DOCUMENT RESUME

ED 275 426 PS 016 128

AUTHOR Ritter, Jean M.; Langlois, Judith H.
TITLE The Role of Physical Attractiveness in the Observation of Adult-Child Interactions: Eye of the Beholder or Behavioral Reality?
SPONS AGENCY National Inst. of Mental Health (DHHS), Bethesda, Md.
PUB DATE Apr 86
GRANT NIMH-MH-31881
NOTE 34p.; Paper presented at the Biennial International Conference on Infant Studies (5th, Los Angeles, CA, April 10-13, 1986).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Bias; Experimenter Characteristics; Females; *Observation; *Physical Characteristics; *Researchers; *Research Methodology; Videotape Recordings
IDENTIFIERS *Adult Child Relationship; *Physical Attractiveness

ABSTRACT

The susceptibility of observations of adult-child interactions to bias due to the physical attractiveness of target persons was examined. Facial features of target persons were occluded in one version of a videotape and unoccluded in another, otherwise identical version. Using a global rating system and a molecular coding strategy, 38 trained observers coded occluded and unoccluded versions of four videotapes, two of which were of attractive targets and two of unattractive targets. Findings indicated that the observers were significantly and favorably biased toward attractive women, when making global judgments about behavioral interactions. No bias was obtained, however, when observers used a molecular coding strategy. The results demonstrate the need to be cautious about employing and interpreting global ratings and suggest that molecular coding techniques may insulate observers from bias due to the attractiveness of the observed person. (Author/RH)

The Role of Physical Attractiveness in the Observation of Adult-Child Interactions: Eye of the Beholder or Behavioral Reality?

Jean M. Ritter and Judith H. Langlois

University of Texas at Austin

Paper presented at the fifth biennial International Conference on Infant Studies, Los Angeles, California, April, 1986.

Author Notes: This research was supported by a grant from the National Institute of Mental Health (MH31881) to Judith H. Langlois. We gratefully acknowledge the assistance of Jimmy Danner in the production of the stimulus materials and of Laura Wolfe and Daniel Berezin in the data collection. We also wish to thank Robert Woodson for comments on a draft of this article as well as the undergraduate observers for the many hours they spent coding and rating videotapes. Requests for reprints should be addressed to Judith H. Langlois at the Department of Psychology, 330 Mezes Hall, The University of Texas, Austin, Texas, 78713-7789.

Abstract

The susceptibility of observations of adult-child interactions to bias due to the physical attractiveness of target persons was examined. Facial features of target persons were occluded in one version of a videotape and unoccluded in another, otherwise identical version. Using a global rating system and a molecular coding strategy, 38 trained observers coded occluded and unoccluded versions of four videotapes, two of which were of attractive targets and two of unattractive targets. When making global judgments about behavioral interactions, the observers were significantly and favorably biased toward attractive women. No bias was obtained, however, when observers used a molecular coding strategy. The results demonstrate the need to be cautious about employing and interpreting global ratings and suggest that molecular coding techniques may insulate observers from bias due to the attractiveness of the observed person.

The Role of Physical Attractiveness in the Observation of Adult-Child Interactions: Eye of the Beholder or Behavioral Reality?

Behavioral observation is considered to be one of the most reliable and objective methods of data collection in developmental psychology. Indeed, it is the most direct methodological technique available to those who study the development of social behavior and social interaction (Lytton, 1973). Behavioral observation has been used to study altruism (e.g., Zahn-Waxler, Radke-Yarrow & King, 1979), peer interaction (e.g., Blurton-Jones, 1972), adult-child interaction (e.g., Lytton, 1973), moral development (e.g., Sears, Rau & Alpert, 1965), sex-typed development (e.g., Langlois & Downs, 1979) and many other social behaviors and social relationships. Observational techniques allow the investigator to amass a large body of information on infants and young children for whom the collection of self-report or other types of data is impossible or undesirable.

As Sackett, Ruppenthal, and Gluck (1978) point out, observational methods involve several levels of abstraction ranging from molecular coding strategies to global rating systems. Molecular systems, which define categories precisely and record very specific activities and behaviors, are highly objective. Yet, molecular coding is difficult to employ because it involves extensive observer training and the use of videotapes to capture the complexity of behavior for analysis. Often, molecular systems cannot be used when interpretations of the context or the intent of the subject are relevent. Global rating systems, on the other hand, are often easier than molecular systems to employ and can have greater heuristic utility (Lytton, 1974). Yet, global rating systems tend to be less

reliable and require a high degree of interpretation and judgment from the observer. Clearly, measurement decisions define the character of the data produced (Kent & Foster, 1979).

