Operant Conditioning and Attention to a Novel Toy).

2. Visual fixation time was less (eg. Audio-Visual Integration).

3. Infants cried and fussed more (probably a result of 1 and 2 above).

These points are made more explicitly in the following paragraphs.

1. The lack of a clear learning effect in the Operant Conditioning task may have been caused by the 5 month old infants being more interested in touching the felt toys on the mobile than in making them move. All of the infants at 5 months reached for the mobile at some stage in this task, whilst none apparently did so at 3 months. 5 infants' data were eliminated from this task at 5 months due to "equipment failure": this was in fact due to the mobile being actually pulled off the supporting ribbon. If touching the mobile became reinforcing throughout this task by 5 months, then infants exhibiting a large number of kicks were perhaps less likely to achieve this goal than infants who produced few kicks but kept their foot in the air, thus putting the mobile toys within reaching distance. The possible change in reinforcing activity, is likely to affect the meaning of the learning measures and their potential predictability at 5 months.

The lack of reliability of fixation time and number of fixations to the toy in the Attention to the Novel Toy task, can perhaps in part be explained by the change in the infants' involvement with the task. Since all of the infants at 5 months and none of those at 3 months, actually manipulated the object, and since the infants at 5 months in many cases were observed to manipulate the toy actively without fixating the object, it is not too suprising that reliabilities for measures of visual fixation are not in evidence at this later age.

2. The only task which allowed direct comparison from 3 to 5 months in terms of visual fixation time was the Audio-Visual Integration task. T-tests across age, were performed for equivalent measures, on the mean of each measure over the two sessions within each age, from those infants who successfully completed any given task at both ages. These revealed that for the audio-visual integration task the total fixation duration was significantly less at 5 months than at 3 months ($t=4.91$ $p < .01$) and mean longest fixation duration was also lower at 5 months ($t=3.8$ $p < .01$) whilst average fixation duration just missed significance ($t=3.3$ NS). These results suggest that at 5 months infants were fixating the stimuli less, and giving shorter average fixation times towards the stimuli than they were at 3 months.

Unfortunately, data from the Visual Preference and Concept Acquisition tasks are not directly comparable across age due to the changes in the experimental designs (see Procedure sections -

6.3). However, the measure of total fixation time towards the mobile, from the Operant Conditioning task, also produced a significant decline in duration over age ($t=2.21$ $p < .05$).

3. The numbers of infants successfully completing the tasks at 5 months was considerably lower than at 3 months. Between 21 and 29 infants completed each of the tasks at 3 months, as opposed to between 12 and 16 at 5 months. The decrease in subject was primarily due to infant distress.

With all of these factors considered, it would seem that the tasks chosen here are not as suitable for 5 month old infants as one might have hoped, and since at the age of 5 months virtually all infants are beginning manual exploration of objects, a different set of tasks at this age, requiring responses of a more active nature from the infant, may have produced a somewhat more fruitful pattern of results at this age.

At present the literature offers little in the way of alternative tasks for use with infants at 5 months of age, and many of the studies reviewed in Chapter 2 have reported successful employment of 5 month old infants in tasks which require visual responses. However, since reliability of measures from these tasks has not previously been systematically studied, the difficulties with using such measures as potential predictors are only now surfacing. The search for more fruitful alternative measures from infants at this age remains a topic for further research.

The next phase of this research involved the testing of the infants at 9 months of age. The next chapter deals with the choice of tasks for the 9 month testing sessions, the measures taken from each task, and the ways they correlate with each other.

CHAPTER 7

THE TESTING SESSIONS AT 9 MONTHS

7.1 INTRODUCTION

This Chapter is concerned with the choice of tasks for the 9 month testing sessions, the testing procedures adopted, and analyses of the data. The choice of tasks is considered first, with a brief description of how they were used in the 9 month study. Full details of the testing procedures and analyses follows in subsequent sections of this Chapter.

7.2 THE TASKS AT 9 MONTHS

Two tasks were presented to the infants at 9 months of age. These were Object Sorting and Object Search. Both of them meet the criteria necessary for measures from them to constitute potential predictors: each show changes in performance over age and each can be considered to reflect some degree of cognitive performance. The two tasks are discussed briefly below.

7.2.1 Object Sorting

The procedure for this task was similar to that reported by

Starkey (1981). Starkey described a longitudinal study to determine how early children will categorise groups of small objects. He presented infants at 6, 9 and 12 months, with a tray containing 8 objects. The objects formed two groups of 4, which were identical or made up an easily recognisable category. Starkey used a great variety of objects for this study, for instance 4 red hook-shaped objects made out of clay paired with four red squares of hardened clay, and four bottle-tops painted black and white paired with four yellow cubes. Over a 3 minute period a number of infant behaviours were monitored, including sequential touching, defined as the extent to which they touched in sequence similar objects, and object grouping which was a measure of the extent to which infants placed together similar objects. These measures are defined in detail in Section 7.5. Starkey found that both frequency of touching objects overall increased with age, as did the frequency of sequential touching and object grouping.

The ability of infants to recognise and categorise object sets has been considered important in terms of the onset of language use, and therefore performance of individuals on this task may have potential predictive power.

In his study, Starkey used many types of stimuli and for the purpose of this project, the stimulus types chosen were those which Starkey reported to elicit most touching and grouping behaviours. Details of these are given in the Method sections below. Test-retest reliability for measures on this task was attempted, and different stimulus types were used on each testing occasion. Unfortunately this was unsuccessful, since on the second testing session, one of the stimulus sets consisted of small clay balls, and virtually all of the infants put them immediately into their mouths! The resulting numerous interferences from E meant that test-retest data from this task was not meaningful. In the methodology sections, only the initial testing session is described.

7.2.2 Object Search

The performance of 9 month old infants in object search tasks has stimulated a large variety of experimental manipulations and theories, in attempts to explain the Piagetian "stage IV error". Essentially, an infant aged approximately 8-9 months, can successfully retrieve an object hidden in one (A) of two locations. However if the object, in full view of the infant, is then hidden in the second location (B), the infant will typically persevere with search at location A. Piaget (1954) argued that infants fail to search at B because they perceive the object in terms of a framework in which the object is itself only related to a particular action. Attempts to define the ways in which infants search for hidden objects have since escalated (see review by Harris, 1983). In spite of what may cause infants to err, relevant here is the fact that, for whatever reason, the error does appear to occur, and what is more performance on the task shows developmental progression in a number of respects.

The "stage IV error" typically occurs at approximately 9 months of age. Gratch and Landers (1971) reported a study in which they assessed developmental changes over age, from making the error through the ability to search successfully throughout all phases of the task. They tested subjects, from the age of 6 months, at 2 weekly intervals, and reported that at some point virtually all of their subjects erred by searching at location A on the B trials, having previously searched correctly at A. Also, by the time they were just over 10 months, nearly all infants searched successfully at B, on the B trials. Furthermore, Gratch and Landers reported an intervening 'mixed' behaviour phase, which appeared to be consistent for most subjects, and reflected a conflict over which side to visually focus on. This conflict was apparent in two respects.

The first source of conflict was apparent in terms of touching behaviour, such that infants would momentarily touch or reach towards one location, before actually pursuing search at the other location. This is interesting, since infants appeared to be affected by their observation of the hiding location (B), but also of their previous successful search behaviour at A.

The second source of conflict was apparent in terms of infants' fixation patterns throughout the sequence. After hiding the object at either location, a delay of 3 seconds was included, prior to infants being given the opportunity to search. Patterns of fixations toward the two locations throughout this period in B trials showed that initially infants would fixate B but then look towards and maintain fixation on position A. At a later 'mixed' developmental stage, infants fixated across from A to B repeatedly, whilst the period of successful search was characterised by continual fixation towards the location B. Furthermore the patterns of fixation at specific stages affected the resultant pattern of search. If subjects were at the error stage, then even if they fixated location B at some stage in the delay, they still typically showed search either at A or B. If infants produced a 'mixed' visual response, then they tended to search at either A or B, but if infants maintained fixation on B throughout the delay period they would also search there.

In view of the developmental trends reported above, this task was considered potentially useful for providing predictive measures at 9 months. An object search task was therefore included in this study, and the procedure adopted was similar to that used by Gratch and Landers (1971). Essentially infants were presented with a search task, with the object hidden consistently at location A until the infant retrieved it successfully on two consecutive trials. Following this a series of 5 trials were presented with the object hidden at location B. Test-retest reliabilities were not established for the measures taken from this task, since the initial testing phase may provide some understanding of how to perform the task. This possibility makes any lack of reliability, which might have been found, insufficient grounds for eliminating measures.

For neither task then was reliability assessed at 9 months of age, but, as discussed later in the Results sections of this Chapter, a number of high correlations between measures, both within and across tasks, indicated that some aspect(s) of performance at 9 months was reliable.

The following sections describe the procedures and designs used in the 9 month testing phase of the longitudinal study. Before administering the tasks to the main subject group, each was piloted with five 9 month old infants, who had not taken part in the major longitudinal study. This was to enable E to become familiar with the testing and videoing procedures involved. The data from these five are not presented here.

7.3 THE 9 MONTH TESTING SESSIONS - METHOD

7.3.1 Introduction

Two tasks were used for the 9 month testing phase, and the procedures adopted and analyses are presented in the following sections.

7.3.2 Subjects

24 subjects were visited in their own homes and administered the tasks at 9 months. All but one of the infants had taken part in the 5 month testing session and all had taken part in the 3 month

study. Unfortunately, one female infant failed to complete the 9 month testing within the required time period (see below).

7.3.3 Design

Within one week prior to their 9 month 'birthday' each infant was visited in the home on one occasion. The two tasks were always administered in a fixed order, with the Object Sorting task always preceding the Object Search task. Throughout the testing period the mother remained in the testing room, but seated herself out of sight of the infant. Mothers were also requested to remain silent throughout the testing periods.

7.3.4 Stimuli and Equipment

Object Sorting

The method adopted for this task was similar to that reported by Starkey (1971). The stimuli used were 4 yellow plastic 'ovals', measuring 5cm. by 3cm. across the major axes, and 4 brightly coloured plastic figurines, each measuring approximately 5 cm. when standing upright. These were presented to infants on a rectangular brown plastic tray measuring 48 by 32 cm. The figures were placed in specific positions on the tray which were marked by small dots. The position of the objects was the same for all infants, and they were placed on a fairly random basis, such that each object was approximately equi-distant from its neighbouring objects, and such that objects of the same category were not all positioned next to each other. The position of the items on the tray can be seen in the photograph, Figure 7.1.

A stop-watch was used to monitor the period of time infants were allowed to manipulate the objects, and a portable Panasonic video recorder with camera was used to video infants whilst performing the task.

Object Search

Presentation of the search task was done by means of a purpose-built ply-wood tray. The tray was rectangular in shape and measured 91.5 cm. by 30.5 cm. and was painted grey. On one side two extruding "wells" were attached, both measured 9.5 cm. by 9.5 cm. by 5 cm. These were postioned 29.5 cm. apart and symmetrical about the centre point of the tray. Two 15 cm. square white cotton handkerchiefs were used as well covers. The object for retrieval was one of three (for which infants showed a preference for - see Procedure section below). The items were a red rubber ball, a dumb-bell shaped rattle, and a silver bell, all of which were taken from the Bayley Scales of Infant Development. A photograph of the tray is presented in Figure 7.2, and the stimuli are shown in Figure 7.3.

Delay periods throughout the task (see below) were measured by means of a stop-watch, and the infants' performance on the task was videod by means of a portable Panasonic video recorder and camera.

FIGURE 7:1 STIMULI USED FOR THE OBJECT SORTING TASK

The photograph shows the tray with positioned objects, as used for the Object Sorting task.

FIGURE 7:2 THE TRAY USED FOR THE OBJECT SEARCH TASK

The photograph shows the tray used for the Object Search task. (The photograph, showing one well covered and one of the 'to be retrieved' items placed in the other, is for illustrative purposes only.)

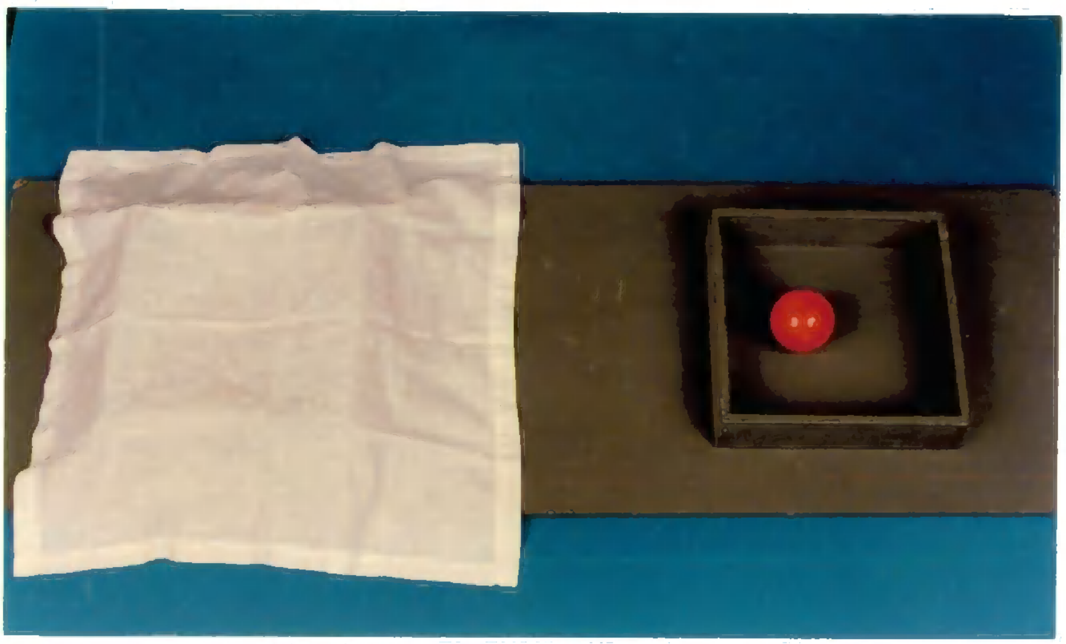


FIGURE 7:3 STIMULI USED FOR THE OBJECT SEARCH STIMULI

The photograph shows the three stimuli used for the Search task. Only one of these (that preferred by each infant) was used in the task.

7.3.5 Procedure

Object Sorting

The infants were placed in their own infant chair, and positioned in front of a table of a suitable height. As soon as the infant appeared content in this position, the tray of 8 objects was placed before them and left for a period of 3 minutes, during which time all manual and visual activity was videod. If the infant accidently knocked, or threw any object onto the floor (or even, as on two occasions, the complete tray!), then the displaced object(s) were replaced in their original position. When the 3 minute period had finished all objects and the tray were removed from the infant.

Object Search

The infants were tested in a similar position to that used in the Object Sorting task. Initially, infants were presented with the three items (the ball, rattle and bell) merely to determine which of them each infant was most interested in. Essentially, the object which was first picked up and played with interest acted as the "to be retrieved" object. All objects were then removed and the testing procedure began, commencing with A trials.

A Trials

When the infant was attending, E placed the preferred toy into well A. The toy was then raised and lowered into well A 3 times, while the infant was still attending. Both wells were then covered with the white cloths, and after a delay of 3 seconds, the tray was moved across the table towards the infant, hence providing opportunity to search for the toy. If the infants searched successfully on their first attempt, they were allowed to play with the toy for a few seconds before it was retrieved by E for the next trial. If the infant searched incorrectly (ie. at location B), or failed to search (ie. had not made any attempt to remove either cover after a 30 second period), then the tray was removed and the toy was exposed and shown to the infant by E. The next trial then commenced. A trials continued until the infant had searched successfully on two consecutive trials, at which point B trials followed thereafter.

B Trials

A total of 5 B trials was presented to each infant and these were given in the same way as the A trials, except of course for the hiding position which became position B.

For all subjects, position A was the well to the right of the infants' mid-line and position B was the well to the left.

The stop-watch was left running for the complete duration of the task, and positioned such that it was out of sight of the infant, but in sight of the E. Its sole purpose was to enable E to time accurately the 3 second delay periods.

The video recorder was positioned such that all of the infants' responses and visual attention to the task could be recorded.

7.3.6 Measures

A number of measures were derived from each task and these are listed below. Many of the measures for Object Sorting and Object Search tasks were taken from Starkey (1981) and Gratch and Landers (1971), respectively.

Object Sorting

1. Sequential touching. This was a measure of the extent to which infants touch and group objects in sequence. Starkey defined four levels of sequential touching: 1) All four of one type of object are touched, followed by all four of the other type; 2) All four of one type of object are touched, followed by three of the other kind; 3) All four of one type of object are touched in sequence; 4) Three objects of the same kind were touched in sequence.

A composite score of sequential touching was computed for each infant which was derived in the following manner: 4 points were awarded for each instance of level 1 above, 3 points for level 2, 2 points for level 3, and 1 point for level 4. The total number of points was summed for each infant over the 3 minute testing period. The higher the number of points an individual achieves, the more 'mature' his or her performance on this measure.

2. Object grouping. This is a measure reflecting the extent to which infants group together similar objects. Again the 4 levels as defined by Starkey were used: 1) Three or four objects of each kind are constituted as a group, and physically separated from other objects; 2) All four objects of one kind are constituted as a group and separated from other objects (this differs from 1 above, in that only one set of objects may be grouped and not both); 3) incomplete or partially correct groups are formed, such as groups of two or three similar objects separated from remaining objects; 4) two similar objects are held together and separated from other objects for at least 15 seconds, with the infant having looked back and forth between them at least twice.

As for the sequential touching measure, the number of instances of the levels were allocated points, with 4 points being scored for level 1, regressing down to 1 point for each incidence of level 4. Again the points were added and the more 'mature' infants scored the greater number of points.

3. Total fixation duration. This was a measure of the total amount of time infants spent fixating the tray or objects, throughout the 3 minute period.

4. Total manipulation time. This was a measure of the total amount of time infants spent actually manipulating object(s) throughout the 3 minute period.

5. Frequency of touching. This was a measure of the total number of times that infants touched any object. If the infant touched an object then withdrew, but touched the same object again later in the testing period, this was recorded as two touches. Hence this score could exceed 8, and was designed to reflect overall activity level of the infants.

Object Search

1. The total number of trials required for the infant to search successfully at location A on two consecutive trials, at the start of this task. This number should be fewer for the more 'advanced' infants.

2. The 'error run' score. This was a measure of the number of B trials which occurred before the infant successfully searched at location B. For an infant who searched at B on the first B trial, the error run score was zero, for an infant who failed to search, or who searched at A on the first B trial, but who searched correctly at B on the second B trial, the error run score was one, and so on. Again, the lower this score the more 'advanced' the infant.

3. The total number of correct searches made on the five B

trials.

4. The number of searches made at the incorrect location A during B trials.

5. The number of failures to search in B trails. One would assume here that the fewer failures to search, the more 'mature' the infant.

