

37

AN EXPERIMENTAL INVESTIGATION OF OBSERVER BIAS
IN THE SCORING OF BOYS' AND GIRLS' AGGRESSION

Judith A. Lyons

A Thesis
in
The Faculty
of
Psychology

Presented in Partial Fulfillment of the Requirements
for the degree of Master of Arts at
Concordia University
Montréal, Québec, Canada

November 1981

© Judith A. Lyons, 1981

ABSTRACT

AN EXPERIMENTAL INVESTIGATION OF OBSERVER BIAS IN THE SCORING OF BOYS' AND GIRLS' AGGRESSION

Judith A. Lyons

Much of the support for the idea that boys are more aggressive than girls comes from observational research. In order to examine the possibility that the attribution of more aggression to boys may be in part a function of observer bias, 40 adult observers recorded the degree of aggression in a series of line drawings showing children interacting. It was found that when observers scanned a drawing showing a dozen children playing and then recorded whether or not any aggression occurred in the scene, both male and female observers recorded significantly more aggression for pictures of boys than for the identical drawings of girls ($p < .01$). When observers were shown a sequence of drawings featuring two children and rated the children's degree of aggression on a 9-point scale, it was found that males rated boys as significantly more aggressive ($p < .05$) while females rated boys and girls the same. Subjects' responses on the Attitudes toward Women Scale

were unable to predict level of bias in scoring aggression, but Personal Attributes Questionnaire scores significantly predicted which male observers rated boys as more aggressive ($p < .05$). These findings challenge the validity of observational studies of sex differences in aggression, and they suggest a possible screening procedure for the selection of observers in future studies. It is possible that significant adults in a child's life may also overlook or underestimate girls' aggression while attending and reacting to boys' aggressive behavior, thereby contributing to the child's sex-role socialization.

Acknowledgements

Dr. Lisa Serbin guided me throughout the preparation of this thesis. I am very grateful for her generous assistance and for her constant support and encouragement.

I would like to thank Drs. David Andres and Jane Stewart for their help in the design and implementation of this study and for their invaluable editorial advice. My deep appreciation is also extended to Dr. F. F. Strayer for suggesting the use of sequential drawings similar to those used in ethological studies for training observers, and to Theresa Blicharski and Mrs. Shekhar Bhandari for designing and drawing the stimuli used in this study. I am very grateful also to Nicole Portelance for the hours she devoted to the typing of this thesis.

Finally, I very much want to thank my family for standing behind me, for encouraging me, and--most of all-- for having faith in me and in my goals.

Table of Contents

	page
Introduction	1
Method	12
Results	19
Discussion	27
Reference Note	35
References	36
Appendices	
A. Drawings Used in Part A	41
B. Drawings Used in Part B	46
C. Answer Booklet	77
D. Statistical Tables for Part A	90
E. Statistical Tables for Part B	94

List of Tables

Part A	page
A. Median Number of Times Each Behavioral Category was Reported for Girls and for Boys	91
B. Wilcoxon T Values for the Difference Between the Number of Times Each Behavior was Recorded for Boys and for Girls	92
C. Point Biserial Correlations Between Presence/Absence of Bias in Recording Aggression and Questionnaire Responses	93
 Part B	
A. Source Table for Analysis of Variance Examining Effect of Presentation Order on Degree of Bias in Rating Aggression	95
B. T-tests Comparing Ratings Given to Boys with Ratings Given to Girls on Each Scale	96
C. Source Table for Analysis of Variance Examining Effect of Sex of Observer on Degree of Bias in Rating Aggression	97
D. Source Table for Analysis of Variance Examining Effect of Sex of Observer on Degree of Bias in Rating Dominance	98
E. Source Table for Analysis of Variance Examining Effect of Sex of Observer on Degree of Bias in Rating Goodness of Behavior	99

F.	Source Table for Analysis of Variance Examining Effect of Sex of Observer on Degree of Bias in Rating Activity Level	100
G.	Source Table for Analysis of Variance Examining Effect of Sex of Observer on Degree of Bias in Rating Passivity	101
H.	Source Table for Analysis of Variance Examining Effect of Sex of Observer on Degree of Bias in Rating Submission	102
I.	Step-wise Multiple Regression of Degree of Males' Bias in Rating Aggression on Attitudes toward Women Scale and Personal Attributes Questionnaire.	103
J.	Step-wise Multiple Regression of Degree of Females' Bias in Rating Aggression on Attitudes toward Women Scale and Personal Attributes Questionnaire	104
K.	Pearson Correlations among Questionnaire Scales ..	105

An Experimental Investigation of Observer Bias
in the Scoring of Boys' and Girls' Aggression

As recently as ten years ago, it was widely assumed that boys and girls were innately different in many behaviors, attitudes, and skills. In 1974, however, Maccoby and Jacklin published an extensive review of the literature on sex differences, in which they concluded that four differences between boys and girls were reported relatively consistently: girls excelled in verbal ability, boys excelled in mathematics, boys had greater visual-spatial ability, and boys were more aggressive. Other attitudes, behaviors and abilities that were commonly thought to differentiate the sexes such as differences in timidity, anxiety, activity level, competitiveness, dominance, and nurturance yielded only ambiguous or inconsistent data when studied empirically. Purported sex differences in sociability, suggestibility, self-esteem, and achievement motivation were no more than myths, Maccoby and Jacklin concluded.

Of the four sex differences consistently found by Maccoby and Jacklin, verbal ability and aggressiveness received much of their support from observational studies and global ratings by significant others, and are thus heavily dependent on the validity of such research procedures. This is particularly true in regard to

aggression, since there have been relatively few studies employing other than observer reports as a dependent measure that support the existence of sex differences in aggressive behavior. (Sex differences in verbal ability, on the other hand, have been demonstrated repeatedly on numerous standardized tests as well as in observational studies.)

Many methodological difficulties are entailed in conducting observational research, however. Numerous studies have pointed out that high interobserver agreement can only be maintained in observational studies if observers are aware that either continuous overt or periodic covert reliability checks are being made (O'Leary & Kent, 1973; Reid, 1970; Romanczyk, Kent, Diamant, & O'Leary, 1973; Skindrud, 1973). Periodic overt checks yield spuriously high reliability coefficients, while absence of any checks allows observers to become careless in their recording. Romanczyk et. al. (1973) found that when observers were unaware that their agreement was being checked, interobserver agreement dropped significantly and observers failed to record approximately 25% of the disruptive classroom behaviors they had taken note of when they had been aware that their coding was being monitored. Skindrud (1973) suggests that unless observers are very highly trained and closely monitored, interobserver agreement is apt to be relatively low and observers' expectations are

apt to introduce bias. Published reports of studies examining sex differences in aggression rarely provide the information necessary to assess the level of observer training used, and reports generally do not indicate that covert/continuous monitoring of interobserver agreement was employed.

In addition, O'Leary and Kent (1973) stress the fact that even if observational data are collected with high levels of interobserver agreement, they are still not necessarily valid data. Romanczyk et. al. (1973) found that observers will actually sacrifice validity to increase interobserver agreement.

Along with the reliability and validity problems that must be considered when interpreting the results of an observational study on any topic, studies focusing on sex differences in aggression are hampered by two other specific problems. One possible confounding variable in naturalistic studies is the fact that parents, teachers, and other adults show a different response to the aggressive behavior of girls than to that of boys (Hatfield, Ferguson, & Alpert, 1967; Serbin, O'Leary, Kent, & Tonick, 1973).

Thus there may be differential levels of observational discriminability such that aggression in boys is more obvious because adults present on the scene, and possibly other children as well, respond openly (through scolding, physical intervention, etc.) to boys' aggressive behavior.

Such responses would tend to draw the observer's attention to the aggressive incident, minimizing the chance that boys' aggression would go unnoticed.

A second potential confound in observational studies of aggression is the fact that sex stereotypes are so deeply engrained in society that it is extremely difficult to avoid observer bias (Maccoby & Jacklin, 1974). Observers may have the expectation that boys are more aggressive, which may lead observers to perceive and/or record aggression differently for boys than for girls.

The possibility of observer bias is a major concern not only in the study of aggression but in all areas of sex-typed behavior. To examine the possibility of biased observations across a broad range of behaviors generally thought of as sex-typed, Meyer and Sobieszek (1972) had 85 adult, middle-class observers rate videotapes of two 17-month-old children. Each of the videotaped children was introduced to half of the observers as a girl, and presented to the other half as a boy. Although no overall tendency was displayed to rate the tapes stereotypically, it was found that male observers with little experience with children did attribute more supposedly sex-appropriate behaviors to the children, while female raters--especially those with a great deal of previous contact with children--showed a marginally significant bias in the opposite direction, attributing more stereotypically masculine

characteristics to children described as girls than to children described as boys.

In 1976, Condry and Condry conducted a similar study using 204 middle-class university-student observers and a videotape of a 9-month-old child. A significant interaction was found between sex of observer, observer's experience with children, and the attributed sex of the child, although the nature of the interaction differed from that found by Meyer and Sobieszek (1972). According to Condry and Condry's data, males with extensive prior experience with children were most prone to rate the child differently as a function of the child's attributed sex.

In 1978, Sobieszek conducted a modified version of the 1972 Meyer and Sobieszek study. Sixty-nine university students rated two 17-month-old children shown in videotapes. As in the 1972 study, female raters with prior experience with children responded the least stereotypically and were biased in the direction opposite to male observers. Male subjects attributed more stereotypically masculine characteristics to children identified as boys, while female subjects attributed more stereotypically masculine characteristics to children identified as girls. As in the 1972 study, the degree of bias displayed did not reach statistical significance, even among male subjects.

Although Condry and Condry (1976), Meyer and

Sobieszek (1972), and Sobieszek (1978) included aggression in the list of dimensions they asked observers to rate, there was really little opportunity for observers to see aggression in the videotapes shown. Condry and Condry used a 10-minute tape of a 9-month-old infant seated in an infant seat. The child was shown "responding to several emotionally arousing stimuli.... a teddy bear, a jack-in-the-box, a doll, and a buzzer" (Condry & Condry, 1976, p. 813). They do not report having analyzed observers' responses on the aggressive-passive dimension by itself, only in conjunction with responses on a number of other scales. In addition, they do not state the extent to which subjects reported seeing any aggression in the tapes. Meyer and Sobieszek (1972) and Sobieszek (1978) used two 10-15-minute tapes, each showing a 17-month-old child "in a fairly wide range of play activities" (Meyer & Sobieszek, 1972, p. 44). They do not report the extent to which observers recorded seeing aggression occur, but they do report that there was no significant difference in the amount of aggression (or any other individual item rated) seen as a function of the attributed sex of the child.

Thus, despite the fact that Condry and Condry (1976), Meyer and Sobieszek (1972), and Sobieszek's (1978) studies were not designed to examine ratings of aggression specifically, the suggestion in their results that

observers may rate behaviors differently when they believe they are watching a girl than when they believe they are watching a boy is relevant. It is important to note that in the observational studies of aggression cited by Maccoby and Jacklin (1974), the sex of the child being rated was usually apparent to the observers. Therefore, the existence of sex differences in aggression may not be as solidly demonstrated as Maccoby and Jacklin believed.

The present study, therefore, examines the extent to which observer bias is present in the behavioral coding of children's aggression. Rather than presenting the observers with videotapes of sex-ambiguous children, the present study employs sets of line drawings depicting pairs of children interacting. Ten sets of drawings were shown to each subject, each set consisting of a sequence of three sketches portraying an aggressive episode. Five of the sets showed interactions between two boys, while the other five sets showed the same interactions occurring between two girls. No cross-sex interactions were presented. After viewing each set of three drawings, subjects were asked to rate the two children in the drawings on several dimensions, including level of aggression. It was hypothesized that the ratings subjects gave would vary as a function of the sex of the child. In addition, it was anticipated that observers who tend to be biased in their data recording might differ from

non-biased observers on other important dimensions. Thus each observer's sex, experience with children, attitude toward sex roles and identification with sex-typed attributes was also examined.

Line drawings rather than videotapes were employed in this study for a number of reasons. First, by the time children are old enough to engage in convincing acts of interpersonal aggression toward peers, they are usually too old to appear sex-ambiguous. Even if their physical appearance does not make their gender obvious to the observer, their stance, mannerisms, toy choice, peer affiliation, etc. generally will. Second, it is not known whether observers respond to children who appear androgynous and whose sex is deduced solely from their attributed name in the same way the observers would respond to children whose appearance is very sex-typed. It might be the case that children with sex-typed clothing and hairstyles would elicit sex-biased responses from observers while androgynous-looking children would not. Third, aggression is a relatively infrequent behavior among children. Additionally, past studies suggest that children tend to inhibit their aggression when being observed or filmed (Durrett, 1959; Levin & Sears, 1956; Sears, 1951). Thus many hours of videotape would have to be compiled and edited to focus on aggressive incidents. Finally, the option of training young children to follow a prescribed script for videotaping

would not allow unquestionable equating of such factors as apparent strength of blow and degree to which the contrived scenes appeared convincing when acted out by each sex.

The use of line drawings avoided such procedural stumbling blocks and allowed much greater experimental control. However, in doing so the reality of naturalistic observations was sacrificed. This raised two important questions concerning the validity of the design: (a) When isolated from context and presented in such exaggerated focus, would aggressive behaviors become so obvious that there was no room left for observer bias? and (b) If bias was displayed, would it operate in the same direction as it might under more natural conditions?

In order to determine whether the bias that might be found was of the same sort that might operate under more natural conditions, additional drawings were also presented. These drawings showed scenes in which approximately one dozen children were playing either by themselves or in groups of two or three. The activities pictured in each drawing included two children involved in an aggressive interaction. Observers were asked to check from a list of possible behaviors those behaviors that were observed in each scene. Two versions of each aggressive interaction were presented, one showing girls and the other showing boys. Thus the hypothesis that, when aggressive behavior is embedded in a scene showing several types of interpersonal

interaction, subjects are more likely to overlook girls' aggression than boys' could be tested.

The two types of stimulus presentation used actually parallel the two most common forms of naturalistic observation--that of the "focal child", in which a given child and his/her immediate companions are observed over a continuous period of seconds or minutes, and that of a "scanning" procedure in which the observer scans an area and records any relevant behavioral episode by any child present in the group. It was hypothesized that the two types of stimuli (the sequences of three drawings showing pairs of children and the drawings of many children in playroom scenes) would evoke different forms of observer bias.

In the playroom scenes, aggression was only one of several activities pictured. It was hypothesized that the aggression shown in the scenes was subtle enough and sufficiently embedded among other activities that it would be possible for observers to overlook the girls' pushing and kicking, or for observers to interpret girls' pushing and kicking as a form of play rather than aggression. A finding that observers record more aggression for boys than for girls, yet record more vigorous play or playful physical contact among girls would suggest that observers interpret girls' pushing and kicking as play while labelling it as aggression among boys. A finding that observers record

more aggression among boys than among girls, with no difference in the amount of vigorous play or playful physical contact reported for each sex of child, would indicate that observers either noticed aggression less when it occurred among girls, or that observers were less apt to remember girls' aggression long enough to record it.

In each of the sequences of drawings showing two focal children, one or both children were acting aggressively. The aggression was therefore more salient than it was in the playroom scenes. In addition, observers were asked to rate the children's aggression on a 9-point scale rather than merely recording whether or not aggression occurred. Thus, it would have been more difficult for subjects to totally overlook girls' aggression in the sequences of drawings than it was in the playroom scenes.

There were two ways in which subjects could differentially respond to boys' and girls' aggression in the sequences of drawings. Subjects, considering aggression a masculine tendency, could rate the boys as more aggressive than the girls even when the aggression was very salient. Alternatively, subjects could find the girls' aggression too salient to overlook. Unable to ignore the girls' aggressive behavior, subjects might then rate the girls as more aggressive than the boys, since the girls' aggression deviates more from what is generally considered normative behavior for that sex.

Method

Subjects

Twenty male and 20 female volunteers were drawn from the student population and/surrounding community (i.e., staff, friends and family of students and staff) of Concordia University. Male subjects ranged in age from 20 to 30, with a mean age of 24.8 years. The mean age for female subjects was 24.1 years, with subjects ranging in age from 19 to 35 years old.

The purpose of the research was described to volunteers as follows:

The study focuses on adults' perception of children's body language. I am trying to determine how capable adults are of judging from children's nonverbal cues (such as facial expression and body posture) what feelings and intentions the children are expressing. I also want to find out what variables influence adults' judgments.

Apparatus

Drawings for Part A. Two scenes each showing 12 or 13 children playing were used in Part A of the testing. Each scene showed several typical playroom activities going on, with some children playing alone and others interacting in groups of two or three. Among the activities occurring in the two different scenes was an aggressive interaction. Embedded in one scene was one child pushing another away

from an easel; embedded in the other scene was one child kicking another. Two versions of each of these scenes were used--one copy showing two girls involved in the aggressive interaction, the other showing two boys.

Facial expressions and body postures were identical across the two versions, only clothing and hairstyles were modified to identify the two children involved in the aggressive interaction as being boys or girls. All other activities and the sex of the other children remained the same across both versions of each scene.

Behavior checklist for Part A. Following the observation of each of the scenes, subjects were asked to check which of the following behaviors they had seen occurring: solitary play, group play, cooperation, playful physical contact, concentration, vigorous play, dependency, nurturance, aggression, and relaxation.

Drawings for Part B. Ten sets of drawings were used in Part B. Each set consisted of three pictures depicting an aggressive interaction between two same-sex children. Of the 10 sets of drawings, five showed pairs of girls interacting while five parallel sets showed boys engaging in the same behaviors. The only difference between male sequences and female sequences was in hairstyle and clothing. All facial expressions and body postures were identical across sexes. The drawings used depicted the following scenes: (a) one child refusing to share a ball

and demonstrating a facial threat, the other child turning away in a sulk, (b) one child pushing another who refuses to respond to the push, (c) one child hitting another, with the second child turning away in tears, (d) one child building a puzzle, another child knocking the puzzle apart, and the first child jumping up and hitting the second, and (e) two children playing with clay, one child snatching the second child's clay, and the second child running to tell the teacher.

Drawings in both Part A and Part B were photographed and shown to subjects using a standard carousel slide projector.

Rating scales for Part B. Following each set of three drawings, observers were asked to rate each of the two children shown on the following dimensions using a 9-point scale: (a) not at all dominant--very dominant, (b) not at all passive--very passive, (c) not at all aggressive--very aggressive, (d) not at all submissive--very submissive, (e) not at all active--very active, and (f) very poorly behaved--very well behaved.

Procedure

All testing was conducted by the present author during the spring of 1981. Subjects were tested in small groups of no more than four people. All subjects completed Part A before Part B. Within each part, however, the presentation of stimuli was counterbalanced across subjects with four

different presentation orders being used.

Subjects were seated several feet apart facing a projection screen at the front of the room. Each subject was given a coding booklet in which to record his/her answers. The booklet was constructed in such a way that the two different versions of each drawing or set of drawings were scored on different pages and subjects were instructed to not refer back to their answers on previous pages in the booklet as they responded to subsequent drawings.

Beginning Part A, subjects were instructed as follows:

I am going to show you several slides picturing typical playroom scenes with approximately a dozen children in each scene. I will project each slide for 10 seconds only, so look at it carefully but quickly. After the 10 seconds are up and the slide is no longer visible, I will ask you to check which of the 10 behaviors listed in your answer booklet you saw occurring in the slide. For example, if a scene includes a child feeding a bottle to a doll, you would probably consider that a nurturant behavior so you would mark a check to the right of the word "Nurturance". If, on the other hand, no nurturant behaviors were shown, you would leave that space blank. Answer according to your own impressions of what constitutes nurturance, dependency, etc. You

can check more than one behavior for each scene.

