INFANT PHYSICAL ATTRACTIVENESS, AFFECT, TEMPERAMENT, AND GENDER IN
RELATION TO TESTER BEHAVIOR

by

Andrea D. Hart

A thesis submitted in partial fulfillment
of the requirements for the degree
of
MASTER OF SCIENCE
in
Family and Human Development
Copyright © Andrea D. Hart 1996

All Rights Reserved
ABSTRACT

Infant Physical Attractiveness, Affect, Temperament, and Gender in Relation to Tester Behavior

by

Andrea D. Hart, Master of Science
Utah State University, 1996

Major Professor: Dr. Lori A. Roggman
Department: Family and Human Development

Easily observable infant characteristics have been shown to influence others' perceptions of infant competence. This study examined the relation between infant characteristics and a tester's willingness to repeat opportunities for the infant to pass items during administration of a cognitive test. Results showed that infant physical attractiveness was related to lower elicited infant performance (the ratio of items initially failed but later passed to the number of items that were initially failed). Positive affect was related to higher test scores. Because first impressions are likely to contribute to future relationships, it may be important to educate adults who interact with infants about the effects of stereotyping infants based on first impressions.
ACKNOWLEDGMENTS

I would like to thank Dr. Lori A. Roggman for making available to me the data set for the research presented in this thesis and for her support and encouragement in her role as my major professor. I would like to thank my committee members, Drs. Ann M. B. Austin and E. Belen Berry, for their insight and support. I would also like to thank Dr. Randall M. Jones for his encouragement and for giving me the incentive I needed to complete this project. I would also like to thank my family and friends for their support, as well as their patience, while I completed this project.

Andrea D. Hart
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>PROBLEM STATEMENT</td>
<td>1</td>
</tr>
<tr>
<td>LITERATURE REVIEW</td>
<td>4</td>
</tr>
<tr>
<td>Physical Attractiveness</td>
<td>7</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>9</td>
</tr>
<tr>
<td>Temperament</td>
<td>10</td>
</tr>
<tr>
<td>Gender</td>
<td>11</td>
</tr>
<tr>
<td>METHOD</td>
<td>14</td>
</tr>
<tr>
<td>Subjects</td>
<td>14</td>
</tr>
<tr>
<td>Testers</td>
<td>15</td>
</tr>
<tr>
<td>Setting</td>
<td>16</td>
</tr>
<tr>
<td>Procedure</td>
<td>16</td>
</tr>
<tr>
<td>Training Observers</td>
<td>17</td>
</tr>
<tr>
<td>Measures</td>
<td>18</td>
</tr>
<tr>
<td>RESULTS</td>
<td>23</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>29</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>38</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>46</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Summary Statistics of All Variables</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>Univariate Correlations and Standard Errors</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>Regression Analyses: Tester Encouragement and Elicited Performance in Relation to Infant Characteristics</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>Regression Analyses: Tester Encouragement and Elicited Performance by Male and Female Subjects</td>
<td>28</td>
</tr>
</tbody>
</table>
PROBLEM STATEMENT

Research has shown that first impressions, the "perceiver's perceptions," may contribute to the development of later relationships, even in mother-infant relationships that would seem to be developed over years (Hales, Lozoff, Sosa, & Kennell, 1977). In this way, individual characteristics that an infant exhibits may influence the quality of his/her environment. Previous studies have shown that physically attractive children tend to receive more attention from caregivers in child-care situations, at least at first (Hildebrandt & Cannan, 1985), and to be rated more positively by research observers (Ritter & Langlois, 1988). Three other readily perceived infant characteristics that contribute to first impressions and subsequent social interactions are gender (Baugh, Hoffman, & Cowan, 1980), expression of positive emotion (Denham, McKinley, Couchoud, & Bolt, 1990), and infant temperament (Caspi & Silva, 1995).

Due to increased maternal employment, a growing number of infants are in child-care settings (Bofferth, Brayfield, Deich, & Holcomb, 1990) where, at least initially, the adult is not familiar with the infant and is interacting with the infant in a setting that requires a certain level of competence. Because the template of the future relationship may come from the initial first impression the adult has of the infant and from early interactions based on those first impressions (Caspi & Silva, 1995; Hildebrandt & Fitzgerald, 1983), it is important to examine infant characteristics that are readily perceived and have been shown in other research to influence
others’ expectations. Infant attractiveness, affect, temperament, and gender are four such characteristics.

Another context in which infant characteristics may influence an adult’s expectations of an infant is a testing situation designed either to assess the infant for possible developmental delays or to provide data for research on infant development. The purpose of this study was to examine the interrelations of physical attractiveness, affect, infant temperament, and gender in relation to the willingness of a tester to repeat items on a developmental test in order to elicit a passing response from the infant. This testing situation was similar to any assessment, research, or new day-care setting. The adult did not know the infant, was assessing the infant’s competence, and could have been influenced by the characteristics of the infant.

Data for this study were derived by coding behavior from videotapes from a previous study of infants observed while playing and being tested. From a set of these videotapes that occluded the infant’s face to reduce the salience of attractiveness, trained observers coded: (a) infant vocal affect during the testing situation, (b) how many times the tester allowed the infant repeated opportunities to pass an item on the test, (c) the number of items that the infant eventually passed after initially failing, and as a comparison assessment, (d) the cognitive level of infant play in an unstructured free-play situation. Information was also available from temperament questionnaires completed by mothers as part of the original study. Still-shots of infant faces were copied from these
videotapes and rated for attractiveness by independent raters. Attractiveness, affect, temperament, and gender were tested in relation to tester encouragement and the elicited performance of the infant. The primary hypothesis was that if the infant was attractive, displayed a lot of positive affect, adapted to new situations easily, approached new things easily, and was male, there would be more tester encouragement, as shown by more test items on the Bayley Scales of Infant Development Mental Development Index (Bayley, 1969) for which the tester allowed the infant repeated opportunities to pass, and also a higher level of elicited performance, as shown by more items that the infant eventually passed after initially failing.
The examination of individual infant characteristics and how those characteristics may influence the infant's environment benefits from a theoretical perspective that is contextual in nature. Because broad perspectives of human development guide the study of infants, it is valuable to examine the history of these perspectives to explain why contextual theories may offer more than conventional unidirectional approaches to our understanding of how infant individual differences may influence how others respond to them.