6. Delay period score. This measure was taken from each of the five B trials and the scores summed. Infants scored: 3 points if they actively leaned towards the tray in the 3 second delay period; 2 points if they maintained fixation on the tray during the delay period; and 1 point if they looked away from the tray for some of the delay period.

7. Presentation period score. Again a number of points were awarded and summed for each of the five B trials. The points were awarded as follows: 3 points if the infant removed the cover as soon as the tray was within reach; 2 points if the infant was hesitant in removing the cover; and 1 point if the infant was very slow to remove the cover, or refused to remove the cover at all. The greater number of points scored the more 'mature' the infant.

8. Total number of fixations towards location A, on B trials. This measure should reflect the extent to which infants portrayed the 'mixed' pattern of responding (see above). The fewer looks to location A (from location B) the more 'mature' the infant.

7.4 RESULTS

7.4.1 Object Sorting

All 24 infants completed this task. Descriptive statistics for all measures are presented in Table 7.1.

Cross correlations of the measures are presented in Table 7.2. The only significant correlations to emerge were between the number of points obtained in sequential touching and object grouping, and between both of these measures and the frequency of manipulation measure. This is not suprising since in order for infants to touch sequentially and group objects of a similar kind, they must touch a large number of objects. None of the remaining measures correlated with each other significantly.

However, the measures of number of points sequential touching, and the number of points object grouping, may not provide a clear indication of how individual infants performed on this task. To illustrate this, supposing 2 infants scored a total of 6 points on the sequential touching measure, one may have achieved this score by scoring once on level 3 (2 points) and four times on level 4 (1 point each time), whilst a second infant might have scored one instance of level 1 (4 points), and one instance of level 3 (2 points). One might assume in this case that the latter infant is more 'advanced' than the former, since the latter infant showed

TABLE 7:1

OBJECT SORTING 9 MONTH DATA

DESCRIPTIVE DATA N=24

MEASURE	MEAN	SD	RANGE
NUMBER OF POINTS SEQUENTIAL TOUCHING	3.25	2.63	0 to 10
NUMBER OF POINTS OBJECT GROUPING	2.79	2.99	0 to 9
TOTAL FIXATION TIME	128.3	32.2	85 to 180
TOTAL MANIPULATION TIME	164.2	18.2	109.4 to 180
FREQUENCY OF TOUCHING	12.54	6.29	2 to 23

TABLE 7.2

OBJECT SORTING 9 MONTH DATA

CROSS CORRELATIONS OF MEASURES N=24

MEASURE	1	2	3	4	5
1. POINTS SEQ. TOUCH	..				
2. POINTS OBJECT GROUP	.78 **	..			
3. TOT. FIX TIME	.07	.23	..		
4. TOT MANIP TIME	-.12	.17	.39	..	
5. FREQ. TOUCH	.72 **	.65 **	.36	.19	..

** $p < .01$ (2-tailed)

more 'mature' behaviour patterns. In view of this, measures achieved on each of the 4 levels for both measures of sequential grouping and object sorting (see measures sections) were examined. Unfortunately, however, the descriptive statistics revealed that for all of these measures there was very little variation. The only measure to produce a mean of greater than 1 was level 4 of the object grouping, and the standard deviations for all of these measures were close to zero. Further, when the individual measures from sequential touching and object grouping were correlated with the other measures from this task, the only significant correlations to emerge, as one would expect, were those with the measures of sequential touching and object grouping, and the frequency of touching measure. It seems likely then that these measures taken independently could contribute little in the way of potential predictors.

These data produce a problem in that the measures of sequential touching and object grouping are directly related to the frequency of touching measure. Causal relationships cannot be inferred from correlational data, so at this stage one cannot determine whether touching a greater number of objects increased the likelihood of obtaining more points on the sequential touching and object grouping measures. One possibility related to this is that those subjects who touched very few items may merely serve to add noise to the data, since then this would prohibit them from scoring a large number of points on the sequential touching and object grouping measures. However, when the data from 5 infants, who only touched five or fewer objects, were omitted from the correlations no significant effect on the magnitude of the resulting r's was found.

The measures of sequential touching and object grouping from this task appeared to be confounded by the number of objects that infants touched throughout the task, and these measures may be reflecting some aspect of overall activity level, as well as or instead of the ability to recognise different groups of objects. Partial correlations in the within and across age task analyses have been used to help clarify this position, in the 9 month across task analyses (see below).

Conclusion

All the measures from this task were forwarded to the longitudinal and within age analyses. Partial correlations were also used to try and assess the effect of frequency of touching on the measures of sequential touching and object grouping.

7.4.2 Object Search

22 infants completed this task, and the data from 2 infants were eliminated due to fussing and failing to search.

Descriptive statistics from the measures taken from this task are presented in Table 7.3. Unfortunately a number of the standard deviations are necessarily small and this may distort any potential predictive power they have.

TABLE 7:3

OBJECT SEARCH 9 MONTH DATA

DESCRIPTIVE DATA N=22

MEASURE	MEAN	SD	RANGE
NO OF TRIALS TO SUCCEED AT A	3.14	1.32	2 to 7
'ERROR RUN' SCORE	0.77	1.15	0 to 4
NO OF SEARCHES AT B ON B TRIALS	3.32	1.43	0 to 5
NO OF SEARCHES AT A ON B TRIALS	1.14	1.13	0 to 4
NO OF FAILURES TO SEARCH ON B TRIALS	0.45	0.8	0 to 3
SCORE IN DELAY PERIOD	10.41	2.84	6 to 15
SCORE IN PRESENTATION PERIOD	11.18	3.65	5 to 15
NO OF LOOKS AT A ON B TRIALS	1.95	1.56	0 to 5

A total of 9 infants actually made the search error by searching at location A on the first B trial, while 13 infants failed to make the error and searched successfully at B on the first B trial.

The correlations across measures within this task are presented in Table 7.4. The number of trials to succeed at location A (#1), correlates positively with the number of failures to search on the B trials (#5) and negatively with scores in the delay period. This is as one would expect: those infants who attended more in the delay period took fewer trials to first search on A trials; and those infants who took more trials to succeed on A trials showed more failures to search on B trials. Measures 2, 3, and 4 reflect aspects of infants' search behaviour, and these also inter-correlate in the way one would expect: those infants who had a low error run score also made fewer searches at A on B trials, and more searches at the correct location B. The more infants failed to search on B trails (#5) the fewer correct B searches were made (#3), and this may have been caused in part by the lack of reinforcing successful search, since the number of failures to search was not related to the number of incorrect searches at location A (#4). The scores obtained in the delay and presentation periods (#6 and #7, respectively) were correlated highly with each other (.61), and these both correlated negatively with the number of failures to search (#5), and to a lesser extent, with the number of incorrect searches at A on B trials (#4). Further, since the scores of performance in the

TABLE 7.4

OBJECT SEARCH 9 MONTH DATA

CROSS CORRELATIONS OF MEASURES N=22

MEASURE	1	2	3	4	5	6	7
1.NO TR. SUCC. AT A	..						
2.ERROR RUN SCORE	.05	..					
3.SEARCHES B ON B TR.	-.33	-.74 **	..				
4.SEARCHES A ON B TR.	.05	.79 **	-.83 **	..			
5.FAILS SEARCH ON B TR.	.52 *	.22	-.51 *	-.02	..		
6.SCORE IN DELAY	-.52 *	-.35	.70 **	-.45 *	-.53 *	..	
7.SCORE IN PRES.	-.41	-.18	.62 **	-.31	-.52 *	.61 **	..
8.LOOKS A ON B TR.	.54 **	.31	-.42 *	.36	.32	-.37	-.43 *

* p < .05 (2-tailed)
** p < .01 (2-tailed)

presentation and delay periods correlated positively with the number of correct searches at B on the B trials, it appears that measures of attention on the task are strongly related to search performance. The measure of the number of looks at A on B trials (#8) correlated negatively with the correct searches on B trials (#3) and also with the score on the presentation period.

In view of the very high correlations between the search measures (#2, #3 and #4) these measures were combined to form one composite 'search' measure for the longitudinal analyses, in the following way. For each measure individual scores were converted to z-scores, with the scores for measure 3 (number of correct searches at location B on the B trials) reversed. The three z-scores for each subject were added and this score was forwarded as the 'search' measure for remaining analyses.

Conclusion

It is apparent that performance on this search task is correlated with aspects of infants attention to the task.

The measures from these two tasks were correlated with each other, and the results of this analysis are considered next.

7.4.3 Across Task Correlations

22 infants successfully completed both of these tasks, and the scores for each measure on the Object Sorting task were correlated

with measures from the Object Search task. The results of these analyses are presented in Table 7.5.

The only measures from the Object Sorting task to produce significant correlations with measures from the Object Search task are: number of points sequential touching (#1), number of points object grouping (#2) and the frequency of touching (#5). All of these measures correlated significantly with the scores in both the presentation and delay periods (#5 and #4, respectively), in the search task. These correlations are in the expected direction: the greater the attention in the delay and presentation periods of the Search task, the more points achieved in the touching and grouping behaviours and the more objects touched in the Object Sorting task.

Number of points sequential touching (#1) correlated negatively with the search measure (#2), and negatively with the number of looks toward location A on B trials (#6), in the Search task. These correlations are in the expected direction: the more points of sequential touching achieved the more 'mature' infants' search performance, and the fewer looks towards location A on the B trials.

Also, the measure of frequency of touching (#5), in the Object Sorting task, correlated negatively with the search score (#2), from the Search task, indicating that the 'brighter' infants with the lower composite search score are also those who touched more objects on the Search task. Frequency of touching also correlated

TABLE 7.5

OBJECT SORTING AND OBJECT SEARCH 9 MONTH DATA

CORRELATIONS OF MEASURES ACROSS TASKS N=22

MEASURE OBJECT SEARCH	OBJECT SORTING 1 NO PTS SEQ TOUCHING	2 NO PTS OBJECT GROUPING	3 TOT FIX TIME	4 TOT MANIP TIME	5 FREQ TOUCHING
1 NO TR. SUCC. AT A	-.21	-.27	-.17	.01	-.37
2 SEARCH SCORE	-.48 *	-.37	-.15	-.30	-.56 **
3 FAILURES TO SEARCH ON B TR	-.21	-.25	-.04	-.01	-.56 **
4 SCORE IN DELAY	.47 *	.60 **	.31	.12	.54 **
5 SCORE IN PRES.	.46 *	.54 **	.14	.06	.63 **
6 LOOKS A ON B TR.	-.44 *	-.33	-.07	-.01	-.63 **

* p < .05 (2-tailed)
** p < .01 (2-tailed)

negatively with the number of failures to search on B trials (#3) and with the number of fixations toward location A on B trials (#6). These correlations again are in the expected direction.

The effect of this pattern of correlations was examined further, by looking at the correlations between the object touching and grouping measures, and Search task measures, with the measure of frequency of touching partialled out. Each subject was ranked in terms of his or her score on both sequential touching and object grouping measures, and for each subject, the two rank scores combined (since these measures are strongly correlated, r=.78). These scores were then correlated with the measures from the Search task but with the ranked frequency of touching measures partialled out. The results of this analysis yielded correlations all of which were approximately zero. When the partial correlation was done in reverse (ie. with the the effect of the combined measure of points of sequential touching and object grouping, partialled out of correlations between the Object Search measures and frequency of touching), then all of the significant correlations remain. This suggests that the frequency of touching measure is producing the correlations above, rather than the sequential touching and object grouping measures. In view of this, the measures of the number of points sequential touching and object grouping were not used in further analyses.

7.5 9 MONTH TASKS - SUMMARY AND CONCLUSIONS

Some of the data from the 9 month tasks unfortunately failed to produce a wide range of scores among infants.

Test-retest reliabilities were not calculated for any of these measures for either task, although some aspects of performance appeared to remain stable both within and across tasks. By and large, this was particularly true of those measures reflecting manual activity in, and attention to, tasks. These include sequential touching, object grouping and frequency of touching objects in the Object Sorting task, and the search measures and measures of attention in the delay and presentation periods of the Object Search task.

A list of all measures forwarded to the longitudinal analyses is presented in Table 7.6. There is no obvious reason to eliminate any measure at this stage.

The next Chapter deals with the infant scales presented to the infants at 18 months of age, the measures from which consistute the criterion measures for the longitudinal analyses, which are presented in Chapter 9.

TABLE 7.6

MEASURES CONSIDERED POTENTIAL PREDICTORS AT 9 MONTHS

OBJECT SORTING

1. Total fixation time
2. Total manipulation time
3. Frequency of touching

OBJECT SEARCH

4. Number of trials to suceed at location A
5. Search score
6. Number of failures to search on B trials
7. Score in delay period
8. Score in presentation period
9. Number of looks at location A on the B trials

CHAPTER 8

THE 18 MONTH STUDY

8.1 INTRODUCTION

The testing of infants at 3, 5 and 9 months has revealed a number of measures which meet the criteria for being potentially predictive. The next phase of the research involved collecting criterion measures from the subjects to determine whether the measures already collected are predictive of later intelligence. This Chapter is concerned with the the choice of the scales adopted for the 18 month study. Data from the scales are also presented and discussed.

8.2 THE CHOICE OF SCALES AT 18 MONTHS

Ideally, data from the subjects on scales of childhood intelligence would constitute the most suitable criterion measures, since these measures correlate highly with adult intelligence levels. For instance, Bayley (1949) reported r's ranging from .62 to .96 between intelligence scores taken between three and a half and eighteen years. Unfortunately it was not possible to collect childhood intelligence measures from the subjects here, within the time limitations of this research. In view of this, one is restricted to the use of infant intelligence scales, which do appear to start increasing in terms of predictive validity at the age of 18 months to 2 years (see Chapter 1).

A number of the studies reported in Chapter 2, which claim to have found significant correlations between aspects of performance in infancy and subsequent intellectual abilities, have used scores on infant scales as their criterion measures. These include; the work by Miller and co. workers (Miller, Ryan, Short, Reis, McGuire and Culler, 1977) who have used the Uzgiris and Hunt Scales at 15 months: Bornstein and Ruddy (1984) and Lewis and Brooks-Gunn (1981) who used the Bayley Scales of Infant Development at 12 and 24 months of age, respectively: and Kopp and Vaughn (1982) who used the Gesell scales at 2 years.

There has also been considerable emphasis on the use of scales and measures reflecting linguistic and language related abilities as criterion measures. In particular these include the work by: Fagan and McGrath (1981); Roe and co workers, (Roe, 1978; Roe and McClure, 1980); Rose, Slater and Perry (1986); Ruddy and Bornstein (1982). These studies have been reviewed in Chapter 2.

There is then sufficient evidence in the literature to warrant justification for the use of some measures from current infant intelligence scales, and in particular those reflecting language and linguistic abilities, in the final study of the present research. Furthermore language is of particular importance here from a theoretical standpoint. This has been discussed in Chapter

1, and is referred to in Chapter 10.

The measures chosen for criterion measures for this research then included; the Bayley Scale of Infant Development (Mental Developmental Index only): the Reynell Developmental Language Scales (Revised version): and the Griffiths Mental Development Scales (Hearing and Speech sub-scale only). Each of these is discussed in turn below.

8.2.1 The Bayley Scales of Infant Development (BSID)

A number of predictive studies to date which have concentrated on infants of up to 2 years of age have favoured the Bayley Scales of Infant Development (BSID). Seigel (1981), Messer, McCarthy, MacQuiston, MacTurk, Yarrow, and Vietze (1986), Lewis and Brooks-Gunn (1981), and Ruddy and Bornstein (1982), all report significant correlations between measures in infancy and subsequent BSID scores, and further, some of these studies have also assessed infants on the BSID at both testing ages and have found that their cognitive measures predict the later BSID score to a greater extent than do the earlier BSID scores (Messer, McCarthy, MacQuiston, MacTurk, Yarrow, and Vietze (1986); Lewis and Brooks-Gunn, 1981).

Hence the BSID was adopted in this research. Further, since aspects of motor performance in infancy predict very little (see Chapter 1) only the Mental Developmental Index of the BSID was included. Bayley (1969) reports a corrected split-half reliability of .89 for the MDI at 18 months, although test-retest reliabilities at this age are not reported.

A maximum of 39 items were presented to the infants from the scale, commencing with item number 117 and continuing with subsequent items until the infant had failed at least 10 successive items on the scale. Item number 117 was considered a suitable starting point, since this item is passed by infants aged just over fifteen months and it could reasonably assumed that all infants tested here would pass all items below it (see Bayley, 1969 for procedural details). In fact, no infant passed more than 23 presented items (see Results Section below).

The actual items on the MDI can be divided into at least two categories, one of which is those which appear to measure essentially motor based performance (such as placing shaped blocks into similar shaped locations on a board, placing pegs in holes, building towers with blocks, etc), the other being those items constituting essentially language based items (such as "says two words", "names objects", "follows directions" etc). Each of the 39 items presented to the infants were placed into one of the categories. Appendix 1 lists each item and the subsequent category into which it was placed.

Three measures were derived from the Bayley MDI; the MDI Raw score which was a measure of the total number of items passed (including all items below the infant's basal rate - see Bayley, 1969): MDI Language score which was the number of items placed in the language category (taken from a maximum of the 18 items - see Appendix 1), and the MDI Performance score which was the total number of items passed in the performance category (taken from a maximum of 21 items - see Appendix 1).

8.2.2 The Reynell Developmental Language Scales - Revised (RLDS)

The assessment of language development in the present research was achieved by adopting the Reynell Developmental Language Scales (RDLS). The test is comprised of 2 scales; Expressive Language and Verbal Comprehension. The Expressive Language scale is further divided into 3 subscales; Language Structure, Vocabulary, and Content. The Content sub-scale was not appropriate for 18 month olds, but the Structure and Vocabulary sub-scales were scored independently.

Reynell (1978) fails to report the test-retest reliabilty of these scales, although the split-half reliabilities are reported as .93 and .82 for the complete Expressive Language and Verbal Comprehension scales, respectively, at the 18 month age. Further, Silva, Bradshaw and Spears (1978) examined the concurrent and predictive validity of the scales by correlating measures of these with 6 other tests, and studying the correlations between the RLDS at 4 years and reading ability and other intelligence scales at 7 years. The data were factor-analysed and the conclusion reached by Silva et al. was that the RLDS correlated highly with other measures taken at the same age and with tests at the later age, in particular those measures forming a language factor, but also with those measures weighting on a non-verbal factor. Hence the RLDS appears to be characterised to some degree by both concurrent and predictive validity, and the scales appear to tap not only verbal ability, but also some aspects of non-verbal performance as well.

In conclusion, although the RLDS has been researched little (especially the late infancy period), it was included in the current research in view of the apparent importance of language measures in the prediction literature. 18 months is the earliest age for which the measures have been fully standardised.

8.2.3 Griffiths Mental Development Scales

The Griffiths Mental Development Scales are made up of 5 sub-scales; Locomotor scale, Personal-Social scale, Hearing and Speech scale, Hand and Eye Coordination scale, and the Performance scale.