As soon as Part A was completed, subjects were instructed to go on to Part B. Subjects were read the following instructions for Part B:

I will be showing you several sequences of drawings, each of which will show two children interacting. There will be three slides per sequence and each slide will be shown for five seconds. The three slides in each sequence form a story, so you will see a general theme running through the three slides. After all three slides in the sequence have been shown, you will mark your responses in your answer booklet. Note that you must remember which child was on your left and which was on your right. You will see in your booklet that each scale goes from 0 to 8, with 0 representing one extreme and 8 representing the other extreme of the scale. The numbers 1 through 7 represent progressive steps between the two extremes. In the spaces to the right of each scale, mark the number that best describes each of the children shown. Only mark one number per space. Do not leave any spaces blank. You must assign every child a number on every scale, even though in some cases you may feel the drawings do not give you enough information to make a judgment. In such cases, just give the best answer you can based on what you

saw in the drawings.

Please mark your judgments, as quickly as you can; do not ponder over them. There are no real right or wrong answers so just give your own impressions.

After subjects had rated all 10 sequences of drawings, they were asked to report their degree of past experience with children by responding to the question "How much contact have you had with children?". A 6-point scale was used, in which 0 represented "No contact" and 5 represented "A great deal of contact". This simple self-report procedure was examined by Sobieszek (1978) and found to be as good a measure as more elaborate and time-consuming questionnaires were in determining level of experience. An additional checklist of types of experience was included for descriptive purposes.

A 24-item short form of the Personal Attributes Questionnaire (Spence & Helmreich, 1978) was also administered. The PAQ yields scores on a Masculinity, a Femininity, and a bipolar Masculinity-femininity scale. In addition, subjects completed an abbreviated 15-item version of the Attitudes toward Women Scale (Spence & Helmreich, 1978). Spence (Note 1) describes the AWS as "a questionnaire designed to measure beliefs about the rights and roles that women ought to have or be permitted, vis à vis men, in educational, vocational, and social spheres". The AWS yields a total score indicating how

traditional vs. pro-feminist the respondent's attitudes are.

The total testing lasted approximately one hour. Rather than debriefing subjects at the conclusion of the testing session and thereby risking discussion of the purpose of the study with future subjects, subjects were later debriefed by mail as to the purpose and results of the experiment.

ResultsPart A

The subjects' task in Part A was to score the behavioral categories they observed occurring in a particular slide. Since there were two slides showing boys pushing or kicking and two slides--identical to those of the boys--showing girls pushing or kicking, each subject could code each of the 10 behavioral categories zero, one, or two times for the slides of boys and zero, one, or two times for the slides showing girls. The degree to which each subject differentiated his or her recording of aggressive behavior as a function of the sex of the child being observed was determined by subtracting the number of times each subject reported seeing aggression performed when girls were shown pushing and kicking from the number of times that subject reported aggression when boys were shown pushing and kicking. Bias scores were similarly computed for each of the other categories of behavior (group play, solitary play, cooperation, playful physical contact, concentration, vigorous play, dependency, nurturance, and relaxation).

To insure that bias did not vary as a function of presentation order, a Kruskal-Wallis one-way analysis of variance by ranks was performed using the degree of bias in the reporting of aggression as the dependent variable. The degree of bias was found to not vary significantly across the four presentation orders, $H(3) = 7.54$, n.s.

Wilcoxon matched-pairs signed ranks tests were performed to examine the significance of the difference between the number of times aggression was recorded for pictures of boys and the number of times aggression was recorded in response to pictures of girls. Six of the 20 male subjects and five of the 20 female subjects reported seeing aggression when shown a boy but did not report aggression when shown a girl performing the same behavior; no subject showed a bias in the opposite direction, $T(11) = 0$, $p < .01$. Similarly, Wilcoxon tests were calculated on the remaining nine behavioral categories to examine the difference between the number of times each category was recorded when boys were shown pushing and kicking and the number of times it was coded in response to pictures of girls pushing and kicking. The only category, other than aggression, for which subjects' responses differed significantly as a function of the sex of the child shown was relaxation. Five males and one female reported relaxation in slides showing aggressive girls without reporting relaxation in response to the comparable male picture, $T(6) = 0$, $p < .05$.

In order to determine whether subjects' descriptions of themselves on the PAQ or the traditionality of their responses on the AWS could predict which subjects would exhibit a sex-bias in recording aggression, point biserial correlation coefficients were calculated between the

presence/absence of bias and the subject's score on the AWS, as well as between the presence/absence of bias and each of the three PAQ scales. Using a two-tailed test of the significance of the correlations, no correlation had a probability of less than .05, although male subjects' responses on the Feminine scale of the PAQ correlated marginally with whether the subject showed bias in reporting aggression, $r_{pb} = +.37$, $p < .11$. Feminine males tended to report more aggression in pictures of boys while less feminine males reported equal amounts of aggression for boys and girls. For female subjects, the correlation of the PAQ Feminine scale with bias did not approach significance, $r_{pb} = +.18$, n.s.

In summary, a significant sex bias in the predicted direction was found in subjects' recordings of aggression, with 25% of female observers and 30% of male observers scoring more aggression for boys than for girls. Only the PAQ Feminine scale was able to even marginally predict which subjects would show bias, and it only predicted among male subjects in suggesting that feminine males are the men most likely to show bias in their responses.

Part B

In Part B, subjects were shown slides of 20 children interacting in pairs. Subjects rated each of the 10 boys and 10 girls on six rating scales--aggression, dominance, passivity, submission, activity level, and goodness of behavior. The degree to which each subject rated children's aggression differently as a function of the child's sex was determined by subtracting the sum of the aggression ratings each subject gave pictures of girls from the sum of the aggression ratings that subject gave pictures of boys. A one-way analysis of variance on the degree of bias shown by subjects in each of the four presentation orders demonstrated that the degree of bias did not vary significantly across presentation orders, $F(3,36) = .95$, n.s.

As described above, a positive aggression bias score indicated that more aggression was attributed to boys than to girls, while a negative score indicated that girls were rated as more aggressive. The average bias score among male subjects was 2.9, with scores ranging from -4 to 21. Of the 20 male subjects, 14 rated boys as more aggressive, five rated girls as more aggressive, and one rated the two sexes exactly the same. The standard deviation of the aggression bias scores was 5.3 for male subjects.

Among female subjects the average aggression bias score was 0.1, with scores ranging from -7 to 12. Of the

20 female subjects, nine rated boys as more aggressive, nine rated girls as more aggressive, and two attributed the same amount of aggression to both sexes. The standard deviation among aggression bias scores was 5.3 for female subjects, the same as that for males.

In order to test the primary hypothesis that subjects would rate aggression differently as a function of the sex of child, planned comparisons were conducted using two-tailed t -tests for dependent samples and employing the nominal .05 alpha level (Hays, 1973). The sum of the aggression ratings given to girls and the sum of aggression ratings given to boys constituted the repeated dependent variable. Data from male and female subjects were analyzed separately. It was evident that while female subjects showed no consistent tendency to rate one sex as more aggressive than the other, $t(19) = 0.08$, n.s., male subjects quite consistently rated boys as more aggressive than girls, $t(19) = 2.45$, $p < .05$.

As predicted, differences between ratings given to boys and those given to girls did not approach significance on any of the five remaining scales.

Two-way repeated measures analyses of variance were performed to examine possible interactions between sex of child and sex of observer on each of the six rating scales. The sum of ratings given to boys on a particular scale and the sum of ratings given to girls on that scale constituted the

two levels of the repeated-measures factor. The sex of the rater provided the two levels of the between-subjects factor. In the analysis of the aggression ratings, the main effect of sex of child approached significance, boys having been rated as more aggressive than girls, $F(1,38) = 3.19$, $p < .09$. The main effect of sex of subject was nonsignificant, $F(1,38) = .03$, n.s. The interaction between sex of subject and sex of child approached significance, $F(1,38) = 2.77$, $p < .11$. On none of the remaining five rating scales did either the main effects or the interaction of sex of child with sex of subject approach significance.

In order to evaluate the ability of PAQ and AWS scores to predict which subjects would rate children's aggression differently as a function of the child's sex, step-wise multiple regressions were performed. For male subjects, the PAQ Feminine scale entered first and correlated significantly with degree of bias, $R = -.50$, $p < .05$. The feminine scale thus explained approximately 25% of the variance in the bias scores. In contrast to Part A, it was primarily the men who described themselves as low in femininity who rated boys as more aggressive, while highly feminine men did not show bias consistently in either direction. The PAQ Masculine scale entered the regression second, showing a marginally significant tendency

for less masculine men to rate boys as more aggressive while more masculine men exhibited no consistent bias, simple $R = -.40$, $p < .09$. Together, the Feminine and Masculine scales of the PAQ explained 37% of the variance in bias, multiple $R = +.61$, $p < .05$. Thus it was seen that men who rated themselves as either low in masculinity and/or low in femininity were the subjects most likely to rate boys as more aggressive than girls, while males who described themselves as very masculine or very feminine showed no consistent bias. Inclusion of the PAQ Masculine-feminine scale and the AWS did not appreciably increase predictability.

For female subjects, none of the measures significantly predicted the degree of bias in the aggression ratings. The PAQ Feminine scale entered the regression first, but explained less than 10% of the variance, $R = +.31$, n.s.

To determine whether subjects who showed bias in Part A also showed bias in Part B, a point biserial correlation coefficient was calculated between subjects' bias score on Part B and the presence/absence of bias in their responses on Part A. The correlation was non-significant for subjects of both sexes. For females, $r_{pb} = +.23$; n.s.; for males, $r_{pb} = -.20$, n.s.

Although subjects' self-ratings of their degree of experience with children were collected, they were excluded from the analysis for two reasons. First, several subjects

overlooked the question during the testing session and had to be contacted by phone to obtain their responses, thus causing the data from these subjects to be collected under very different and nonanonymous conditions. Second, individual subjects seemed to base their ratings on very different criteria. For example, some subjects listed many capacities in which they had interacted extensively with children yet rated themselves as inexperienced, while other subjects--particularly females--appeared to rate themselves as quite experienced in spite of the fact that they could list very few actual sources of contact they had with children. The meaning and validity of the experience self-report was therefore felt to be questionable.

In summary, only male subjects were found to be significantly biased in their ratings of boys' and girls' aggression in Part B. It had been hypothesized that the aggression shown in Part B was so salient that subjects might be unable to overlook girls' aggression and hence might rate girls as more aggressive than boys. The data, however, showed bias to be in the same direction as in Part A, with more aggression being attributed to boys than to girls. It was found that the subjects who showed bias on Part A were not necessarily those who showed bias on Part B. On Part A, the subjects most likely to show bias were feminine males, while on Part B males low in femininity and/or low in masculinity were most likely to show bias.

Discussion

In the literature, it has generally been reported that while global ratings based on retrospective memory are often subject to bias in accordance with the rater's expectations, immediate observational coding of behavior tends to be more resilient against such bias (Kent, O'Leary, Diament, & Dietz, 1974; Shuller & McNamara, 1976). The finding that fully one quarter of both male and female subjects in the present study were biased in coding the occurrence of aggression in the playroom scenes in Part A is somewhat surprising. It is particularly striking in light of the fact that not one subject responded in the opposite direction by reporting more aggression in girls than in boys.

It had been hypothesized that, in Part A, subjects might see girls' pushing and kicking and relabel it as vigorous play or playful physical contact. Had subjects reported more aggression for boys while reporting more vigorous play or playful physical contact for girls, the hypothesis of relabeling would have been supported. However, subjects actually reported a greater amount of playful physical contact and vigorous play for boys than for girls, although this difference was not statistically significant. Thus, rather than seeing girls' pushing and kicking and relabeling it as play, it appears that subjects either did not notice the girls' behavior or

did not recall it seconds later when asked to record their observations. Since the aggressive behavior of the girls in these scenes is inconsistent with cultural sex-role norms, it would appear that in concordance with past research, expectation-consistent behaviors are more likely to be perceived and remembered than behaviors that are inconsistent with the observer's expectations (Darley & Fazio, 1980; Snyder & Frankel, 1976; Zadny & Gerard, 1974).

In the ratings of children's behavior in Part B, the overall level of bias found was less than in Part A. This was due to the fact that 14 subjects (nine women, five men) rated girls as more aggressive, thereby somewhat offsetting the responses of the 23 subjects (nine women, 14 men) who rated boys as more aggressive. Some subjects' ratings of boys and girls differed by only a small margin which could doubtless be attributed to error.

The finding that it was male raters who most consistently rated boys as more aggressive is concordant with the literature which shows that more men than women tend to be stereotyped in their perception of and response toward children (Biller, 1971; Langlois & Downs, 1980; Rubin, Provenzano, & Luria, 1974). It could be argued, on the basis of such findings, that bias could be minimized by using only female observers in future studies. As seen in Part A, however, female observers also show

a significant bias under some conditions.

It had been hypothesized that in Part B--where aggression was much more salient than in Part A and where subjects were required to focus on each child's level of aggression in order to rate it numerically--subjects would be less able to ignore or relabel girls' aggression and therefore might overcompensate by rating girls as more aggressive than boys. The bias shown in Part B did not support this hypothesis, however. Boys, were rated as more aggressive than girls even when the aggression was very salient.

It is not possible to specify in the present study why the bias found in Part A is stronger than that found in Part B. Different observational methods and recording procedures were used in the two conditions in order to make aggression relatively subtle in Part A and very salient in Part B. In addition, Part A was always administered prior to Part B to prevent the obvious focus on aggression in Part B from affecting subjects' responses to the aggression in Part A. Additional studies would be necessary to isolate which of the three variables--method of observation ("scanning" versus "focal child"), recording task (scoring occurrence versus rating degree of aggression), or presentation order--is primarily responsible for the greater bias found in Part A.

In light of the literature, it is not surprising that

the PAQ and especially the AWS were not stronger predictors of which subjects would show biased responses. Previous studies have shown that although subjects interact in sex-typed ways with children, verbal indices do not show evidence of sex-typing (Fagót, 1978; Seavey, Katz, & Zalk, 1975; Will, Self, & Datan, 1976). Frisch (1977) found that the AWS predicted sex bias only on selected behavioral categories. Will et. al. (1976) reported that subjects were unaware of their sex-biases so that an instrument such as the AWS, which requires rather blatantly sex-typed responses to yield a score categorized as traditional, may be unable to tap such biases to the extent that they actually manifest themselves in a person's behavior. The PAQ, being somewhat more subtle and focusing on whether or not the subject identifies with sex-typed qualities, appears somewhat more useful in predicting which subjects will exhibit sex-typed biases. The PAQ was unable to predict which females would show bias, but was able to identify the feminine men who tended to report more aggression for boys in Part A and was able to point to the two groups of men--men low in masculinity and/or low in femininity--who tended to rate boys as more aggressive than girls in Part B.

Attention should be drawn to one nonsignificant result obtained in the present study--the absence of a significant interaction between sex of subject and sex of child on

ratings of goodness of the child's behavior. Past research has been contradictory on this point. Some studies report that subjects rate opposite-sex children more favorably (Gurwitz & Dodge, 1975; Rothbart & Maccoby, 1966) while others report more favorable ratings for same-sex children (Meyer & Sobieszek, 1972). The present study supports neither of these positions.

Observers in the present study were neither pretrained nor supplied with explicit coding definitions for the behaviors they were asked to score. Two lines of reasoning led to the decision to test bias under these conditions. First, had observers been carefully trained and supplied with rigid definitions for the behavioral categories used, the probability of eliciting bias would most likely have been reduced. If no bias was found, all that could be concluded would be that, given intensive training and clear coding definitions for each behavior, observers do not show a significant sex bias in their responses. This is in fact what Horn and Haynes (1981) concluded from their study on sex-typed observer biases in which they gave observers several hours of supervised practice using explicit coding definitions and in which they disqualified from further testing the 10% of observers who failed to reach 80% agreement with the criterion scores used during training. Such a finding offers little information regarding the possibility of observer bias in

research using less stringent methodology. Second, in addition to determining the potential for observer bias in observational research, it is important to learn the extent to which untrained adults' observations may be biased in their daily encounters with children, since such biases might be important factors in the socialization of sex differences in children.

Having demonstrated that many observers do indeed respond in a sex-biased manner under at least some conditions, it is now important for future studies to determine the ability of specific training procedures to eliminate such biases. Possibly the bias is transient enough so that something as simple as instructing subjects to try to use the same standards in responding to the behavior of both sexes could alleviate subjects' tendencies to respond in a biased way. On the other hand, it may be found that nothing short of the extensive training procedures used by Horn and Haynes (1981) is powerful enough to overcome some observers' tendency to respond in a biased fashion. If the latter was found to be the case, it would cast serious doubt on much of the aggression literature since few of the older studies on which the literature relies heavily used training procedures as elaborate as those of Horn and Haynes.

Additionally, in attempting to determine the level of training necessary to guard against biased observations,

it is important to keep in mind that numerous studies have shown that even a high level of initial training is not sufficient to maintain a high level of reliability in either observational coding or in the use of behavioral rating scales (O'Leary & Kent, 1973; Reid, 1970; Romanczyk et. al., 1973; Skindrud, 1973; Weinrott, 1977). Without continued monitoring and feedback, observers' expectations increasingly influence their coding and ratings over time (Skindrud, 1973; Weinrott, 1977). Thus it is possible that bias may reappear even after extensive observer training.

There is a possibility that tasks similar to those used in the present study could be employed as a screening device for selecting observers. If it could be demonstrated that the subjects who show bias on an analogue task such as those used in the present research are the same observers who show bias in actual classroom observations, it would be possible to pretest potential observers quite easily to select those observers who are least prone to bias. In implementing such a screening procedure, it would be important that the screening task used approximate the conditions of the observational measure as closely as possible, since it appears that individual subjects may be prone to bias on some tasks but not on others.

In the larger context of the development of sex-typed behavior in children, the fact that a large subset

of the observers studied showed a significant bias in their responses to children's aggression under some conditions is very important. Rosenthal (1966, 1969) has repeatedly suggested that the expectations of an observer can significantly influence the behavior of the person being observed. Zanna and Pack (1975) demonstrated that female undergraduates' attitudes and achievement level varied as a function of the type of woman they were led to believe the male confederate in the experiment preferred. If university students adapt their behavior so significantly to conform to the supposed likings of a total stranger, we must be careful not to underestimate the role parents and other adults play in molding their children to fit sex-typed expectations regarding level of aggression.

In sum, although Eme (1979, p. 583) in his literature review of sex differences in psychopathology describes "the male preponderance in aggressive behavior" as "probably the most unequivocal sex difference in the literature", it appears that such a sex difference may be less solidly established than has been thought.

Reference Note

1. Spence, J. T. Traits, rôles and the concept of androgyny. Paper presented at the Conference on Perspectives on the Psychology of Women, Michigan State University, May 1977.