The study of infants' individual differences has evolved from two views of child development. The first view of the developing child emphasizes the inherent qualities that a child brings to its environment (nature) (e.g., Thomas, Chess, & Birch, 1968). The second view emphasizes the malleability of the child due to environmental influences (nurture) (e.g., Skinner, 1978). These views may seem diametrically opposed, but many developmental theorists consider the effects of nature to be inseparable from the effects of nurture (Ford & Lerner, 1992). The nature-nurture issue has evolved into a more contextual approach to development. A contextual approach examines the influence an individual has on the environment, the influence the environment has on the individual, and the unique interaction between the individual and the environment. A contextual perspective of human development has been explained by Ford and Lerner (1992) in the following way:
The individual as a dynamic unit is to be understood as a complex, multilevel organization of biological structures plus biological and psychological/behavioral processes embedded and fused in dynamic interaction with multilevel environments. (p. 88)

This fusion suggests that variables should be evaluated within the context of a particular situation. This situational context may include the physical environment, the social environment, and also what individual characteristics the child brings to the environment.

Ford and Lerner (1992) assumed that genetic and environmental variables are inseparable in their influence on behavior, that individuals influence their own development by being self-organizing and self-constructing, and that individuals establish and maintain person-context patterns of organization. Contextual theory provides an excellent structure for examining infant characteristics such as physical appearance and behavioral expression that may influence a specific situation such as interaction with a strange adult.

An example of a theory from a contextualist approach is behavioral confirmation theory. It supports this notion of context-specific behavior by framing the question, "Can a perceiver's perceptions of another individual channel social interaction in ways that actually cause the target individual's actions to provide behavioral confirmation of the perceiver's beliefs?" (Snyder & Swann, 1978, p. 149, emphasis in original). Behavioral confirmation theory examines individual characteristics in a context of expectations based on those characteristics and the social interactions and implications that arise from that context.
A model of the process of behavioral confirmation in relation to individual characteristics is described by Langlois and Stephan (1981). First, the perceiver sees the subject, and stimulus information is processed (e.g., level of attractiveness); then the stimulus information is compared to an existing stereotype held by the perceiver (e.g., beauty-is-better), and the information is processed as a function of present stimuli, existing stereotypes, and characteristics of the perceiver. Differential expectations based on stereotypes then lead to differential treatment by reinforcing expected behaviors and discouraging unexpected behaviors. Differential treatment then may shape differential behaviors that confirm the stereotype. This circular interaction, in the beginning, may be based on erroneous assumptions, which then evoke new behavior, and the new behavior confirms those erroneous assumptions (Merton, 1957). Several studies have been done that show that initial social interactions may indeed be guided by first impressions, thereby supporting behavioral confirmation theory (Coffman, Levitt, Deets, & Quigley, 1991; Curtis & Miller, 1986; Snyder & Swann, 1978). Because initial social interactions with an infant may elicit differential treatment from adults that may influence behavioral development, it is important to examine variables that may contribute to the development of first impressions.
Physical Attractiveness

First impressions and resulting expectations are likely to be influenced by visible characteristics, such as attractiveness, expression of affect, certain temperamental characteristics, and gender. Physical attractiveness is immediately apparent to an unfamiliar adult and has been shown to elicit differential expectations that are relatively consistent among perceivers. There is remarkable reliability in people's ratings of others' attractiveness across age, race, and culture (Cunningham, 1986; Kleck, Richardson, & Ronald, 1974; Langlois, Ritter, Roggman, & Vaughn, 1991; Langlois, Roggman, & Rieser-Danner, 1990; Langlois & Stephan, 1981). These judgments are usually obtained by asking adults to view standardized pictures of people and then rate the pictures on a Likert-type scale as to their attractiveness. Beauty is apparently not only in the eye of the beholder but in everyone's eyes. Not only do adults reliably rate physical attractiveness, but infants also look longer at photographs of faces previously rated as attractive by adults (Langlois et al., 1990; Samuels & Ewy, 1985). This implies that the notion of physical attractiveness is not taught through cultural transmission but may be innate. Therefore, physically attractive individuals may elicit an automatic positive response from others.

How does this automatic response to physical attractiveness influence our expectations of others? Stephan and Langlois (1984) found that adults were more likely to rate attractive infants, across a mix of three different ethnic groups (Black, Caucasian, and
Mexican American), as good babies and as smart/likable babies. The unattractive infants were seen as causing their parents more problems than attractive infants. Other studies have also shown strong positive correlations between the level of perceived physical attractiveness and the level of perceived competence (Adams, Hichen, & Salehi, 1988; Keanealy, Frude, & Shaw, 1988; Langlois & Vaughn, 1983). In a more recent study, it was found that college students were more favorable in their ratings of attractive infants, infants showing positive affect, as well as male infants (Karraker & Stern, 1990). Ritter, Casey, and Langlois (1991) found that while mothers rating unfamiliar infants expected unattractive infants to be able to do more specific developmental tasks than attractive infants, they nevertheless rated the general competence of the attractive infants to be higher than that of the unattractive infants. This research clearly shows the power of physical attractiveness as a stimulus that elicits differential expectations. Differential expectations often lead to differential treatment. Social interaction experienced by babies also seems to be influenced by their appearance. In one study, cute babies had more opportunity to engage in interaction and experienced more positive exchanges than unattractive babies (Hildebrandt & Fitzgerald, 1983). Later in development, physical attractiveness has been correlated with peer status and social competence (Langlois & Vaughn, 1983; Lerner & Lerner, 1977) and with teachers' ratings of children's social skills, popularity, cognitive skills, confidence, and qualities of leadership (Keanealy et al., 1988).
Positive Affect

A second individual characteristic that has been shown to elicit differential expectations is positive affect. Affect is defined as the expression of emotion. The quality of affect is likely to influence a first impression and thereupon the continuing interactions of an adult with an infant. Power, Hildebrandt, and Fitzgerald (1982) have shown that both men and women look longer at photographs of infants who are smiling versus those who are crying. A recent study by Tronick (1989) has shown that infants express a variety of affective expressions that are appropriate to the context in which they find themselves. It seems that the emotional expression of the infant and the caregiver allow both to regulate their interactions.