Test-retest and split-half reliabilities for the Griffiths scales have not been systematically studied, although 60 children were tested twice on the full scale with between session periods ranging between 7 and 70 weeks and the test-retest correlation was .87. Griffiths (1976) reported that this correlation increased to .92 when the data from 8 infants were eliminated due to their unusual circumstances (such as prematurity, jaundice, feeding difficulties and malnutrition). Although there is an apparent increase of .05 this new value is well within the confidence limits of .87 with a sample size of 60.

For the purposes of this research, only the Hearing and Speech scale from the Griffiths was used. This was due to 2 reasons; first, the apparent irrelevence of the use of motor behaviours as predictors of later intelligence, and second, the apparent importance of language based measures as predictors of later performance.

8.2.4 Summary

To conclude, the scales administered to the infants at 18 months of age were; the Bayley Scales of Infant Development (Mental Developmental Index only, which was also divided into components of Verbal and Performance Scales): the Reynell Developmental Language Scales (including the Comprehension and Productive scales, the latter of which was divided further into Structure and Vocabulary): and the Griffiths Developmental Scale (Hearing and Speech Subscale only).

8.3 DESIGN

8.3.1 Subjects

All of the 24 infants who had been tested at 9 months were tested at 18 months.

8.3.2 Procedure

All of the infants were visited on one occasion in their homes, within one week either side of their 18 month 'birthdays'. Prior to the start of the testing session, E played with the infant in order to establish a good rapport. When the infant seemed relaxed with the E, testing began. The scales were always administered in the same order, starting with the Bayley MDI, followed by the Reynell Comprehension and Production scales, respectively, and finally the Griffiths Hearing and Speech scale. A number of the items for two, or occasionally all three scales were identical (eg. 'uses (a particular number of) clear words', 'word combinations' - Reynell and Griffiths; and 'identifies (a particular number of) objects in a box' - Bayley MDI, Reynell and Griffiths). Infants were credited for these items on each scale if they were successful on at least one occasion. This clearly will have the effect of increasing the correlations between measures, although all scales also had a number of items specific to themselves.

8.4 MEASURES

7 measures were extracted from the data;

1. The Bayley MDI
2. The Bayley MDI - Verbal measures
3. The Bayley MDI - Performance measures

4. The Reynell Comprehension
5. The Reynell Production - Structure
6. The Reynell Production - Vocabulary
7. The Griffiths - Hearing and Speech

8.5 RESULTS

All 24 infants completed the scales within the specified time period. Two infants were visited twice due to fussing on the initial testing period.

8.5.1 Descriptive Statistics and Correlations Between Measures

Descriptive statistics for the 7 measures are presented in Table 8.1. All scales produced a reasonable range of scores in this sample, which is a pre-requisite for the measures to consitiute potentially useful criterion measures.

The mean Bayley MDI raw score was 133.62, which when converted to the MDI standardised score (see Bayley, 1969) gives an average of 117. This is slightly above average, and corresponds to the average performance of infants aged approximately 20 months. For all infants in this sample, the scores on the Bayley MDI range between 129 and 140 points, which one would expect approximately for 19 month olds, and 23 month olds, respectively. Hence the

TABLE 8.1

18 MONTH DATA

DESCRIPTIVE DATA N=24

MEASURE	MEAN	SD	RANGE
BAYLEY MDI	133.62	3.46	129 to 140
BAYLEY MDI VERBAL	7.92	3.32	3 to 14
BAYLEY MDI PERFORMANCE	10.17	2.79	4 to 16
REYNELL COMPREHENSION	14.08	3.45	7 to 20
REYNELL PRODUCTION STRUCTURE	11.08	2.3	6 to 16
REYNELL PRODUCTION VOCABULARY	5.58	2.32	0 to 9
GRIFFITHS HEARING AND SPEECH	38.87	3.4	35 to 48

Note - The mean scores from the measures of Bayley MDI verbal and Performance do not sum to the total mean score for the Bayley MDI, since the verbal and performance measures were taken from only a sub-set of items from the complete MDI scale. The complete MDI scale includes scores for all measures below the basal level (see Bayley, 1969) infants were scoring slightly above average on this scale.

No statistics are available to assess performance separately on the Verbal and Performance sub-scales of the Bayley MDI as defined here (see Appendix 1).

The mean score for the Reynell Comprehension scale was 14.08 which is the approximate equivalent score for infants aged 19 months. The range of scores varied between 7 and 20 for this scale, and these scores are equivalent scores for one year and two years, respectively. The Language Production scale (Production Structure and Production Vocabulary scales combined) produced a mean score of 16.67, which corresponds to the equivalent of infants aged 21 months in the standardisation sample. The range of scores on this scale varied between 9 and 25, scores which correspond to 12 month and 26 month old infants, respectively.

As they had on the Bayley MDI, infants on the Reynell scales on average are scoring slightly above what is expected of their age group.

The average score on the Griffiths Hearing and Speech sub-scale was 38.87, which is fractionally below the expected standardised score for 18 month olds, with their average score lying between the expected score for 17 and 18 month olds.

8.5.2 Inter-Correlations of 18 Month Measures

The inter-correlations between all of the 18 month measures are presented in Table 8.2. All of the correlations are Pearsons.

It can be seen that all of the measures correlate highly with one another, with the exception of the 'performance' measures of the Bayley MDI. This suggests that the performance measures are picking up some ability which is distinct from all of the language-based subscales. The language sub-scales, on the other hand all correlate highly with each other and therefore one can assume that they are all reflecting the same sorts of abilities. This pattern of results in itself adds weight to the case for dividing the Bayley MDI into the sub-scales as defined here (see Appendix 1), since if, for example language based measures are found to be predicted by the earlier measures, but performance measures are not, then addition of the performance measures into the correlations may only serve only to add 'noise' to the data.

In view of this very clear pattern of inter-correlations between the language based scores, a composite score was forwarded for the longitudinal analyses. For each of the five language based measures (ie. the Bayley MDI Verbal, the Reynell Comprehension, the Reynell Production Structure and Vocabulary, and the Griffiths Hearing and Speech measures) scores were ranked and and the rank scores summed for each infant.

TABLE 8.2

18 MONTH DATA

CROSS CORRELATIONS OF MEASURES N=24

MEASURE	1	2	3	4	5	6
1.BAYLEY MDI	..					
2.BAYLEY MDI VERBAL	.86 **	..				
3.BAYLEY MDI PERFORMANCE	.37	.06	..			
4.REYNELL COMPREHENSION	.80 **	.80 **	.17	..		
5.REYNELL PRODUCTION STRUCTURE	.64 **	.79 **	-.06	.78 **	..	
6.REYNELL PRODUCTION VOCABULARY	.69 **	.69 **	.09	.77 **	.66 **	..
7.GRIFFITHS HEARING AND SPEECH	.65 **	.68 **	.09	.65 **	.80 **	.67 **

** $p < .01$ (2-tailed)

The Bayley MDI Performance score was forwarded as a separate criterion measure as was the complete Bayley MDI.

8.6 18 MONTH DATA - SUMMARY AND CONCLUSIONS

For all scales, with the exception of the Griffiths Hearing and Speech scale, infants performed slightly above the expected standard for their age at the time of testing. This apparently slightly advanced performance of the subjects may be due to either the E having been lenient on scoring the Scales, or due to the infants themselves being above the norm. The latter in particular seems likely, since all of the infants came from middle class homes. With respect to the former, if the E had been lenient on scoring it was felt that scoring had been consistent for all subjects.

The Griffiths sub-scale was the final one to be administered and the slightly poorer performance on this scale may have been due to fatigue or boredom. Alternatively, the scale may merely differ from the others in terms of difficulty.

Two types of abilities appear to have emerged from the scales administered. The first reflecting language based skills from the Bayley MDI 'verbal measures', the Reynell Comprehension scale, the Reynell Production scales (both structure and vocabulary), and the Griffiths Hearing and Speech subscale. The second ability is reflected only by 'performance measures' from the Bayley MDI. A composite 'verbal' measure, and the performance measure were used as criterion measures in the longitudinal analyses, as was the Bayley MDI full scale measure.

The next stage of the analyses was to determine the ways in which measures, from infants at all of the ages tested, inter-relate, in an attempt to illustrate any continuities which emerge from the measures taken over this period. The next Chapter is concerned with these analyses.

CHAPTER 9

PREDICTION OF PERFORMANCE ACROSS AGES: LONGITUDINAL ANALYSES OF 3, 5, 9 AND 18 MONTH DATA

9.1 INTRODUCTION

One of the major concerns of this research project is to assess whether any of the reliable measures of performance within testing ages, also produced significant correlations across age. It was hoped that these analyses would contribute to our understanding of the nature and consistency of intellectual growth in infancy.

This Chapter is concerned with how measures from each testing age intercorrelate across ages. Only those measures which have been considered potentially predictive are included in these analyses. These measures have been listed in Tables 5.17, 6.14 and 7.6, for the 3, 5 and 9 month ages, respectively. At 18 months, the measures used were the Bayley MDI, the verbal composite measure, and the performance measure derived from the Bayley MDI (see Chapter 8). To aid clarification of the importance of the Verbal and Performance based measures, the correlations between the Bayley MDI Verbal score and potentially predictive measures are also presented.

For those measures from tasks at 3 and 5 months which were administered on 2 testing occasions (ie. Concept Acquisition, Audio-Visual Integration, DVR and Attention to a Novel Toy), the individual's average score across the two testings was used in the longitudinal analyses. For those measures which were administered on one testing session only (ie. Operant Conditioning measures at 3 and 5 months, and all 9 and 18 month measures), single scores were used in the longitudinal analyses. Also, some measures were combined to form composite scores for the longitudinal analyses. For each measure which contributed to a composite score, an individual's score was converted to a standard score (its deviation from the mean was expressed in terms of standard deviation units), and the composite score was calculated by summing the separate standard scores. These have been described where appropriate in the preceeding experimental Chapters.

All correlations in this Chapter have been calculated such that 'mature' performers on each measure have the higher scores, at least for those measures on which this direction can be determined (this, for example, does not include the measure of percentage fixation to the video on the left, from the Audio-Visual Integration task). This means that scores for some of the measures, where low scores indicate 'mature' performance (for example measures of aspects of fixation duration), have been reversed. Hence, throughout, positive correlations indicate that infants obtaining 'good' scores on one measure also do so on the other measure.

Both Pearson and Spearman correlations are presented for these analyses, and the two-tailed significance levels are indicated. Spearmans correlations are presented in parentheses. The reasons for presenting Spearmans correlations here were to determine whether any significant Pearsons correlations were affected by a few extreme scores, rather than a general trend in the data. However, since those measures producing Pearson correlations significant to the .01 level of significance, produced in all instances Spearmans correlations to at least the .05 level, and vice versa, it would appear that the significant correlations are caused by general trends in the data rather than by one or two extreme scoring subjects. The pattern of correlations resulting from the Pearsons correlations is given greater weight in the discussion sections.

Sections 9.2 through 9.5 present the cross-correlations between measures from tasks at each combination of the different ages. Within each task at each age, many of the potentially predictive measures correlate significantly with each other, often because of some degree of statistical dependence (eg. the measures of fixation in the Concept Acquisition task). In some cases, where measures within a task were statistically dependent and were found to correlate very highly with each other, composite measures have already been derived. These include the learning measures from the Operant Conditioning task (at 3 months), and the search measures (from the 9 month Object Search task). Details of these composites have been presented in the preceeding experimental chapters. However, as will become apparent from the following longitudinal analyses, some measures within age, which correlate with some later measure, fail to correlate with each other, or do so only moderately. From these patterns of data it would be of interest to assess the maximum prediction which can be obtained from the combined effects of the independent measures. Multiple correlations and regression analyses would be appropriate for this, although their use is precluded by the low subject numbers and large number of measures, which makes any interpretation of such analyses very tentative (see Chapter 4). However, correlations between a number of combined measures at one age and one criterion measure from a later age are presented here, in order to illustrate the predictive ability of combinations of such measures. In these cases, all of the composite measures were created by adding the standard scores for subjects on all of the measures included (see above and subsequent sections). These analyses provide an indication of the maximum prediction one would expect from a combination of equally weighted individual measures which each correlate with the criterion measure. These are presented in the Discussion sections, for each age comparison, below. Full discussions of the patterns of correlations across age are presented in Chapter 10.

Those correlations which include the 3 month measures are considered first starting with the correlations with measures from tasks at 5 months (section 9.2), followed by correlations with measures on tasks at 9 months (section 9.3). The correlations between 3 and 18 month measures are considered in section 9.4. Then the 5 month measures and the correlations with 9, and 18 month measures are considered in sections 9.5 and 9.6, respectively, followed by inter-correlations between the 9 and 18 month measures (section 9.7).

The 3 month measures are considered first.

9.2 CORRELATIONS BETWEEN 3 AND 5 MONTH MEASURES

All of the correlation matrices involving measures from the 3 and 5 month testing sessions produced a total of just 5 significant Pearson correlations out of a total of 136 (this approximates 3.6 percent). Four of the Spearman correlations reached significance and these were four of those which also reached significance as Pearsons. There are clearly no more significant correlations than would be expected by chance at the 5 percent level. The conclusion reached here was that from the measures adopted in this research, there is no evidence of stability of performance from 3 to 5 months of age. A brief discussion of these data is presented in section 9.6 when all the predictive data involving 5 month measures has been presented.

9.3 CORRELATIONS BETWEEN 3 AND 9 MONTH MEASURES

9.3.1 Introduction

These sections are presented in the following format. The 3 month tasks are considered in the sequence used in Chapters 5 and 6 (ie. Concept Acquisition, Audio-Visual Integration, Operant Conditioning, Differential Vocal Responsiveness, and Attention to a Novel Toy). In Sections 9.3.2 through 9.3.6 the correlations between each of these 3 month tasks in turn, and measures from the 9 month Object Sorting task are presented, and this is followed by each 3 month task in turn and the correlations with the 9 month Object Search task (sections 9.3.7 through 9.3.11). First then, the 3 month Concept Acquisition task and the 9 month Object Sorting task.

9.3.2 Concept Acquisition at 3 Months and Object Sorting at 9 Months

22 infants completed these tasks.

The correlations between the measures from these tasks are presented in Table 9.1.

The only correlation to reach significance was that between the measures of average fixation duration (from the 3 month Concept Acquisition task) and the frequency of touching measure (from the Object Sorting task at 9 months). This correlation reaches significance only on the Pearson test.

TABLE 9.1

CORRELATIONS OF MEASURES ACROSS TASKS 3 - 9 MONTHS

CONCEPT ACQUISITION (3 MONTHS) AND OBJECT SORTING (9 MONTH) N=22

MEASURE OBJECT SORTING (9 MONTHS)	CONCEPT ACQUISITION (3 MONTHS) TOTAL FIXATION DUR	AVERAGE FIXATION DUR	MEAN LONGEST FIXATION DUR	SINGLE LONGEST FIXATION DUR
TOTAL FIXATION DURATION	-.07 (-.05)	-.24 (-.23)	-.14 (-.16)	.00 (.07)
TOTAL MANIPULATION DURATION	.11 (.01)	.23 (.32)	.22 (.24)	.15 (.11)
FREQUENCY OF TOUCHING	.32 (.21)	.48 * (.41)	.39 (.31)	.41 (.32)

* $p < .05$ (2-tailed)

Spearmans rank correlations are shown in parentheses.

9.3.3 Audio-Visual Integration at 3 months and Object Sorting at 9 months

16 infants completed these tasks.

The correlations between the measures from these tasks are presented in Table 9.2.

The Spearmans correlations produce a significant relationship between the percentage fixation duration to the stimuli on the left of the infants mid-line (3 months) and total fixation duration (at 9 months). There is no obvious logical explanation for this, and since the Pearson correlation misses significance it is not discussed further.

The only Pearson correlation to reach significance was that between the measures of average fixation duration (at 3 months) and frequency of touching (at 9 months). The equivalent Spearmans correlation fails to reach significance.

9.3.4 Operant Conditioning at 3 months and Object Sorting at 9 months

21 infants completed these tasks.

The correlations between measures from these tasks are presented in Table 9.3.

TABLE 9.2

CORRELATIONS OF MEASURES ACROSS TASKS 3 - 9 MONTHS

AUDIO-VISUAL INTEGRATION (3 MONTHS) AND OBJECT SORTING (9 MONTH) N=16

MEASURE OBJECT SORTING (9 MONTHS)	AUDIO-VISUAL INTEGRATION (3 MONTHS)				
	% FIXATION SYNCHRONISED	% FIXATION LEFT	TOTAL FIX DURATION	AVERAGE FIX DURATION	FIRST FIX DURATION
TOTAL FIXATION DURATION	.03 (-.07)	.49 (.53 *)	.36 (.21)	-.19 (-.15)	-.16 (.10)
TOTAL MANIPULATION DURATION	.13 (-.08)	.42 (.27)	.02 (.08)	.18 (-.15)	-.07 (-.05)
FREQUENCY OF TOUCHING	-.12 (-.21)	.24 (.30)	.07 (-.13)	.53 * (.36)	.48 (.30)

* p < .05 (2-tailed)

Spearmans rank correlations are shown in parentheses

TABLE 9.3

CORRELATIONS OF MEASURES ACROSS TASKS 3 - 9 MONTHS

OPERANT CONDITIONING (3 MONTHS) AND OBJECT SORTING (9 MONTHS) N=21

MEASURE OBJECT SORTING (9 MONTH)	OPERANT CONDITIONING (3 MONTH)		
	LEARNING MEASURE	TOT FIX DURATION	TOT NO KICKS
TOTAL FIXATION DURATION	-.11 (.13)	.28 (.31)	.31 (.22)
TOTAL MANIPULAUION DURATION	.06 (.04)	-.06 (.08)	.19 (-.07)
FREQUENCY OF TOUCHING	-.09 (-.10)	.51 * (.56 **)	.54 ** (.57 **)

* p < .05 (2-tailed)
** p < .01 (2-tailed)

Spearmans correlations are shown in parentheses

The only measures from the 3 month Operant Conditioning task to correlate significantly with measures from the 9 month Object Sorting task, were total fixation duration and total number of kicks. Both of these measures correlated with the measure of frequency of touching. This pattern of correlations is similar for both types of correlation.

9.3.5 Differential Vocal Responsiveness at 3 months and Object Sorting at 9 months

23 infants completed these tasks.

The correlations between measures from these tasks are presented in Table 9.4.

The vocal differential responsiveness measure (DVR) failed to correlate significantly with any of the measures from the Object Sorting task. The only measure to do so was the measure of total vocalisation duration, which correlated with the frequency of touching measure. This correlation reached significance on both the Pearson and Spearman correlations.