References

- Billler, H. B. Father, Child and Sex Role: Paternal Determinants of Personality Development. Lexington, Mass.: Heath Lexington, 1971.
- Condry, J., & Condry, S. Sex differences: A study in the eye of the beholder. Child Development, 1976, 47, 812-819.
- Darley, J. M., & Fazio, R. H. Expectancy confirmation processes arising in the social interaction sequence. American Psychologist, 1980, 35, 867-881.
- Durrett, M. E. The relationship of early infant regulation and later behavior in play interviews. Child Development, 1959, 30, 211-216.
- Eme, R. F. Sex differences in childhood psychopathology: A review. Psychological Bulletin, 1979, 86, 574-595.
- Fagot, B. I. The influence of sex of child on parental reactions to toddler children. Child Development, 1978, 49, 459-465.
- Frisch, H. L. Sex stereotypes in adult-infant play. Child Development, 1977, 48, 1671-1675.
- Gurwitz, S. B., & Dodge, K. A. Adults' evaluation of a child as a function of sex of adult and sex of child. Journal of Personality and Social Psychology, 1975, 32, 822-828.

- Hatfield, J. S., Ferguson, L. R., & Alpert, R. Mother-child interaction and the socialization process. Child Development, 1967, 38, 365-414.
- Hays, W. L. Statistics for the Social Sciences (2nd ed.). New York: Holt, Rinehart, and Winston, 1973.
- Horn, W. F., & Haynes, S. N. An investigation of sex bias in behavioral observations and ratings. Behavioral Assessment, 1981, 3, 173-183.
- Kent, R. N., O'Leary, K. D., Diament, C., & Dietz, A. Expectation biases in observational evaluation of therapeutic change. Journal of Consulting and Clinical Psychology, 1974, 42, 774-780.
- Langlois, J. H., & Downs, A. C. Mothers, fathers, and peers as socialization agents of sex-typed play behaviors in young children. Developmental Psychology, 1980, 51, 1237-1247.
- Levin, H., & Sears, R. R. Identification with parents as a determinant of doll play aggression. Child Development, 1956, 27, 135-153.
- Maccoby, E. E., & Jacklin, C. N. The Psychology of Sex Differences. Stanford, Cal.: Stanford University, 1974.
- Meyer, J. W., & Sobieszek, B. I. Effects of a child's sex on adult interpretations of its behavior. Developmental Psychology, 1972, 6, 42-48.

- O'Leary, K. D., & Kent, R. Behavior modification for school action: Research tactics and problems. In L. A. Hamerlynck, L. C. Handy & E. J. Mash (Eds.), Behavior Change: Methodology, Concepts and Practice. Champaign, Ill.: Research Press, 1973.
- Reid, J. B. Reliability assessment of observation data: A possible methodological problem. Child Development, 1970, 41, 1143-1150.
- Romanczyk, R. G., Kent, R. N., Diament, C., & O'Leary, K. D. Measuring the reliability of observational data: A reactive process. Journal of Applied Behavior Analysis, 1973, 6, 175-184.
- Rosenthal, R. Experimenter Effects in Behavioral Research. New York: Appleton-Century-Crofts, 1966.
- Rosenthal, R. Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal & R. L. Rosnow (Eds.), Artifact in Behavioral Research. New York: Academic Press, 1969.
- Rothbart, M. K., & Maccoby, E. E. Parents' differential reactions to sons and daughters. Journal of Personality and Social Psychology, 1966, 4, 237-243.
- Rubin, J., Provenzano, F., & Luria, Z. The eye of the beholder: Parents' views on sex of newborns. American Journal of Orthopsychiatry, 1974, 44, 512-519.

- Sears, P. S. Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence. Psychological Monographs, 1951, 65, No. 6.
- Seavey, C. A., Katz, P. A., & Zalk, S. R. Baby X: The effect of gender labels on adult responses to infants. Sex Roles, 1975, 1, 103-110.
- Serbin, L. A., O'Leary, K. D., Kent, R. N., & Tonick, I. J. A comparison of teacher response to the preacademic and problem behavior of boys and girls. Child Development, 1973, 44, 796-804.
- Shuller, D. Y., & MacNamara, J. R. Expectancy factors in behavioral observations. Behavior Therapy, 1976, 7, 519-527.
- Skindrud, K. Field evaluation of observer bias under overt and covert monitoring. In L. A. Hamerlynck, L. C. Handy & E. J. Mash (Eds.), Behavior Change: Methodology, Concepts and Practice. Champaign, Ill.: Research Press, 1973.
- Snyder, M. L., & Frankel, A. Observer bias: A stringent test of behavior engulfing the field. Journal of Personality and Social Psychology, 1976, 34, 857-864.
- Sobieszek, B. I. Adult interpretations of child behavior. Sex Roles, 1978, 4, 579-588.
- Spence, J. T., & Helmreich, R. L. Masculinity and Femininity: Their Psychological Dimensions, Correlates and Antecedents. Austin, Texas: University of Texas Press, 1978.

Weinrott, M. R. Improving the validity of global ratings.

Journal of Abnormal Child Psychology, 1977, 5, 187-197.

Will, J., Self, P., & Datan, N. Maternal behavior and
perceived sex of infant. American Journal of

Orthopsychiatry, 1976, 46, 135-139.

Zadny, J., & Gerard, H. B. Attributed intentions of

informational selectivity. Journal of Experimental

Social Psychology, 1974, 10, 34-52.

Zanna, M. P., & Pack, S. J. On the self-fulfilling nature
of apparent sex differences in behavior. Journal of

Experimental Social Psychology, 1975, 11, 583-591.

Appendix A

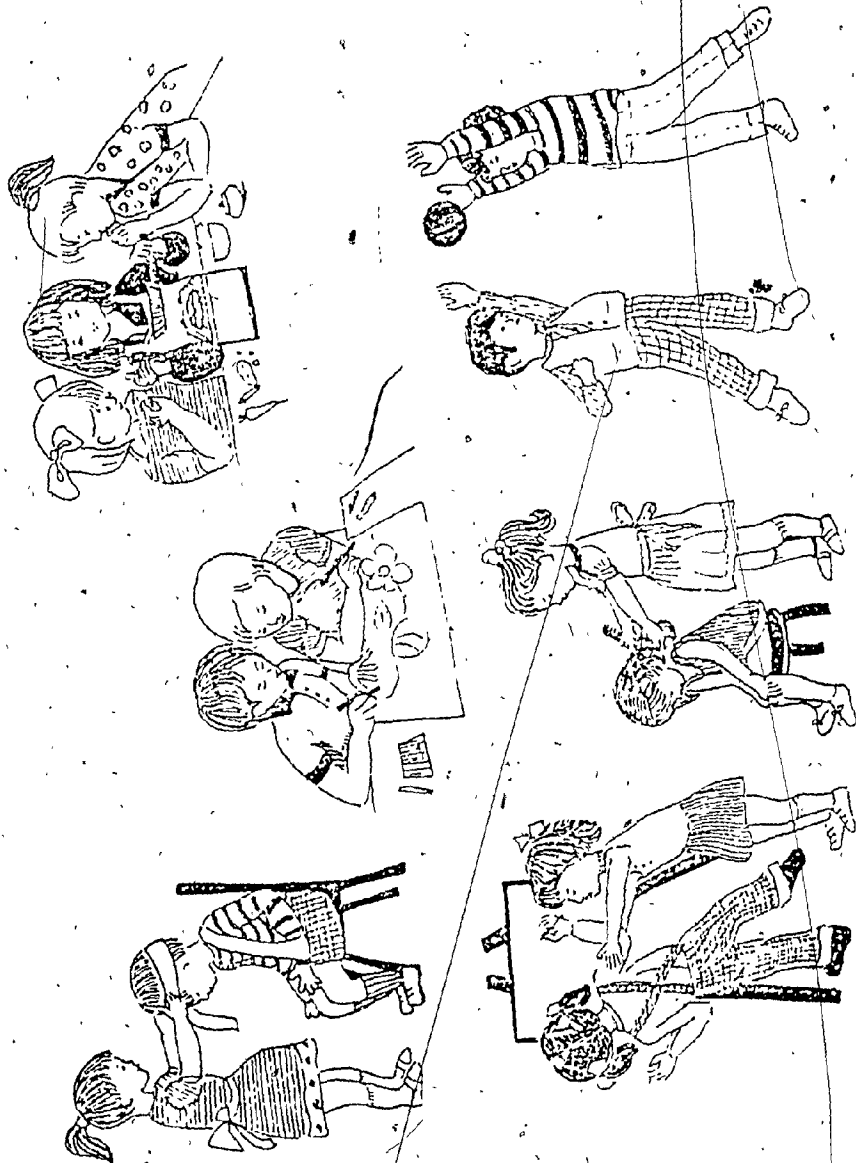
Drawings Used in Part A.