With regular monitoring of reliability, it is commonly assumed that observers of behavior are accurate and unbiased in collecting data (Kent & Foster, 1979) whether global or molecular systems are used. However, observers can be highly reliable, yet inaccurate and biased. Research in the area of expectancy effects illustrates just such a phenomenon: Expectations about the behavior of subjects can influence the results of the investigation such that observers reliably see what they expect to see (e.g., Rosenthal, 1976). For example, Langlois and Prestholdt (1977) showed planaria worms to observers who were instructed to count the number of body contractions and head turns made by the planaria. Observers were told that they would record the behavior of two planaria: an experimental worm which had received electric shock and should thus exhibit high levels of activity and a control worm which had not received any shock and should therefore show a low response rate. In reality, all subjects saw a videotape of the same planaria. The observers' expectations about the behavior of their subject strongly influenced what they saw and consequently recorded.

Similarly, Rapp (1965) had eight pairs of observers describe the behavior of a child in nursery school where one member of each observer pair was told that the child was not feeling well and the other member was given the opposite expectation. For seven of the eight pairs, a recording bias was obtained that mirrored the individual observer's expectation. In another study, Kent, O'Leary,

Diament, and Dietz (1974) told two groups of observers that they would be making detailed behavioral recordings as well as global evaluations of baseline and treatment videotapes of a classroom of disruptive children. One group was told to expect improved behavior from baseline to treatment, the other group was told to expect no change. The observers were found to be biased in the direction of their expectations for the global ratings, but not for their molecular behavioral recordings. Shuller and McNamara (1976) report similar findings for both molecular and global observations of behavior.

There is, therefore, evidence that expectation bias can occur in observation data. Most expectancy effect studies, however, have examined the influence of artificially induced expectancies on data collection rather than expectancies inherent in the observer. One such inherent expectancy is the now well-known physical attractiveness stereotype. Adults as well as children use similar standards in evaluating the attractiveness of others; they also attribute positive qualities and abilities to attractive individuals and negative ones to unattractive individuals (see Adams, 1977; Berscheid & Walster, 1974; Langlois, in press; Langlois & Stephan, 1981; Maruyama & Miller, 1981; or Sorell & Nowak, 1981 for reviews). The data on appearance-based stereotypes are one of the most consistent and robust findings in social and social-developmental psychology and have been found in at least three different ethnic groups in the United States (Kleck, Richardson, & Ronald, 1974; Langlois & Stephan, 1977; Stephan & Langlois, 1984).

The purpose of this study was to investigate whether observers of behavioral interactions between adults and children are influenced by the

facial attractiveness of the observed target person. Such attractiveness-based expectations on the part of observers may influence their recording of behavior, and demonstration of an attractiveness bias would have significant implications for the validity of research using this method of data collection.

To test for bias due to attractiveness, trained observers coded behavioral interactions from videotapes in which the facial attractiveness cues of the subject were visible. They also coded the same interaction with facial attractiveness occluded. Both global and molecular coding schemes were examined for susceptibility to bias. Based on previous research (Kent et al., 1974; Shuller & McNamara, 1976), it was predicted that a global rating strategy would be susceptible to appearance-related bias but that expectation effects would not be demonstrated when a molecular coding strategy was used. Based on the attractiveness literature, it was predicted that observers would give more favorable global judgments when the facial features of attractive targets were present than when they were occluded and, conversely, that observers would be more positive about unattractive targets when their facial features were occluded than when they were visible.

Method

Subjects

Thirty-eight upper-division undergraduate psychology majors were recruited to serve as observers for the study. This sample was selected because it is the same pool from which many researchers in academic settings draw observers for behavioral assessment. Twenty-eight of the observers were female, and 10 were male.

Stimulus Materials

The stimulus materials were four 10-minute videotapes of interaction between an adult female and an infant. The interactions occurred in the same room with the same 6-month-old infant to ensure that the infant's physical attractiveness would not differentially influence the observers. Adult and infant played on a blanket on the floor next to a box of toys with the adult facing an observational mirror. A videocamera behind the mirror recorded the interaction. A second version of the four videotapes was produced from copies of the original video record by occluding the facial features of the female adult with an editing spot covering the face and hair. These eight tapes, four unoccluded and four occluded, comprise the test tapes.

Two of the women in the test tapes were unattractive, and two were attractive. The two unattractive women were recruited from photography files from a previous research project. The photographs were rated for physical attractiveness by 98 undergraduates. Based on these ratings, several women who received low ratings were asked to participate in a study designed to train psychology students in behavioral observation. The women agreed to allow the videotaped interaction to be used with the understanding that their names would not be identified with the tapes. The tapes of the two women who appeared least attractive were chosen as the test tapes. The two women in the attractive test tapes were research assistants. All four of the target adults were in their twenties and each had a slender body build.

Because the observers coded tapes in both the occluded and unoccluded conditions, it was necessary to ensure that they not remember how they coded

the test tapes from one time to the next. Thus, sixteen additional videotapes of women interacting with infants were produced to serve as intermediate tapes. These tapes were similar to the test tapes: Half had the adult's face occluded, half did not. Some had actual mother/infant pairs, others included the same infant who served in the test tapes. Because the conversation or sounds made by the adult or infant could cue the observers that they had seen or heard the tape previously, no audio was available to the observers on any tape.