9.3.6 Attention to a Novel Toy (3 months) and Object Sorting (9 months)

23 infants completed these tasks. None of the 6 correlations between measures on these tasks reached significance. The

TABLE 9.4

CORRELATIONS OF MEASURES ACROSS TASKS 3 - 9 MONTHS

DIFFERENTIAL VOCAL RESPONSIVENESS (3 MONTHS) AND OBJECT SORTING (9 MONTHS) N=23

MEASURE

OBJECT SORTING (9 MONTH)	DIFFERENTIAL VOCAL RESPONSIVENESS (3 MONTH)		
	DVR	TOTAL VOCALISATION MOTHER	TOTAL VOCALISATION
TOTAL FIXATION DURATION	.01 (.01)	.06 (.10)	-.08 (-.10)
TOTAL MANIPULATION DURATION	.18 (.11)	.06 (.01)	.03 (.00)
FREQUENCY OF TOUCHING	.19 (.16)	.38 (.38)	.43 * (.45 *)

* $p < .05$ (2 tailed)

Spearmans rank correlations are shown in parentheses

correlations ranged between -.09 and and .23 and between -.11 and .24 for the Pearson and Spearman correlations, respectively.

9.3.7 Concept Acquisition (3 months) and Object Search (9 months)

21 infants completed these tasks.

The correlations between the measures on these two tasks are presented in Table 9.5.

In general, the fixation measures from the Concept Acquisition task correlated significantly with two of the measures from the Object Search task at 9 months. These are: first, the number of trials to search successfully at location A on the initial A trials; and second, the number of failures to search on B trials. While all of these correlations are significant with the Spearman technique, there are two exceptions with the Pearsons which are as follows: first, mean fixation duration on the Concept Acquisition task failed to correlate significantly with the number of trials to succeed at A; and second, the correlation between single longest fixation duration and the number of failures to search just misses significance.

None of the other measures across the tasks correlate significantly.

TABLE 9.5

CORRELATIONS OF MEASURES ACROSS TASKS 3 - 9 MONTHS

CONCEPT ACQUISITION (3 MONTHS) AND OBJECT SEARCH (9 MONTHS) N=21

MEASURE OBJECT SEARCH (9 MONTH)	CONCEPT ACQUISITION (3 MONTH)			
	TOTAL FIXATION DUR	AVERAGE FIXATION DUR	MEAN LONGEST FIXATION DUR	SINGLE LONGEST FIXATION DUR
NO TRIALS TO SUCCEED AT A	.46 * (.49 *)	.38 (.49 *)	.49 * (.57 **)	.58 ** (.56 **)
SEARCH SCORE	.21 (.26)	.32 (.22)	.27 (.25)	.17 (.25)
NO OF FAILURES TO SEARCH ON B TRIALS	.54 * (.54 *)	.62 ** (.56 **)	.56 ** (.54 *)	.41 (.49 *)
SCORE IN DELAY PERIOD	.39 (.29)	.40 (.33)	.37 (.31)	.35 (.22)
SCORE IN PRESENTATION PERIOD	-.13 (.22)	.29 (.26)	.22 (.23)	.15 (.23)
NO. LOOKS AT A ON B TRIALS	.09 (.06)	.07 (-.02)	.08 (.05)	.18 (.14)

* $p < .05$ (2-tailed)
** $p < .01$ (2-tailed)

Spearmans rank correlations are shown in parentheses

9.3.8 Audio-Visual Integration (3 months) and Object Search (9 months)

15 infants completed these tasks.

The correlations between measures on these tasks are presented in Table 9.6.

The measures of percentage fixation time to the synchronised video, percentage fixation duration to the video on the left of the infants' mid-line and the first fixation duration from the 3 month Audio-Visual Integration task, failed to correlate with any of the 9 month measures from the Object Search task.

Average fixation duration (3 months) correlated significantly with the number of failures to search on the B trials (9 months) by both methods of correlation. The Pearson (but not Spearman) correlation between average fixation duration and the composite search measure also reached significance, as did the Spearman (but not the Pearson) correlation between total fixation duration (3 months) and the number of looks at location A on the B trials (9 months).

TABLE 9.6

CORRELATIONS OF MEASURES ACROSS TASKS 3 - 9 MONTHS

AUDIO-VISUAL INTEGRATION (3 MONTHS) AND OBJECT SEARCH (9 MONTHS) N=15

MEASURE OBJECT SEARCH (9 MONTH)	AUDIO-VISUAL INTEGRATION (3 MONTH)				
	% FIXATION SYNCHRONISED	% FIXATION LEFT	TOTAL FIX DURATION	AVERAGE FIX DURATION	FIRST FIX DURATION
NUMBER TRIALS TO SUCCEED AT A	.03 (-.09)	-.13 (-.14)	.43 (.48)	.40 (.51)	-.02 (-.01)
SEARCH MEASURE	-.04 (-.04)	.40 (.21)	.03 (.16)	.58 * (.47)	-.03 (-.05)
NO FAILURES TO SEARCH ON B TRIALS	-.11 (-.03)	.11 (-.07)	.36 (.50)	.75 ** (.62 *)	.07 (-.16)
SCORE IN DELAY PERIOD	.30 (.17)	-.31 (-.22)	.09 (.05)	.49 (.31)	.02 (-.06)
SCORE IN PRESENTAT. PERIOD	.22 (-.09)	.18 (.27)	-.01 (.38)	.45 (.36)	.17 (.02)
NO LOOKS AT A ON B TRIALS	-.03 (-.07)	-.29 (-.40)	.35 (.55 *)	.37 (.49)	.29 (-.04)

* p < .05 (2-tailed)
** p < .01 (2-tailed)

Spearmans rank correlations are shown in parentheses

9.3.9 Operant Conditioning (3 months) and Object Search (9 months)

19 infants completed these tasks.

The correlations between the measures are presented in Table 9.7.

The composite learning measure failed to correlate with any of the measures from the 9 month task. The only Operant Conditioning measures to produce significant Pearson correlations with the 9 month measures from this task, was total fixation duration. This measure correlated with the following 9 month measures: the number of trials to succeed at A (Pearson only); the number of failures to search on the B trials (Pearson only); and the scores obtained in both the delay and presentation periods (Pearson and Spearman).

The Spearman rank correlations produced one significant correlation which the Pearson correlations did not, and that was between the total number of kicks (3 months) and the number of failures to search on B trials (9 months).

9.3.10 Differential Vocal Responsiveness (3 months) and Object Search (9 months)

21 infants completed these tasks.

TABLE 9.7

CORRELATIONS OF MEASURES ACROSS TASKS 3 - 9 MONTHS

OPERANT CONDITIONING (3 MONTHS) AND OBJECT SEARCH (9 MONTH) N=19

MEASURE OBJECT SEARCH (9 MONTH)	OPERANT CONDITIONING (3 MONTH)		
	LEARNING MEASURE	TOTAL FIXATION DURATION	TOTAL NUMBER KICKS
NUMBER TRIALS TO SUCCEED AT A	.33 (.20)	.47 * (.29)	.25 (.17)
SEARCH MEASURE	.19 (.20)	-.17 (-.17)	-.06 (.01)
NO FAILURES TO SEARCH ON B TRIALS	.19 (-.01)	.55 * (.28)	.42 (.48 *)
SCORE IN DELAY PERIOD	-.14 (-.16)	.68 ** (.66 **)	.39 (.33)
SCORE IN PRESENTATION PERIOD	-.16 (-.10)	.57 ** (.53 *)	.02 (.00)
NO LOOKS AT A ON B TRIALS	-.12 (-.20)	.41 (.25)	.19 (.18)

* $p < .05$ (2-tailed)
** $p < .01$ (2-tailed)

Spearmans rank correlations are shown in parentheses

None of the measures from these tasks produced significant correlations with one another. The 18 correlations in the matrix ranged between -.11 and .17, and between -.32 and .29 for the Pearsons and Spearmans correlations, respectively.

9.3.11 Attention to a Novel Toy (3 months) and Object Search (9 months)

21 infants completed these tasks.

None of the measures between these tasks reached significance, the Pearson correlations ranged between .28 and -.42 (which just misses significance), and the Spearman correlations ranged between .23 and -.40.

9.3.12 Correlations between 3 and 9 month measures - Discussion and Composite Correlations

It is apparent, from the analyses described above, that a number of measures from the 3 month testing sessions are predicting some of the 9 month measures. Since there are no major differences between the size of the correlations from both Pearson and Spearman methods, the correlations which do reach significance seem to reflect general trends in the data overall, as opposed to one or two subjects having consistent extreme scores.

In any situation where a considerable number of correlations are calculated, some will reach significance by chance. At the 5 per cent level, approximately 5 per cent of any given number of correlations will reach significance by chance. 153 correlations in all were calculated between measures from tasks at 3 and 9 months, and just 17 (11.1 percent) reached significance. The actual measures themselves which reach significance are fairly consistent in their nature (this will become more apparent when the other across age correlations have been presented).

The 17 significant correlations ranged from .43 to .75, and the confidence intervals can give some indication as to how reliable this range of correlation may be in terms of the normal population. The correlation of .43 emerged between the measures of total vocalisation duration from the DVR task at 3 months and the frequency of touching measure from the Object Search task at 9 months. The confidence intervals for this correlation with an n of 23 are .02 and .71. The correlation of .75 emerged between the measures of average fixation duration from the Audio-Visual task at 3 months and the number of failures to search on the B trials in the Object Search task. The confidence intervals for a correlation of this magnitude with an n of 15 are .37 and .91. This means that the higher correlations emerging from this data are likely to be at least moderate in the normal population, although inevitably the best estimates one has for the magnitude of any correlation is the actual coefficient which the data produces. It is also worth pointing out however that two-tailed correlation tests are reported here, and this is somewhat conservative (since virtually all of the of the significant correlations are in the expected predictive direction).

Furthermore, it is apparent that the size of the correlations emerging from these predictive data is in many cases almost as high as one could expect given the separate reliabilities of the measures involved. For instance, the measure of total fixation duration, from the Concept Acquisition task at 3 months, had a test-retest reliability correlation of .67, while this measure correlates with measures from 9 month Object Search task to the magnitude of .46 and .54 with the measures of the number of trials to successfully search at location A, and the number of failures to search on the B trials, respectively. Mean longest fixation in the 3 month Concept Acquisition task, had a test-retest reliability of .56, and correlated with the number of trials to successfully search on the A trials .49, and .56 with the number of failures to search on the B trials (both measures again from the Object Search task at 9 months). In conclusion it seems that given the reliabilities obtained for the 3 month measures, and given that a measure cannot predict anything better than it predicts itself, then the correlations obtained appear to be as high as one could expect.

The reliability of a particular measure does limit the extent to which it can predict, or be predicted by, any other measure. Many of the 3 month measures showed indications of being moderately as opposed to highly reliable. Possible sources of variation affecting the reliability of measures have been discussed in

Chapter 4, but a number of points are relevent here. Generally, the mean score over a series of testing sessions (or, as used here, the mean of the scores across the two sessions) is considered to be the closest estimate of an individual's 'true' score. However, it is possible that scores on the first testing session are better indicators of the subject's 'true' scores due to practice effects distorting scores on the second session. Alternatively, scores on the second testing session may be considered to be closer to the 'true' scores since the infants have already had experience of the tasks and hence give a 'truer' pattern of performance on the second testing occasion merely because the testing situation is no longer novel. Taking these possibilities into consideration, correlation matrices were produced separately with the 3 month scores on session 1 with all 9 month measures, and likewise with the scores on session 2 at 3 months and the 9 month measures. These data (not presented here), showed that scores from neither session were better predictors than each other, nor were they better than the measures from the two sessions combined. Further discussion relating to the magnitude of prediction from the 3 month measures is presented in Chapter 10, although to conclude, given the moderate reliabilities of the 3 month measures (see Chapter 5), the prediction to performance at 9 months seems to be as high as one could expect.

It is of interest to consider the nature of the measures which correlate from 3 to 9 months. One of the most noticeable factors here is that the predictive 3 month measures all reflect, what can perhaps be best described as, quantitative aspects of performance (eg. total and average fixation duration from both the Concept Acquisition and Audio-Visual Integration tasks, total fixation duration to the mobile from the Operant Conditioning task and total vocalisation duration in the DVR task at 3 months, all reflect measures of the extent to which infants respond to various types of stimuli, and not the extent to which they actually achieve the expected performance in the individual tasks). All of the measures mentioned above (with the exception of total fixation duration in the Concept Acquisition and Audio-Visual Integration tasks) correlated significantly with the frequency of touching measure from Object Sorting task at 9 months. The predicting 3 month measures here were all combined to form a composite score (by adding together the standard scores for individuals, with the score for total fixation to the mobile reversed) and this composite score correlated with the frequency of touching measure .56, which is slightly higher than the correlations produced from any of the single predictors (see Tables 9.1, 9.2, 9.3 and 9.4).

Further, three of the measures at 3 months (namely, average fixation duration from the Concept Acquisition and Audio-Visual Integration tasks, and total fixation time to the mobile from the Operant Conditioning task) which correlate with the frequency of touching measure at 9 months , all reflect aspects of visual fixation, whilst the remaining two measures (total number of kicks from the Operant Conditioning task, and total vocalisation from the DVR task) do not. The three fixation measures were combined, again by adding the standard scores, and this composite measure produced a correlation of .83 with the frequency of touching measure, from the 9 month Object Sorting task. This correlation is particularly high, and indicates that the variance on the composite fixation score (from the 3 month measures) accounts for almost 69 percent of the variance on the 9 month frequency of touching score. Further, since the addition of the two remaining 3 month measures which also predict the 9 month frequency of touching measure, to the composite measure, has the effect of reducing the correlation (see above) then it can be tentatively concluded that the fixation measures in particular at 3 months, reflect the extent to which infants will touch objects at 9 months.

The 3 month measures also produced some interesting correlations with measures from the 9 month Object Search task. These correlations again reflect the importance of the visual fixation measures at 3 months. In general the fixation measures from the Concept Acquisition task at 3 months, correlated significantly with the number of trials to search successfully at location A, and the number of failures to search at 9 months. Average fixation duration from the Audio-Visual Integration task (3 months), also correlated with the number of trials taken to search successfully at location A, as well as the composite search measure. (Virtually all of the 3 month measures mentioned above do in fact correlate significantly with each other.) Fixation duration to the mobile in the Operant task (3 months), correlated with the number of failures to search and the scores achieved in the presentation and delay periods in the Object Search task at 9 months. The combined standard scores from total fixation duration, average fixation duration, average longest fixation duration, and single longest fixation duration (Concept Acquisition at 3 months), and the fixation measure from the Operant Conditioning task at 3 months, increased the prediction of the number of trials to succeed at location A to .73. The individual measures produce lower correlations than this, and range from .46 to .58 (see Tables 9.5 and 9.7). Hence the composite score from the 3 month measures appears to account for more of the variance on the 9 month score than any individual 3 month measure.

A composite measure was derived from the scores of the 3 month measures which correlated with the 9 month measure of number of failures to search on B trials (including total fixation duration, average fixation duration and average longest fixation duration from Concept Acquisition, average fixation duration from Audio-Visual Integration, and the total fixation measure from the Operant Conditioning task). This composite measure produced a correlation of .61 with the 9 month measure of the number of failures to serch on the B trials. Hence, the composite score predicted the 9 month score virtually as well as did the individual 3 month measures.

As was the case for the 9 month Object Sorting task, it seems that measures of visual attention at 3 months are the best predictors of measures on the Object Search task at 9 months. The two 9 month Object Search measures which are most consistently predicted are the number of trials to search successfully at location A, and the number of failures to search on the B trials. These measures can perhaps be best considered to reflect aspects of attention to the Search task. The other significant correlations to emerge between the 3 month measures and this 9 month task were as follows: average fixation duration from the Audio-Visual Integration task (3 months), and the search measure (9 months); and the total fixation duration from the Operant Conditioning task (3 months), and the scores in both the delay and presentation periods from the Object Search task (9 months). While the scores in the presentation and delay periods can fit into the category of measures of attention in the Search task, the search measure itself seems to reflect the qualitative aspect of how well the infants performed on the task.

In conclusion, it appears that some aspect of performance does show a reasonable degree of reliability from 3 to 9 months of age. Further, these measures at 3 months seem predominantly to reflect aspects of visual attention, while at 9 months the measures seem to reflect measures of exploration or activity (in terms of frequency of touching from the Object Sorting task) and essentially attentive behaviour (from measures on the Object Search task). The nature of the predictive and predicted measures is discussed further in Chapter 10.

A number of measures at both ages failed to correlate with any other measure across age. These include: percentage fixation duration to the synchronised video and first fixation duration, from the 3 month Audio-Visual Integration task; the learning measure from the 3 month Operant Conditioning task; the 3 month measures from the DVR tasks (with the exception of total vocalisation duration); and measures from the Attention to the Novel Toy task; and the 9 month measures of manipulation and fixation duration, from the Object Sorting task.

9.4 CORRELATIONS BETWEEN 3 AND 18 MONTH MEASURES

9.4.1 Concept Acquisition (3 Months) and 18 Month Scores

23 infants completed these tasks. The correlations between these scores are shown in Table 9.8.

All of the fixation measures from the Concept Acquisition task correlated significantly with the composite verbal measure at 18 months, whilst none of the 3 month measures correlated significantly with the performance measures. This pattern of correlations emerged for both the Pearson and Spearman correlations, with the exception of the Spearman correlation between total fixation duration (3 months) and the verbal comopsite measure (18 months) which failed to reach significance.

The correlations between the Concept Acquisition measures and the Bayley MDI are all of a magnitude greater than those with the 18 month performance measure, and smaller than those with the composite verbal measure. All are positive and some reach significance. The importance of the 18 month verbal measures as

TABLE 9.8

CORRELATIONS OF MEASURES ACROSS TASKS 3 - 18 MONTHS

CONCEPT ACQUISITION (3 MONTHS) AND INFANT SCALES (18 MONTHS) N=23

MEASURE 18 MONTH MEASURES	CONCEPT ACQUISITION (3 MONTH)			
	TOTAL FIXATION DUR	AVERAGE FIXATION DUR	MEAN LONGEST FIXATION DUR	SINGLE LONGEST FIXATION DUR
BAYLEY MDI	.30 (.25)	.43 * (.43 *)	.43 * (.42 *)	.47 * (.39)
VERBAL MEASURE (COMPOSITE)	.47 * (.39)	.48 * (.43 *)	.49 * (.51 *)	.53 ** (.42 *)
PERFORMANCE MEASURE (MDI)	.25 (.19)	.24 (.24)	.34 (.39)	.40 (.34)
VERBAL MEASURE (MDI)	.35	.40	.43 *	.47*

* p < .05 (2-tailed)
** p < .01 (2-tailed)

Spearmans correlations are shown in parentheses criterion measures is emphasised by the greater magnitude of the correlations between the 3 month measures and the Verbal MDI measure, than the 3 month measures and the Bayley Performance measure. This pattern is consistent throughout the predictive data and the implications are discussed in Chapter 10.

9.4.2 Audio-Visual Integration (3 months) and 18 month scores

17 subjects completed these tasks. The correlations between the measures are presented in Table 9.9.