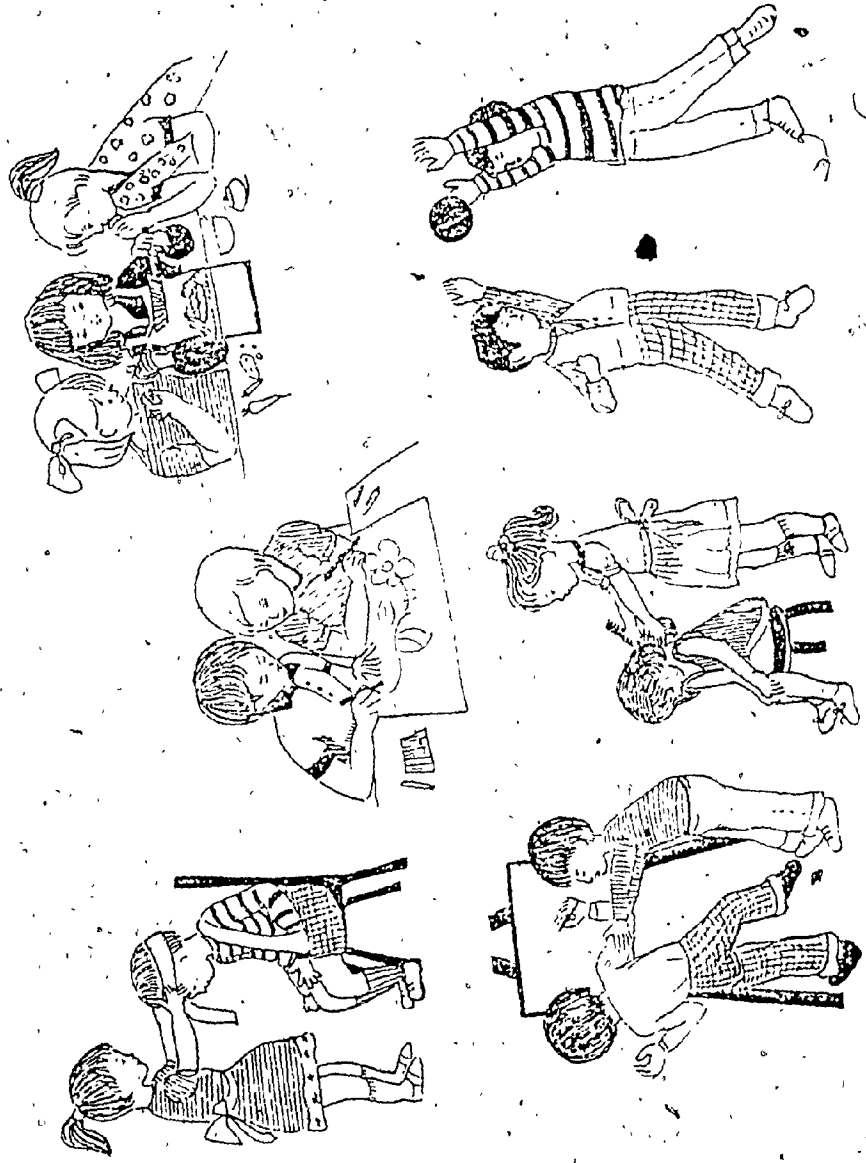
Scené A, Girls



Scene A, Boys



Scene B, Girls

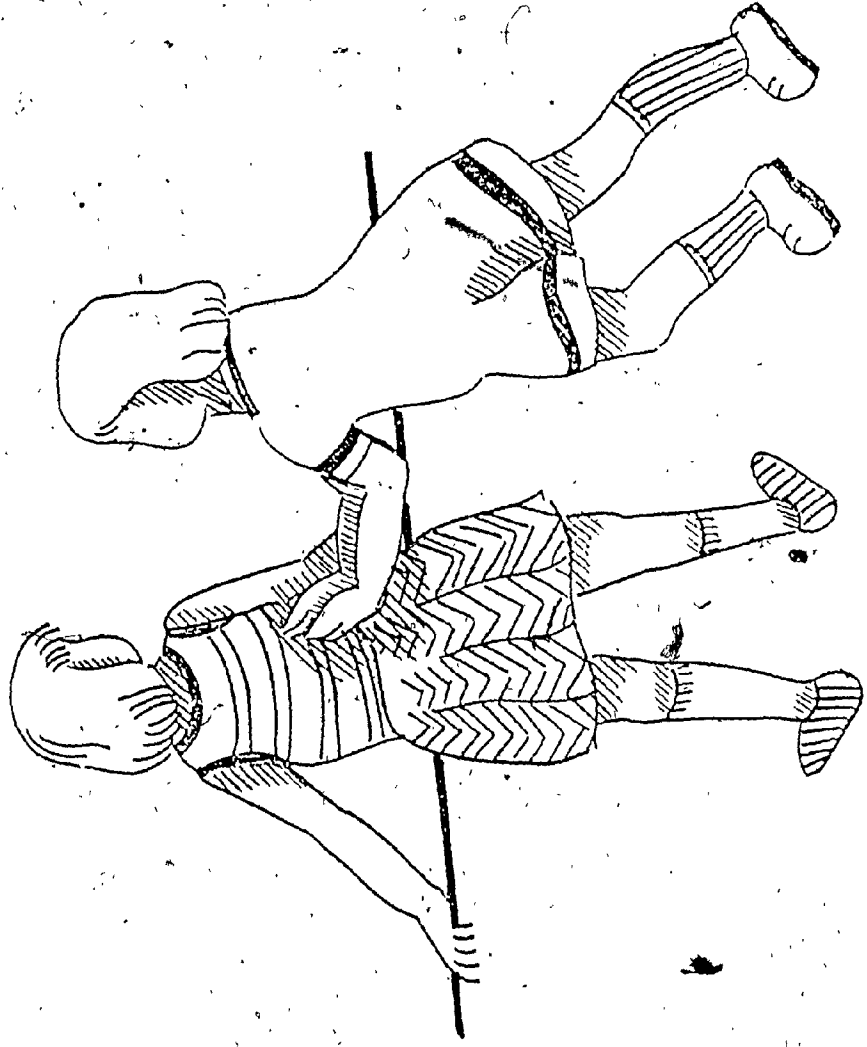


Scene B, Boys

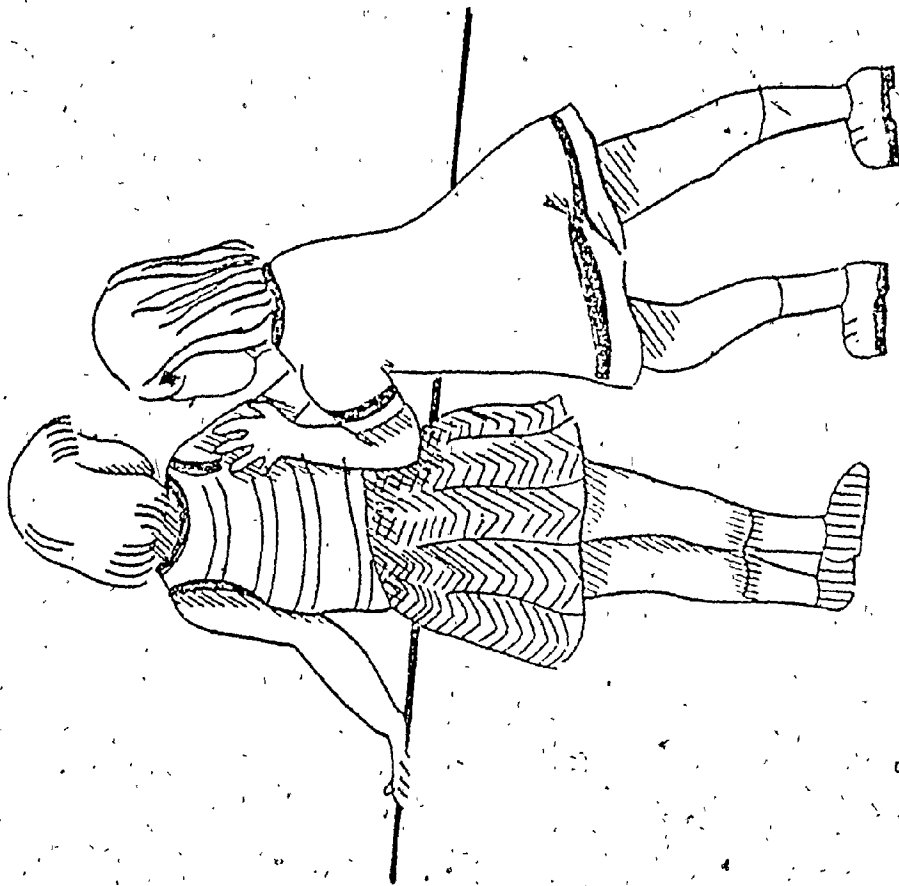
6

Appendix B

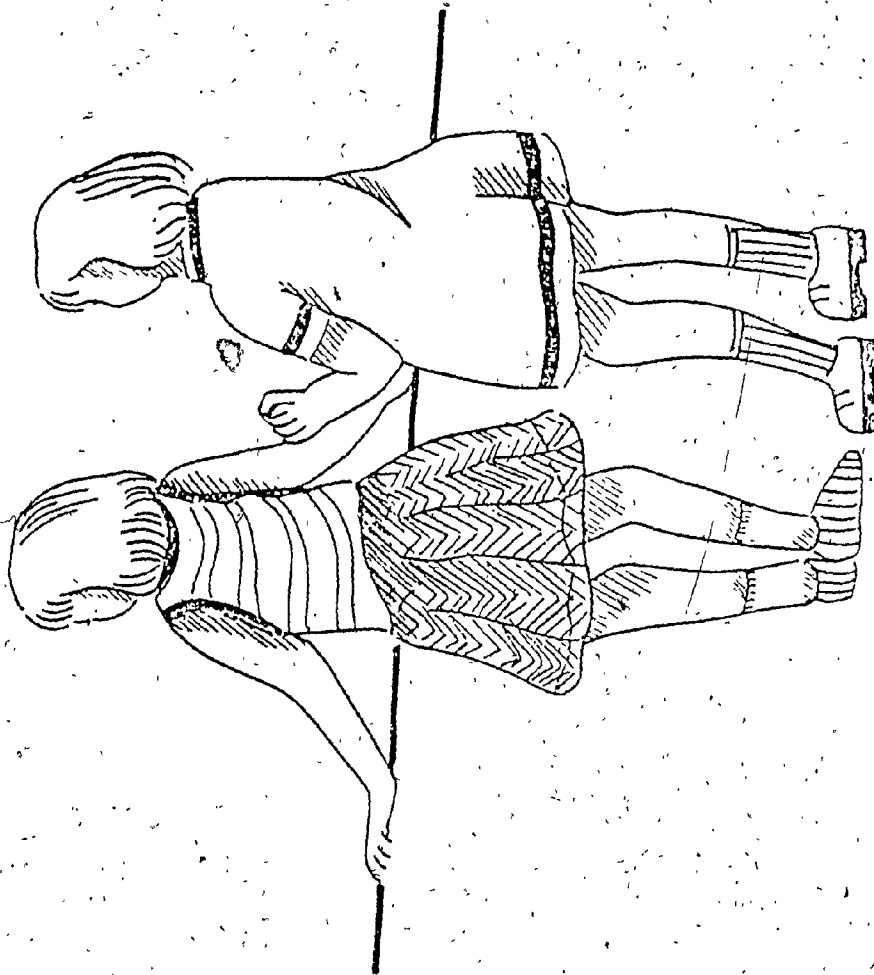
Drawings Used in Part B



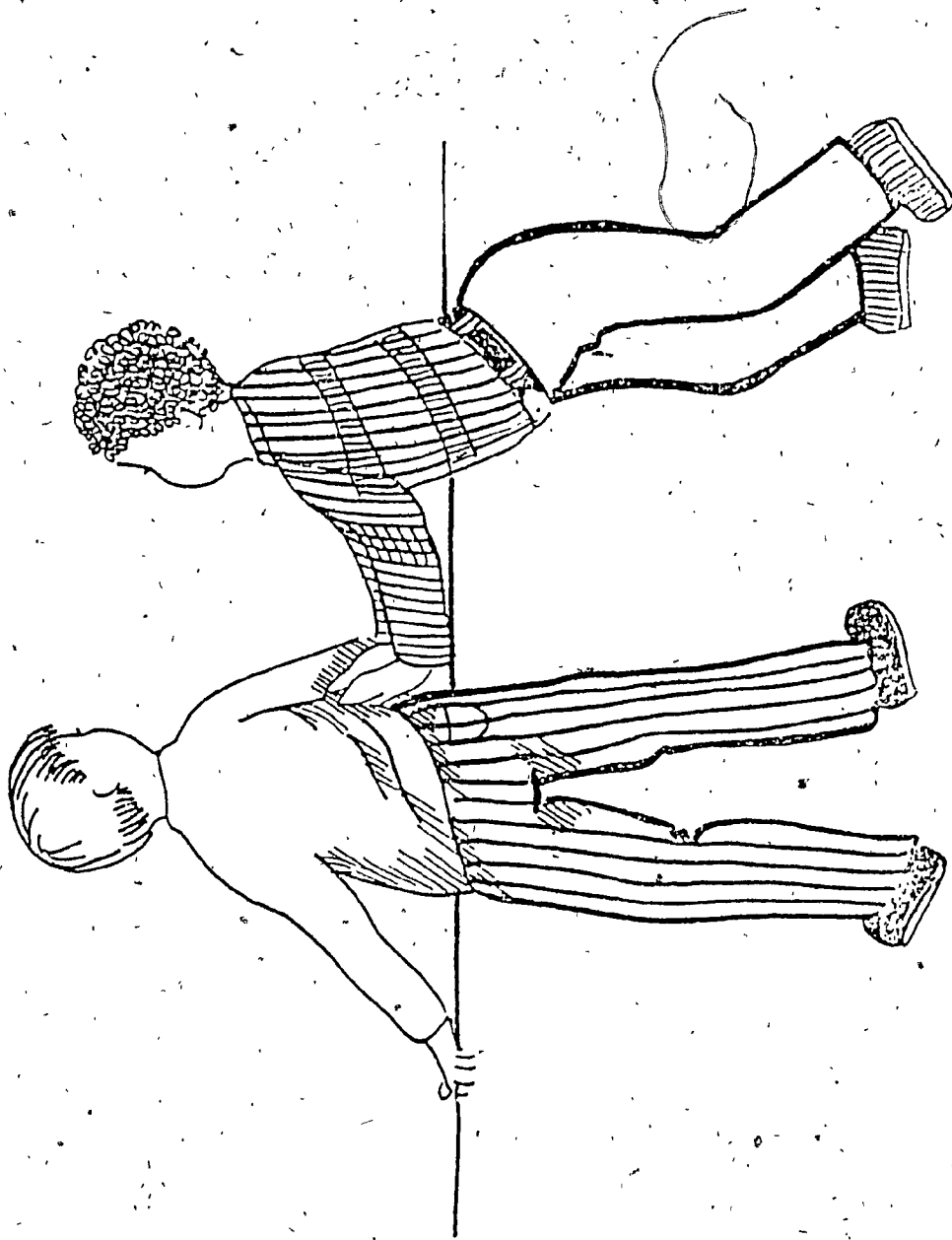
Sequence A, First Drawing, Girls



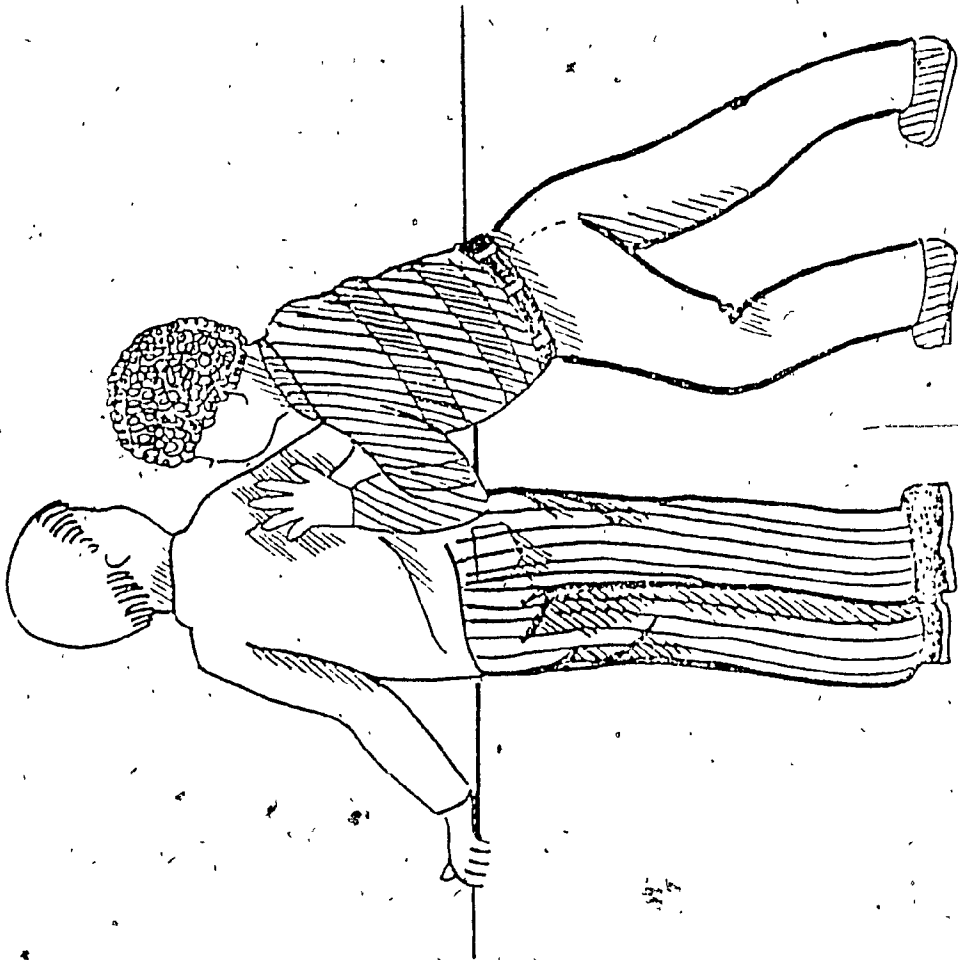
Sequence A, Second Drawing, Girls.