Attractiveness Assessment

As a manipulation check of the attractiveness of the target women, the women in the four test tapes and four randomly chosen women from the intermediate tapes were rated for physical attractiveness from the tapes themselves by a group of 33 independent judges. The videotape was paused at a place where the physical attractiveness of the target was clearly visible; the judges rated attractiveness from the still screen using a 1-5 Likert-type scale. The reliability of those ratings was .98 as assessed by coefficient alpha. The mean ratings for the attractive targets were 3.97 and 3.52, the means for the unattractive targets were 1.39 and 2.42, and the means for the women from the intermediate tapes ranged from 2.52 to 2.97. There was a statistically significant difference between the average rating of the attractive targets, $M = 3.74$, and the average of the unattractive targets, $M = 1.91$, $F(1,32) = 193.6$, $p < .001$.

Coding Schemes

After they had viewed the entire 10-minute interaction, observers used a 1-5 Likert-type scale to make their global ratings. Observers judged how relaxed the interaction was, how competent the caregiver was, how happy the

caregiver and infant seemed, and how much fun the pair seemed to be having together.

The molecular coding schedule was adapted from a coding scheme developed by Parke and Sawin (1975) and modified by Sawin, Langlois, and Leitner (1977). The modified schedule was then further adapted because of the loss of cues caused by the occluding of facial features (e.g., visual activity, smiling, etc.). The revised scheme (see Table 1) assessed the following caregiver behaviors: (a) holding patterns, (b) tactual and kinesthetic stimulation, (c) toy/object play, and (d) caregiving and socializing activities. The individual measures have previously been shown to be reliable in live observations of mother-infant pairs (Langlois, Sawin & Stephan, 1981). The tapes in the study were coded using time sampling of 15-second intervals; the individual codes were listed as acronym stems across the top of the coding sheet, and the presence or absence of each behavior was recorded.

Training

The observers were trained to use the molecular coding scheme on training videotapes that were more complex than the test tapes (Mash & McElwee, 1974; Mash & Makohoniuk, 1975). Training was complete when each observer attained reliability of at least .80 assessed against predetermined criterion protocols. Three experimenters served as the criterion coders who established the protocols: each coded every training tape independently and then established a consensus for every 15-second interval. Reliability for each observer was assessed using Kappa, a chance-corrected statistic (Bartko &

Carpenter, 1976). Seventy-four percent of the observers achieved reliability of at least .90; 26% achieved a reliability of at least .80 before coding the test tapes.

Procedure

Two groups of 9 and two groups of 10 observers met twice a week for one hour. During each session they coded two 10-minute videotapes. The first six to eight sessions were spent learning the coding schemes and establishing reliability. After each observer had reached the reliability criterion the test sessions began. A within-subjects design was employed in which every observer coded both tapes in both attractiveness levels in both the occluded and unoccluded conditions (total of eight test tapes). All observers coded 16 intermediate tapes between the two conditions.

Each group viewed two test tapes per day for 2 days at the beginning of the test sessions and saw the complements of those tapes, two a day for 2 days, at the end of the study. To control for potential order effects, the tapes were presented to the groups in the following manner at the beginning of the study: (a) group 1 first viewed attractive-occluded tapes and then unattractive-unoccluded tapes each day, (b) group 2 first viewed unattractive-unoccluded tapes and then attractive-occluded tapes each day, (c) group 3 first viewed unattractive-occluded tapes and then attractive-unoccluded tapes each day, and (d) group 4 first viewed attractive-unoccluded tapes and then unattractive-occluded tapes each day. The presentation at the end of the semester was the same for each group except that the occlusion condition was switched.

The observers were blind to the specific hypotheses of the study. They were told that we were investigating whether lack of cues due to occlusion

influences behavioral coding. The two research assistants who conducted the coding sessions were not informed about which sessions were test sessions and which were not.

After the data had been collected, the observers wrote short papers about the project. An examination of the papers revealed that none of the observers guessed the hypotheses. Several of them reported that they recognized the tape duplication, but, since a month or more intervened between the occlusion conditions, they did not remember how they had previously coded the tapes.

Data Analysis

Reliability

Observers. Because observers who believe their reliability will be randomly and covertly assessed tend to maintain high agreement with criterion ratings, the observers were told that their reliability would be assessed in this fashion during the course of the study (Taplin & Reid, 1973). On four separate occasions during the coding of the intermediate tapes, each observer's reliability with criterion protocols on the molecular coding schedule was reassessed using Kappa. Intermediate tapes containing the most complex interactions were chosen for covert reliability assessments. On three of those tapes, the median reliability was .83 or better, with 100% of the observers scoring above .80 on one tape, 78% of the observers scoring above .80 on one tape, and 96% scoring above .80 on a third tape. On the fourth tape, the median reliability score was only .79, but 79% of the observers scored .75 or above. Eighty-seven percent of the observers had a median reliability across tapes of .80 or better. The median reliabilities across tapes ranged from .77 to .88.