An example of this type of regulation is demonstrated by the research showing that positive affect is related to more joint attention behavior, which is a major communicative skill in preverbal infants (Mundy, Dasari, & Sigman, 1992). Much of the research on infant emotion expression is discussed in these contextual terms. Fogel and Reimers (1989) discussed how emotional expression may be linked to emotional experiences and how experience interacts with emotional development. A recent study demonstrates this contextual interaction. Mothers' and infants' affective expressions were examined within the context of the "Strange Situation," a measure of an infant's style of attachment (Goldberg, MacKay-Soroka, & Rochester, 1994). It was found that attachment styles are related to a unique pattern of interaction with the
mother. For example, mothers of secure infants respond to both positive and negative affect displayed by the infant. However, mothers of avoidant infants are particularly unresponsive to negative infant affect. And mothers of resistant infants are unresponsive to both negative and positive infant affect. It has also been shown that when a child is a participant in a mother-child dyad that displays a lot of mutual positive affect, that child is much more likely to be compliant when asked to complete a task (Kochanska & Aksan, 1995). These findings bolster the argument that positive affect is an interactive indicator of a positive adult-child relationship.

Temperament

In related studies, various aspects of children's temperament have also been shown to be related to the social interactive behavior between children and adults (Caspi, Bem, & Elder, 1989). Temperament is linked to affect because the temperament construct includes characteristics of affective expression, such as how cheerful the child is in general or how fearful the child may be in a new situation. These aspects of temperament are stable characteristics that are expressed by the child affectively and thereby influence other's interactions with the child (Chess & Thomas, 1987). Other dimensions of temperament that may influence an adult in an assessment situation are the infant's ability to adapt to new situations and the infant's tendency to approach new stimuli. A type of child described as "slow to warm up" is slower
to adapt, withdraws from new situations, and "may not receive persistent effort [from others] ...to maintain positive interactions" (Black, Puckett, & Bell, 1992, p. 161). Together, these studies suggest that the expression of affect, both in the immediate context and as part of stable temperament characteristics, may play a part in shaping the context of an infant's world.

**Gender**

A fourth individual characteristic that has been shown to elicit differential responses is gender. In an observational study by Condry and Condry (1976), it was shown that an observer's report of affect can be influenced by the perceived sex of an infant, regardless of whether the infant is actually a girl or a boy. A 9-month-old infant was videotaped playing with a jack-in-the-box. If observers were told the infant was a boy, they were more likely to label a strong reaction to the jack-in-the-box as anger. If other observers were told the same infant was a girl, they were more likely to label the same reaction as fear. Another study by Condry and Ross (1985) again found that gender labels significantly changed adult ratings of aggressiveness.

Adults have also been shown to attend more to girls' less intense communication attempts and to boys' more intense attempts (Fagot, Hagan, Leinbach, & Kronsberg, 1985). More specifically, gender has been shown in many studies to be a factor in differential expectations of competence. Preschool children have been shown to expect more competent behavior from males than females. When asked
to point to the smarter of two infants shown on a videotape, 3- and 5-year-olds typically chose the "boy" infant (Haugh et al., 1980).

Not only do children and adults rate unfamiliar boys as more intelligent or capable, but even parents expect more from their sons than their daughters (Parsons, Adler, & Kaczalla, 1982; Phillips, 1987). Teachers also not only expect more competent behavior from boys than from girls, but when boys succeed, teachers respond in ways that reinforce boys' sense of innate intellectual ability, while when girls succeed, teachers reinforce them for nonintellectual abilities like neatness (Dweck, Davidson, Nelson, & Enna, 1978). In a more recent study by Murphy (1991), gender differences showing that boys at ages 11 and 13 outperform girls in science and mathematics may have been due to differential classroom feedback for boys versus girls and gender-biased forms of task presentation, for example, building a story problem around a situation more familiar to boys than girls. Although these studies do not use the terminology of behavioral confirmation theory, they all reveal a cycle of differential treatment that is likely to lead to different and confirming behavior. These studies emphasize the need to examine gender in the context of the testing of infant cognitive ability.

As evidenced by the previous discussion, individual characteristics have been shown to influence others' responses and expectations of competence. A beautiful, smiling, temperamentally easy male seems to have the advantage of greater expectations of competence. If it is true that immediately observable individual
characteristics produce biased treatment due to differing expectations, then these variables should be further explored in the context of adult responses to infants.

An example of a situation that is vulnerable to bias is one in which an infant is being assessed for cognitive skills using a standardized test such as the Bayley Scales of Infant Development (Bayley, 1969). The instructions for administration of the Bayley items may actually encourage tester bias. For example, the instructions state:

[The examiner's efforts should be directed toward obtaining the best test performance possible by adapting the order of presentation of items to the child's responsiveness. Rapport with the infant is essential....Frequent smiles and approving murmurs at appropriate behaviors will help maintain the child's test-oriented interest. (p. 25)]

Testers who administer such a test may be like many adults that infants meet for the first time. Testers are unfamiliar with the infant and are interacting with the infant in a setting that requires a certain level of performance from the infant. All of the reviewed studies indicate that adult-infant interactions change according to the level of physical attractiveness, affect, temperament, and gender of the participants. The objective of this study was to explore these four infant characteristics in relation to (a) tester encouragement, as shown by the tester's willingness to allow repeated opportunities for the infant to pass the items on the test, and (b) elicited performance by the infant, as shown by a higher proportion of passed items that were initially failed.
METHOD

The data for this study were obtained from videotapes from a previous research project that included three scenarios: (a) the infant playing while seated in a high-chair next to the mother, (b) the infant playing on the floor next to the mother in an unstructured free-play situation, and (c) the infant seated near the mother while being tested by trained testers with the Mental Development Index of the Bayley Scales of Infant Development (Bayley, 1969). From the videotaped images, raters scored physical attractiveness from a still shot of the infant in scenario a, trained observers coded the cognitive level of infant free play in scenario b, and independent, trained observers coded infant affect and testing interactions in scenario c.