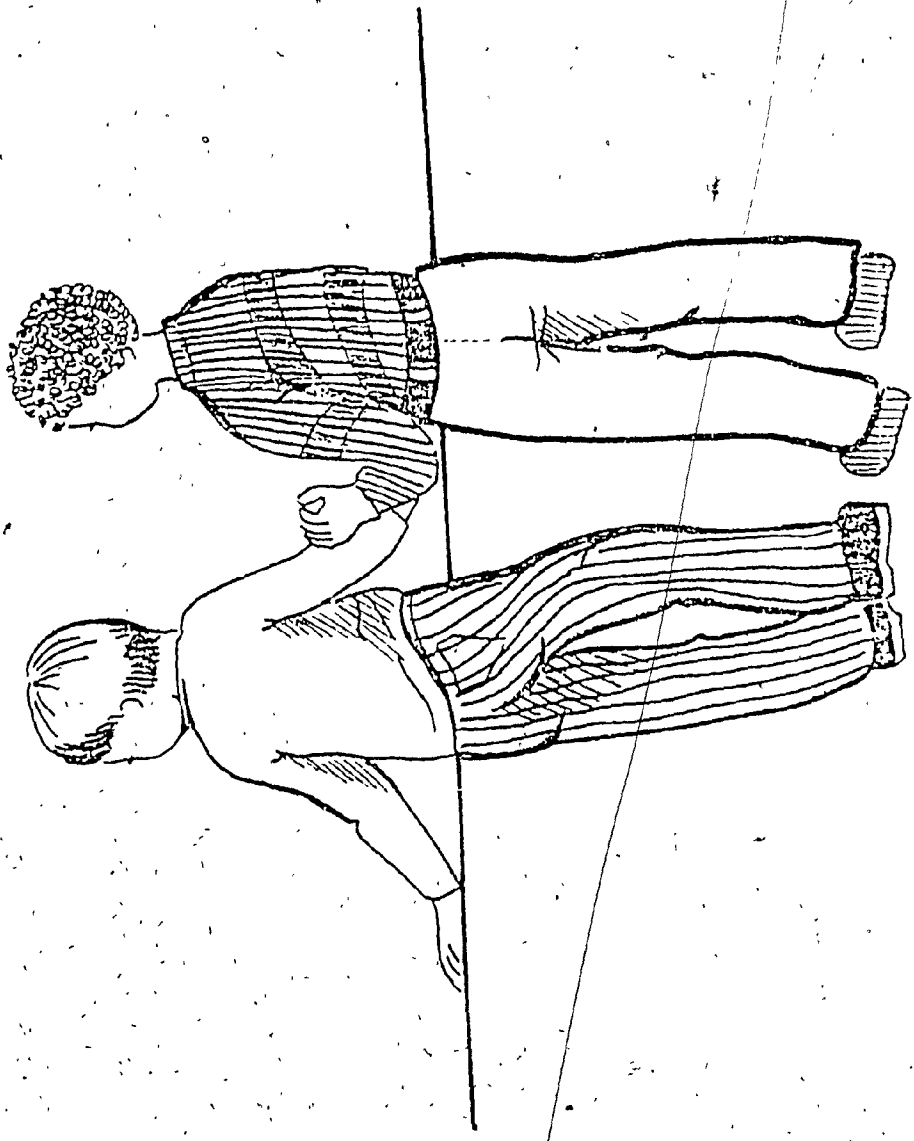
Sequence A, Third Drawing, Girls



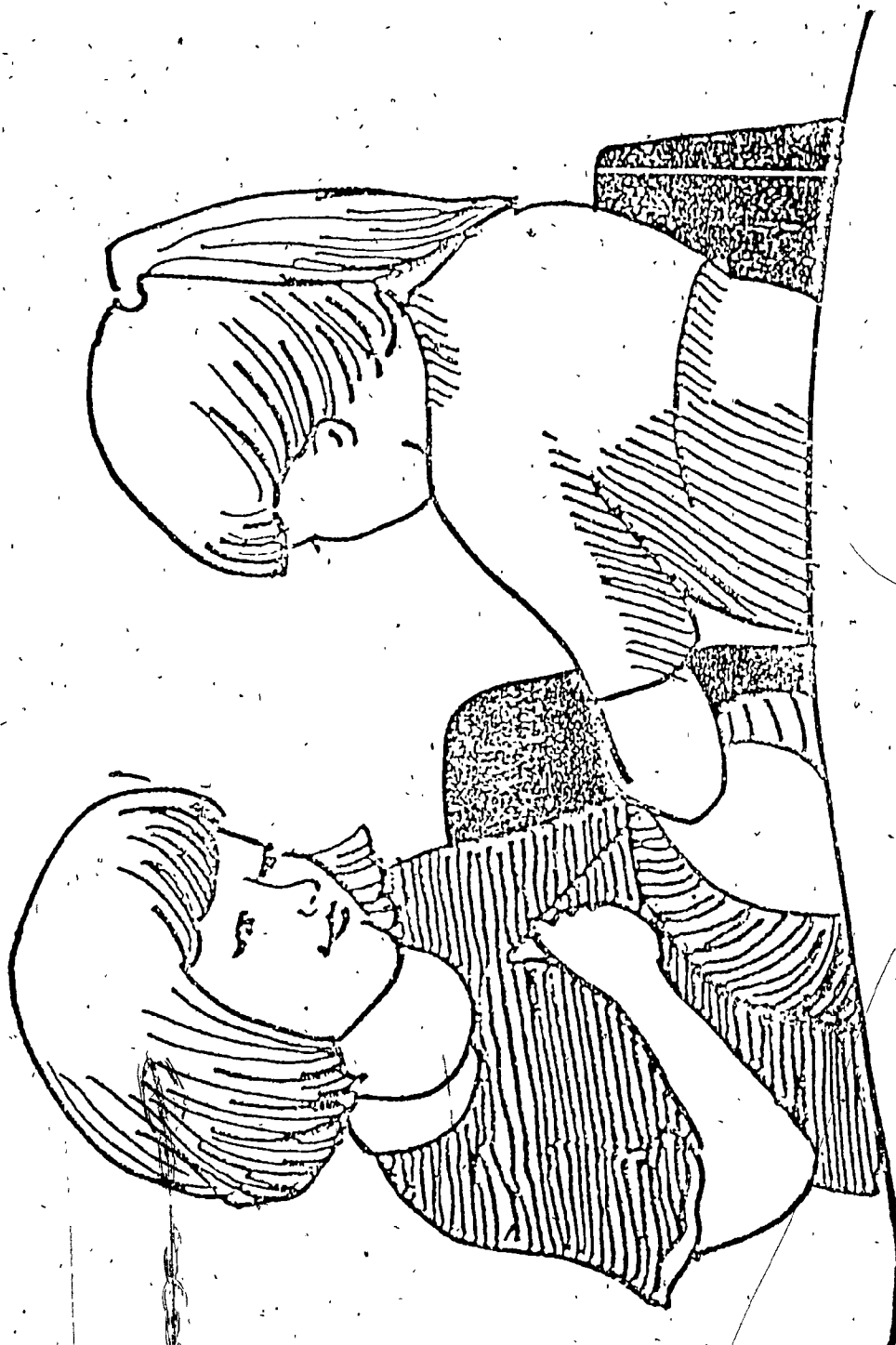
Sequence A, First Drawing, Boys



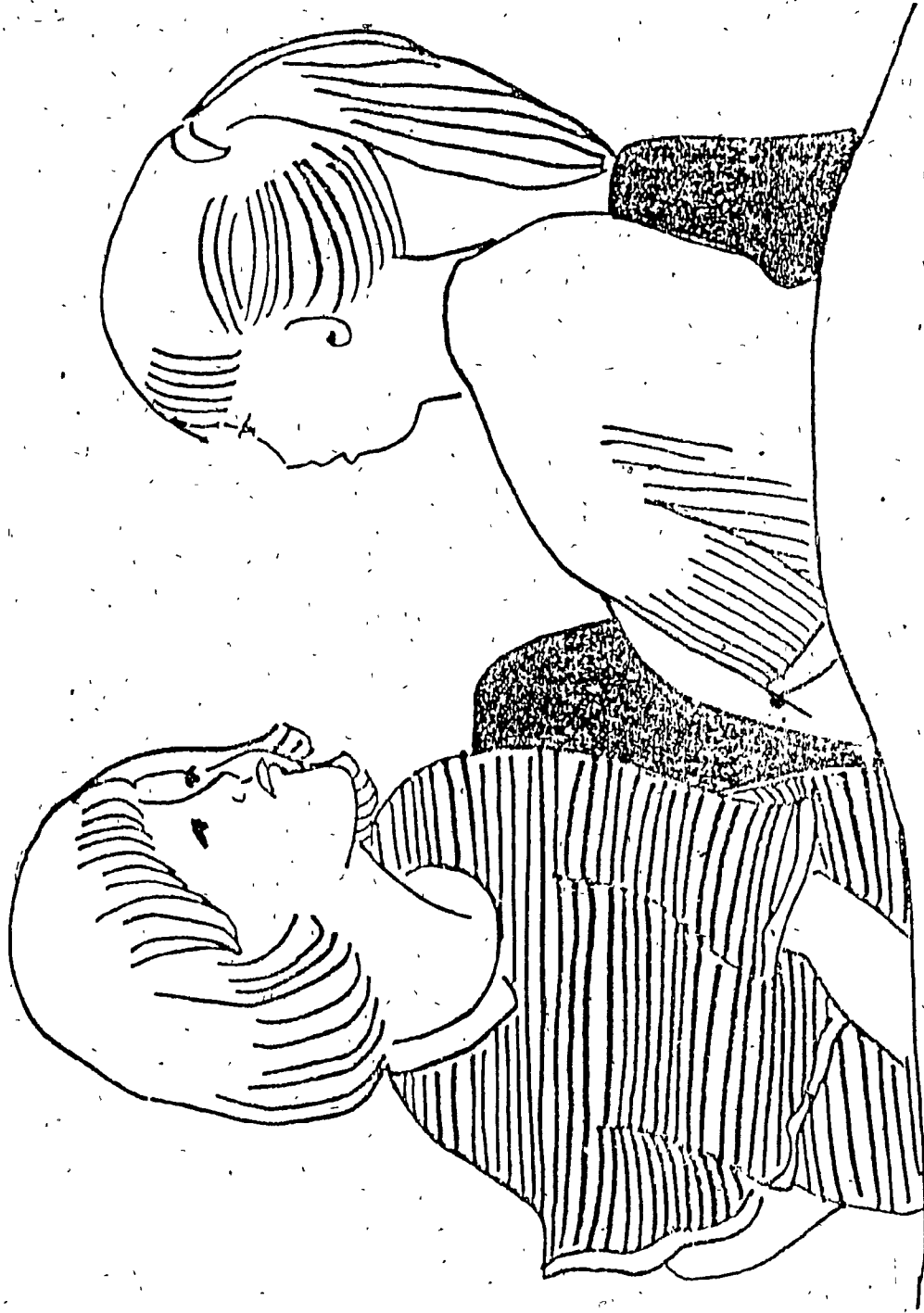
Sequence A, Second Drawing, Boys



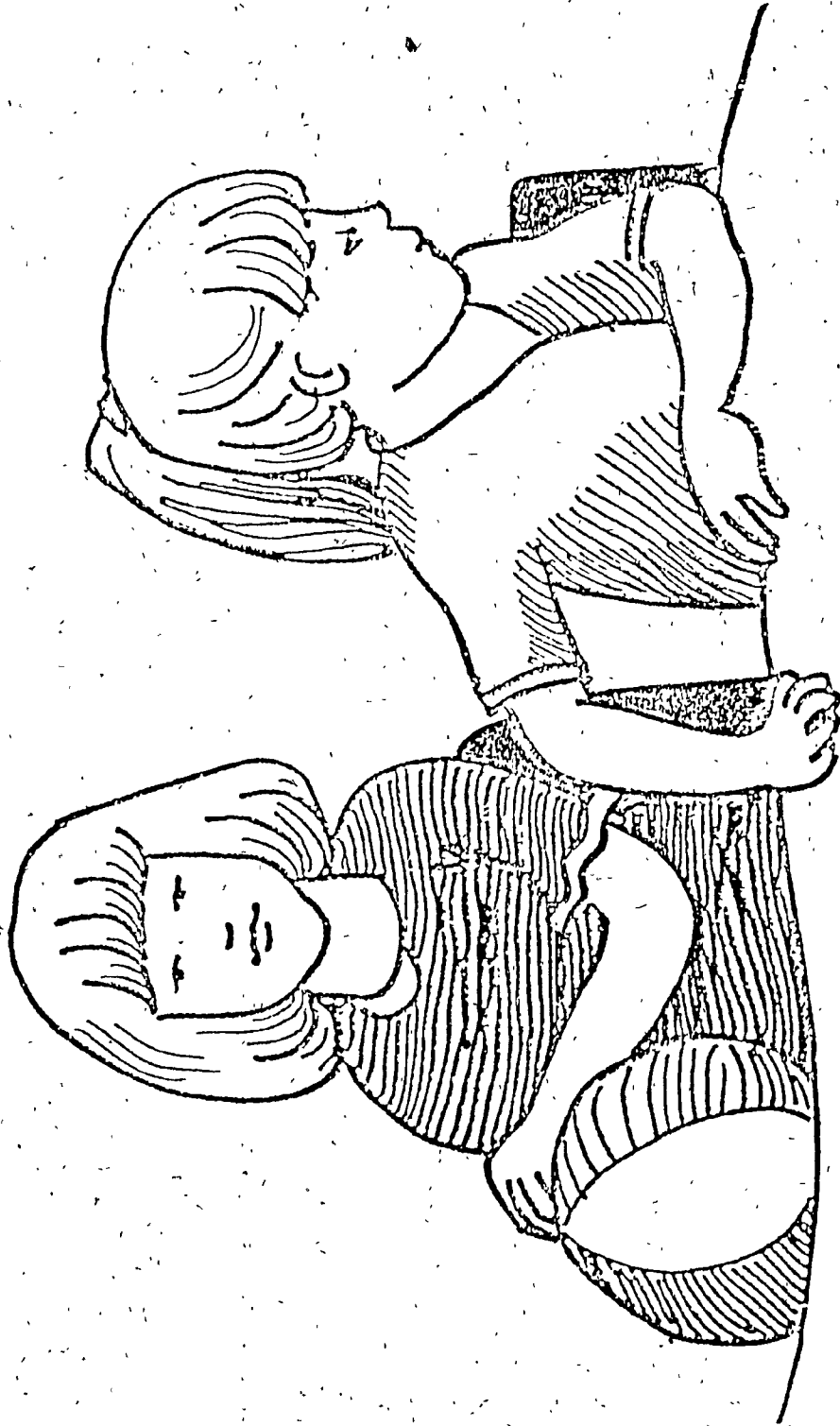
Sequence A, Third Drawing, Boys



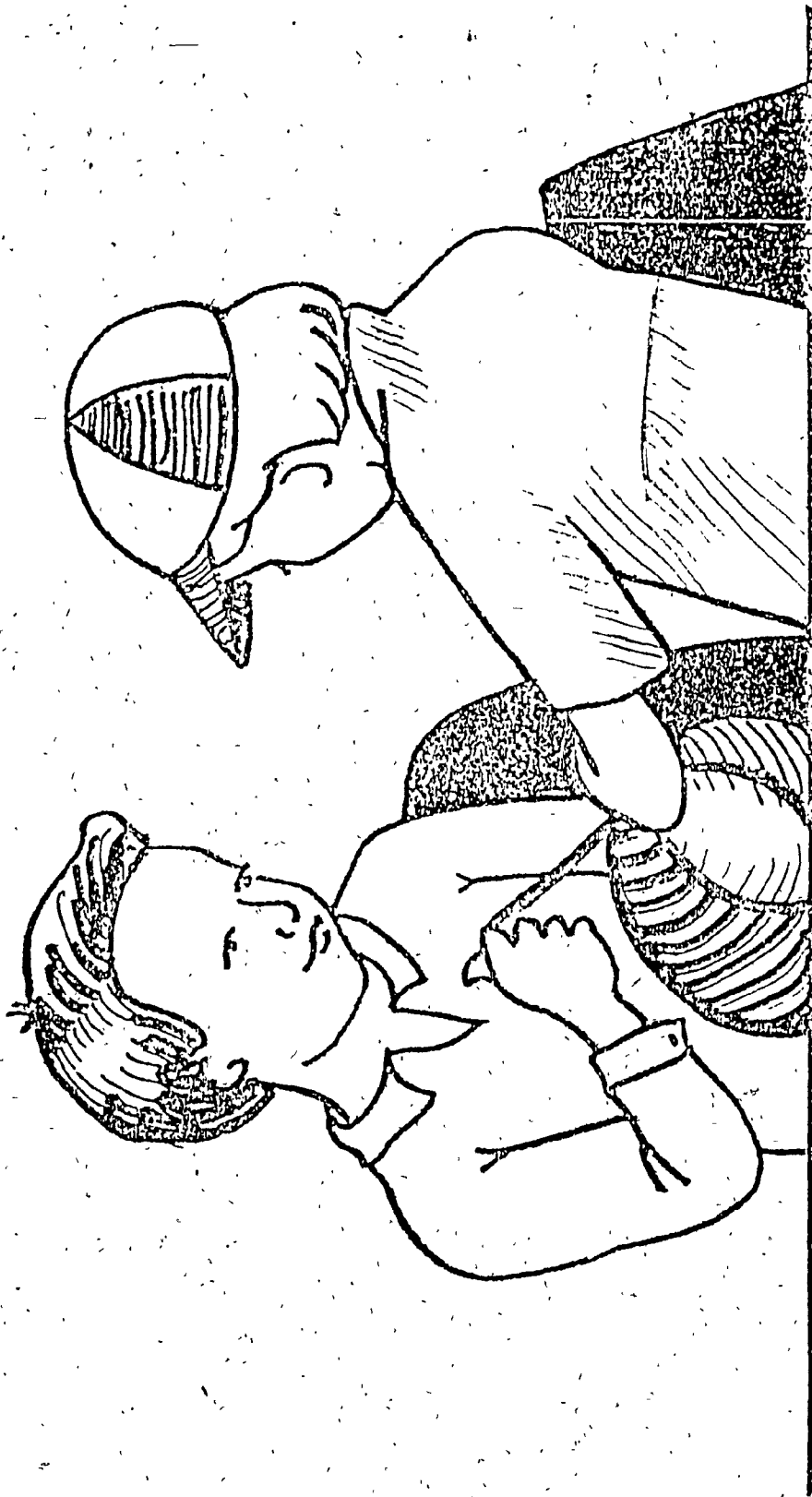
Sequence B, First Drawing, Girls



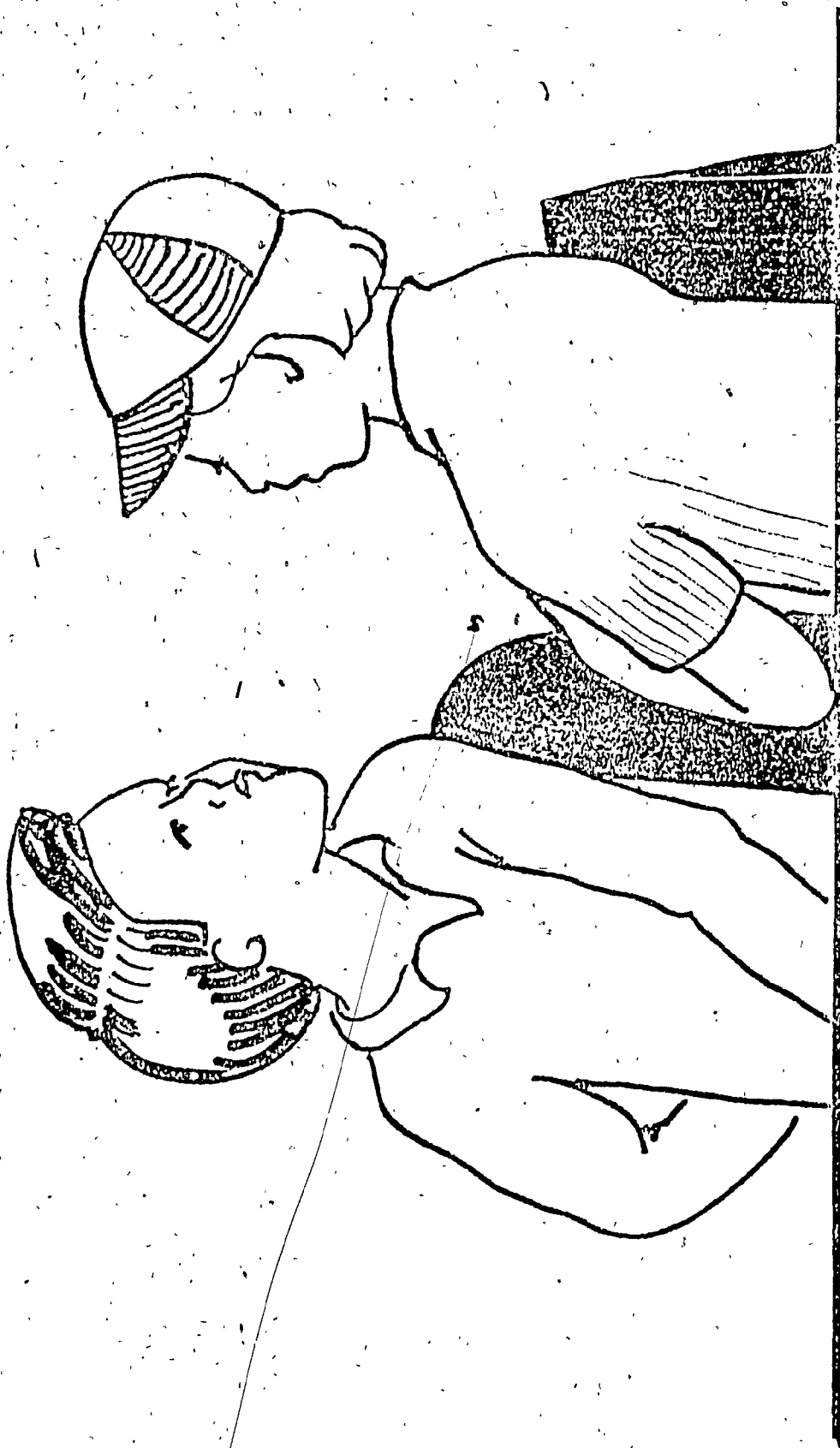
Sequence B, Second Drawing, Girls



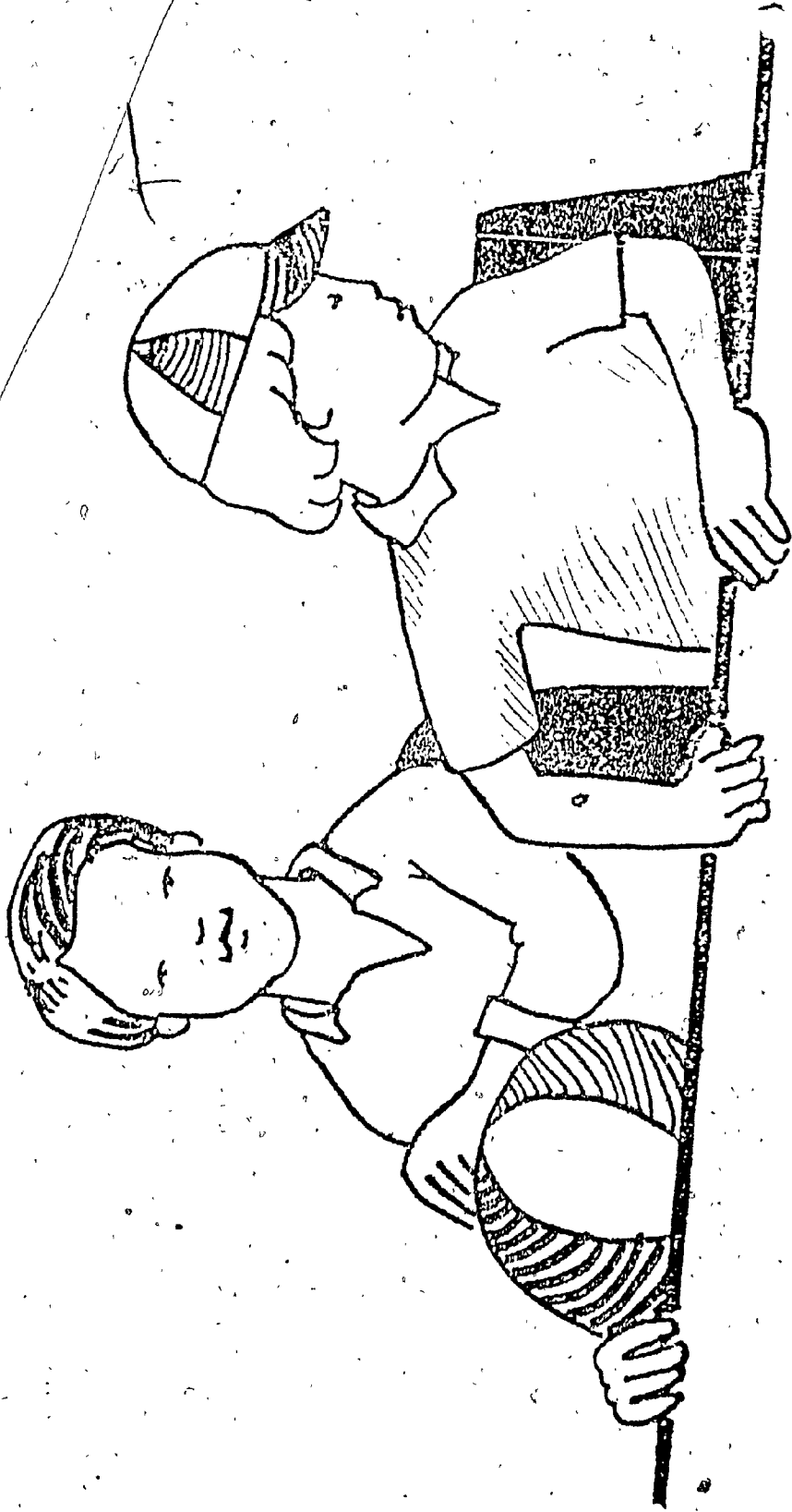
Sequence B, Third Drawing, Girls