For each of the test tapes, the observers achieved reliability of .99 for the coding of the molecular variables as assessed by coefficient alpha: Thus, all 38 observers were used in the molecular analyses. Twenty-one of the observers, 15 females and 6 males, were reliable on the global ratings on the test tapes; coefficient alpha was .88 for these observers as assessed over all test tapes. The data from the 17 observers who failed to achieve satisfactory reliability on the global ratings were excluded from further analysis.

Variables. Test-tape reliability for each global variable in each occlusion condition was calculated with the 21 reliable observers representing randomly selected trained raters observing four randomly selected tapes. Reliability for each molecular variable in each occlusion condition was calculated in the same manner using all 38 observers. The resulting intraclass correlations for the global and molecular variables are presented in Table 1. Because checks and/or adjusts was not reliable, it was not included in the analyses.

Insert Table 1 about here

Global Ratings

To provide generalizability beyond single tapes, the values from both tapes representing each level of attractiveness were averaged within occlusion condition. These data were initially subjected to a one-between, three-within multivariate analysis of variance. Observer gender, the between-subjects factor, was not significant nor did it interact with any other factors; thus, the remainder of the analyses collapsed across gender. The three within-subjects

factors were occlusion condition (unoccluded and occluded), attractiveness level (attractive and unattractive) and a measure factor to determine whether there were differences between the dependent measures. The dependent measures were relaxed, happy, competent, and fun.

When univariate effects merited evaluation, a protection procedure developed by Ramsey (1980, 1982) was employed in order to maintain an appropriate Type 1 error rate. Because there were four dependent variables the initial criterion of significance for the F ratio was $p < .0125$. This criterion changed, however, depending on the number of significant univariates obtained (see Ramsey, 1982 for an explanation of this procedure).

Molecular Coding

Because the frequencies for some of the molecular variables were small, the data were transformed using a formula provided by Winer (1971). This transformation consisted of adding .5 to each variable and taking the square root of that sum. The transformed variables were used in the analyses. Similar to the procedure for the global variables, the values from both tapes representing each level of attractiveness were averaged within occlusion condition.

The molecular variables were grouped into three conceptual categories for analysis: holding patterns, tactual/kinesthetic stimulation and toy play, and caregiving and socializing activities (Parke & Sawin, 1975). Each category was separately subjected to a one-between, three-within multivariate analysis of variance. Gender of the observer was the between-subjects factor. Gender was not significant nor did it interact with any other factor for any of the

categories of behavior; therefore, the remainder of these analyses were conducted collapsing across gender. The three within-subjects factors were occlusion condition (unoccluded and occluded), attractiveness level (attractive and unattractive) and measure. The holding pattern measures were not holding, holds far, holds close, and adjusts holding. The stimulation and toy play measures were touch, pat, manipulate; rock, bounce, jostle; shows toy/object and takes toy/object. The caregiving and socializing measures were wipes, grooms, cleans; physical restraint; and encourages motor response. Again, where appropriate, protected univariate analyses were conducted (Ramsey, 1980, 1982).

Results

Global Ratings

Bias due to the attractiveness of the target was apparent when observers made their global ratings: Attractive targets were rated more favorably when their facial features were present than when those features were occluded. Further, when appearance cues were available, attractive targets were rated more favorably than unattractive targets. These results were revealed by a significant MANOVA main effect for occlusion condition, $F(1,20) = 9.68$, $p < .01$, and a significant occlusion by attractiveness condition interaction, $F(1,20) = 38.67$, $p < .001$. An analysis of simple effects was performed to further examine this interaction.

As expected, for the attractive targets, the ratings obtained in the unoccluded condition were higher than those obtained in the occluded condition, $F(1,20) = 56.07$, $p < .001$. A protected univariate analysis revealed that

each dependent variable was rated higher for these attractive targets in the unoccluded than the occluded condition: relaxed, $F(1,20) = 64.51$, $p < .001$; happy, $F(1,20) = 20.08$, $p < .001$; competent, $F(1,20) = 24.32$, $p < .001$; fun, $F(1,20) = 37.65$, $p < .001$(Table 2). For the unattractive targets, the difference between the unoccluded and the occluded measures was not significant, and none of the univariate analyses were significant although, in unprotected analyses, relaxed, $F(1,20) = 4.03$, $p < .06$, and fun, $F(1,20) = 4.10$, $p < .06$, approached significance with the occluded condition rated more favorably.