The measures of percentage fixation to the synchronised video and to the video on the left of the infant's mid-line (3 months), failed to predict performance on any of the 18 month scores, as indeed they had so failed with the 9 month scores. The measures of total fixation duration and average fixation duration (3 months) both correlated with the verbal, but not the performance measures at 18 months. This pattern was consistent for both Pearson and Spearman correlations. The Bayley MDI score produced a significant correlation with the average fixation duration from the Audio-Visual Integration task, although the correlations with the Bayley Verbal measure suggest that this was due to the Verbal rather than the performance items.

TABLE 9.9

CORRELATIONS OF MEASURES ACROSS TASKS 3 - 18 MONTHS

AUDIO-VISUAL INTEGRATION (3 MONTHS) AND INFANT SCALE MEASURES (18 MONTHS) N=17

MEASURE 18 MONTH MEASURES	AUDIO-VISUAL INTEGRATION (3 MONTHS)				
	% FIXATION SYNCHRONISED	% FIXATION LEFT	TOTAL FIX DURATION	AVERAGE FIX DURATION	FIRST FIX DURATION
BAYLEY MDI	.15 (-.04)	-.06 (.16)	.18 (.34)	.61 ** (.49 *)	-.05 (-.11)
VERBAL MEASURE (COMPOSITE)	-.08 (.11)	.02 (.08)	.50 * (.62 **)	.69 ** (.58 *)	.19 (.12)
PERFORM MEASURE (MDI)	.25 (.22)	.25 (.28)	.06 (.13)	.16 (.03)	.07 (.04)
VERBAL MEASURE (MDI)	-.07	-.13	.32	.67 **	.14

* p < .05 (2-tailed)
** p < .01 (2-tailed)

Spearmans rank correlations are shown in parentheses

9.4.3 Operant Conditioning (3 months) and 18 month scores

20 infants completed these tasks.

None of the 3 month Operant Conditioning task scores, (the composite learning measure, the fixation duration and the total number of kicks), produced significant correlations with any of the 18 month scores. The nine correlations in the matrix ranged between -.25 and .27 (Pearsons) and between -.31 and .32 (Spearmans).

9.4.4 Differential Vocal Responsiveness (3 months) and 18 month scores

23 infants completed these tasks. All of the 9 correlations between the measures on these tasks approximated zero. (Correlations ranged between -.04 and .19, and .07 and .20 for Pearson and Spearman correlations, respectively).

9.4.5 Attention to a Novel Toy (3 months) and 18 month scores

23 infants completed these tasks. All of the 6 correlations between measures on these tasks approximated zero.

9.4.6 Correlations between the 3 and 18 month measures - Discussion

A number of measures at 3 months correlated significantly with the 18 month scores. There were no major differences between the Pearson and Spearman correlations which again suggests that the significant correlations reflect general trends rather than a small number of subjects having extreme scores.

51 correlations were calculated in all between the 3 month and the 18 month measures, and 10 (ie. 19.6 percent) reached significance. Also, as was the case for the correlations emerging between the 3 and 9 month measures, the correlations here are of the magnitude that one would expect given that neither measure will be perfectly reliable.

If we next consider the actual measures producing the significant correlations, one of the things which is immediately striking is that none of the 3 month measures correlated with the 18 month performance measure from the Bayley MDI. Although a number of 3 month measures correlated significantly with the Bayley MDI scale, the highest correlations emerged between the 3 month measures and the composite verbal measure. (The significant correlations which emerged between the 3 month measures and the Bayley MDI in most instances seemed to be due to the high correlations with the verbal component within the Bayley MDI Scale.)

The 3 month measures which predict the 18 month verbal scores, are in many cases the same ones which predicted performance at 9 months. These include all of the fixation measures from the Concept Acquisition task at 3 months, and total fixation duration, and mean fixation duration from the Audio-Visual Integration task at 3 months. All of these measures again reflect aspects of patterns of fixation duration, whilst the qualitative 3 month measures (such as the percentage fixation to the synchronised video from the Audio-Visual Integration task, the learning measure from the Operant Conditioning task, and the DVR measure from the DVR task), fail to predict any of the 18 month scores.

The 3 month measures (listed above) which correlate with the 18 month verbal score were combined into a composite measure (by adding individual subjects' standard scores), and this measure correlated with the 18 month verbal measure with an r of .7. This accounts for almost half of the variance of the 18 month verbal measure, and is just higher than the correlation from any single 3 month measure. This correlation of .7 probably reflects the most accurate prediction of the 18 month score from these 3 month measures.

In conclusion, measures at 3 months reflecting aspects of fixation toward visual stimuli, appear to correlate specifically with subsequent verbal ability at 18 months of age. Furthermore, this verbal ability includes measures of both comprehension and language production.

Consistency in the Data

With three exceptions, the same measures from the 3 month tasks which predicted measures at 9 months, also correlated with verbal measures at 18 months. The exceptions are: the Operant Conditioning measures of total fixation duration and total number of kicks; and total vocalisation duration from the DVR task. None of the 3 month measures which failed to predict 9 month measures significantly predicted the measures at 18 months.

9.5 CORRELATIONS BETWEEN 5 MONTH MEASURES AND 9 AND 18 MONTH MEASURES

Out of the 63 correlations which made up the correlation matrices between tasks at 5 and 9 months, only 4 (6.34 percent) were significant. This number of significant correlations would be expected by chance, and it was therefore concluded that the measures at 5 months fail to predict performance on the 9 month tasks.

None of the 21 correlations between measures from the 5 month tasks and the 18 month measures reached significance.

Brief discussion of the 5 month measures is presented in the following section.

9.6 DISCUSSION-5 MONTH CORRELATIONS WITH 3, 9, AND 18 MONTHS

From the research carried out here, it has become apparent that the tasks chosen to produce potentially predictive measures at 5 months of age, have not been successful. The tasks produced considerably fewer reliable measures than they did at 3 months, and of those measures which were reliable and those for which reliability coefficients were not meaningful, none clearly predict performance on tasks at 9 months or on measures from infant scales at 18 months. This finding is not apparent from the literature to date. To re-iterate the issues discussed in Chapter 6, it seems that at 5 months tasks requiring predominantly passive responses from the infants may be poor indices of subsequent performance. It may be that alternative tasks requiring manual activity and exploration from the infants would constitute more valid measures at this age. Further discussion on the 5 month data is presented in Chapter 10.

9.7 CORRELATIONS BETWEEN 9 AND 18 MONTH MEASURES

9.7.1 Object Sorting (9 months) and 18 month scores

23 infants completed these two tasks.

The correlations between measures are presented in Table 9.10.

Total manipulation duration (at 9 months) correlated significantly with the Bayley MDI measure, but not with the performance or the verbal measures. This pattern of results is consistent for both the Pearson and Spearman correlations and is unusual since total manipulation duration also failed to correlate with any other measure across age, hence the correlation possibly occurred by chance.

Frequency of touching at 9 months, on the other hand, produced a more consistent pattern of intercorrelations. This measure correlated significantly with the Verbal composite measure and with the Bayley MDI at 18 months (reflected by both Pearson and Spearman correlations). The measure of frequency of touching was also the measure with which a number of 3 month measures correlated significantly (see Tables 9.1, 9.2, 9.3, and 9.4 above).

TABLE 9.10

CORRELATIONS OF MEASURES ACROSS TASKS 9 - 18 MONTHS

OBJECT SORTING TASK (9 MONTHS) AND INFANT SCALE MEASURES (18 MONTHS)
N=23

MEASURE 18 MONTH MEASURES	OBJECT SORTING (9 MONTH) TOTAL FIXATION DURATION	 TOTAL MANIPULATION DURATION	 FREQUENCY OF TOUCHING
BAYLEY MDI	.30 (.34)	.45 * (.51 *)	.65 ** (.65 **)
VERBAL MEASURE (COMPOSITE)	.21 (.18)	.21 (.30)	.58 ** (.57 **)
PERFORMANCE MEASURE (MDI)	.12 (.16)	.26 (.27)	-.18 (.23)
VERBAL MEASURE (MDI)	.10	.31	.61 **

* p < .05 (2-tailed)
** p < .01 (2-tailed)

Spearmans rank correlations are shown in parentheses

The Bayley MDI, on this occasion, strangely produced a higher r (.65) with frequency of touching at 9 months, than did the composite verbal measure or the Verbal measure from the Bayley MDI.

9.7.2 Object Search (9 months) and 18 month scores

21 infants completed these tasks.

The correlations between the measures from these tasks are presented in Table 9.11.

All of the 9 month Object Search task measures, with the exception of the score in the presentation period (which just misses significance on both types of correlation), correlate significantly with the 18 month composite verbal measure.

All of the measures from the Object Search task (with the exception of the number of looks at location A on the B trials, which just misses significance with the Pearsons correlation, although does reach significance on the Spearmans), correlate significantly with the Bayley MDI.

The number of trials to succeed at location A, and the score in the presentation period also correlate significantly with the 18 month Performance score (reflected by both Pearson and Spearman correlations). This is odd since no other measures correlate with the performance measure from any testing age.

TABLE 9.11

CORRELATIONS OF MEASURES ACROSS TASKS 9 - 18 MONTHS

OBJECT SEARCH TASK (9 MONTHS) AND INFANT SCALE MEASURES (18 MONTHS) N=21

MEASURE 18 MONTH MEASURES	OBJECT SEARCH (9 MONTHS)					
	NO TRIALS SUCCEED AT A	SEARCH SCORE	NO FAILURES TO SEARCH ON B TR.	SCORE DELAY PERIOD	SCORE PRES. PERIOD	NO LOOKS AT A ON B TR
BAYLEY MDI	.48 * (.49 *)	.59 ** (.61 **)	.45 * (.47 *)	.57 ** (.54 *)	.54 * (.49 *)	.43 (.50 *)
VERBAL MEASURE (COMPOSITE)	.48 * (.52 *)	.50 * (.49 *)	.48 * (.49 *)	.55 ** (.51 *)	.40 (.42)	.54 * (.56 **)
PERFORM MEASURE (MDI)	.56 ** (.48 *)	.06 (.13)	.00 (-.01)	.10 (.00)	.50 * (.46 *)	.05 (.13)
VERBAL MEASURE (MDI)	.47 *	.57 **	.46 *	.50 *	.37	.47 *

* p < .05 (2-tailed)
** p < .01 (2-tailed)

Spearmans rank correlations are shown in parentheses

9.7.3 9 and 18 Month Correlations - Discussion

As with the other two age comparisons, measures from the 9 and 18 month testing sessions are producing measures which correlate significantly. The pattern of inter-correlations was similar for the Pearson and Spearman correlations, as it was for all other across age correlation matrices (see above).

From 9 to 18 months the correlation matrices were made up of 27 correlations, and 14 of these (over 50 percent!) reached significance. Between these two testing ages then, emerged a strong indication that at least some of the correlations were not likely to have occurred by chance.

Further, if the measures which do inter-correlate are considered, some interesting patterns emerge. The frequency of touching measure from the Object Sorting task (9 months) correlates with the Verbal composite and the Bayley (MDI) scores at 18 months.

Virtually all of the measures from the 9 month Object Search task correlate with the Bayley MDI and the composite verbal measures at 18 months. In addition the 9 month measures of the number of trials to search successfully at location A, and the score in the presentation period also correlate significantly with the 18 month performance measure from the Bayley MDI. It is not clear why these two 9 month measures should be related to the performance

MDI measure, especially since all other correlations between the Object Search task and this 18 month measure approximate zero. These two correlations may have occurred by chance, although one can never assume this to be the case. Indeed one alternative possibility is that the measures from the Object Search task at 9 months are reflecting abilities of a wider range, than merely verbal ability, which include performance based abilities. Indeed the Bayley MDI score is predicted as well as the verbal composite measure by measures on this 9 month task.

To illustrate the best estimate of the relationship between the measures of 9 months and the performance on verbal ability, a composite score, was derived from those 9 month measures which correlated with the 18 month verbal composite measure (ie. the frequency of touching from the Object Sorting task, and the number of trials to succeed at location A, the search measure, the number of failures to search on the B trials, the score in the delay period and the number of looks at location A on the B trials, from the Object Search task). This composite score was correlated with the 18 month Verbal score, and the resulting r was .72. When this was done in a similar way for the Bayley MDI, the resulting correlation was also .72. Hence in general the composite scores constitute better predictors than any individual predictor.

In conclusion, the prediding measures at 9 months appear to reflect aspects of attention to and activity within the tasks. These measures appear to correlate substantially with the verbal composite measure, and the Bayley MDI, at 18 months.

Consistency in the Data - and Conclusions

The four 9 month measures (ie. frequency of touching (Object Sorting), the number of trials to successfully search at location A, the combined search measure, and the number of failures to search at location B (Object Search)), which were predicted by fixation measures at 3 months, all correlate with the measure of Verbal performance at 18 months. Further, the 3 month fixation measures also correlate with the 18 month Verbal composite score. This suggests that there is some aspect of ability over these age periods which is characterised by a fair degree of continuity.

9.8 CONCLUSIONS FROM THE LONGITUDINAL ANALYSES

The patterns of inter-correlations across the data are shown in Table 9.12. Some degree of continuity across the ages reported here was apparent. In general the patterns of the inter-correlations have led to suggestion that the early measures reflect verbal ability as opposed to ability on the performance items at 18 months.

It is important to emphasise however that the findings of this research cannot be considered conclusive, particularly in view of the small sample size employed.

The possible implications of this research for theoretical and

TABLE 9:12

The table shows all significant correlations between measures across age.

Each measure is abbreviated (to the left) and the abbreviation is presented on the right of each measure with which it correlates significantly.

	3 MONTH	
	Concept acquisition	
TFC	Total Fixation Duration	TSA. FSB. VM
AFC	Average Fixation Duration	FT. FSB. MDI. VM
MLF	Mean Longest Fixation Duration	TSA. FSB. MDI. VM
SLF	Single Longest Fixation Duration	TSA. MDI. VM
	Audio-visual Integration	
TFA	Total Fixation Duration	VM
AFA	Average Fixation Duration	FT. SS. FSB. MDI. VM
	Operant Conditioning	
TFM	Total Fixation to Mobile	FT. TSA. FSB. SID. SIP.
TNK	Total Number of Kicks	FT.
	Differential Vocal Responsiveness	
TVD	Total Vocalisation Duration	FT.
	9 MONTH	
	Object Sorting	
FT	Frequency of Touching	AFC. AFA. TVD. TFM. TNK. MDI. VM
TMD	Total Manipulation Duration	MDI.
	Object Search	
TSA	Number of Trials to Succeed at A	TFC. MLF. SLF. TFM. MDI. VM. PM
SS	Search Score	AFA. MDI. VM
FSB	Number of Failures to Search on B Trials	TFC. AFC. MLF. AFA. TFM. MDI. VM
SID	Score in Delay	TFM. MDI. VM
SIP	Score in Presentation	TFM. MDI. PM
LAB	Number of looks at A on B Trials	VM
	18 MONTH	
MDI	Bayley M.D.I.	AFC. MLF. SLF. AFA. FT. TMD. TSA. SS. FSB. SID. SIP.
VM	Verbal Measure (composite)	TFC. AFC. MLF. SLF. TFA. AFA. FT. TSA. SS. FSB. SID. LAB.
PM	Performance Measure M.D.I.	TSA. SIP.

practical issues are discussed in the following final Chapter.

CHAPTER 10

DISCUSSION AND CONCLUSIONS

10.1 INTRODUCTION

This chapter is divided into seven major sections. The second section presents a review of the thesis, considering each chapter in turn. Section 3 considers whether the data are best interpreted in terms a continuous or discontinuous model of intellectual growth. Section 4 presents a discussion of the nature of the stable and predictive measures at each testing age and speculates as to the nature of the underlying continuity. Section 5 discusses the data in terms of the theoretical positions described in Chapter 1, and in section 6 tentative links are drawn with current conceptions of intelligence in adulthood. Section 7 concludes the thesis with some implications of this work for future research.

It is at the outset important to point out that any interpretation of the data reported in this thesis is necessarily speculative, particularly in view of the small sample size used.

10.2 REVIEW OF THESIS

Chapter 1

This research is concerned with the stability and continuity of cognitive performance in infancy. The first chapter presented a brief history of intelligence testing, the failure of the infant scales to predict subsequent childhood levels of IQ, and the possible theoretical reasons for the failure. Emphasis was placed on the suggestion that the type of ability tested by the standardised infant scales was generally motoric (as opposed to cognitive) in nature, and that recently developed techniques which appear to reflect more cognitively based behaviours in infancy may well constitute better predictors.

Chapter 2

In this chapter a review of predictive studies adopting these relatively new techniques for assessing ability in infancy was presented. It is concluded that measures from these tasks do appear to produce some degree of prediction, although the area is severely lacking in terms of psychometric grounding. The aims of the thesis were presented. These were essentially to assess the test-retest and split-half reliabilities of a variety of measures in infancy, and to examine the ways in which these measures inter-correlate within age at 3, 5 and 9 months, then to examine intercorrelations across these ages, and also their correlations with a series of measures from infant scales at 18 months.

Chapter 3

This chapter described the Pilot Studies for each of the tasks chosen for use in the 3 and 5 month testing ages. Five tasks were administered in the laboratory, and two were administered in the children's homes. The laboratory based tasks were: Visual Preferences; Visual Habituation; Concept Acquisition; Audio-Visual Integration; and Operant Conditioning. The two home based tasks were Differential Vocal Responsiveness and Attention to a Novel Toy. The test-retest and split-half reliabilities for dependent measures from a number of these tasks were examined, and all were piloted for procedural and related problems which were likely to arise. As a result of these studies a number of changes were made to the design and nature of the tasks used for the major longitudinal study.

Chapter 4

In this chapter a number of issues relating to the statistical analyses used for the longitudinal research were discussed. The rationale underlying the use of test-retest and split-half reliabilities were discussed, and also the reasons as to why multivariate procedures were not appropriate for subsequent data analyses. Correlation matrices were presented as the major form of data analyses, for both the within session, and the between session longitudinal analyses, and justification for their use was presented.

Chapter 5

This chapter presented the design and results of the 3 month testing age. Four tasks in the laboratory and two in the home were presented to the infants at 3 months, and these tasks are the same as those used in the Pilot studies (Chapter 3) with the exception of the Visual Preference task. The results include the test-retest and split-half reliabilities for a number of measures from each task, and the inter-correlations of reliable measures across tasks. A brief discussion of the results was also presented, and the conclusions reached were that a number of measures from the tasks at 3 months are reliable both within and across testing sessions. Further, many of these measures were of length of fixations toward stimuli, and these measures are also related (correlated) across tasks.

Chapter 6

This Chapter was concerned with the testing sessions at 5 months. Those children who had been tested at 3 months were tested again at 5 months of age on the same tasks which had been used at 3 months. This allowed further reliabilities of the measures from the tasks to be assessed at a different age, as well as stability in performance on the tasks across age to be examined. However, the measures at 5 months showed little evidence of test-retest reliability, and those few measures which did emerge as reliable were not related to (correlated with) equivalent measures at 3 months. It was concluded that at 5 months, the tasks chosen here are not useful for providing potentially predictive measures. Possible reasons for this were discussed.