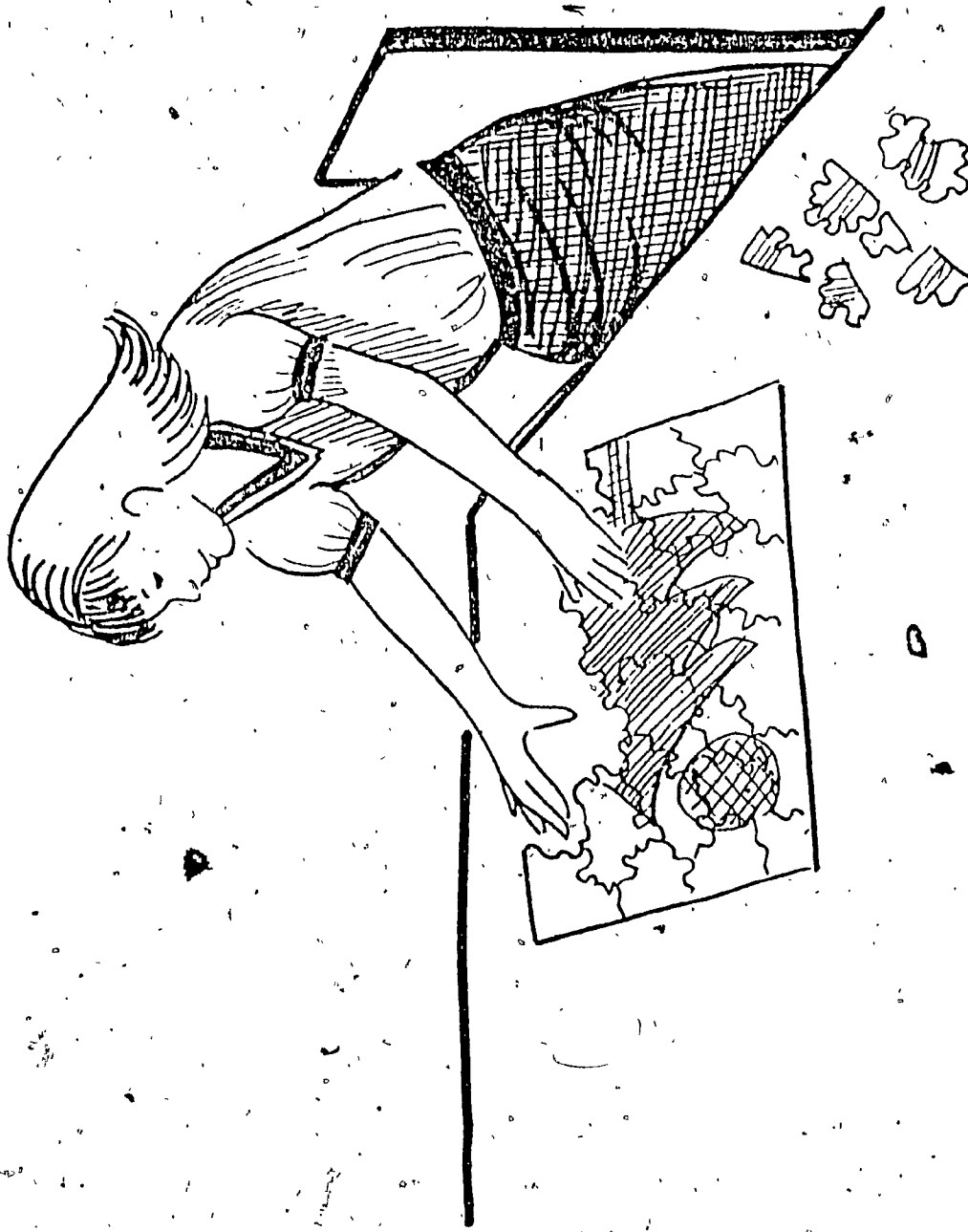
Sequence B, First Drawing, Boys



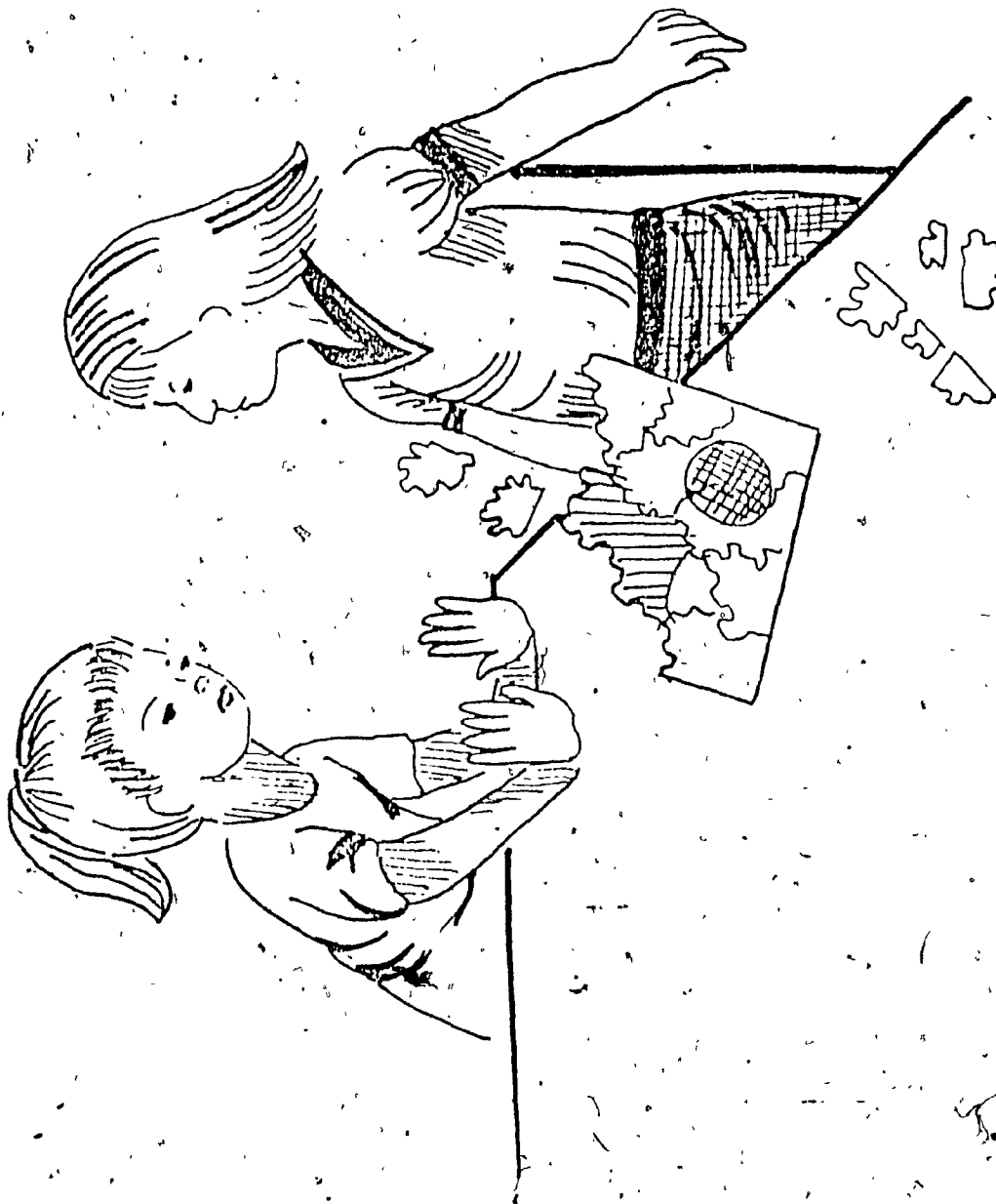
Sequence B, Second Drawing, Boys



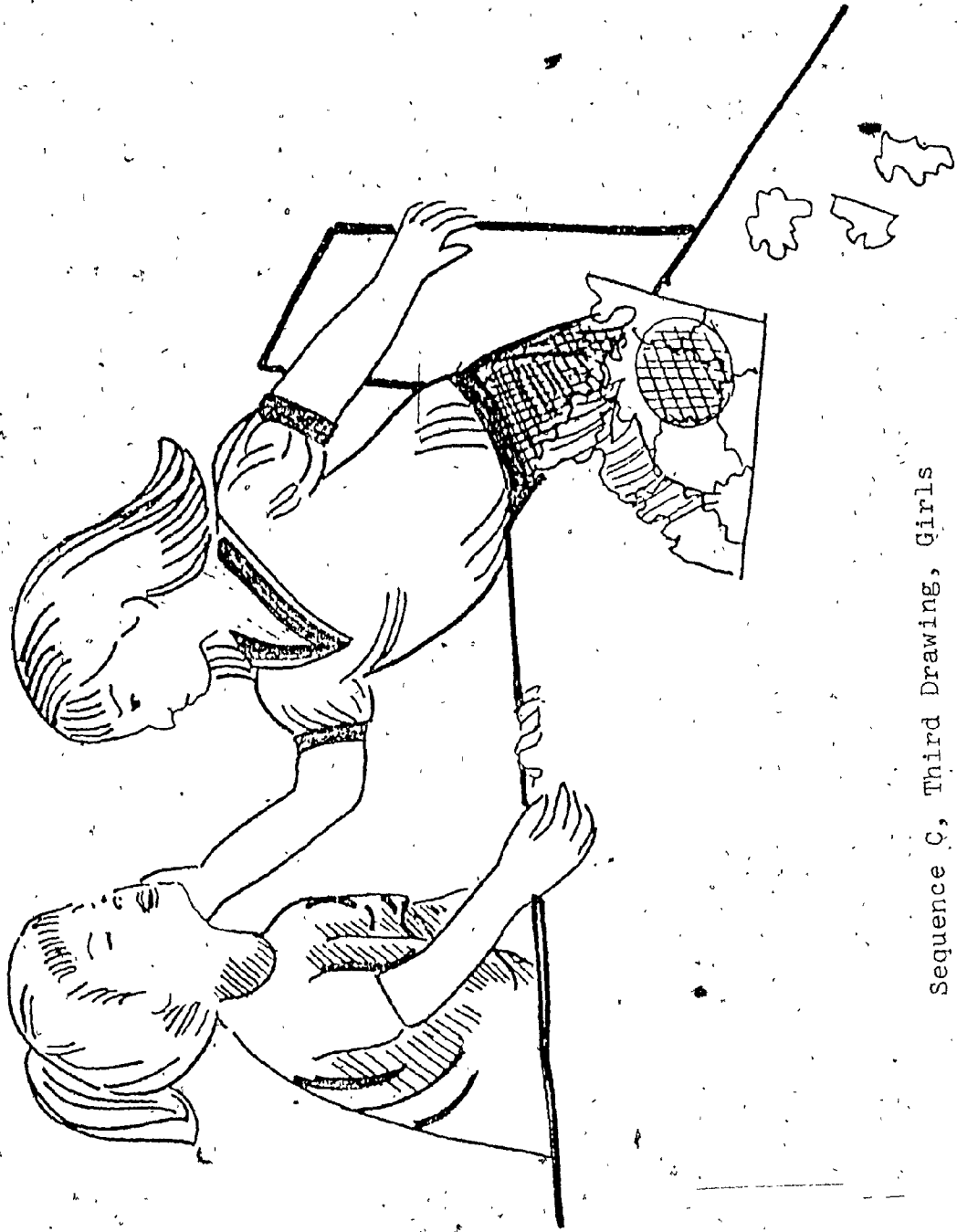
Sequence B, Third Drawing, Boys



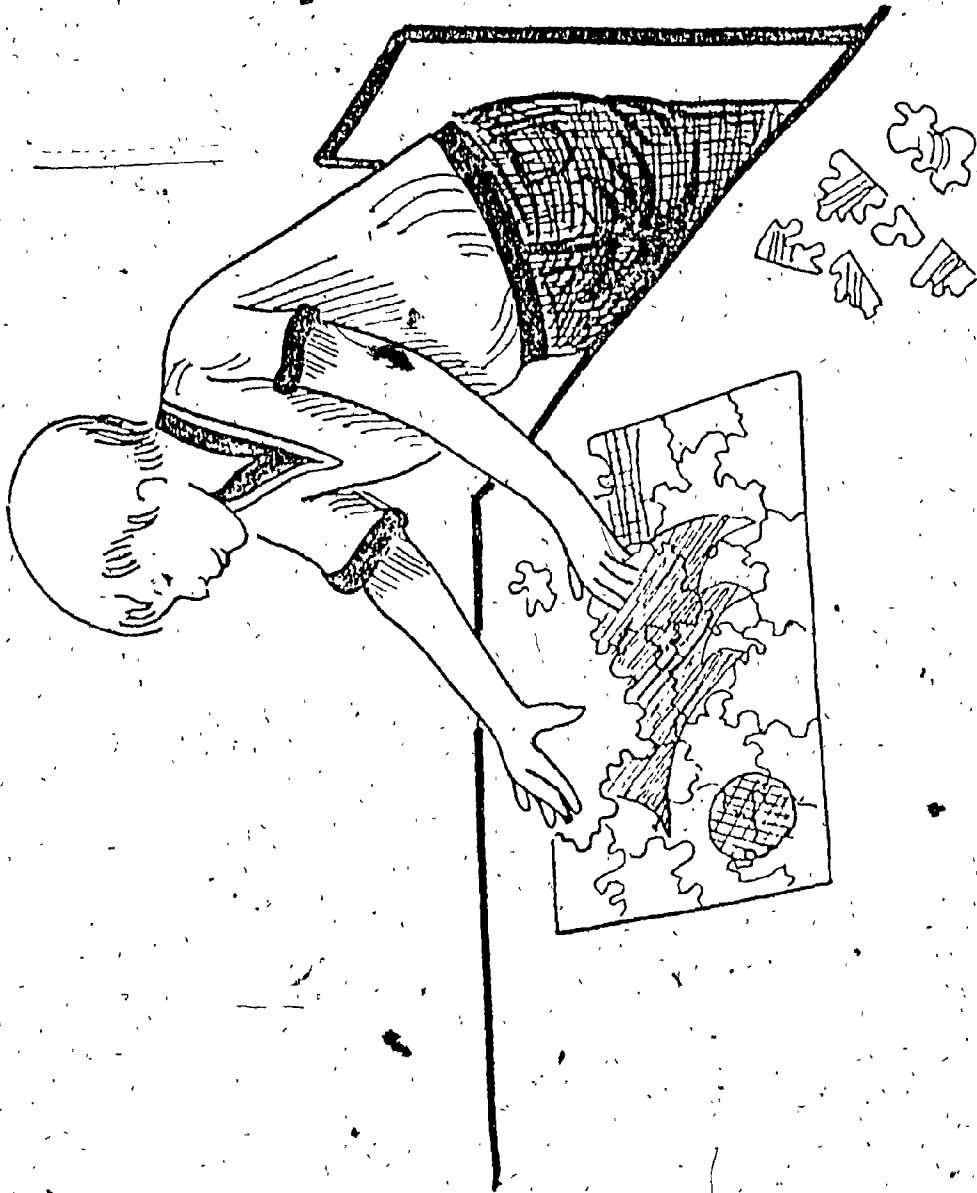
Sequence C, First Drawing, Girls



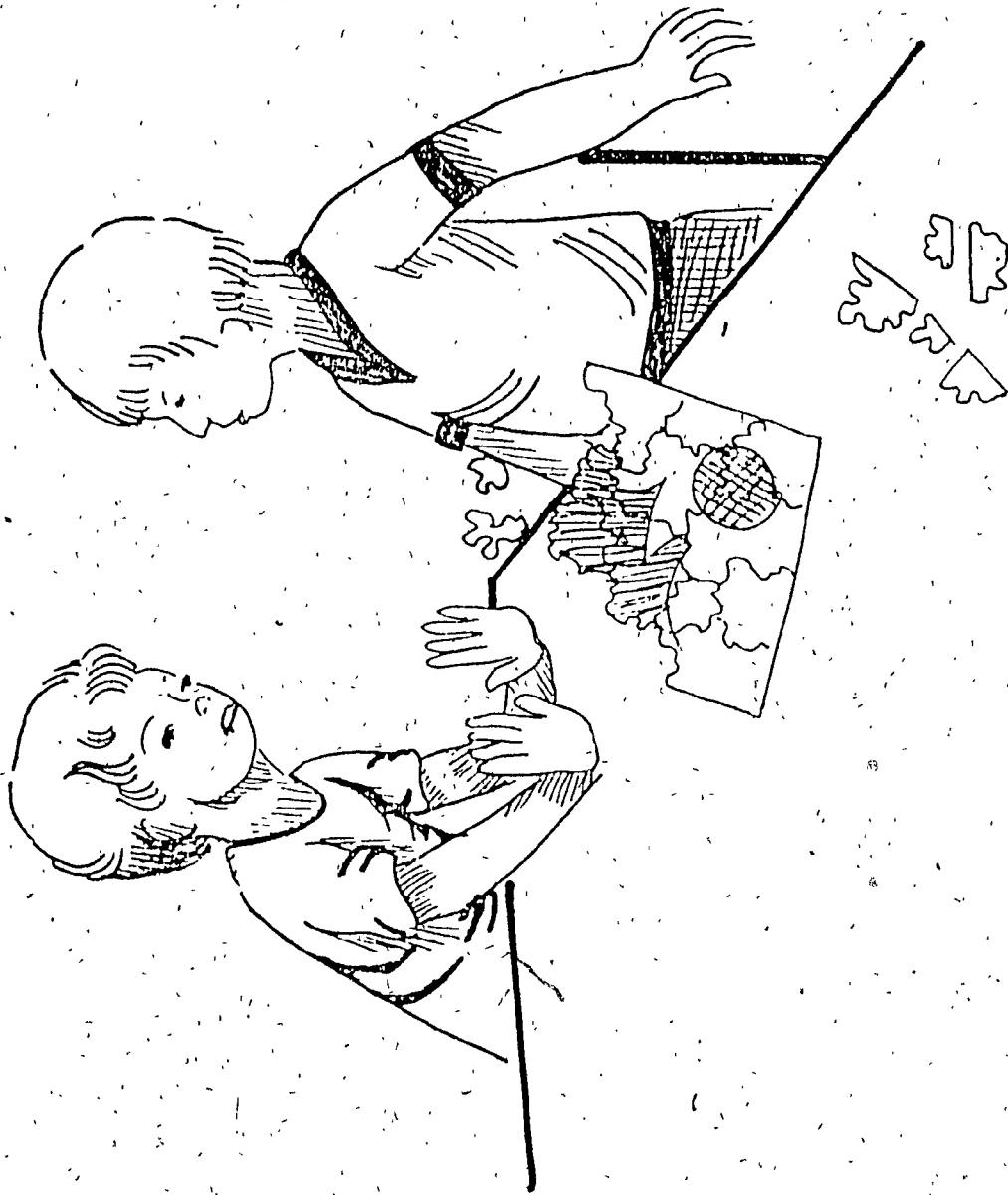
Sequence C, Second Drawing, Girls



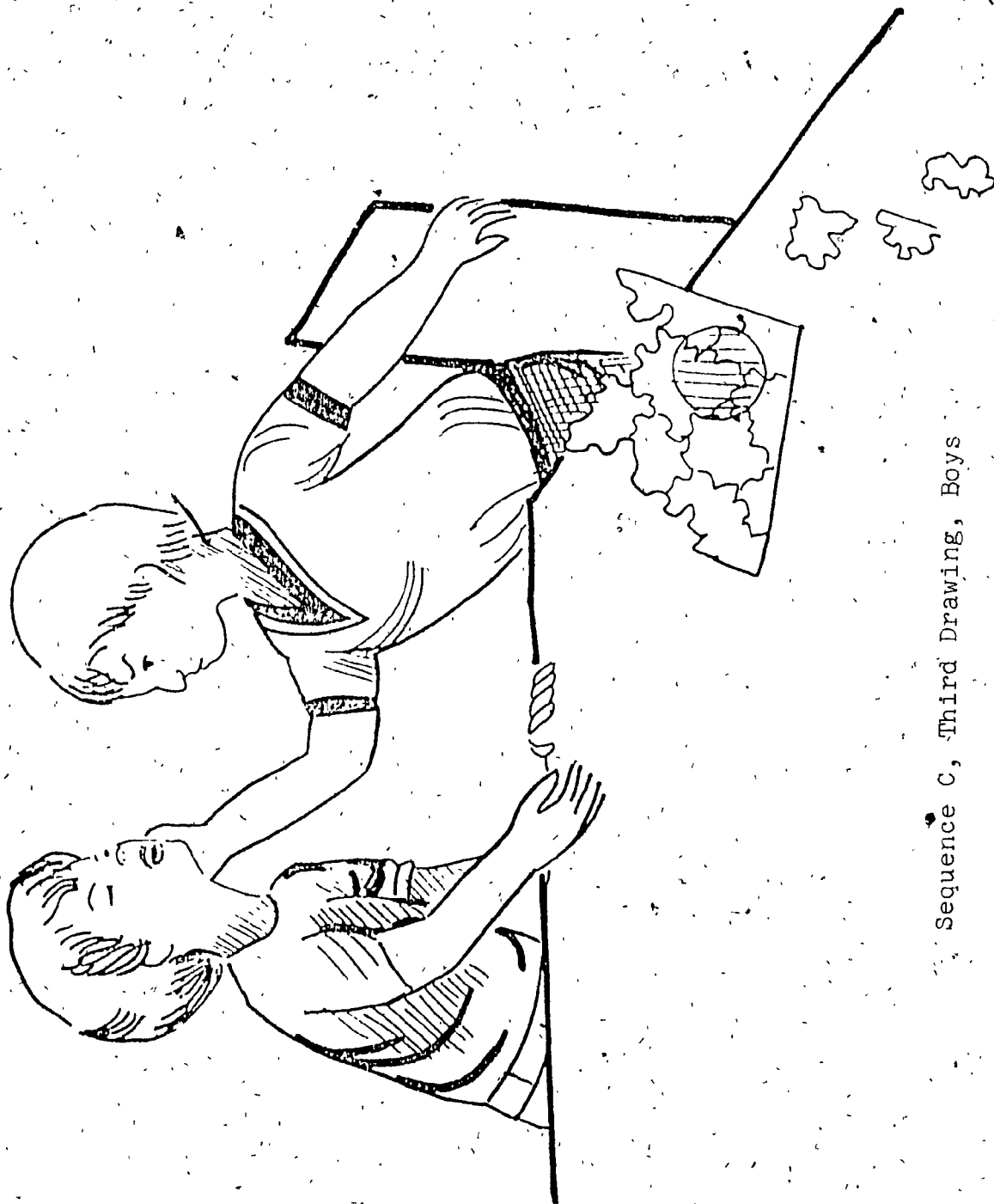
Sequence C, Third Drawing, Girls



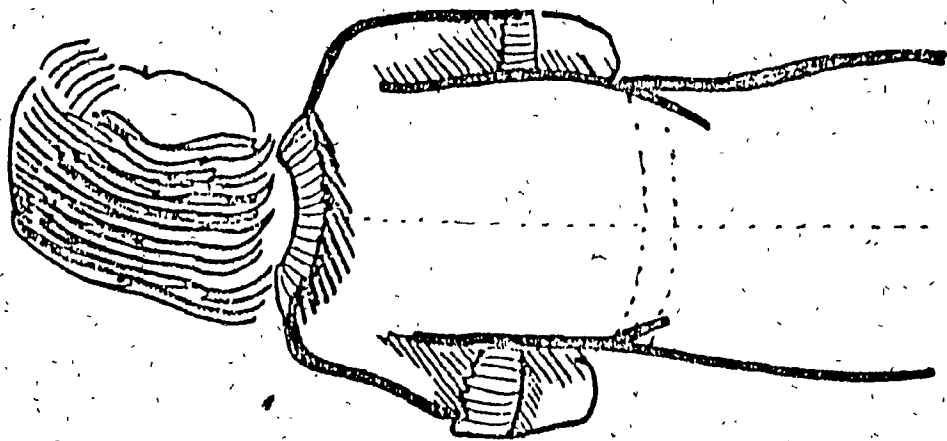
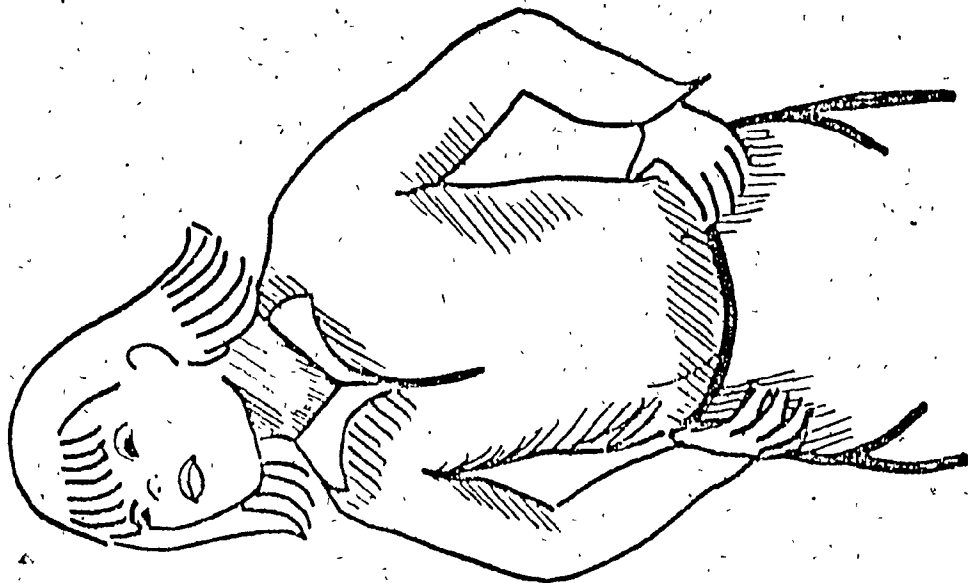
Sequence C, First Drawing, Boys