Subjects

The sample for the original study was a convenience sample identified through a record of birth announcements in local newspapers. As part of that study, mother-infant pairs were contacted by telephone and asked to participate in a research project that examined play, attachment, and cognitive skill. Mothers were offered various coupons provided by local merchants as an incentive to participate in the research project. An appointment was scheduled at the convenience of the subjects. The original study included 166 14-month-olds, the target group for the present study.
All subjects who participated in the study were Caucasian and lived in a rural area. No information was collected as to the family level of income. For the present study, the target age for the infants at the time of testing was 14 months. The number of subjects used for this study was based on a power analysis using a convention stated by Cohen (1965), which assumes alpha at .05, nondirectional power at .80, and a medium effect size. This power analysis directed by Cohen's standards was based on a table in Kraemer and Thiemann (1988). Based on the power analysis, 51 subjects were selected from the 14-month-old group of infants. An attempt was made to select an equal number of males and females (males n = 28, females n = 23). However, the total number of available subjects available was restricted by the subjects I had tested at 14 months, and therefore there is some inequality in gender representation.

Testers

The five testers, who in the original study tested the subjects with the Bayley Scales of Infant Development (Bayley, 1969), were all female students (graduate and undergraduate) in the department of Family and Human Development at Utah State University. They were trained in standard procedures for testing infants with the Bayley Scales of Infant Development Mental Development Index (Bayley, 1969). Role-plays were done after the testers had reviewed printed instructions for the testing procedures and viewed and scored videotapes of other trained testers administering the test.
Testers did not see or meet the subjects until the mother and infant arrived at the laboratory for the observation and testing session. All testing situations were videotaped for the purpose of reliability assessment.

**Setting**

All observation and testing was conducted in a research room in the Family Life building on the campus of Utah State University. This provided a standardized setting for the study. The days and times that testing was done were flexible, not only to encourage working mothers' participation but also to get the best results from the infants. Nap and feeding times were explicitly discouraged as testing times.

**Procedure**

During the first part of the testing session, mothers completed the Toddler Temperament Scale (Fullard, McDevitt, & Carey, 1978) while their infants played nearby during a 10-minute structured play session with the infant in a high chair and a 10-minute free-play session with the infant on the floor. After the play sessions, the infant was administered the Bayley Scales of Infant Development (Bayley, 1969) by a trained tester while the mother sat nearby completing another assessment measure. The administration of the Bayley differed from the protocol, which is specified in the manual that accompanies the developmental test. The Bayley recommends administering the Bayley in the following manner:
The basal level is the item preceding the earliest failure (earliest in terms of age placement) and the ceiling is the item representing the most difficult success. The examiner should...continue testing until he feels certain that the child will not pass further items...but he should be sure to test forward for items at least a month above the apparent ceiling. (p. 29)

All Bayley items (in terms of age placement), which ranged from 8 months to 20 months, were administered to all children, regardless of the child's basal level or apparent ceiling. The proximity of the mother helped reduce stress on the part of the infants. Infants were videotaped during all parts of the session, with a timer displayed on the screen.

Training Observers

Observers who coded infant affect, tester and infant responses to the Bayley items (Bayley, 1969), and cognitive level of infant play were recruited from undergraduate human development classes and were naive to the hypotheses of this study. All observers coded modified versions of the videotapes of the testing situation in scenario a. These videotapes were modified so that the face of the baby was occluded. Not allowing observers to see the infants' faces helped eliminate the salience and potentially biasing effect of infant physical attractiveness. Again, to alleviate any biasing influence of affect, observers who coded tester encouragement were instructed to turn the audio off while observing the videotape of the testing session (except for the last three test items that dealt with language development).

All observers participated in approximately 3 hours of training. This instruction included a description of the variables
they were responsible for coding, a discussion of agreement and reliability, and procedures to ensure confidentiality. Four videos that were not used to derive data for this study were used as training tapes. After being trained in standard coding procedures, observers coded the tapes separately. During training, agreement was calculated interval by interval by using both percent agreement and also by using Kappa, a measure of the proportion of possible agreement achieved beyond the estimated chance agreement. Observers then reviewed the tapes together to clarify discrepancies and coding definitions. Agreement was checked at regular intervals throughout coding.

Measures

Tester and infant responses during the Bayley testing were coded in the following manner. Observers recorded the number of initially failed items that the tester allowed the infant repeated opportunities to pass. The definition of an opportunity to pass an item was when the tester either asked the infant to do something, or demonstrated the item and then allowed the infant to try to complete the item. Any repeat of this procedure after the infant failed to pass the item was classified as a repeated opportunity to pass the item. For example, a tester may demonstrate that a doll squeaks by squeezing the doll several times before even allowing an infant the opportunity to try to make the doll squeak by him-/herself. A multiple demonstration was not classified as a repeated opportunity for the infant until the infant failed the item, and then the tester
actually allowed the infant another chance to complete the test item.

The measure of tester encouragement was calculated as the number of repeated opportunities in proportion to the total number of items initially failed. The same observers also recorded items the infant initially failed but eventually passed, as a measure of infant elicited performance. This measure was calculated as the number of items that were initially failed but later passed in proportion to the number of items that were initially failed.