Insert Table 2 about here

Also as expected, within the unoccluded condition, the attractive targets were rated higher on the measures than the unattractive targets, $F(1,20) = 22.09$, $p < .001$. A protected univariate analysis revealed that each dependent variable was rated higher for the attractive targets than for the unattractive targets: relaxed, $F(1,20) = 29.41$, $p < .001$; happy, $F(1,20) = 16.84$, $p < .001$; competent, $F(1,20) = 6.81$, $p < .02$; fun, $F(1,20) = 14.90$, $p < .001$ (Table 2). Within the occluded condition, however, the unattractive targets were rated higher than the attractive targets, $F(1,20) = 4.65$, $p < .05$. A univariate analysis showed that the measure relaxed, $F(1,20) = 10.44$, $p < .005$, contributed most strongly to the pattern ($M = 3.36$ for the unattractive and $M = 3.07$ for the attractive). The other dependent measures did not differ significantly as a function of attractiveness level in the occluded condition.

Finally, the MANOVA also revealed a significant measure factor, $F(3,18) = 3.62$, $p < .04$, indicating that there were differences between the dependent measures as would be expected. None of the possible interactions with the measures factor were significant, however.

Molecular Coding

In contrast to the global ratings, no evidence of bias due to the attractiveness of the target was found for the molecular coding scheme.

Holding. For holding patterns, the main effect for occlusion condition was not significant and none of the possible interactions with occlusion condition were significant. The analysis did reveal a significant main effect for attractiveness, $F(1,37) = 61.24$, $p < .001$, and a significant measures factor, $F(3,35) = 961.30$, $p < .001$. A significant attractiveness level and measure interaction emerged, $F(3,35) = 193.67$, $p < .001$, indicating that the frequency with which a particular behavior was performed depended on the level of attractiveness of the target.[1] The means for the holding patterns are presented in Table 3.

Insert Table 3 about here

Stimulation and toy play. This analysis revealed significant main effects for occlusion, $F(1,37) = 6.77$, $p < .02$, attractiveness, $F(1,37) = 711.32$, $p < .001$, and the measures factor, $F(3,35) = 8166.91$, $p < .001$. Although occlusion condition did not interact with attractiveness or with attractiveness by measure, two significant interactions did emerge: occlusion condition with the measures factor, $F(3,35) = 6.32$, $p < .002$, and attractiveness level with the measures factor,

$F(3,35) = 26.15$, $p < .001$. An analysis of simple effects was conducted to uncover the source of the occlusion condition by measures factor interaction.

The only dependent measure for which there was a significant difference by occlusion effect was touch, pat, manipulate, $F(1,37) = 17.47$, $p < .001$, with a centroid of 2.74 in the unoccluded version and a centroid of 2.57 in the occluded version. Neither of the protected univariate analyses were significant using the Ramsey criteria. This occlusion difference, however, was an artifact of the design rather than evidence of observer bias. In one of the tapes, the infant was kissed several times; in the unoccluded condition this behavior was observed and coded as touch, pat, manipulate. In the occluded condition, however, the kisses were hidden by the occlusion device, and the observers could not see the behavior to code it. The means for the stimulation and toy play behaviors are in Table 3.

Caregiving. The results of the analysis of the caregiving and socializing activities showed a significant main effect for attractiveness, $F(1,37) = 52.25$, $p < .001$ and a significant effect for the measures factor, $F(2,36) = 55.99$, $p < .001$. Occlusion condition was not significant nor did it interact with any of the other factors. There was a significant interaction between attractiveness level and the measures factor, $F(2,36) = 14.60$, $p < .001$. The means for the caregiving and socializing activities are found in Table 3.

Discussion

The results of the study are clear, straightforward, and have important implications for research: Observers are significantly and consistently biased by the attractiveness of the target when making global ratings but not when

using a molecular scheme of coding behavior. The results for the occlusion manipulation showed that when observers could see the facial features of an attractive woman interacting with a baby, they rated her as more happy and competent at caregiving and they rated the interaction as more fun and relaxed than when they could not see the target's facial features. The observers did not significantly differ in their overall global ratings between occlusion conditions for the unattractive women although there was a nonsignificant trend in the predicted direction.

When the observers' global ratings are evaluated for differences due to attractiveness, the results revealed a typical "beauty-is-good" bias: When appearance cues were visible, the interactions between attractive women and the infant were rated more positively than were interactions between unattractive women and the infant on all the global measures. Conversely, when attractiveness cues were not available, unattractive women received higher ratings overall, with one of the four measures contributing most strongly to this pattern. Taken together, the results indicate that facial features, particularly those of an attractive woman, release an inherent set of expectations from observers that influence their global judgments. These judgments reflect a positive bias toward attractive targets rather than a negative prejudice against those who are less attractive.