Chapter 7

This chapter presented details of and data from the two tasks presented to the same infants when they reached 9 months of age. The tasks used at this age were, an Object Sorting task, which Starkey (1981) suggests reflects aspects of concept acquisition in infants, and an Object Search task designed to examine Piaget's 'stage IV error'. The measures of performance on these two tasks are inter-correlated, and they reveal a reasonable degree of both within and across task stability.

Chapter 8

This chapter described the final stage of the longitudinal testing, which occurred when the infants reached 18 months of age. At this age the infants were administered a series of scales and sub-scales from current infant intelligence scales. Many of these reflected aspects of language development. The inter-correlations of these measures were presented and it was apparent that all of the measures reflecting aspects of language development correlated very highly with each other, and that these measures did not correlate with the performance measure at 18 months.

Chapter 9

One of the main aims of this thesis was to assess the inter-relationships between measures across ages, and Chapter 9 was specifically concerned with the relationship between the measures across age. Correlations between each reliable measure across age were presented and the patterns of continuity in performance across age were discussed. With one or two exceptions the pattern of inter-correlations across age produced a fairly consistent picture. Essentially, measures reflecting aspects of fixation at 3 months correlated with measures at 9 months which appeared to reflect measures of attention towards, and accuracy in performance on, tasks. Furthermore, most of these measures (at 3 and 9 months) correlated significantly with measures of verbal ability at 18 months. Hence it is concluded that some aspect of behaviour, from the measures used here, did show stability at least over the ages of 3, 9 and 18 months.

10.3 CONTINUITY VERSUS DISCONTINUITY HYPOTHESES

The data reported in this thesis appear to indicate that some aspect of performance in infancy is, to some degree, characterised by stability. However, to what degree are models of stability supported, and can models advocating the discontinuous nature of intelligence be rejected? In general, the reliability coefficients and predictive correlations reported in this thesis are moderate. This can either allow one to reject the null hypothesis and argue for some thread of continuity, or equally to emphasise the inherent instability of intellectual growth in the infant. These data do not allow rejection of discontinuity hypotheses, and like all other data in this area, are equivocal.

However, a model of total discontinuity is not appropriate for the data, since moderate correlations were found between different measures across age, it would seem that intellectual development is characterised by some degree of stability. Further, the continuity hypothesis is emphasised here for several other reasons (a number of these issues are discussed in more detail later in this chapter). For example, better measures than the ones taken here may subsequently be discovered. Clear measures of intellectual performance in infancy have not yet been identified, and future research in this area may produce higher correlations. Perfect correlations from data of this nature are never likely to be found due to a large number of factors both intrinsic and external to individuals. The question of how high predictive correlations of this nature need to be before they can be considered useful is not easily answered, although in terms of practical usage, it is those children who are likely to suffer cognitive deficits in childhood that ideally we would like to identify in infancy. A sample of infants covering a range of "at risk" infants may have increased the correlations considerably. This is potentially one area for future investigation. In this study the sample of infants who acted as subjects was fairly homogeneous. The infants were all from the middle-classes, and to the author's knowledge none are suffering from any developmental difficulties.

10.4 THE RELIABILITY AND PREDICTIVE ABILITY OF MEASURES AT EACH TESTING AGE

In these sections, further discussion is presented concerning the nature of those measures which showed stability and predictive ability at each age. Also an attempt is made to theorise as to the nature of the predictive measures at each age. The 3 month measures are considered first.

10.4.1 3 Month Measures

Tasks involving visual fixation measures are considered first.

At 3 months the measures from the tasks requiring visual fixation as a response (Concept Acquisition and Audio-Visual Integration), which were reliable and predictive are very consistent in nature. They are all measures of visual fixation length. This perhaps seems at first somewhat unusual, since for predictive purposes, many previous studies have emphasised the importance of the measures reflecting the extent to which infants produce an expected pattern of performance (see Chapter 2). For instance, if we consider first the Concept Acquisition task, for the purpose of prediction emphasis has been placed on the measures of speed of decline in fixation to the concept exemplars, and amount of recovery to the non-concept exemplars (see Chapter 2). However, this expected pattern of attentional decline during the familiarisation, followed by a recovery of looking to the novel stimuli was not found in this research. Therefore, it is perhaps a little unlikely that these measures would have predicted anything. The failure of the infants to produce the expected pattern of responding may be due to particular constraints of the experimental design adopted here (eg. type of stimuli, use of the fixed trial procedure). Furthermore, there are a number of other methodological difficulties associated with the habituation procedure which reduce the possibility of measures of rate of decline and amount of recovery to novelty being potentially predictive. Many of these have been discussed in Chapter 3, although a few are worth further consideration here.

Measures of rate of decline and amount of recovery to novelty are necessarily affected by the length of fixations exhibited by an infant at the start and end of the habituation sequence. One cannot assume that an infant who fixates the first stimulus for, say, 10 seconds, will as easily show a decrement over trials as a second infant, who fixates the first stimulus for, say, 30 seconds. Similarly, at the end of the habituation sequence, an infant who fixates the stimulus for 2 seconds will perhaps more easily show a subsequent increment in fixation than an infant who fixates the final familiar stimulus for 15 seconds.

In view of these factors (discussed above), it is argued here that although the ability of infants to recognise familiar and novel stimuli may be important for cognitive development, habituation procedures will often not provide clear measures of this.

The actual nature and the reliability of measures used for any predictive study are clearly important considerations for future research. There appears to be some justification in supposing that measures of visual decline and subsequent recovery may not be predictive. However, in the research described in this thesis measures of fixation time did correlate with later performance, a finding which is consistent with others (see Fenson, Sappir and Minner, 1974; Pecheux and Lecuyer, 1983; Rose, Slater and Perry, 1986). What might these measures reflect? It is suggested here that these measures reflect information processing speed in infancy, and the justification for this is presented next.

This suggestion was derived from a discrepancy in the literature between, on the one hand, Fagan and McGrath (1981) who reported that the extent of the novelty preference from the novelty preference testing technique predicts later intelligence (see Chapter 2), and on the other hand from Rose, Slater and Perry (1986) who show that with the use of an infant controlled habituation procedure, whilst novelty preferences failed to predict later intelligence, measures of fixation length did. Both of these experimental designs incorporated the use of paired stimuli (one novel paired with the familiar) in the post-familiarisation test trials.

There is one fundamental difference between these two testing procedures which is the amount of familiarisation time infants were given. What seems to be the case is that given a short familiarisation period novelty preferences predict, and given a lengthy one they do not. If, in the novelty preference technique, after the short familiarisation period, some infants have failed to encode the stimulus, then they may continue to fixate this stimulus (as opposed to the novel one) in order to complete the process. The 'bright' infants will have successfully encoded at least some parts of the stimulus in the given time and will therefore show a novelty preference on the test trial. The novelty preference then may reflect the amount of processing infants achieve in the familiarisation period, rather than the extent to which infants can detect and respond to discrepancies in visual stimuli (as suggested by Fagan, 1984).

This suggestion is compatible with Rose et al's findings. They failed to find novelty preferences predictive, from the use of the infant control procedure. In this procedure, since all infants are in theory given the opportunity to encode the stimulus fully, then all infants are likely to show some degree of preference for novelty. The actual extent of this novelty preference may be affected by other factors, such as stimulus preferences (rather than a reflection of how well infants have encoded the familiar stimulus), and hence is predictive of nothing. In line with this reasoning, it seems reasonable to assume that the length of time infants spend fixating stimuli may be a reflection of the speed at which they are able to process stimuli. Measures of fixation time then, may be good reflectors of speed of processing or speed of memory formation in infancy.

Accommodating the fixation measures from the Audio-Visual Integration task into the notion of speed of processing is at first sight a little more difficult. These measures correlated significantly with the measures of fixation from the Concept Acquisition task (see Chapter 5), and also correlated significantly with performance at 9 and 18 months. The measure from the Audio-Visual Integration task which one might have expected to be predictive is the percentage fixation time to the synchronised video (this measure gave test-retest reliability but was not predictive). However, from a cognitive perspective, what may be important from this task is the actual time taken for individual infants to establish the fact that one of the videos was synchronised with the sound while the other was not. There is no obvious way of extracting this measure from these data. If the 'brighter' infants had established the difference very early on in the task, one cannot assume that they will necessarily continue to fixate the integrated film. Indeed the time-lag between the two videos may have become more interesting. The problem arises that in the Audio-Visual Integration task, the visual behaviours measured fail to produce a clean measure of the extent to which infants comprehended the cross-modal component of the task. However, the measures of fixation time from this task which showed test-retest reliability were related to measures of fixation from the Concept Acquisition task, and were related to later measures. One possible explanation is that the faster lookers require less time to detect the synchrony. At the very least it would seem that if an infant can process stimuli rapidly, then different parts of a new environment are likely to be fixated and explored in quick succession.

The only other task at 3 months to involve measures of visual fixation of a similar nature to those from the Concept Acquisition and Audio-Visual Integration tasks was Attention to a Novel Toy. The measures from this task failed to predict anything. This task differed from the others, however, since it was administered in the infant's own home and the visual environment was not restricted from external stimulation in the same way as it was in the laboratory tasks. This factor alone may have affected the fixation patterns of any individual infant. Furthermore, the measures themselves were somewhat crude.

The remaining 3 month tasks (Operant Conditioning and the Differential Vocal Responsiveness task), produced some measures which were predictive of the 9 month, but not the 18 month, measures. The 9 month measures which they did predict were in many cases the same as those predicted by the fixation measures and in view of this and in view also of their failure to predict the 18 month scores, these 3 month measures seem less interesting.

If we consider first measures from the Operant Conditioning task, one thing which was apparent was the failure of the composite learning measure to predict. The measures making up this composite however are vulnerable to the same sorts of problems as the measures of decline and recovery in the Concept Acquisition task. Specifically, the number of kicks made by an infant in the initial base-line period is likely to affect the extent to which infants can demonstrate learning ability. For example, those infants who kick constantly in the base-rate phase may have been physically unable to show a learning affect because they could not kick any faster, in spite of actually having learnt the relationship between kicking and mobile movement. Alternatively, those infants who produce very few foot kicks may have little chance to learn the contingency. Furthermore, the actual nature of the reinforcement may have differed for individual infants, and this may have affected their rate of kicking (for instance, in some cases it appeared that infants, having learned the response, played games with the mobile, by not kicking until the mobile virtually stopped moving and then fast bursts of many kicks, amidst excited screaming and laughing). It would appear that the measures from this task which one would expect to be the most likely predictors are obscured by other factors influencing performance.

The measures from the Operant Conditioning task which predicted aspects of performance at 9 months are total fixation to the mobile and total number of kicks. These measures were positively related to 9 month performance, such that the longer lookers and the more active infants at 3 months had the more mature 9 month scores. These measures were related to each other (r=.51), but were not related to other predictive measures at 3 months. These measures may reflect activity level at 3 months. It seems plausible to suppose that those infants who produced more mobile movement were also likely to fixate the mobile for longer periods.

It does appear that whatever these measures do reflect, it is not related to the fixation measures from the Concept Acquisition and Audio-Visual Integration tasks.

Total vocalisation duration was the only measure from the Differential Vocal Responsiveness Task at 3 months to predict performance at 9 (but not at 18) months. Roe and McClure (1980) suggested that the DVR measure is more important for prediction than measures of amount of vocalisation, but this was not the case for the present sample.

Conclusion

It has been suggested above that the fixation measures from the Concept Acquisition and Audio-Visual Integration tasks can be seen as measures of processing speed in infancy. These interpretations are speculative, and there are alternatives. For instance, one could suggest that all of the predictive measures at 3 months are reflecting general activity levels of infants, although such an interpretation seems unlikely to be the case, since one might expect a greater variety of measures (including, for example, number of kicks in the Operant Conditioning task and total vocalisation duration in the Differential Vocal Responsiveness task) to have been related.

In conclusion, from the data collected here the 3 month measures appear to reflect at least two types of ability, one of which predicts to 9 and 18 months, the other predicting only to measures at 9 months.

10.4.2 5 Month Measures

The measures at 5 months were much less reliable across sessions than the 3 month measures, and the 5 month measures which were reliable were not predictive of either the 9 or the 18 month measures. This leads to several possible interpretations.

It may be the case that at 5 months development is characterised by a period of instability, and that no measures would show potential usefulness as predictors.

Alternatively, it may be the case that the tasks administered are inappropriate for predictive purposes at 5 months. For instance, Kagan (1979) has suggested that infants' attention to visual stimuli wanes considerably from 3 to 7 months, due to infants being able to develop mental schemata with increasing ease. He claims that after 3 months, infants are easily able to assimilate a repeated event, and so it it is possible that infants were performing at the ceiling level on many of these tasks. If this was the case, it may be that tasks of a different nature are more appropriate for prediction at 5 months. In Chapter 6, it was pointed out that the infants at 5 months seemed interested in touching and manually exploring objects (eg. reaching for the mobile in the Operant Conditioning task), and tasks of this nature may have constituted more useful predictors at this age. This suggestion, that measures incorporating manual exploration at 5 months may be more useful indices of later development, is clarified later in this chapter, in relation to the 9 month measures.

Indeed Messer et al. (1986) found that at 6 months, measures reflecting the activities through which infants explored objects were predictive of performance on cognitive scales at 12 months (for details of this study, see Chapter 2).

The one task administered to the 5 month olds which fitted the category of manual exploration was the Attention to the Novel Toy task, and this failed to produce predictive measures. However the measures taken from this task were crude (see Chapter 6); sequences were not videoed, so unfortunately measures of the different ways in which infants explored the toy are not available.

In conclusion, it would seem that measures of visual attention at 5 months are inappropriate indices of later ability, and therefore that the tasks administered to the infants were inappropriate for this age group. Alternative types of task, particularly ones requiring manual exploratory behaviour may provide predictive measures, and this is a consideration for future research.

10.4.3 9 Month Measures

Just one measure from the Object Sorting task at 9 months showed an indication of being both predicted by, and predictive of, the 3 and 18 month measures, respectively. This was the measure of frequency of touching - the number of times infants independently picked up objects throughout the task. As discussed in Chapter 7, this measure confounded two other measures from this task which were, the extent to which infants a) touched sequentially, and b) grouped together, objects of a similar kind. Hence, as was the recurring case for the 3 month tasks, it seems that the measures of performance which one may have expected to be predictive do not emerge clearly from the test because of the presence of confounding variables.

The Object Search task at 9 months produced a variety of measures which were predicted by, and predictive of, the 3 and 18 month measures, respectively. This is particularly the case for the following measures: the number of failures to search on the B trials; the number of trials taken to search successfully at location A; and the composite search score. In Chapter 7 it was pointed out that, from the inter-correlations between measures on this task, it was clear that measures of attention to the task (reflected by scores in the presentation and delay periods), and the likelihood of search, and success in search, were highly related. Those infants who attended to the task throughout the delay periods, and those infants who reached towards the tray in the presentation periods were not only more likely to search, but also were more likely to search in the correct location.

The actual cause and effect relationship between attention to the task and successful search is difficult to establish, for instance, is it that those infants who attend to the task also tend to find the toy because of their mature pattern of attention, or do infants who have seen the toy hidden, realise that in order to find it easily they must maintain fixation towards the location at which the object was hidden?

The measure of frequency of touching (from the Object Sorting task) also correlated with many of the measures on the Object Search task. The measure of failures to search on the B trials (from the Object Search task) was the measure which appeared to be most consistently predicted by, and predictive of, the 3 month fixation measures and the 18 month composite verbal measure, respectively. Measures of ability in task performance seem also to be important at 9 months (as reflected by the search measure from the Object Search task). Since many of the 3 month measures of fixation were related to a number of 9 month measures (in particular frequency of touching from the Object Sorting task, and number of trials to succeed at location A, and the number of failures to search on the B trials, from the Object Search task) it does seem feasible to suggest that those infants who process information at a faster rate at 3 months also exhibit greater manual exploration of stimuli presented to them. The rate of processing therefore appears to determine the extent to which infants will explore their immediate environment. Those infants who explore their environments to a greater extent are also more likely to search in the appropriate place in the Object Search task.

10.4.4 18 Month Measures

Three measures were extracted from the 18 month measures, these were: the Bayley MDI; a composite Verbal measure (made up from verbal measures from the Bayley MDI, the Reynell Language Development Scales including both the comprehension and production sub-scales, and the Hearing and Speech Scales from the Griffiths Developmental Schedules); and a Performance measure from the performance based items on the Bayley MDI. Further details of these measures are given in Chapter 8.

Of these measures, the verbal composite measure was the most consistently predicted by the earlier measures (see Chapter 9). This finding is not unusual, and many authors have emphasised the importance of verbal measures as criterion measures in predictive studies (see Chapter 2, and review by Bornstein and Sigman, 1986).

The interpretation of what is actually being reflected by language development differs depending on the way in which one perceives intelligence (see Chapter 1). Many psychologists have suggested that intelligence is a multi-dimensional construct consisting of many different types of ability, although most theorists have acknowledged the existence of a general factor ('g') as an underlying, all-embracing component of intelligence. Charles Spearman for instance suggested that intelligence is made up of a general factor and also that individual mental tasks load to some degree on 'g' while each also has a loading on other factors specific to each task. Burt (1949) and Vernon (1950) suggested a hierarchical structure of intelligence which incorporated 'g' and also two major abilities, namely verbal and visual-spatial abilities, each of which sub-divides into minor group factors. Thurstone also identified independent verbal and spatial factors. Hunt, Lunneborg and Lewis (1975) has pointed out that verbal ability always emerges as a strong factor in Intelligence tests, encompassing measures of English usage, spelling, reading comprehension and vocabulary. It may be the case therefore, that while the measures adopted in this research appear to be related to specifically verbal ability, there may be other tasks in infancy which reflect subsequent performance on spatial tasks.

Intelligence has however been conceived of in other ways. For instance, Cattell (1963) differentiated two types of intelligence, namely fluid and crystallised intelligence. Fluid intelligence is the hereditary component of intelligence, while crystallised intelligence is a measure of what one has learnt from one's environment. Fluid and crystallised intelligence are not entirely independent of one another, since the former reflects the intellectual potential, and the latter reflects the realised potential, and Eysenck (1981) has pointed out that:

> "Where the acquisition of knowledge is reasonably standard, the amount of knowledge acquired might be considered a direct measure of intelligence." (page 22).

In relation to this, if we consider that all infants experience verbal stimulation directed towards them, then those infants who acquire both a greater understanding and appropriate use of language, will be those who can process information more rapidly. This measure of verbal ability may be the most accurate measure of fluid intelligence that can be obtained. The performance measure at 18 months, on the other hand, may be more directly infuenced by infants' prior experience on the individual task items, being determined by the provision of play materials in the home, and therefore reflect crystallised intelligence to a greater extent.