Overall item-by-item agreement for Bayley coding on 10% of the subjects was .81; Kappa was .70. Observers did not know in advance which subjects would be checked for agreement. Reliability across 34 subjects was estimated using coefficient alpha. Alpha was .82 for tester encouragement and .74 for elicited performance. The Bayley score that the tester originally gave the infant was also used as a measure because the items the tester scored the infant as having passed on the first try restricted the number of items the infant could have had multiple attempts to pass. The tester's total Bayley score was used to estimate reliability of the observer's total Bayley score; alpha was .87.

Due to technical difficulties, on 16 of the 51 videotapes the audio track did not record. Because the infants' faces were occluded on all 51 of the videotapes, the audio portion of the videotape was critical in coding infant affect. Therefore, infant quality of affect was assessed by observers who viewed a subset of 35 modified videotapes (with infants' faces occluded) of the testing
session and coded their general impression every 15 seconds as to the quality of infant affect on a Likert scale from 1 being very negative to 5 being very positive. Definitions of these ratings are as follows:

1. Very negative--Child is having a bad time, not very interested in the toys, vocalizations are usually very negative.

2. Slightly negative--Child is not having a very good time, seems bored and disinterested, is contacting toys but shows little play, vocalizations are slightly negative.


4. Slightly positive--Child is having an enjoyable time, seems to be enjoying toys and interactions.

5. Very positive--Child is having an exceptionally great time, shows enthusiastic playfulness, frequent positive vocalizations, and laughter.

The amount of positive affect was measured by obtaining the ratio of positive intervals (intervals scored 4 or 5) to the total number of intervals. One observer coded all 35 of the videotapes and another observer coded 18 of those videotapes. Interval-by-interval agreement and Kappa were calculated between observers for 10% of the subjects as a reliability check during coding. Interval-by-interval percent agreement was .69, but Kappa was only .41. Observers did not know in advance which subjects would be checked for agreement. After coding was completed, reliability between that pair of raters was estimated across the 18 subjects using Cronbach's coefficient alpha. Alpha for affect was .73. Infant adaptability and ease with
which the infant approaches new stimuli were assessed by maternal
report using the Toddler Temperament Scale (Fullard et al., 1978).
Alpha for adaptability was .52. Alpha for ease of approach was .81.

Complexity of toy play, a separate assessment of cognitive
ability, was coded from the infant's spontaneous display of
competence with toys in the free-play situation, scenario b.
Complexity of play was coded using the "Infant Play with Objects"
measure, a sequence of play levels representing increasing levels of
cognitive development (Belsky & Most, 1981). Definitions of these
play levels can be seen in the appendix. For the 10-minute free-
play session, the observer recorded the highest level of play that
was displayed in every 15-second interval. The highest level of
play recorded in the entire free-play situation was included in data
analyses to control for the infant's spontaneous level of competence
with toys (Brncir, Speller, & West, 1985). A primary observer coded
all of the videotapes and a reliability observer coded 17 videotapes
for which interval-by-interval agreement was .59 and Kappa was .42.
Reliability was estimated by Cronbach's alpha, which equaled .73.

Physical attractiveness was assessed by obtaining ratings of
each infant viewed in a 10-second video still-shot showing the
infant facing forward with a neutral expression. To provide a
consistent situation, this still-shot was copied from the part of
the videotape when the infant was sitting in the high-chair,
scenario a. The videotape was modified through the use of a
computer-aided video editing program. Most of the environment
around the infant was occluded. These video stills were shown to 81
psychology students, at another university in another region of the country, who did not know the infants and who rated the infants' physical attractiveness on a Likert scale from 1 (unattractive) to 5 (attractive). Raters did not see any infant behavior or any of the testing situation. These ratings were assessed for reliability using coefficient alpha, which equaled .96.
RESULTS

To ensure the accuracy of the data, all data were entered twice and one file was subtracted from the other. Any result from this subtraction other than zero was checked and compared to the original data to ensure accuracy. Frequencies were examined as an additional accuracy check. Summary statistics are shown in Table 1.

Table 1

Summary Statistics of All Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tester's Bayley Score</td>
<td>24.04</td>
<td>3.93</td>
<td>14.00</td>
<td>34.00</td>
<td>51</td>
</tr>
<tr>
<td>Observer's Bayley Score</td>
<td>23.43</td>
<td>3.86</td>
<td>15.00</td>
<td>33.00</td>
<td>51</td>
</tr>
<tr>
<td>Tester Encouragement</td>
<td>.80</td>
<td>.11</td>
<td>.53</td>
<td>1.00</td>
<td>51</td>
</tr>
<tr>
<td>Elicited Performance</td>
<td>.25</td>
<td>.10</td>
<td>.05</td>
<td>.50</td>
<td>51</td>
</tr>
<tr>
<td>Physical Attractiveness</td>
<td>2.83</td>
<td>.55</td>
<td>1.86</td>
<td>4.15</td>
<td>51</td>
</tr>
<tr>
<td>Complexity of Play</td>
<td>6.86</td>
<td>1.58</td>
<td>4.00</td>
<td>9.00</td>
<td>51</td>
</tr>
<tr>
<td>Ease to Approach New Things</td>
<td>3.03</td>
<td>.85</td>
<td>1.08</td>
<td>5.17</td>
<td>51</td>
</tr>
<tr>
<td>Adaptability</td>
<td>3.45</td>
<td>.67</td>
<td>1.78</td>
<td>5.14</td>
<td>51</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>.70</td>
<td>.14</td>
<td>.47</td>
<td>.96</td>
<td>35</td>
</tr>
</tbody>
</table>
Gender differences were examined using one-way ANOVA, and no significant differences between male and female infants were found for any of the variables considered. Because there were several original testers, differences between testers were also examined for infant physical attractiveness, affect, tester encouragement, and elicited infant performance using one-way ANOVA. No significant differences were found among testers.