Three points should be noted regarding the findings for the global ratings. First, because facial cues may be important in making global judgments, it might be argued that the occlusion condition may have obscured the basis for the judgment. The global judgments, however, concerned the interaction

between the caregiver and the infant which was always visible to the observer. Second, the intraclass correlations showed that the observers were quite reliable in making their global judgments even in the absence of facial cues. Thus, the observers were "reliably biased" and were consistently influenced by the attractiveness of the target, particularly by attractive targets.[2] Finally, the attractiveness of the targets was well within the normal range of individuals likely to be selected in a typical study of adult-child interaction; the targets were neither unusually attractive nor unattractive. Thus, there is no reason to expect that these results would not generalize to most studies of social behavior and interaction.

The developmental psychology literature reveals how frequently global judgments of behavior are used in research. A case in point is the large increase over the last few years in the number of studies of infant temperament that have used observational techniques (see Campos, Barrett, Lamb, Goldsmith, & Stenberg, 1983, for a review of this literature). Because temperament is defined as a disposition and not as specific behaviors, global ratings rather than molecular coding schemes are often employed to assess temperament. Although we examined only observers' perceptions of adult behavior as a function of the adult's attractiveness, it is likely that observers would be similarly biased by the attractiveness of a child or infant (e.g., Dion, 1972; Corter et al, 1975; Stephan & Langlois, 1984), with attractive infants more likely to be assigned positive or "easy" temperament classifications than less attractive infants. Given the frequency with which global observational procedures are employed in this and other areas of research in child development, it is important to directly

examine the possibility that bias due to the attractiveness of the subjects has influenced the outcome of the data collection process and, thus, the conclusions reached by the investigator. In addition, coding systems (e.g., attachment classification abstracted from the scoring of interactive behavior in the Strange Situation) which represent levels of abstraction in between the two extremes selected for study here should be investigated for susceptibility to bias due to attractiveness.

Global ratings of behavior, even if susceptible to bias, can still be useful in developmental research. Not only are such rating systems flexible and economical, but if there is an interest in motivation or intention, global ratings may be the only practical measurement method available. The researcher, however, must control for extraneous bias that is irrelevant to the judgment to be made. For bias due to the attractiveness of the target, several solutions are available to researchers. Videotapes of the target and/or the interaction could be edited to obscure facial cues prior to the behavioral coding in much the same way as was done here. Second, when coding must include facial characteristics (when studying emotional expression, for example), the attractiveness of the subjects could be assessed and used as a covariate in the data analysis of global measures. Finally, some research has shown that information about bias and the role it can play in influencing observations can partially protect observers from bias. Thus, if observers are explicitly instructed in the dangers of bias in judgments, it may intrude less into the observational process (Langlois & Prestholdt, 1977).

In contrast to global judgments, stringently defined, molecular coding schedules seem to insulate observers from bias based on the attractiveness of the target. In this study at least, there were no significant differences between observers' molecular coding when attractiveness cues were avaliable and when they were not. Regardless of the attractiveness of the target's face, the results for the occlusion factor showed that, when the observers saw facial cues, they coded the target person as engaging in essentially the same amount of behavior as when attractiveness cues were unavailable. Operational definitions of specific aspects of behavior seem to eliminate the ambiguity associated with molar schemes and to thus insulate the observer from bias by removing the necessity for subjective interpretation. This insulating effect of molecular coding strategies has been obtained in other research as well (e.g., Kent et al., 1974; Shuller & McNamara, 1976).

Overall, the results of the study indicate that as the degree of judgment and inference required of the observer increases, the more susceptible the results of the observation become to inherent biases of the observer. Because a molecular coding strategy seems to eliminate bias, at least bias due to stereotypes associated with attractiveness, it should be the methodological technique of choice whenever possible. Otherwise, the differences we observe between the sociable and unsociable infant, the sensitive and insensitive parent, and the popular and unpopular child may exist more in their appearance than in their behavior.

References

Adams, G. R. (1977). Physical attractiveness research: Toward a developmental social psychology of beauty. Human Development, 20, 217-323.

Bartko, J. J. & Carpenter, W. T. (1976). On the methods and theory of reliability. The Journal of Nervous and Mental Disease, 163, 307-317.

Berscheid, E. & Walster, E. (1974). Physical attractiveness. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 7, pp. 157-215). New York: Academic Press.

Blurton-Jones, N. (1972). Categories of child-child interaction. In N. Blurton-Jones (Ed.), Ethological studies of child behavior (pp. 97-127). London: Cambridge University Press.

Campos, J. J., Barrett, K. C., Lamb, M. E., Goldsmith, H. H., & Stenberg, C. (1983). Socioemotional development. In M. M. Haith & J. J. Campos (Eds.), Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology (pp. 783-915). New York: John Wiley and Sons.

Carter, C., Trehub, S., Bonkydis, C., Ford, L., Celhoffer, L., & Minde, K. (1975). Nurses' judgments of the attractiveness of premature infants. Unpublished manuscript, University of Toronto.

Dion, K. K. (1972). Physical attractiveness and evaluation of children's transgressions. Journal of Personality and Social Psychology, 24, 207-213.