In conclusion, it is apparent from these data that the scales and subscales administered to the infants at 18 months of age reflect two apparently independent types of ability, namely, verbal and performance ability. Verbal ability is consistently predicted by the earlier measures and performance ability is not. One possibility is that these two types of measure represent independent aspects of intelligence, and that it may be possible to find infant measures which will predict the performance measure. Further, the verbal measures may reflect fluid intelligence to a greater degree than the performance measure, which reflects predominantly crystallised ability. Roe, McClure and Roe (1983) have suggested that intelligence in the infancy period is multi-dimensional in nature, at least in terms of verbal and perceptal-motor based tasks. They report that the Gesell Developmental Schedules predicted visual-perceptual and perceptual-motor (ie. nonverbal) tasks in childhood, whilst Roe and McClure (1980) reported the vocal differential responsiveness measure at 3 months to be predictive of later verbal ability.

The next section is concerned with the nature of the continuity in relation to the three continuity models discussed in Chapter 1.

10.5 THE NATURE OF THE CONTINUITY

Three models of continuity were outlined in Chapter 1. These were: homotypic continuity, or the continuity of identical behaviour; heterotypic continuity, or the continuity of different behaviours, which can be theoretically linked by some common underlying process(es); and continuity of developmental status. The question of which of these three models can best accommodate the data is considered next. The first model of continuity of identical behaviour is considered initially, followed by the third model (continuity of developmental status), and finally the second model of continuity of underlying process is discussed.

The data clearly do not test the first of these models, since at each testing age the measures taken are different in nature.

The third model, concerning continuity of developmental status, can accommodate the data. For example, it is possible that fixation patterns at 3 months, manipulatory and exploratory behaviour at 9 months, and language development at 18 months, all develop as entirely different abilities at different points in time, but each maintaining a similar developmental growth or maturational pattern. If this were the case, we might expect infants who are mature in one of these modalities, also to show maturity in another of these modalities, even though the two develop at different points in time. Also, one might expect a variety of types of measure to follow the same pattern of maturational development and therefore one might expect significant correlations to emerge between a large variety of types of measure. For example, in this research, one might have expected the measure of performance at 18 months to have been related to earlier reliable measures. The failure of the 18 month Performance measure to correlate with remaining measures may be seen to suggest that the model of continuity of underlying status is not supported by these data. However, it was suggested above that scores on the performance measure might be more directly affected by specific types of prior experience. If this were the case, then should underlying maturational pattern of this ability be stable, the current status of an individual may be difficult to assess. In conclusion, from the data reported here, one cannot reject the model of continuity of developmental status.

The second model, incorporating continuity of different behaviours which can be theoretically linked, can also accommodate the data. Indeed, there are a number of ways in which the related measures across age can be theoretically linked, and several possibilities will now be discussed.

In preceeding sections, it was suggested that the fixation measures at 3 months were possibly reflecting some aspect of processing speed. Given this, it seems intuitively plausible that those infants who can process visual information rapidly may also more adequately deal with new information in their environment. Hence these infants will manually explore their environment to a fuller extent at 9 months, and when language production and comprehension begins, these faster processors will be able to interpret and use language appropriately. Further, it does seem plausible that the faster processors are also the infants who are able to interpret and use information appropriately in their environment. The ability of infants to assimilate new information into their already existing knowledge might be important. Kagan (1979), for example, suggests that in the early months of life infants create schemata of objects and events. These schemata change as new discrepant forms of stimulation are assimilated into their already existing schemata. Towards the end of the first year Kagan suggests that infants are able to retrieve and hold in memory a schema for a prior event, and compare it with a current discrepant event. The more mature infants are more likely to be able to assimilate the new discrepant event into their existing schemata. Kagan further suggests that this is dependent on a new memorial competence, and these abilities allow the development of new symbolic categories, and ultimately appropriate language use, to develop. The speed of processing may be linked to the infants' ability to create scemata of stimuli and events in the environment, and may also be directly or indirectly linked to the ability of infants to assimilate new information into the already existing schemas. If the same underlying processes are involved in both schemata formation and assimilation processes this would suggest a direct link between the two. Alternatively, it may be the case that information processing speed determines schemata development, and the quality of these schemata affects infants' ability to assimilate new information. In this latter case, initial schemata formation may be dependent on speed of information processing, while assimilation is more dependent on the organisation of the already developed schemata.

These possible theoretical links are speculative. Nevertheless, it is apparent that theoretical links across the patterns of continuity reported here can be made. The model of underlying continuity cannot be rejected.

In conclusion, it is possible that the patterns of inter-relationships between measures reported in this thesis can be accommodated by at least two models of continuity. These are continuity of developmental status and continuity of an underlying process linking different measures of behaviour. The data presented in this thesis does not offer support for either of these models any more than for the other. Further research would be needed to clarify these models.

10.6 MODELS OF ADULT INTELLIGENCE - TENTATIVE LINKS

Models of intelligence in adulthood have over the past few years taken a turn away from the traditional psychometric approaches, towards information processing based theories of intelligence. The proceedings from a recent conference on human assessment have been published by Newstead, Irvine and Dann (1986), and this provides a good review of current theories of intelligence. A number of these are considered briefly, and their implications for the current research are discussed.

Sternberg (1981a,b) has pursued the idea that intelligence, in post infancy, may in part be understood as the ability to acquire and reason with new conceptual systems. Sternberg emphasises the role of motivation as an important factor in determining ones' interest in, seeking of, and finding novel conceptual systems in which to think, and this ability is of greater importance, in terms of intelligence, than the ability to reason with and learn concepts within a context with which the individual is already familiar.

Sternberg (1981,a) has discussed the implications of this theory in terms of infant intelligence. He suggests that the results reported by Fagan and McGrath (1981) (which indicate that preference for novel stimuli are predictive of childhood intelligence - see Chapter 2) in particular, are consistent with the suggestion that performance on novel tasks are important in terms of intelligence in infancy onward. However, it was argued earlier in this chapter that the novelty preference measure as reported by Fagan and McGrath may be better considered as a measure of processing speed rather than a preference for novelty per se. Nevertheless, processing speed might be directly related to interest in novelty (see Section 10.4.1), and it does seem plausible that individuals who are capable of processing information rapidly will be those who seek out novel forms of stimulation. Hence, Sternberg's views on the relationship between intelligence and interest in novelty can be accommodated by the findings reported here. Earlier in this chapter, the suggestion was made that fixation time at 3 months can be interpretated in terms of processing speed. Since the 'familiar' stimulus is of course novel to the infant initially, the faster processors may exhibit short looks because they are actually looking for novel stimulation in their environment. The faster they habituate to a stimulus the sooner they search for new stimulation. At 9 months, those infants who are more active in their environment (in terms of frequency of touching objects and the tendency to search), are those who process information quickly, and are keen to investigate novel events to their full extent. These infants, at 18 months, are also the ones who are likely to acquire language based skills more rapidly. Throughout infancy there is a vast amount of new concepts and relationships to comprehend, and the interest in novel may well reflect infant processing speed. Sternberg's theory then does appear to be compatible with these data.

Hunt (1978, 1986) bases his theory specifically around verbal ability. He points out that performance on verbal tasks in adulthood is dependent <u>on what an individual already knows</u>. Hunt suggests that verbal intelligence indirectly measures information processing ability by directly measuring the products of information processing. Those individuals who are better at processing information in the early years of life will be verbally more able in later life. Hunt adopts the term 'Current Information Processing' (CIP) to refer to the processes by which people integrate sequences of stimuli over time to develop a stable percept, to coordinate it with stored percepts, and to transform, abstract, and use it during problem solving. Hunt hypothesised that individuals who are able to process information rapidly will also have better verbal ability. In support of this theory, Hunt, Lunneborg, and Lewis (1975) reported that university students who obtained high scores on written tests of verbal ability, also did well on a variety of tasks which require use of CIP (for example, retention of the order of particular stimuli, and rapid manipulation of data in short-term memory as measured by the ability to remember a briefly presented visual stimulus as a word or letter). Further, Hunt also claimed that it is plausible to believe that high verbal subjects know more about linguistic aspects of their culture because they are more adept at CIP tasks, rather than the reverse. Hunt, Frost and Lunneborg (1973) hypothesised that if individuals are given the same exposure time to some specific linguistic information, then those with rapid CIP ability will fix more of the information into their long-term memories than slow processers. If on the other hand all subjects are allowed sufficient time such that they were equated in terms of processing, then long-term retention would be equated. Indeed, Hunt et al reported no correlation between verbal ability and retention of information over a period of weeks when individual learning was equated, but fast processors did remember more after a limited exposure period.

This suggestion is very similar to the argument referring to novelty preferences presented earlier in this chapter (when habituation is equated, novelty preferences are not predictive, whilst under limited exposure novelty preferences do predict), and emphasis on information processing speed in relation to at least verbal intelligence in infancy has been made throughout the chapter. Hence, Hunt's conception of CIP being central to verbal ability in adulthood is interesting, and this is certainly an area for future investigation.

There are links to be made then between current conceptions of intelligence in adulthood in relation to the possible nature of intelligence in infancy as presented in this thesis. These are currently speculative, although there are a number of clear directions emerging for future investigation.

10.7 FINAL CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH

The work described in this thesis has made some contributions to the study of continuity in intellectual development in infancy. It has emphasised the importance of a structured approach to assessing the nature of the infancy measures which have been considered to be potentially predictive of later intellectual ability. The importance of establishing the reliability of measures has been stressed (some measures considered potentially predictive were found not to be reliable). It is also apparent from these results that determining the inter-correlations between measures, both within and across age, aids interpretation of any resulting continuity. Empirical data are essential for theory formation. For the study of the development and continuity of intelligence in infancy this type of study would seem to be crucial, and it may be that substantial developments may only occur in the future if it is possible to carry out this type of study on a larger scale.

What does seem clear from these data is that inconsistencies in the data are as apparent as consistencies. Also, the reliable and predictive measures are in many cases are not the ones which one intuitively may expect to be so. Possible reasons for this have been put forward, and it is clear that theoretical reasons alone will never be sufficient in distinguishing those measures which will be predictive from those which will not. It is apparent that often, in the past, researchers have failed to consider what aspects of underlying ability, specific measures are measuring in infancy.

The data reported here suggest that there is some degree of continuity between measures of fixation at 3 months, measures of manual exploration, attention and successful search at 9 months, and verbal ability at 18 months. It has been suggested that this continuity can be explained in terms of at least two models of continuity. First, that there is continuity in the developmental status of individuals, and second that the measures across age are related in terms of an underlying process such as information processing speed.

A number of directions for future research, for both the applied and theoretical issues, are apparent and several are discussed here.

The correlations between the earlier and later measures reported here are moderate, and it would be of interest to establish the extent to which the inclusion of other measures increases the correlations. A number of studies were reviewed in Chapter 2 which reported significant correlations between aspects of the home environment, and the quality of stimulation from the care-giver, and later intellectual performance. These measures, along with the predictive measures reported in this thesis, may together account for more variance than either type of measure alone.

Also, the sample of infants used in this research were homogenous, and inclusion of 'at risk' children may have also significantly increased the predictive correlations. If this were the case, then one could establish the reliability with which one could identify in infancy, children who will suffer difficulties later in life.

Theoretical issues, relating to the nature and growth of intelligence, may be clarified to some extent by examining the inter-correlations of a wide variety of reliable measures over age in order to establish the extent to which intelligence is multi-dimensional in nature. For example, it may be possible to identify measures in infancy which predict later performance ability. However, if performance ability is related to specific experiences, as suggested earlier in this chapter, then there may be no infancy measures which will predict this ability.

In conclusion, future research concerned with prediction and development of intellectual growth should consider the issues discussed in this thesis. It seems that the question of continuity in intellectual growth in infancy is now an area for expansion; theoretical development and practical assessment may be very close at hand.

REFERENCES

Bakeman, R. and Brown, J.V. (1980) Early interaction: consequences for social and mental development at three years. Child Development, 51 437-447.

Bates, E., Benigni, L., Camaioni, L., Bretherton, I. and Volterra, V. (1979) The emergence of symbols. Cognition and communication in infancy. New York: Academic Press.

Bayley, N. (1933) The California First Year Mental Scale. Berkeley: University of California Press.

Bayley, N. (1949) Consistency and variability in the growth of intelligence from birth to eighteen years. Journal of Genetic Psychology, 75, 165-196.

Bayley, N. (1955) On the growth of intelligence. American Psychologist, 10, 805-818.

Bayley, N. (1969) Bayley Scales of Infant Development. New York: The Psychological Corporation.

Bayley, N. (1970) Development of mental abilities. In P.H. Mussen (ed.) Carmichael's manual of child psychology, Vol 1, (ed3). J. Wiley and Sons.

Beckwith, L. (1979) Prediction of emotional and social behaviour. In J.D. Osofsky (ed.) Handbook of infant development, New York: J. Wiley.

Binet, A. & Simon, T. (1905) Application des methods nouvelles au diagnostic du niveau intellectual chez les infants normaux et anormaux d'hospice et d'ecole primiare. L'Anee Psychologique, 11 245-266

Bornstein, M.H. (1985) Habituation of attention as a measure of visual information processing in human infants: summary, systematization and synthesis. In G. Gottleib and N.A. Krasnegor (eds.) Development of audition and vision during the first year of life: A methdological overview. Norwood, N.J: Ablex.

Bornstein, M.H., and Ruddy, M.G. (1984) Infant attention and maternal stimulation: prediction of cognitive and linguistic development in singletons and twins. In H. Bouma and D.G. Bouwhuis (eds.) Attention and Performance,X, London: Lawrence Erlbaum Associates.

Bornstein, M.H., and Sigman, M.D., (1986) Conuinuity in mental development from infancy. Child Development, 57, 251-274.

Brooks-Gunn, J. & Weinraub, M. (1983) Origins of infant intelligence testing. In M. Lewis (ed.) Origins of intelligence: infancy and early childhood. New York: Plenum.

Brown, R. (1973) A first language: The early stages. Cambridge, Mass: Harvard University Press.

Burt, C. (1949) The structure of the mind: A review of the results of factor analysis. British Journal of Educational Psychology, 19, 100-114, 176-199.

Butterworth, G. (1983) Structure of the mind in human infancy. In L.P. Lipsitt (ed.) Advances in infancy research, Vol II, Ablex publications.

Caron, A.J. and Caron, R.F. (1981) Processing of relational information as an index of infant risk. In S. Friedman and M. Sigman (eds.) Preterm birth and psychological development. New York: Academic Press.

Cattell, R.B. (1963) Theory of fliud and crystallised intelligence: A critical experiment. Cited by H.J. Butcher (ed.) (1968) Human intelligence: Its nature and assessment. London: Methuen and Co. Ltd.

Cattell, R.B. (1966) The meaning and strategic use of factor analysis. In R.B. Cattell (ed.) Handbook of multivariate experimental psychology. Chicago: Rand McNally and Company.

Cattell, P. (1940) The measurement of intelligence of infants and young children. New York: The Psychological Corporation.

Clarke, A.B.D. (1978) Predicting human development: Problems, evidence implications. Bulletin of the British Psychological Society, 31, 249-258.

Cohen, J. and Cohen, P. (1975) Applied multiple regression / correlation analyses for the behavioural sciences. New Jersey: Lawrence Erlbaum Associates.

Cohen, L.B. (1969) Observing responses, visual preferences, and habituation to visual stimuli in infants. Journal of Experimental Child Psychology, 7, 419-433.

Cohen, L.B. (1981) Examination of habituation as a measure of aberrant infant development. In S. Friedman and M. Sigman (eds.) Preterm birth and psychological development. New York: Academic Press.

Cohen, S.E., and Beckwith, L. (1979) Preterm infant interaction with the caregiver in the first year of life and competence at age two . Child Development, 50, 767-776.

Cohen, S.E. and Parmelee, A.H. (1983) Prediction of five-year Stanford-Binet scores in preterm infants. Child Development, 54, 1242-1253.

Cohen, L.B. and Strauss, M.S. (1979) Concept acquisition in the human infant. Child Development, 50, 419-424.

Corman, H.H. and Escalona, S.K. (1969) Stages of sensori-motor development: A replication study. Merrill-Palmer Quarterly, 15, 351-361.

Cornell, E.H. (1974) Infants' discrimination of photographs of faces following redundant presentations. Journal of Experimental Child Psychology, 18, 98-106.

Cornell, E.H. (1979) Infants' recognition memory, forgetting, and savings. Journal of Experimental Child Psychology, 28, 359-374.

Crockenburg, S. (1983) Early mother and infant antecedents of Bayley scale performance at 21 months. Developmental Psychology, 19, 727-730.

Crothers, B. and Paine, R.S. (1957) The natural history of cerebral palsy. Cited by Zelazo, P.R. (1982) An information processing approach to infant cognitive assessment. In M. Lewis and L.T. Taft (eds.) Developmental disabilities: Theory, assessment and intervention. Lancaster: MTP Press.

Darwin, C. (1872) Expressions of the emotions of man and animals. New York: Appleton, 1973 (First published 1872).

Decarie, T.G. (1969) A study of mental and emotional development of the thaledomide child. cited by Kopp, C.B. (1975) op cit.

Escalona, S.K. & Corman H. (1969) Albert Einstien scales of sensori-motor development. New York: Albert Einstien College of Medicine of Yeshiva University.

Estes (1982) Learning, memory and intelligence. In R.J. Sternberg (ed.) Handbook of human intelligence. Cambridge University Press.

Eysenck, H.J. (1981) What are intelligence tests? In H.J. Eysenck versus L. Kamin, Intelligence: A battle for the mind. London: Pan Books.

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

Fagan, J.F. (1972) Infants' recognition memory for faces. Journal of Experimental Child Psychology, 14, 453-676.

Fagan, J.F. (1974) Infant recognition memory: the effects of length of familiarisation and the type of discrimination task. Child Development, 45, 351-356.

Fagan, J.F. (1976, a) Infants' recognition memory for faces. Journal of Experimental Child Psychology, 14, 453-476.

Fagan, J.F. (1976, b) Infants' recognition of invariant features of Faces. Child Development, 47, 627-637.

Fagan, J.F. (1977) An attentional model of infant recognition. Child Development, 48, 345-359.

Fagan, J.F. (1979) The origins of facial pattern recognition. In M. Bornstein and W. Kessen (eds.) Psychological development from infancy. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Fagan, J.F. (1984) The intelligent infant: theoretical implications. Intelligence, 8, 1-9.

Fagan, J.F. and McGrath, S.K. (1981) Infant recognition memory and later intelligence. Intelligence, 5, 121-130.

Fagan, J.F. and Singer, L.T. (1983) Infant recognition memory as a measure of intelligence. In L. Lipsett (ed.) Advances in Infancy Research, Vol II, Ablex Publication.

Fantz, R.L. (1958) Pattern vision in young infants. Cited by Fantz, R.L., Fagan, J.F and Miranda, S.B. (1975) op cit.

Fantz, R.L. (1964) Visual experience in infants: decreased attention to familiar patterns relative to novel ones. Science, 146, 668-670.

Fantz, R.L. (1965) Visual perception from birth as shown by pattern selectivity. Cited by Fantz R.L., Fagan, J.F. and Miranda, S.B. (1975) op cit.

Fantz, R.L. and Fagan, J.F. (1975) Visual attention to size and number of pattern details by term and preterm infants during the first six months. Child Development, 46, 3-18.