Pearson correlations were used to explore the relations of infant characteristics with hypothesized variables of interest (see Table 2). Tester encouragement, or the number of repeated opportunities in proportion to the total number of items initially failed, was positively correlated ($r = .37, p \leq .01$) with elicited performance, or the number of items that were initially failed but later passed in proportion to the number of items that were initially failed. Elicited performance was positively correlated with infant's tendency to approach new things ($r = .28, p \leq .05$), adaptability ($r = .28, p \leq .05$), and the observer's Bayley score ($r = .59, p \leq .01$). Infant adaptability was positively correlated ($r = .38, p \leq .01$) with the infants' tendency to approach new things. Infant positive affect was positively correlated ($r = .37, p \leq .05$) with the original tester's Bayley score as well as the observer's Bayley score ($r = .34, p \leq .05$). The original tester's Bayley score was positively correlated ($r = .77, p \leq .01$) with the observer's Bayley score and with elicited performance ($r = .34, p \leq .05$).
Table 2

Univariate Correlations and Standard Errors

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Tester’s Bayley Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Observer’s Bayley Score</td>
<td>.77**</td>
<td>(.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Tester Encouragement</td>
<td>-.17</td>
<td>.07</td>
<td>(5.25)</td>
<td>(.004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Elicited Performance</td>
<td>.34*</td>
<td>.59**</td>
<td>.37**</td>
<td>(5.14)</td>
<td>(.004)</td>
<td>(.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Physical Attractiveness</td>
<td>-.04</td>
<td>.06</td>
<td>-.16</td>
<td>.26</td>
<td>(1.02)</td>
<td>(.02)</td>
<td>(.03)</td>
<td>(.03)</td>
</tr>
<tr>
<td>6 Complexity of Play</td>
<td>-.05</td>
<td>.04</td>
<td>.15</td>
<td>.13</td>
<td>-.09</td>
<td>(.36)</td>
<td>(.06)</td>
<td>(.01)</td>
</tr>
<tr>
<td>7 Ease to Approach New Things</td>
<td>.01</td>
<td>.03</td>
<td>.05</td>
<td>.28*</td>
<td>.13</td>
<td>.21</td>
<td>(.66)</td>
<td>(.03)</td>
</tr>
<tr>
<td>8 Adaptability</td>
<td>-.06</td>
<td>.07</td>
<td>.04</td>
<td>.28*</td>
<td>.20</td>
<td>.24</td>
<td>.38**</td>
<td>(.84)</td>
</tr>
<tr>
<td>9 Positive Affect</td>
<td>.37*</td>
<td>.34*</td>
<td>-.10</td>
<td>.09</td>
<td>.27</td>
<td>.15</td>
<td>.00</td>
<td>-.13</td>
</tr>
</tbody>
</table>

* p ≤ .05,  ** p ≤ .01

Regression analyses were used to explore the combined effects of the measured infant characteristics on tester encouragement and on elicited performance of the infant. Four regression models were used: two to predict tester encouragement, and the other two to predict elicited infant performance.
The first model included the following independent variables: peak play, which was entered into the regression first as the control variable; infant tendency to approach new things; infant adaptability; and infant physical attractiveness. The second model included all of the previous independent variables as well as positive infant affect. For each of these analyses, peak level of play was entered first as a control for unelicited infant competence, then the measures of infant characteristics (positive affect, tendency to approach new things, adaptability, gender, and physical attractiveness) were added to the models to predict the dependent variables (tester encouragement or elicited performance).

As shown in Table 3, the first two regression equations with tester encouragement as the dependent variable had no statistically significant results. However, in the second two regression models, infant physical attractiveness negatively predicted elicited infant performance. In the regression model predicting elicited performance that included affect as an independent variable, infant adaptability predicted greater elicited performance.

To further explore gender differences, two regression equations were tested separately for male infants and for female infants. The results can be seen in Table 4. Models with positive affect were not tested separately for male and female infants because of limited numbers of subjects with affect data. In the two regression equations with tester encouragement as the dependent variable, no independent variables were significant predictors. For boy infants, the regression model with elicited performance as the
dependent variable had no statistically significant results. However, for girl infants, physical attractiveness negatively predicted elicited performance.

Table 3
Regression Analyses: Tester Encouragement and Elicited Performance in Relation to Infant Characteristics

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Tester Encouragement</th>
<th>Elicited Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Physical Attractiveness</td>
<td>-.16</td>
<td>-.16</td>
</tr>
<tr>
<td>Complexity of Play</td>
<td>.12</td>
<td>.15</td>
</tr>
<tr>
<td>Ease to Approach New Things</td>
<td>.03</td>
<td>-.17</td>
</tr>
<tr>
<td>Adaptability</td>
<td>.03</td>
<td>.11</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>-.07</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>.22</td>
<td>.30</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>-.04</td>
<td>-.07</td>
</tr>
<tr>
<td>Standard Error</td>
<td>.11</td>
<td>.11</td>
</tr>
<tr>
<td>N</td>
<td>51</td>
<td>35</td>
</tr>
</tbody>
</table>

* p ≤ .05
## Table 4

Regression Analyses: Tester Encouragement and Elicited Performance by Male and Female Subjects

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Tester Encouragement</th>
<th></th>
<th>Elicited Performance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Physical Attractiveness</td>
<td>-.27</td>
<td>-.19</td>
<td>-.19</td>
<td>-.53*</td>
</tr>
<tr>
<td>Complexity of Play</td>
<td>.08</td>
<td>.18</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>Ease to Approach New Things</td>
<td>-.33</td>
<td>.44</td>
<td>.24</td>
<td>.23</td>
</tr>
<tr>
<td>Adaptability</td>
<td>.35</td>
<td>-.31</td>
<td>.34</td>
<td>.00</td>
</tr>
<tr>
<td>B</td>
<td>.42</td>
<td>.49</td>
<td>.44</td>
<td>.58</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.03</td>
<td>.07</td>
<td>.06</td>
<td>.19</td>
</tr>
<tr>
<td>Standard Error</td>
<td>.10</td>
<td>.10</td>
<td>.09</td>
<td>.11</td>
</tr>
<tr>
<td>N</td>
<td>28</td>
<td>23</td>
<td>28</td>
<td>23</td>
</tr>
</tbody>
</table>

* $p \leq .05$
DISCUSSION

The purpose of this study was to examine infant characteristics that may influence adult-infant interactions in an evaluative environment. It was expected that infant characteristics and behaviors such as physical attractiveness, the amount of positive affect shown, infant temperament, and gender would be related to tester encouragement, as shown in the proportion of initially failed items for which the tester allowed more than one response opportunity. These same infant characteristics were also expected to influence the elicited performance of the infant, as shown by the proportion of initially failed test items that the infant eventually passed.