Kent, R. N. & Foster, S. L. (1977). Direct observational procedures: Methodological issues in naturalistic settings. In A. R. Ciminero, K. S. Calhoun & H. E. Adams (Eds.), Handbook of behavioral assessment (pp. 279-328). New York: John Wiley and Sons.

Kent, R. N., O'Leary, K. D., Diament, C. & Dietz, A. (1974). Expectation biases in observational evaluation of therapeutic change. Journal of Consulting and Clinical Psychology, 42, 774-780.

Kleck, R. E., Richardson, S. A., & Ronald, C. (1974). Physical appearance cues and interpersonal attraction in children. Child Development, 45, 305-310.

Langlois, J. H. (in press). From the eye of the beholder to behavioral reality: The development of social behaviors and social relations as a function of physical attractiveness. In C. P. Herman (Ed.), Physical appearance, stigma, and social behavior: The Ontario symposium on personality and social psychology. New York: Erlbaum.

Langlois, J. H. & Prestholdt, P. H. (1977). Information: A control for observer bias. The Journal of Social Psychology, 102, 133-141.

Langlois, J. H., Sawin, D. & Stephan, C. (1981, April). Infant physical attractiveness as an elicitor of differential parenting behaviors. Paper presented at the meeting of the Society for Research in Child Development, Boston, Mass.

Langlois, J. H. & Stephan, C. (1977). The effects of physical attractiveness and ethnicity on children's behavioral attributions and peer preferences. Child Development, 48, 1694-1698.

Langlois, J. H. & Stephan, C. (1981). Beauty and the beast: The role of physical attractiveness in the development of peer relations and social behavior. In S. S. Brehm, S. M. Kassin, & F. X. Gibbons (Eds.), Developmental Social Psychology (pp. 152-168). New York: Oxford University Press.

Lytton, H. (1973). Three approaches to the study of parent-child interaction: ethological, interview and experimental. Journal of Child Psychology and Psychiatry, 14, 1-17.

Lytton, H. (1974). Comparative yield of three data sources in the study of parent-child interaction. Merrill-Palmer Quarterly, 20, 53-64.

Maruyama, G. & Miller, N. (1981). Physical attractiveness and personality. In B. A. Maher (Ed.), Progress in experimental personality research, (Vol. 10, pp. 203-280). New York: Academic Press.

Mash, E. J. & Makohoniuk, G. (1975). The effects of prior information and behavioral predictability on observer accuracy. Child Development, 46, 513-519.

Mash, E. J. & McElwee, J. D. (1974). Situational effects on observer accuracy: Behavior predictability, prior experience, and complexity of coding categories. Child Development, 45, 367-377.

Parke, R. & Sawin, D. (1975, April). Infant characteristics and behaviors as elicitors of maternal and paternal responsivity in the newborn period. Paper presented as the meeting of the Society for Research in Child Development, Denver, Colorado.

Ramsey, P. H. (1980). Choosing the most powerful pairwise multiple comparison procedure in multivariate analysis of variance. Journal of Applied Psychology, 65, 317-326.

Ramsey, P. H. (1982). Empirical power of procedures for comparing two groups on p variables. Journal of Educational Statistics, 7, 139-156.

Rapp, D. W. (1965). Detection of observer bias in the written record. Cited in R. Rosenthal, Experimenter effects in behavioral research. New York: Irvington Publishers, Inc.

Rosenthal, R. (1976). Experimenter effects in behavioral research. New York: Irvington Publishers, Inc.

Sackett, G. P., Ruppenthal, G. C. & Gluck, J. (1978). Introduction: An overview of methodological and statistical problems in observational research. In G. P. Sackett (Ed.), Observing behavior: Data collection and analysis methods (pp. 1-14). Baltimore: University Park Press.

Sawin, D. B., Langlois, J. H. & Leitner, E. F. (1977). What do you do after you ay hello? Observing, coding, and analyzing parent-infant interactior Behavior Research Methods and Instrumentation, 9, 425-428.

Sears, R. R., Rau, L. & Alpert, R. (1965). Identification and child rearing. Stanford, California: Stanford University Press.

Shuller, D. Y. & McNamara, J. R. (1976). Expectancy factors in behavioral observation. Behavior Therapy, 7, 519-527.

Sorell, G. T. & Nowak, C. A. (1981). The role of physical attractiveness as a contributor to individual development. In R. M. Lerner & N. A. Busch-Rossnagel (Eds.), Individuals as producers of their development: A life-span perspective (pp. 389-446). New York: Academic Press.

Stephan, C. & Langlois, J. H. (1984). Baby beautiful: Adult attributions of infant competence as a function of infant attractiveness. Child Development, 55, 576-585.

Taplin, P. S. & Reid, J. B. (1973). Effects of instructional set and experimenter influence on observer reliability. Child Development, 44, 547-554.