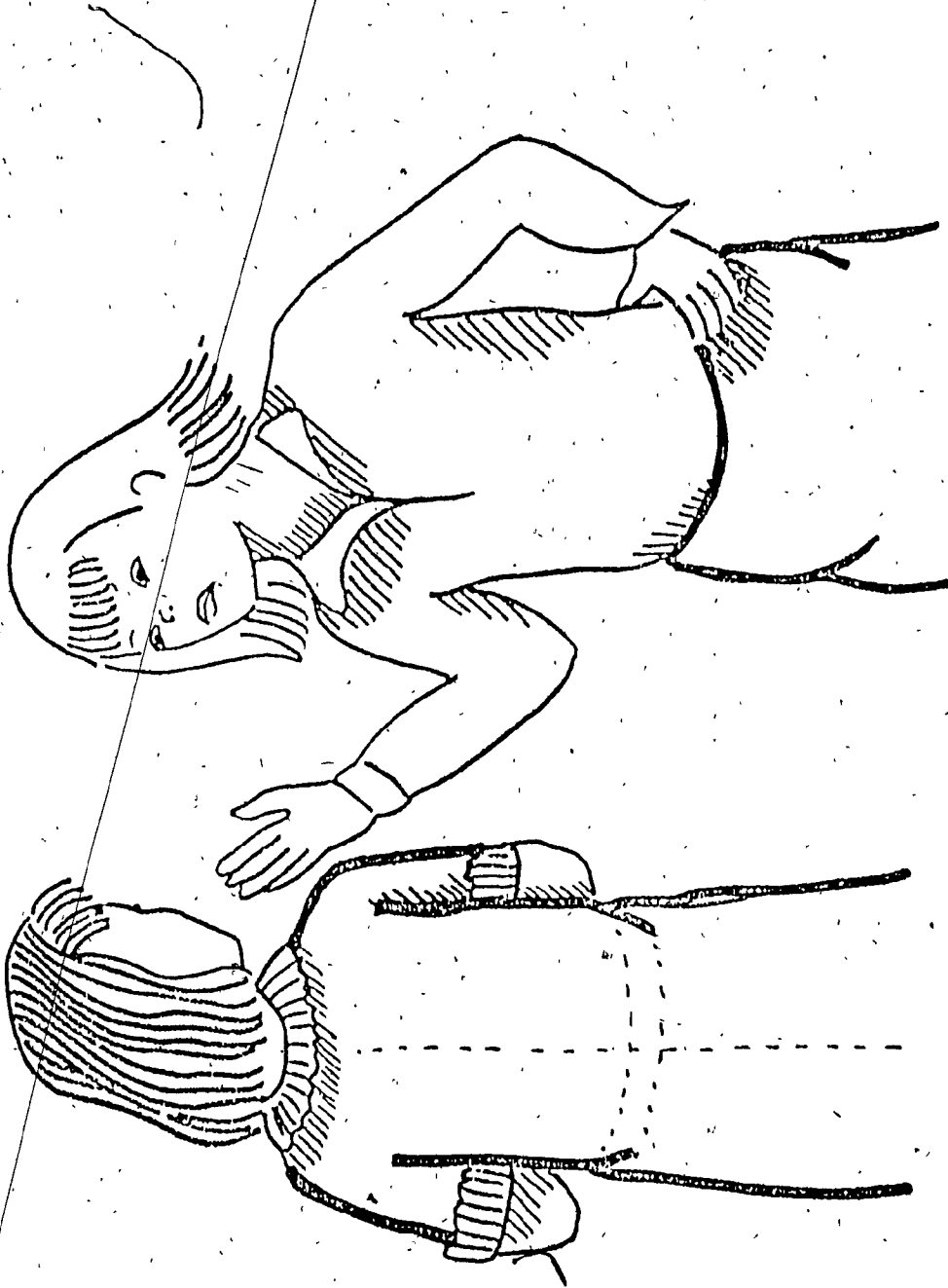
Sequence C, Second Drawing, Boys



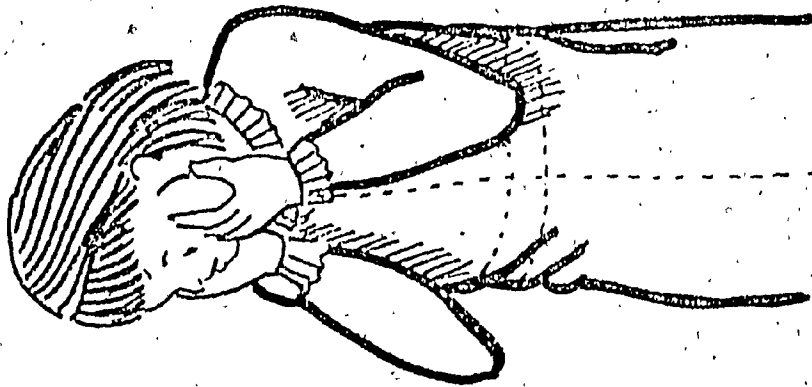
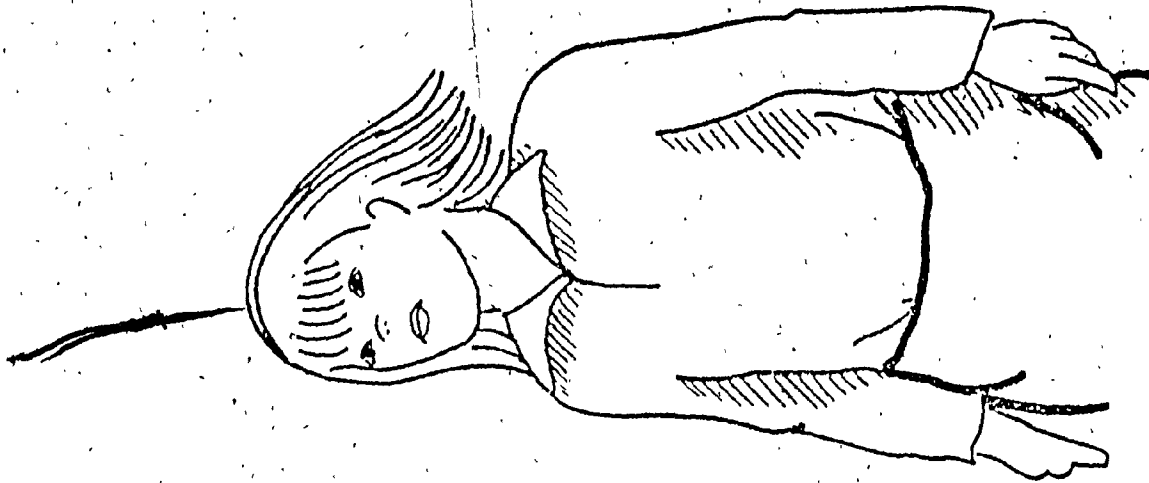
Sequence C, Third Drawing, Boys



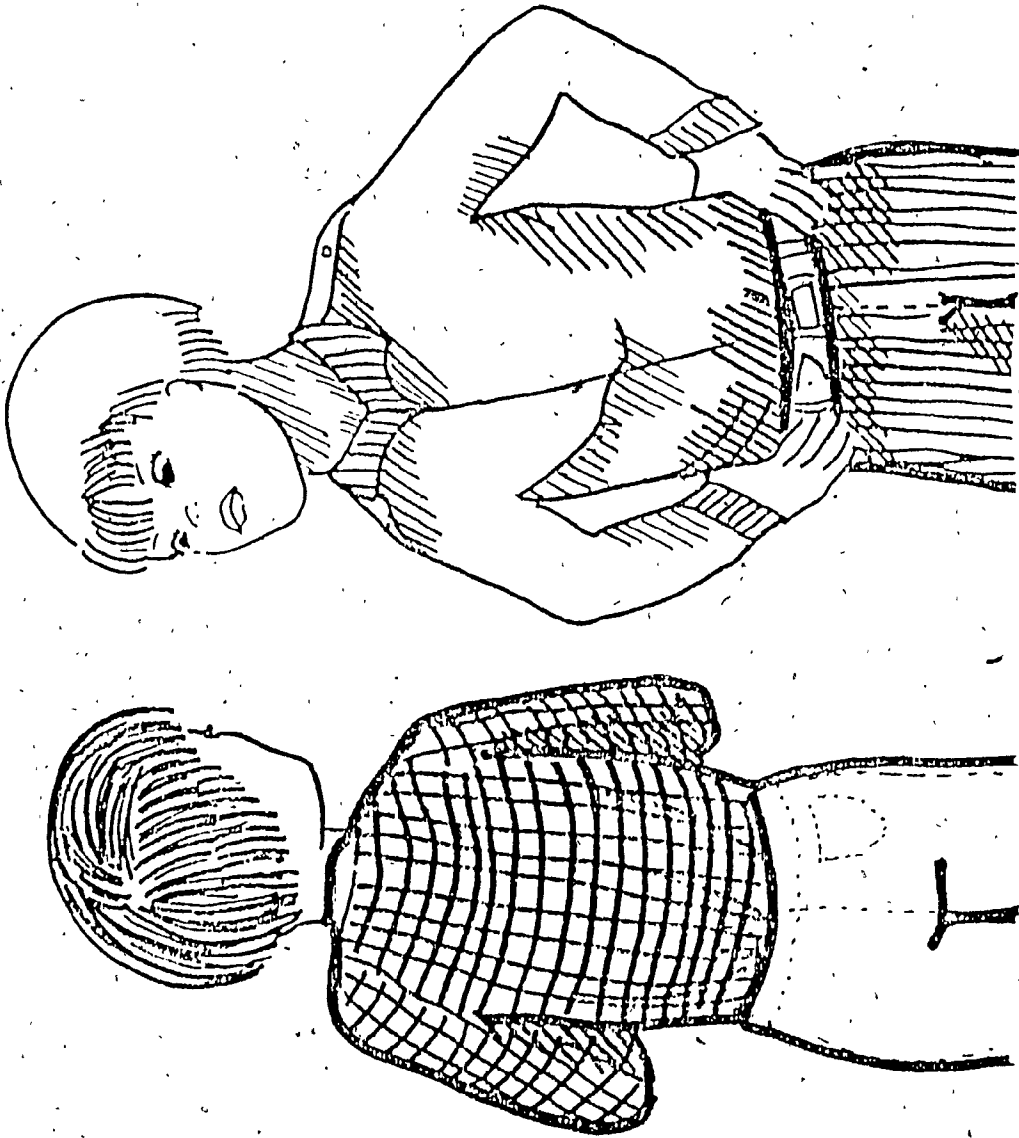
Sequence D, First Drawing, Girls



Séquence D, Second Drawing, Girls

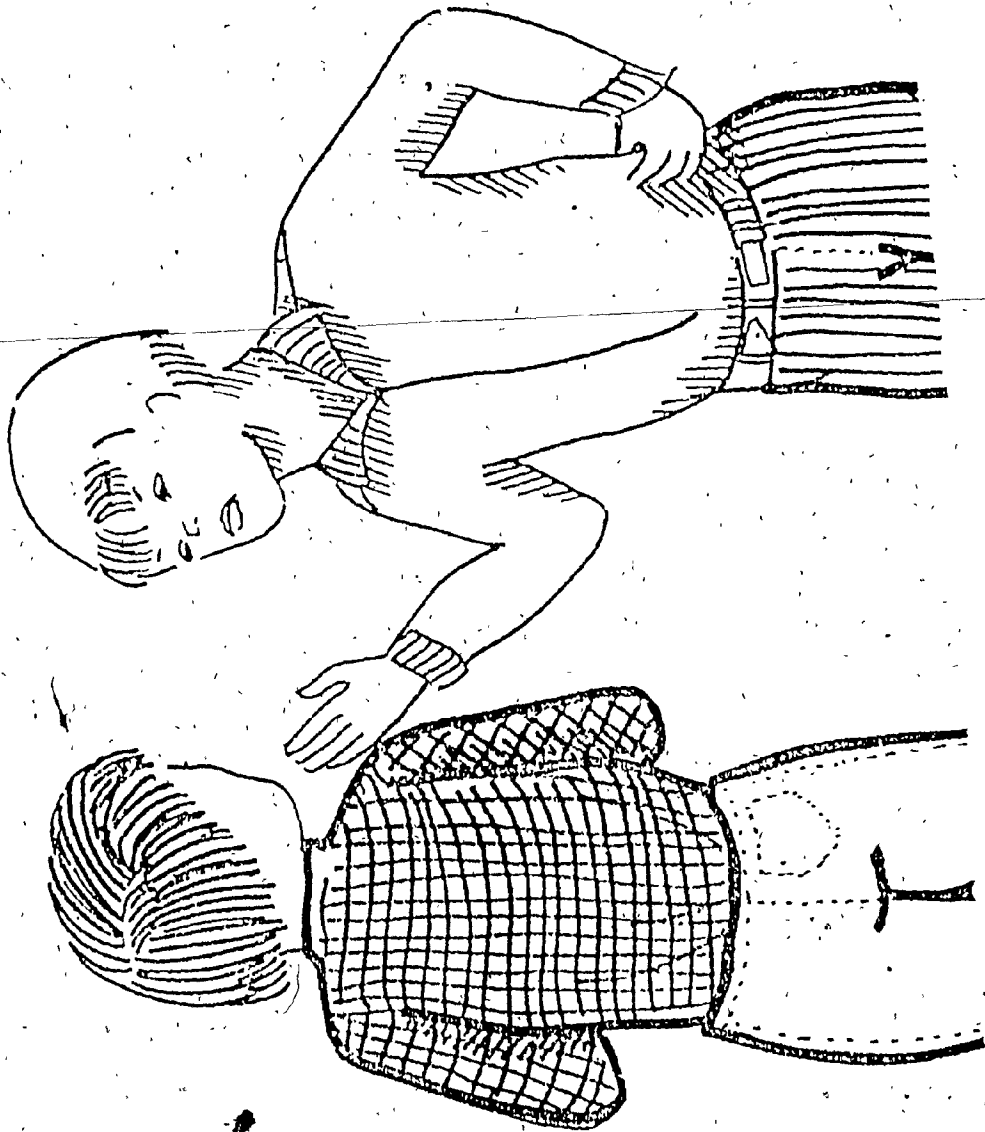


Sequence D, Third Drawing, Girls.

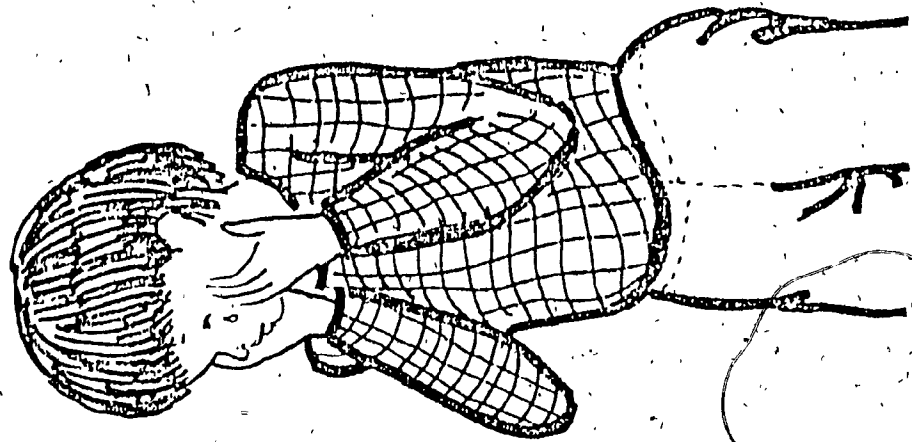
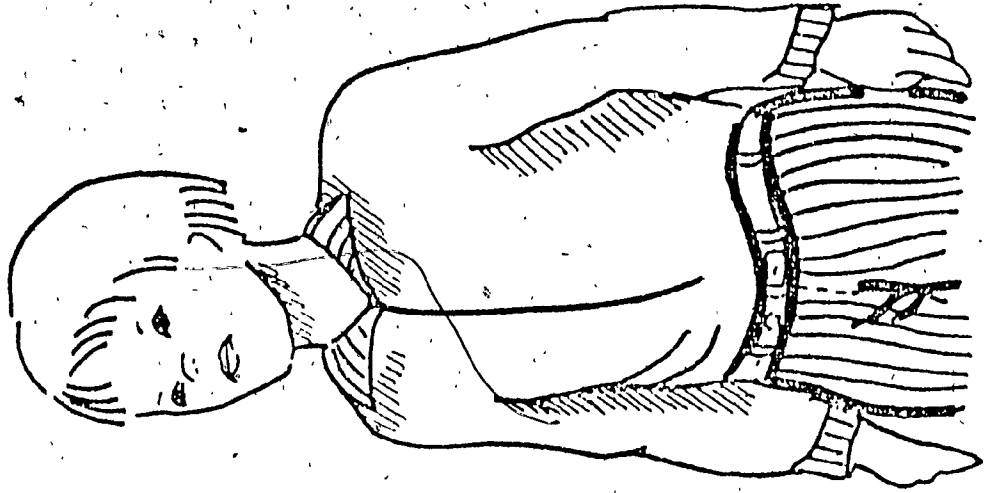