The primary question that this study addressed was whether the tester was influenced by infant physical attractiveness during the testing situation. Does a cuter infant have a better chance at scoring higher on a cognitive test that is subject to tester bias? The results from this study are contrary to the hypothesis that was supported by literature reviewed previously. The finding that physical attractiveness negatively predicted elicited infant performance contradicts the hypothesis that more physically attractive infants would be given more opportunities to pass the items on the test and would have higher elicited infant performance. However, other research has shown that although physically attractive children were attributed with more positive characteristics than unattractive children, when looking at specific bias in scoring verbal subtests of the WISC--R rather than more
global bias, no main effect of attractiveness was found (Wheeler, Adams, & Nielsen, 1987). Furthermore, Ritter et al. (1991) found that although global evaluations of general competence were higher for physically attractive infants, unattractive infants were expected to be more competent in specific behaviors. Also, in the Ritter et al. (1991) study, adult perception of infant age was consistently overestimated for unattractive infants. When the researchers compared the expectations of unattractive/older-looking infants with the attractive-looking infants, they found that the judgments of the abilities of the unattractive/older-looking infants were unrealistic and were not appropriate for that age group. These studies indicate that although general expectations of physically attractive infants may be high, when examining specific behaviors, such as passing Bayley test items, adults' expectations of physically unattractive infants may be too high. Based on the Ritter et al. (1991) study, unattractive infants look older and therefore would be expected to be more competent on specific tasks. Perhaps it is this expectation that leads testers to elicit greater cognitive performance from unattractive infants.

A puzzling result was the lack of any significant relations between infant characteristics and tester encouragement. Perhaps tester encouragement is not an appropriate measure for assessing the effects of the adult-infant interaction. Other measures that may assess this interaction more accurately might include: tester affect, or, the amount of time allowed for an additional response opportunity by the infant.
The next question addressed by this study was whether the amount of positive affect displayed by an infant was related to the tester's behavior. Does a happy infant have a better chance at scoring higher on a cognitive test? As expected, on the basis of previous research, which has indicated that affect is a predictor of positive interpersonal interaction, the results from this study indicate that positive affect is correlated positively with both the original tester's and the observer's Bayley score. This supports the hypothesis that the display of positive emotion is related to the test performance by the infant.

However, when examining the regression equation results that included affect, infant expression of positive affect was not a significant predictor of tester encouragement or elicited performance. This finding, in conjunction with the significant correlations found between positive affect and both the original tester's and the observer's overall Bayley score, indicates that perhaps the original tester was influenced by the positive affect of the baby and may have passed the infant after only one trial, therefore restricting the number of multiple opportunities to pass test items. This early passing of test items would have functionally eliminated the possibility of subsequent tester encouragement or elicited performance. Positive affect seems to be interrelated with the infant's ability to adapt, which in turn is related to a tester's willingness to elicit a higher level of functioning from the infant. These results may be interpreted from a functionalist perspective. The expression of emotion, which may
be facilitated by the ease with which an infant adapts to or approaches a new situation, may be not just an expression of feelings but also an active way of manipulating the environment (Campos, Mumme, Kermoian, & Campos, 1994). An infant is not just a passive recipient of the environment, but actively pursues getting needs met. For example, a major function of crying in infancy is not just to express emotion but also to actively manipulate the environment in order to get fed or to have a diaper changed. Therefore, an expression of positive affect may be the infant's method of obtaining more assistance from the tester.

Another explanation for the correlations found between positive affect and higher Bayley scores could be the pleasure induced by success, or the expression of mastery. An observational study found that infants displayed more positive affect while engaged in mastering new motor tasks (Mayes & Zigler, 1992) than while engaged in well-practiced tasks. The expression of positive affect by those infants who scored higher on the Bayley could be just an expression of satisfaction because they were doing well. This previous research supports the findings in the present study that infants who passed more items on the Bayley test displayed more positive affect in the testing situation.

For an infant to manipulate the environment through the expression of emotion, there must be an acknowledgment or response from the recipient of that expression of emotion. Research on affect has shown that a phenomenon called emotional contagion can happen in social situations similar to the testing situation in this
study. Emotional contagion may be defined as the expression of a particular emotion that generates the same or similar feeling in the perceiver (Campos et al., 1994). This phenomenon could result in an infant's manipulation of the interaction between an infant and a tester, which could, in turn, lead to a higher or lower score on the Bayley test. Also, research on positive mood has shown that positive affect is related to increased helping behavior (Carlson, Charlin, & Miller, 1988). This could explain why positive affect is related to a higher number of passes on the Bayley test.

The final question addressed by this study was the effect of gender on tester bias. Is being a male infant related to getting a better score on a cognitive test? The results from the separate regression models for male infants and for female infants indicate that physical attractiveness is a significant negative predictor of elicited performance for female infants only. This finding is related to other studies that have examined gender and physical attractiveness. In a comprehensive review of the literature on gender and physical attractiveness, Jackson (1992) reported that some studies conducted in naturalistic settings found that attractiveness may be a liability for a female interacting with other females. Jackson (1992) reported some research evidence indicating that attractive females in particular are not selected as same-sex friends when examining actual friendships (McKillip & Riedel, 1983) and naturally occurring social interactions (Reis et al., 1982). Since all testing was done by female testers, previous research could explain why female infants who were more attractive
got lower elicited performance scores. These findings indicate a complex social interaction between females who are attractive and other females. This has been explained through a sociobiological perspective in which "gender differences in the importance of facial attractiveness stem from differences in the reproductive significance of attractiveness" (Jackson, 1992, p. 54). However, this does not seem to be a valid argument within the context of an adult-infant interaction. There would be little competition for a mate between an adult and an infant female. To my knowledge, there is no research specifically addressing the interaction between adult and infant females. This is an area of research that needs to be examined further.