Winer, B. J. (1971). Statistical principles in experimental design. New York: McGraw-Hill, Inc. p. 399.

Zahn-Waxler, C., Radke-Yarrow, M. & King, R. A. (1979). Child-rearing and children's pro-social initiations toward victims of distress. Child Development, 50, 319-330.

Footnotes

[1]This finding, although peripherally interesting, does not bear on the hypothesis under investigation because attractiveness and measure did not interact with occlusion condition. Thus, the specific results pertaining to attractiveness level by measures factor interactions will not be discussed here although the means are reported in Table 3. Readers wishing further information concerning the results of the interactions may contact the authors.

[2]Although the reliabilities for the global variables were slightly lower for the occluded versions of the test tapes than for the unoccluded (see Table 1), the pattern of results obtained across occlusion conditions was different for attractive and unattractive targets within each global variable. Therefore, the slightly lower reliabilities for the occluded versions of the tapes cannot account for the pattern of results obtained.

Table 1

Intraclass Correlations for Global and Molecular Variables on Test Tapes

Variable	Unoccluded	Occluded
Global		
Relaxed	.90	.87
Happy	.94	.73
Competent	.91	.72
Fun	.90	.85
Molecular		
Holding patterns		
Not holding	.99	.99
Holds far	.97	.97
Holds close	.97	.98
Adjusts holding	.99	.99
Tactual/kinesthetic stimulation		
Touch, pat, manipulate	.98	.98
Rock, bounce, jostle	.99	.99
Toy/object play		
Shows toy/object	.99	.99
Takes toy/object	.97	.97

(table continues)

Variable	Unoccluded	Occluded
Caregiving and socializing activities		
Checks and/or adjusts	.78	-.50
Wipes, grooms, cleans	.99	.99
Physical restraint	.98	.94
Encourages motor response	.94	.96

Table 2

Means and Standard Deviations for Global Ratings

Variable	Attractiveness	Condition		
		Unoccluded	Occluded	Differences
Relaxed				
	High	3.62_{d} (.52)	3.07_{a} (.46)	$p < .001$
	Low	3.14_{a} (.60)	3.36_{c} (.39)	NS
Happy				
	High	3.50_{d} (.57)	2.98_{a} (.34)	$p < .001$
	Low	2.93_{a} (.60)	3.14_{a} (.45)	NS
Competent				
	High	3.45_{b} (.45)	2.93_{a} (.40)	$p < .001$
	Low	3.07_{a} (.62)	3.05_{a} (.50)	NS
Fun				
	High	3.33_{d} (.43)	2.76_{a} (.44)	$p < .001$
	Low	2.79_{a} (.62)	3.02_{a} (.49)	NS

Note. Within each variable and occlusion condition, means having different subscripts differ significantly (subscripts a and b at $p < .02$, a and c at $p < .005$, and a and d at $p < .001$).

Table 3

Means and Standard Deviations for Molecular Variables

Variable	Attractiveness	Condition	
		Unoccluded	Occluded
	Holding patterns		
Not holding			
	High	3.26 (.42)	3.27 (.43)
	Low	4.65 (.24)	4.69 (.15)
Holds far			
	High	4.10 (.75)	3.98 (.67)
	Low	4.48 (.36)	4.62 (.24)
Holds close			
	High	2.07 (.98)	2.23 (.98)
	Low	1.39 (.77)	1.23 (.42)
Adjusts holding			
	High	2.04 (.36)	1.86 (.40)
	Low	1.54 (.35)	1.58 (.29)

(table continues)

Variable	Attractiveness	Condition: Unoccluded	Condition: Occluded
Tactual/kinesthetic stimulation and toy play			
Touch, pat, manipulate[a]			
	High	2.40 (.57)	2.25 (.48)
	Low	3.07 (.41)	2.88 (.54)
Rock, bounce, jostle			
	High	1.25 (.23)	1.17 (.28)
	Low	1.31 (.19)	1.33 (.27)
Shows toy/object			
	High	4.97 (.22)	4.92 (.21)
	Low	5.35 (.13)	5.35 (.15)
Takes toy/object			
	High	.93 (.17)	.94 (.19)
	Low	1.25 (.24)	1.31 (.23)
Caregiving and socializing activities			
Wipes, grooms, cleans			
	High	.59 (.07)	.59 (.08)
	Low	.96 (.09)	1.05 (.11)

(table continues)

Variable	Attractiveness	Condition	
		Unoccluded	Occluded
Physical restraint			
	High	.78 (.17)	.81 (.26)
	Low	1.23 (.29)	1.19 (.29)
Encourages motor response			
	High	1.78 (.80)	1.95 (1.02)
	Low	1.70 (.53)	1.81 (1.08)

Note. Table means were calculated from transformed frequencies.

[a]This variable was coded more often in the unoccluded version than the occluded version ($p < .001$). The univariates were not significant.