Fantz, R.L., Fagan, J.F. and Miranda, S.B. (1975) Early visual selectivity as a function of pattern variables, previous exposure, age from birth and conception, and expected cognitive deficit. In L.B. Cohen and P. Salapatek (eds.) Infant Perception: From Sensation to Cognition. Basic Visual Processes. Vol I. New York: Academic Press.

Fantz, R,L., and Nevis, S. (1967) Pattern preferences and perceptual-cognitive development in early infancy. Merrill-Palmer Quarterly, 13, 77-108

Fantz, R.L., Ordy, J.M. and Udelf, M.S. (1962) Maturation of pattern vision in infants during the first six months. Journal of Comparative and Physiological Psychology, 55, 907-917.

Fenson, L., Sappir, V. and Minner, D.G. (1974) Attention and manipulative play in the one-year-old child. Child Development, 45, 757-764.

Friedman, S. and Carpenter, G.C. (1971) Visual response decrement as a function of age of the human newborn. Child Development, 42, 1967-1973.

Friedman, S., Carpenter, G.C. and Nagy, A.N. (1970) Decrement and recovery of response to visual stimuli in the human newborn. Cited by Friedman, S. and Carpenter, G.C. (1971) op cit.

Friedman, S., Nagy, A.N. and Carpenter, G.C. (1970) Newborn attention: differential response decrement to visual stimuli. Journal of Experimental Child Psychology, 10, 44-51.

Garrett, H.E. (1946) Developmental theory of intelligence. American Psychologist, 1, 372-378.

Gesell, A. (1940) The first five years - a guide to the study of the preschool child. New York: Harper Bros.

Gesell, A. (1954) The ontogenesis of infant behaviour. In D. Carmichael (ed.) Manual of Child Psychology. New York: Wiley.

Gesell, A. & Armatruda, C. (1941) Developmental diagnosis. New York: Hoeber.

Gibson, J.J. (1979) The ecological approach to visual perception. Boston, M.A.: Houghton Mifflin.

Goodenough, F.L. (1949) Mental testing. New York: Rhinehart.

Gratch, G. and Landers, W.F. (1971) Stage IV of Piaget's theory of infants' object concepts: A longitudinal study. Child Development, 42, 359-372.

Greenberg, D.J. (1971) Accelerating visual complexity levels in the human infant. Child Development, 42, 905-918.

Griffiths, R. (1954) The abilities of babies. A study in mental measurement. New York: McGraw-Hill.

Guildford, J.P. (1967) The nature of human intelligence. New York: McGraw-Hill.

Holman, L.B. and Freedheim, D.K. (1959) A study of IQ retest evaluation of 370 cases of cerebral palsy. Cited by Zelazo, P.R. (1982) An information processing approach to infant cognitive assessment. In M. Lewis and L.T. Taft (eds.) Developmental disabilities: Theory, assessment and intervention. Lancaster: MTP Press.

Horobin, K. and Acredolo, L. (1986) The role of attentiveness, mobile history, and separation of hiding sites on stage IV search behaviour. Journal of Experimental Child Psychology, 41, 114-127.

Horowitz, F.D., Paden, L. Bhana, K. and Self, P. (1972) An infant-control procedure for studying infant visual fixations. Developmental Psychology, 7, 90.

Hunt, E. (1978) Mechanics of verbal ability. Psychological Review, 85, 109-130.

Hunt, E., Frost, N. and Lunneborg, C. (1973) Individual differences in cognition: A new approach to intelligence. Cited by E. Hunt, C. Lunneborg, and J. Lewis (1975) op cit.

Hunt, E., Lunneborg, C. and Lewis, J. (1975) What does it mean to be high verbal? Cognitive Psychology, 7, 194-227.

Hunt, J. McV. (1970) Attentional preference and experience: I. Introduction. Journal of Genetic Psychology, 117, 99-107.

Hunt, J. McV. & Bayley, N. (1971) Explorations into patterns of mental development and prediction from the early Bayley scales of mental development. Minnesota Symposium on Child Psychology. 5, 52-71.

Jacobson, S.W., Fein, G.G., Jacobson, J.L., Schwartz, P.M. and Dowler, J.K. (1985) The effect of intrautrtine PCB exposure on visual recognition memory. Child Development, 56, 853-860.

Jaffa, A.S. (1934) The California preschool mental scale. Berkeley: University of California Press.

Jordan, N. (1972) Is there an Achilles' heel in Piaget's theorizing? Human Development, 15, 379-382.

Kagan, J. (1971) Change and continuity in infancy. New York: Wiley.

Kagan, J. (1979) Structure and process in the human infant: The ontogeny of mental representation. In M.H. Bornstein and W. Kessen (eds.) Psychological development from infancy: Image to retention. New Jersey: Lawrence Erlbaum Associates.

Karmel, B.Z. (1969) The effect of age, complexity and amount of contour on pattern preference in human infants. Journal of Experimental Child Psychology, 7, 339-354.

Kopp, C.B. (1975) Development of fine motor behaviours: issues and research. In N. Ellis (ed.) Aberrant development in infancy: human and animal studies. New Jersey: Lawrence Erlbaum publications.

Kopp, C.B. and Shaperman, J. (1973) Cognitive development in the absence of object manipulation during infancy. Developmental Psychology, 9, 130.

Kopp, C.B., Sigman, M. and Parmelee, A.H. (1974) Longitudinal study of sensory-motor development. Cited by Cohen, S.E. and Beckwith, L. (1979) op cit.

Kopp, C.B. and Vaughn, B.E. (1982) Sustained attention during exploratory manipulation as a predictor of cognitive competence in preterm infants. Child Development, 53, 174-182.

Kuhlman, F. (1922) A handbook of mental tests. Baltimore: Warwick and York.

Lewis, M. and Baldini, N. (1979) Attentional processes and individual differences. In G.A. Hale and M. Lewis (eds.) Attention and cognitive development. Plenum Press.

Lewis, M. and Brooks-Gunn, J. (1981) Visual attention at 3 months as a predictor of cognitive functioning at two years of age. Intelligence, 5, 131-140.

Lewis, M., Goldberg, S. and Campbell, H. (1969) A developmental study of learning within the first three years of life: response decrement to a redundant stimulus. Monographs of the Society for Research in Child Development, 34, (Serial no. 133).

Lewis, M., Goldberg, S. and Rausch, M. (1967) Attention distribution as a function of novelty and familiarity. Cited by Lewis, M., Goldberg, S. and Campbell, H. (1969) op cit.

Lewis, M. and McGurk, H. (1972) The evaluation of infant intelligence. Science, 178, 1174-1174.

Lewis, M. and Starr, M.D. (1979) Developmental Continuity. In J.D. Osofsky (ed.) Handbook of infant development. New York: J. Wiley and Sons.

Littman, B. and Parmelee, A.H. (1978) Medical Correlates of Infant Development. Cited by Cohen, S.E. and Beckwith, L.(1979) op cit.

Marsden, R.E. (1903) Discussion and apparatus. A study of early colour sense. Psychological Review, 10, 37-47.

McCall, R.B. (1979) The development of intellectual functioning in infancy and the prediction of later IQ. In J.D. Osofsky (ed.) Handbook of infant development. New York: J. Wiley and Sons.

McCall, R.B. (1981) Early predictors of later IQ: The search continues. Intelligence, 5, 141-147.

McCall, R.B. (1983) A conceptual approach to early mental development. In M. Lewis (ed.) Origins of intelligence Infancy and early childhood. New York: Plenum.

McCall, R.B., Hogarty, P.S., Hamilton, J.S. and Vincent, J.H. (1973) Habituation rate and the infant's response to visual discrepancies. Child Development, 44, 280-287.

McCall, R.B., Hogarty, P.S., & Hurlburt, N. (1972) Transistors in sensorimotor development and the prediction of childhood IQ. American Psychologist, 27, 278-307.

McCarthy, D. (1972) Manual for the McCarthy scales of children's abilities. Cited by Messer, D.J., McCarthy, M.E., McQuiston, S., MacTurk, R.H., Yarrow, L.J., and Vietze, P.M. (1986) op cit.

Messer, D.J., McCarthy, M.E., McQuiston, S., MacTurk, R.H., Yarrow, L.J., and Vietze, P.M. (1986) Relation between mastery behaviour in infancy and competence in early childhood. Developmental Psychology, 22, 366-372.

Miller, D.J., Ryan, E.B., Aberger, E., McGuire, M.D., Short, E.J. and Kenny, D.A. (1979) Relationships between assessments of habituation and cognitive performance in the early years of life. International Journal of Behavioural Development, 2, 159-170.

Miller, D.J., Ryan, E.B., Short, E.J., Reis, P.G., McGuire, M.D. and Culler, M.P. (1977) Relationships between early habituation and later cognitive performance in infancy. Child Development, 48, 658-661.

Miller, D.J., Spiridigliozzi, G., Ryan, E.B., Callan, M.P. and McLaughlin, J.E. (1980) Habituation and cognitive performance: relationships between measures at four years of age and earlier assessments. International Journal of Behavioural Development, 3, 131-146.

Miranda, S.B and Fantz, R.L. (1974) Recognition memory in Down's syndrome and normal infants. Child Development, 45, 651-660.

Nelson, V.L. & Richards, T. W. (1938) Studies in mental development: Performance on Gesell items at six months and its predictive value for performance on mental tests at two and three years. Journal of Genetic Psychology, 52, 303-325.

Newstead, S.E., Irvine, S.H. and Dann, P.L. (1986) Human assessment: Cognition and motivation. Dordrecht: Martinus Nijhoff Publishers, NATO ASI Series.

Piaget, J. (1953) The origin of intelligence in the child. London: Routledge and Kegan Paul Ltd.

Piaget, J. (1954) The construction of reality in the child. New York: Basic books.

O'Connor, M.J., Cohen, S., and Parmelee, A.H. (1984) Infant Auditory Discrimination in preterm and full-term infants as a predictor of 5 year intelligence. Developmental Psychology, 20, 159-165

Olson, S.L., Bates, J.E., and Bayles, K. (1984) Mother-infant interaction and the development of individual differences in children's cognitive competence. Developmental Psychology, 20, 166-179.

Pecheux, M-G. and Lecuyer, R. (1983) Habituation rate and free exploration tempo in four-month old infants. International Journal of Behavioural Development, 6, 37-50.

Reynell, J.K. (1978) The Reynell developmental language scales (revised). Windsor: NFER-Nelson publishers.

Richards, T.W. & Nelson, V.L. (1939) Abilities of infants during the first 18 months. Journal of Genetic Psychology, 55, 299-318.

Roe, K.V. (1978) Mother-stranger discriminations at 3 months as a predictor of cognitive development at 3 and 5 years. Developmental Psychology, 14, 191-192.

Roe, K.V. and McClure, A. (1980) Differential vocal responsiveness to mother versus stranger at three months and cognitive-acadamic functioning twelve years later. Paper presented at the International Conference on Infant Studies, New Haven, Conn., April 1980

Roe, K.V., McClure, A., and Roe, A. (1983) Infant Gesell scores vs. cognitive skills at 12 years. Journal of Genetic Psychology, 142, 143-147.

Rose, D.H. and Slater, A.M., (1983) Infant recognition memory following brief stimulus exposure. British Journal of Developmental Psychology, 1, 221-230.

Rose, D.H., Slater, A.M. and Perry, H. (1986) Prediction of childhood intelligence from habituation in early infancy. Intelligence, 10, in press.

Rose, S.A. (1981) Developmental changes in infants' retention of visual stimuli. Child Development, 52, 227-233.

Rose, S.A., Gottfried, A.W., Melloy-Carminar, P., and Bridger, W.H. (1982) Familiarity and novelty preferences in infant recognition memory: Implications for information processing. Developmental Psychology, 18, 704-713.

Rose, S.A. and Wallace, I.F. (1985) Visual recognition memory: A predictor of later cognitive functioning in preterms. Child Development ,56, 843-852.

Rovee-Collier, C.K. and Gekoski, M.J.. (1979) The economics of infancy: A review of conjugate reinforcement. In H.W. Reese and L.P. Lipsitt (eds.) Advances in Child Development, Vol 13, New York: Academic Press.

Rovee-Collier, C.K., Morrongiello, B.A., Aron, M. and Kupersmidt, J. (1978) Tpopgraphical response differentiation and reversal in 3-month-old infants. Infant Bahaviour and Development, 1, 323-333.

Ruddy, M.G. and Bornstein, M.H. (1982) Cognitive correlates of infant attention and maternal stimulation over the first years of life. Child Development, 53, 183-188.

Sameroff, A.J. and Chandler, M.J. (1975) Reproductive risk and the continuum of caretaking casualty. In F.D. Horowitz, M. Hetherington, S. Scarr-Salapatek and G. Siegel (eds.) Review of child development research. Chicago, Ill: University of Chicago Press.

Scarr-Salapatek, S. An evolutionary prespective of infant intelligence: species patterns and individual variations. In M. Lewis (ed.) Origins of intelligence infancy and early childhood. New York: Plenum press.

Schexnider, V.Y.R., Bell, R.Q., Shebilske, W.L. and Quinn, P. (1981) Habituation of visual attention in infants with minor physical anomolies. Child Development, 52, 812-818.

Segalowitz, S.J. (1980) Piaget's Achilles' heel: A safe soft spot? Human Development, 23,137-140.

Siegel, L.S. (1981) Infant tests as predictors of cognitive development at two years. Child Development, 52, 545-557.

Siegel, L.S. (1982) Early cognitive development and environmental correlates of language development at 4 years. International Journal of Behavioural Development, 5, 433-444.

Siegler, R.S. and Richards, D.D. (1982) The development of intelligence. In R.J. Sternberg (ed.) Handbook of human intelligence. Cambridge University Press.

Silva, P.A., Bradshaw, J. and Spears, G.F. (1978) A study of the concurrent and prediction validity of the Reynell developmental language scales. In J.K. Reynell (1978) op cit.

Skodak, M. & Skeels, H.M. A final follow-up study of 100 adopted children. Journal of Genetic Psychology, 75, 85-125.

Sokolov, Ye.N. (1963) Perception and the conditioned reflex. Cited by Lewis, M. and Baldini, N. (1979) op cit.

Spelke, E.S. (1979) Exploring audible and visible events in infancy. In A.D. Pick (ed.) Perception and its development: A tribute to E.J. Gibson. PUB***

Spelke, E.S. and Owsley, C. (1979) Intermodal exploration and knowledge in infancy. Infant Behaviour and Development, 2, 13-27.

Starkey, D. (1981) The origins of concept formation: Object sorting and object preference in early infancy. Child Development, 52, 489-497.

Sternberg, R.J. (1981 a) Novelty-seeking, novelty-finding, and the developmental continuity of intelligence. Intelligence, 5, 149-155.

Sternberg, R.J. (1981 b) Intelligence and non-entrenchment. Journal of Educational Psychology, 73, 1-16.

Stott, L.H. and Ball, R.S. (1965) Evaluation of infant and preschool mental tests. Monographs of the Society for Research in Child Development, 30 (3), Serial No.101.

Strauss, M.S. (1979) Abstraction of prototypical information by adults and 10-month old infants. Journal of Experimental Psychology: Human Learning and Memory, 5, 618-632.

Thomas, H. (1970) Psychological assessment instruments for use with human infants. Merrill-Palmer Quarterly, 16, 179-224.

Thorndike, R.L. (1940) Constancy of the IQ. Psychological Bulletin, 37, 167-186

Uzgiris, I.C. (1983) Organisation of sensorimotor intelligence. In M. Lewis (ed.), Origins of intelligence infancy and early childhood. New York: Plenum press.

Uzgiris, A.C. & Hunt, J. McV. (1975) Assessment an infancy: Ordinal scales of psychological development. Urbana: University of Illinois press.

Valentine, C.W. (1913-1914) The colour perception and colour preferences of an infant during its fourth and eigth months. British Journal of Psychology, 6, 363-386.

VanDoorninck, W.J., Caldwell, B.M., Wright, C. and Frankenburg, W.K. (1981) The relationship between twelve month home stimulation and school achievement. Child Development, 52, 1080-1083.

Vernon, P.E. (1950) The structure of human abilities. London: Methuen.

Vernon, P.E. (1969) Intelligence and cultural environment. London: Methuen, 1969.

Weizmann, F., Cohen, L.B. and Pratt, J. (1971) Novelty, familiarity, and the development of infant attention. Developmental Psychology, 4, 149-154.

Werner, J.S. and Perlmutter, M. (1979) Development of visual memory in infants. Advances in Child Development and Bahaviour, 14, 1-56.

Werner, J.S. and Siqueland, E.R. (1978) Visual recognition memory in the Preterm Infant. Infant Behaviour and Development, 1, 79-94.

Wetherford, M.J. and Cohen, L.B. (1973) Developmental changes in infant visual preferences for novelty and familiarity. Child Development, 44, 416-424

Wohlwill, J.F. (1973) The study of behavioural development. New York: Academic Press.

Zelazo, P.R. (1979) Reactivity to perceptual cognitive events application for infant assessment. In R.B. Kearsley & I.E. Sigel (eds.), Infants at risk: Assessment for cognitive functioning. New Jersey: Lawrence Erlbaum publications.

APPENDIX I

THE BAYLEY SCALES OF INFANT DEVELOPMENT

MENTAL DEVELOPMENTAL INDEX

THE DIVISION OF ITEMS INTO VERBAL AND PERFORMANCE SUB-SCALES. ITEMS 117 to 156. For further details of items, see Bayley (1969).

VERBAL ITEMS

ITEM NUMBER	DESCRIPTION
117	Shows shoes, clothing or own toy
124	Names one object
126	Follows directions with object (eg. brush dolly's hair)
127	Uses words to make wants known
128	Points to parts of doll
130	Names one picture
132	Points to 3 pictures
136	Sentence of 2 words
138	Names two objects
139	Points to five pictures
141	Names three pictures
144	Discriminates 2: cup, plate, box
145	Names watch (4th. picture - incomplete watch)
146	Names three objects
148	Points to seven pictures
149	Names five pictures

150	Names watch (2nd. picture - incomplete watch)
152	Discriminates 3: cup, plate, watch

PERFORMANCE ITEMS

ITEM NUMBER	DESCRIPTION
118	Pegs placed in 70 seconds
119	Builds tower of 3 cubes
120	Pink board: places round block
121	Blue board: places 2 round blocks
122	Attains toy with stick
123	Pegs placed in 42 seconds
125	Imitates crayon stroke
129	Blue board: places 2 round and 2 square blocks
131	Finds 2 objects
133	Broken doll: mends marginally
134	Pegs placed in 30 seconds
135	Differentiates scribble from stroke
137	Pink board: completes
140	Broken doll: mends approximately
142	Blue board: places 6 blocks
143	Builds tower of 6 cubes
147	Imitates strokes: vertical and horizontal
151	Pink board: reversed
153	Broken doll: mends exactly
154	Train of cubes
155	Blue board: completes in 150 seconds
156	Pegs placed in 22 seconds