Do the results from this study bolster the premises of behavioral confirmation theory? Due to limitations in the design of this study, the results are inconclusive. The most serious flaw of this study is the low reliability estimates for several of the measures. Other limitations to the design that need to be addressed are discussed by Jussim and Eccles (1995). When examining studies dealing with behavioral confirmation theory, Jussim and Eccles (1995) identified three variables that must be measured in order to identify the existence of the self-fulfilling prophecy implied by behavioral confirmation theory. The first variable is the perceiver's expectations at time 1, which should correlate with the target's behavior at time 2. Due to the nature of using data that were originally collected for a separate study, this study could not measure perceiver's expectations at the beginning of the testing
situation. The second variable that Jussim and Eccles (1995) discussed is the accuracy of the perceiver's prediction. No self-fulfilling prophecy occurs if the prediction is merely accurate. The prediction must be shown to influence the target's behavior beyond what the target would have done anyway. The most the present study can do is draw inferences; it is based on a correlational design, not a causal experimental design. Perceiver's expectations were never measured, and therefore it is only an assumption that the testers' perceptions mediated the relation between infant affect and test performance. The third factor discussed by Jussim and Eccles (1995) is the confounding issue of the perceiver being the main judge of the targets' behavior. This was not a limitation in this study because there were trained observers who viewed videotapes of the testing situation as well as the testers' judgments of the targets' behavior. Both of these final test scores were highly correlated ($r = .77, p < .05$). This judgment from two sources eliminates the bias of relying on only one perceiver.

Another limitation of this study was the testers' awareness of demographic variables that could influence their interpersonal interactions with the infants. Contextual variables such as mother's and father's age and education, infant birth order, and family income have all been shown to be factors in the development of the infant's socioemotional and cognitive competence (Duncan, Brooks-Gunn, & Klebanov, 1994; Garrett, Ng'andu, & Ferron, 1994; Hashima & Amato, 1994; Huston, McLoyd, & Coll, 1994). Testers obtained
background information from the mothers before testing the infants and could have been influenced by that information.

There are several strengths of this study. First, this design allowed for the infant-adult interaction to occur in a naturalistic setting (by not providing testers with false information prior to the test), which is described by Jussim and Eccles (1995) as a more accurate method for drawing any conclusions about self-fulfilling prophecies. Second, the test occurred in a standardized setting, was consistent across subjects, and is similar to intervention settings. Third, the testers in this study did not know the infants before the test and therefore had no previously formed expectations of the infants. Nevertheless, the evidence collected in this study is insufficient to be more than just descriptive.

Based on the results from this study, characteristics of infants that could influence "perceivers' perceptions" are indeed related to infant performance. Further research should be done, which includes data on adults' prior perceptions of the infants' age and abilities. This area of research could be beneficial in providing information for additional training to professionals who must be as unbiased as possible in situations similar to the testing situation described in this study. Because the adult-infant future relationship may be colored by the initial first impression and from early interactions based upon those first impressions (Hildebrandt & Fitzgerald, 1983), it is important to further examine the ways that infant characteristics influence initial adult-infant interactions.
Research that focuses on exploring infants' ability to actively manipulate their own environment could lead to productive questions of intervention. How actively do infants manipulate their immediate environment? Is the manipulation of their environment productive, that is, do they get what they need? What are mediating or moderating variables for those infants who are getting what they need versus those infants who are not? How can adults be taught to be unbiased in testing or other situations? This type of research would make a significant contribution to our understanding of early adult-infant interactions and what shapes them.
REFERENCES


interactions: Eye of the beholder or behavioral reality?

Developmental Psychology, 24, 254-263.


Definitions of Levels of Play

1. Mouthing--indiscriminate mouthing of toys.

2. Simple manipulation--turning over object, touching object, looking at object, banging, shaking, banging wall or window.

3. Functional--manipulation that is particularly appropriate for a certain object and involves the intentional extraction of some unique piece of information (e.g., rolling or throwing ball, shaking chain, looking at and turning pages of the book).

4. Relational--bringing together and integrating two or more toys in an inappropriate manner, that is, in a manner not initially intended by the manufacturer (e.g., larger cup on top of smaller cup, banging two cups together, holding two toys that do not belong together).

5. Functional relational--bringing together and integrating 2 toys in an appropriate manner, that is, in a manner intended by the manufacturer (e.g., nesting cups, making a tower out of cups, putting the lid on the box).

6. Enactive naming--approximate pretense activity but without confirming evidence of actual pretense behavior (e.g., raise phone receiver in proximity of ear without making talking sounds, touch cup to lip without making drinking sounds).

7. Pretend self--pretense behavior directed toward self in which pretense is apparent (e.g., raise cup to lip and make drinking sounds, put telephone to ear and "talk").
8. Pretend other—pretense behavior directed away from the child toward other (e.g., offer mother a drink, offer mother the telephone to talk into).

9. Substitution—using a "meaningless" object in a creative or imaginative manner (e.g., use the cup as a hat, use the cup as a phone, wear the chain as a necklace).

10. Sequence pretend—repetition of a single pretense act with minor variation (e.g., drink from cup, pour into cup, pour into plate) or linking together different pretense schemes (e.g., stir in the cup, then drink; answer the phone, then give it to mom).

11. Sequence pretend substitution—same as sequence pretend except using an object substitution within sequence (e.g., put doll in cradle, cover with green felt piece as "blanket"; feed self with spoon, then with stick).

12. Double substitution—pretense play in which two materials are transformed, within a single act, into something they are not in reality (e.g., treat peg as doll and a piece of green felt as a blanket, and cover peg with felt and say "night-night"; treat stick as a person and seashell as a cup, and give stick a drink).

Note. Adapted from Belsky & Most (1981).