Sequence D, First Drawing, Boys

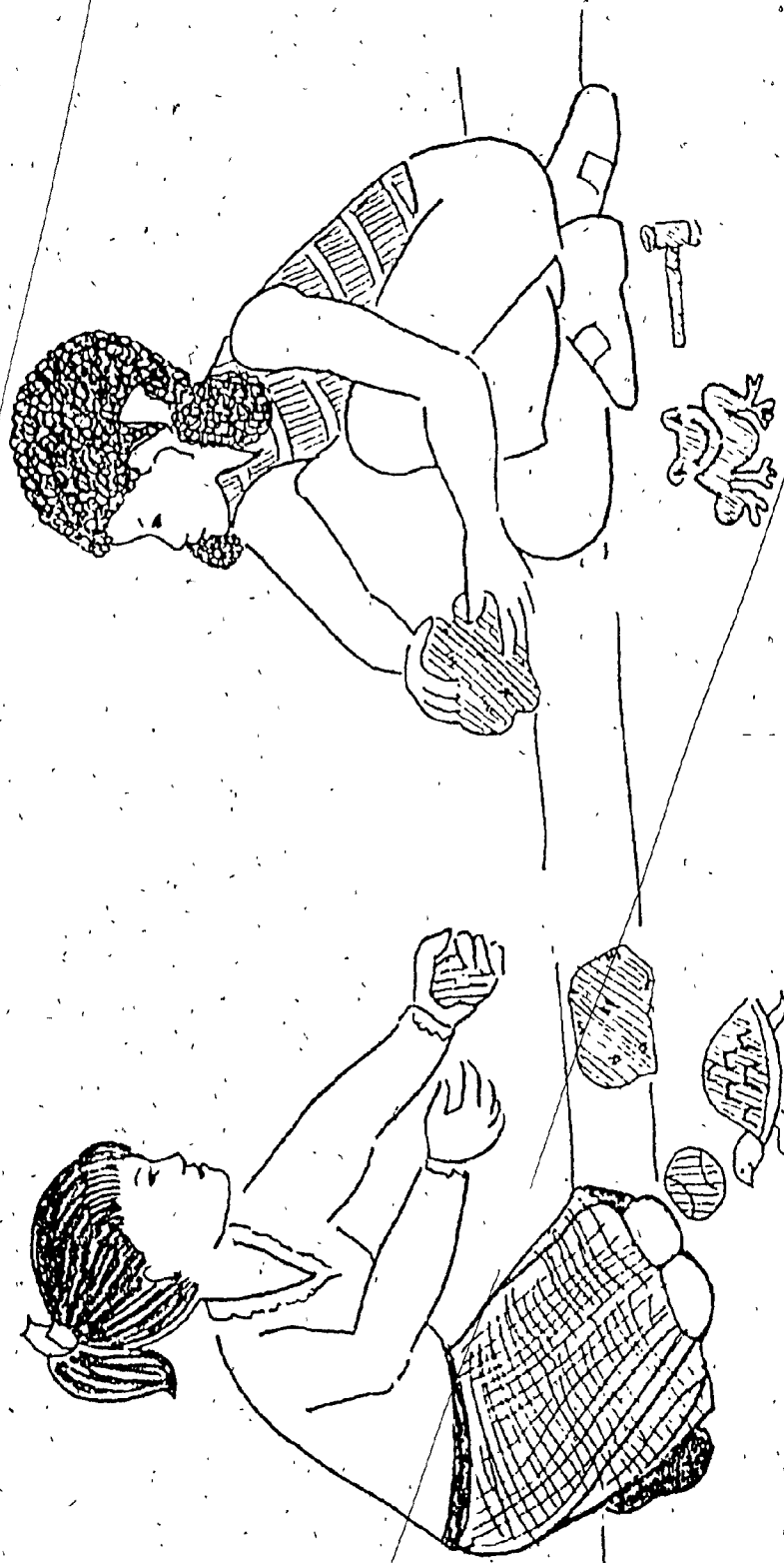
4



Sequence D, Second Drawing, Boys



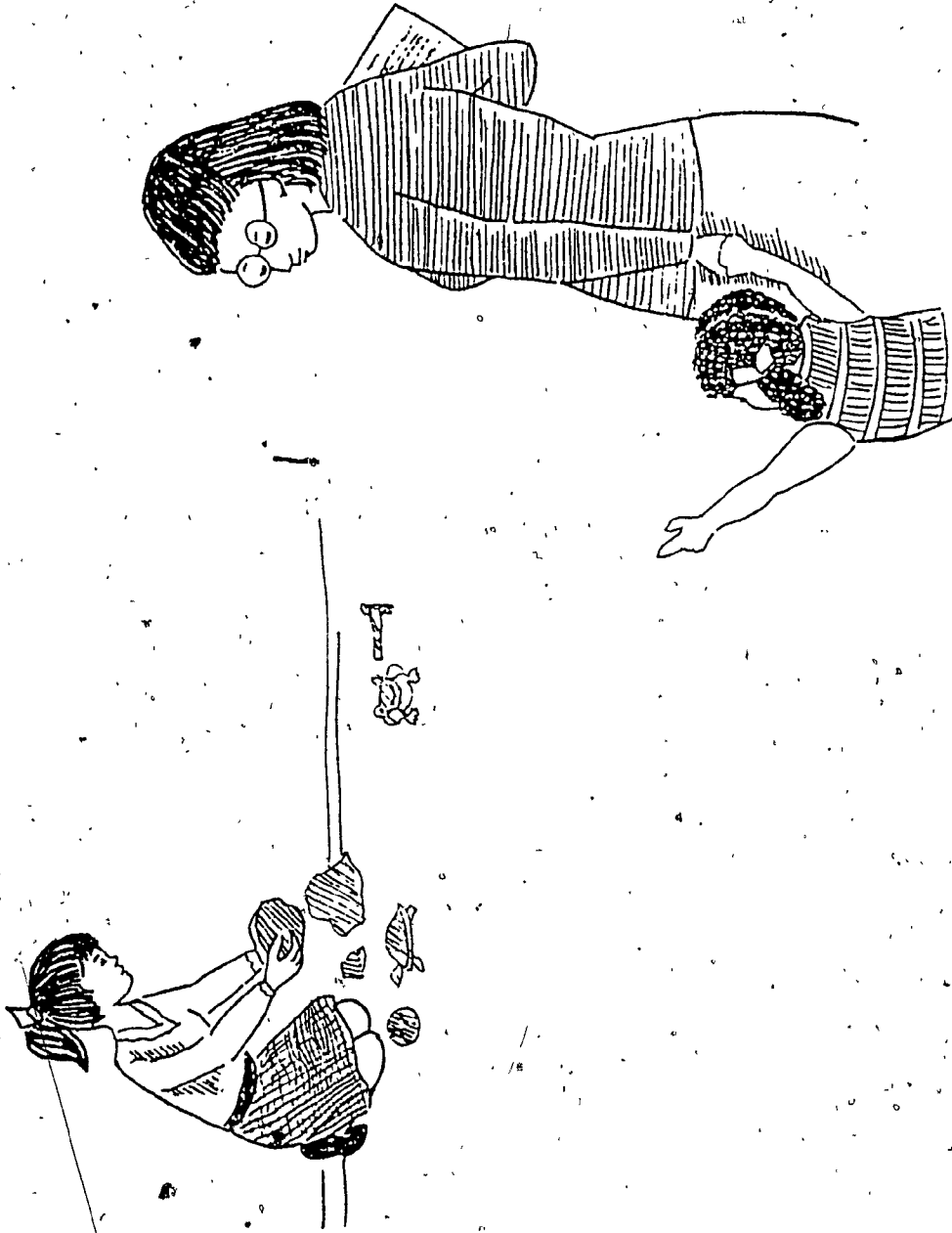
Sequence D, Third Drawing, Boys



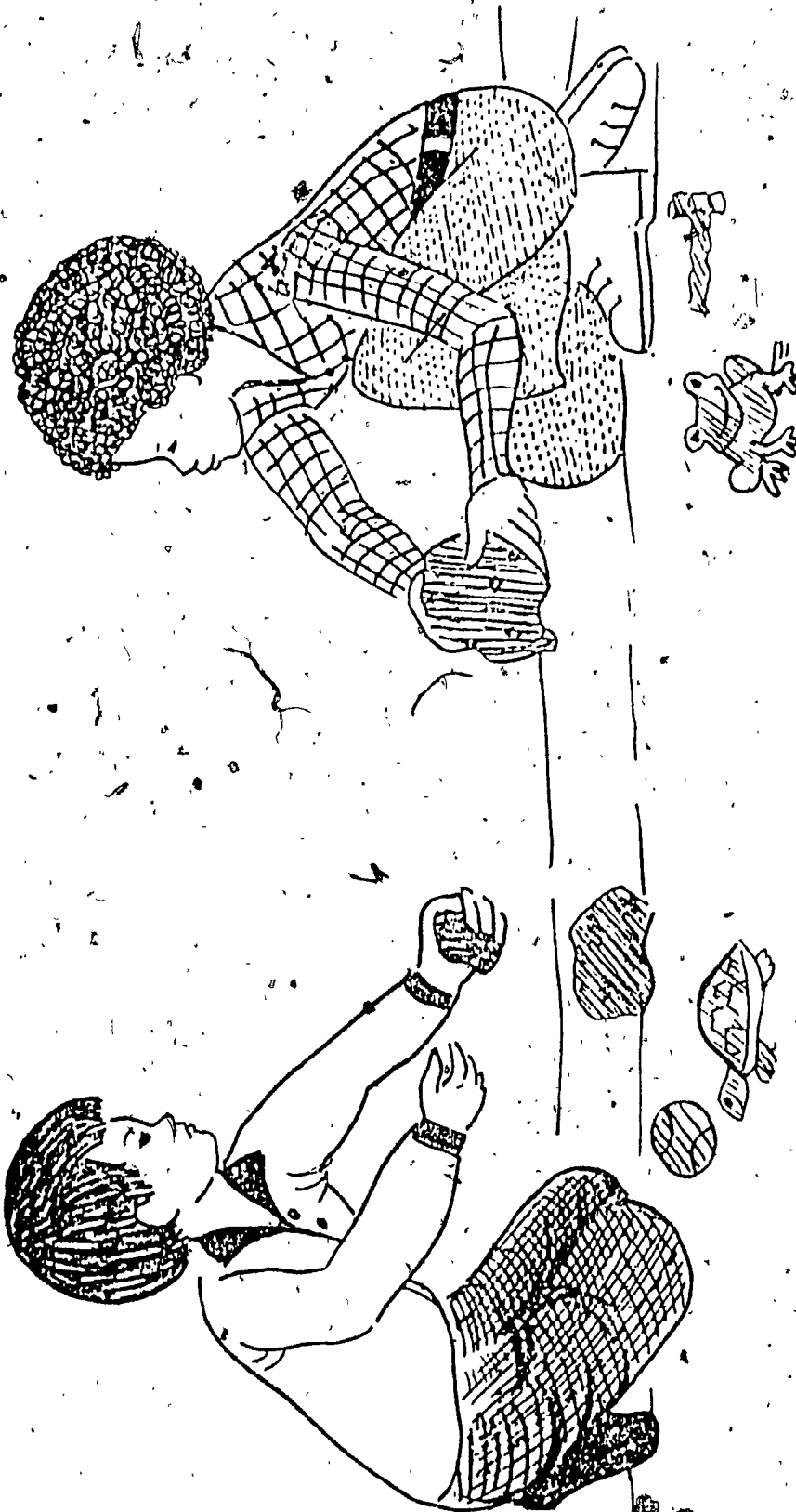
Sequence E, First Drawing, Girls



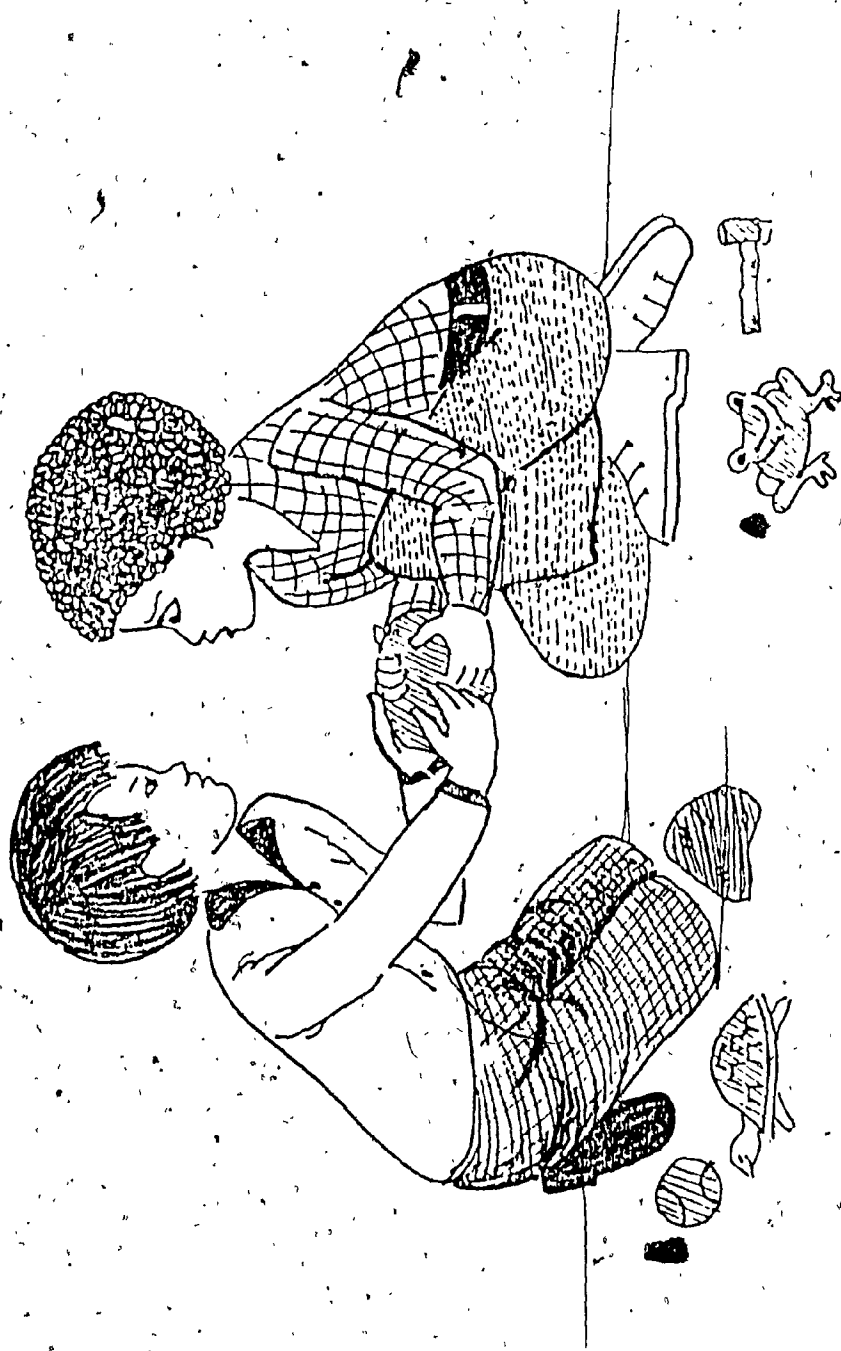
Sequence E, Second Drawing, Girls



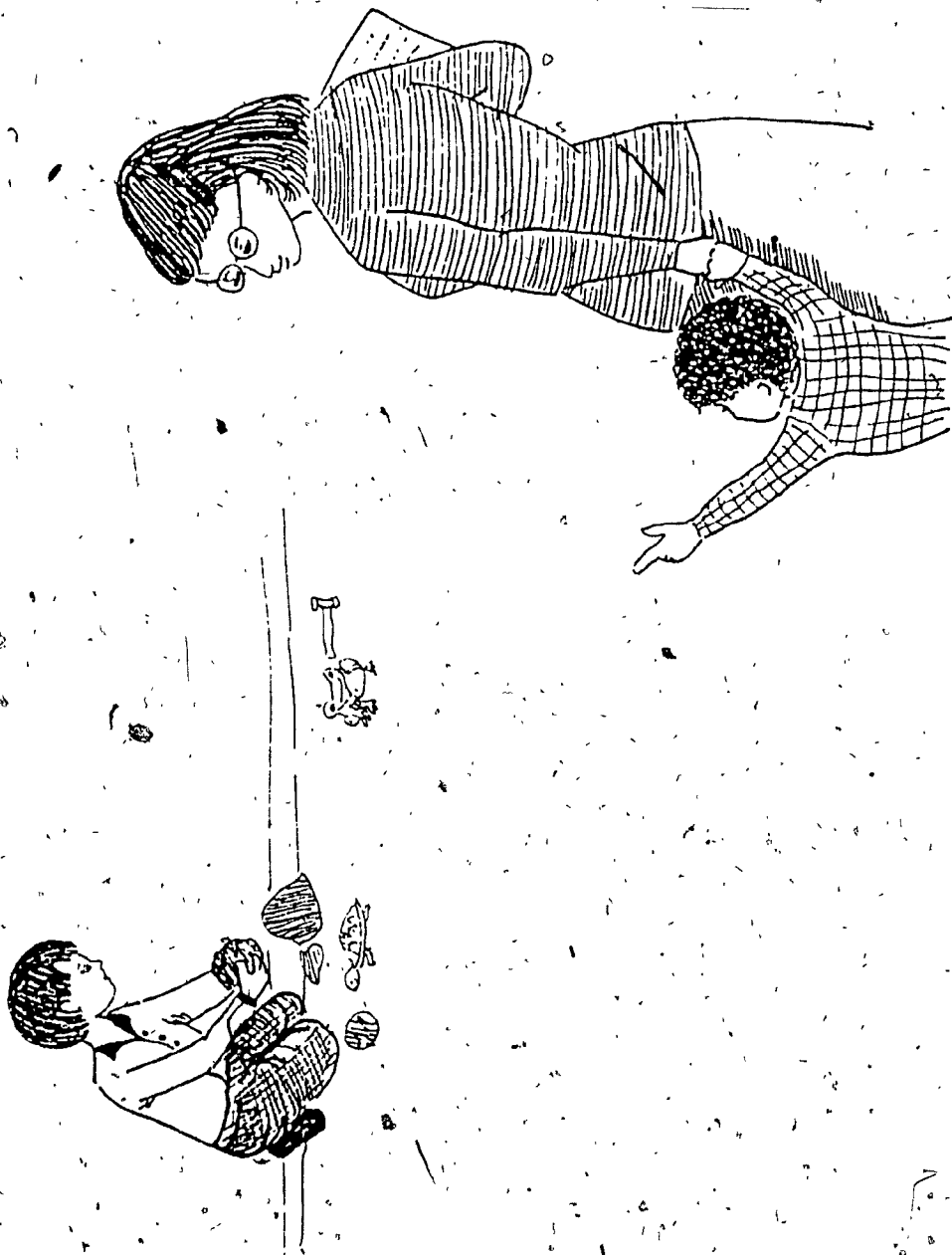
Sequence E, Third Drawing, Girls



Sequence E, First Drawing, Boys



Sequence E; Second Drawing, Boys



Sequence E, Third Drawing, Boys

Appendix C

Answer Booklet

Name _____ Presentation Number _____

Telephone _____ Date _____

Level of Study (e.g., undergraduate, M.A., etc.) _____

Field of Study _____

Sex _____ Age _____

Part A:

Place a check (✓) next to each behavior you saw occurring in the scene shown. If you did not see an example of a particular behavior, leave the space next to that behavior blank.

Scene 1:

Solitary Play _____ Vigorous Play _____

Group Play _____ Dependency _____

Cooperation _____ Nurturance _____

Playful Physical Contact _____ Aggression _____

Concentration _____ Relaxation _____

Scene 2:

Solitary Play _____ Vigorous Play _____

Group Play _____ Dependency _____

Cooperation _____ Nurturance _____

Playful Physical Contact _____ Aggression _____

Concentration _____ Relaxation _____

Scene 3:

Solitary Play _____	Vigorous Play _____
Group Play _____	Dependency _____
Cooperation _____	Nurturance _____
Playful Physical Contact _____	Aggression _____
Concentration _____	Relaxation _____

Scene 4:

Solitary Play _____	Vigorous Play _____
Group Play _____	Dependency _____
Cooperation _____	Nurturance _____
Playful Physical Contact _____	Aggression _____
Concentration _____	Relaxation _____

Part B:

Rate both of the two children shown on each of the following scales. Mark your answers (as a number from 0-8) in the spaces provided to the right of each scale. In the space below the word "Left", mark your rating for the child shown on the left-hand side of the screen. Mark your rating for the child shown on the right-hand side in the space under the word "Right".

Sequence 1:

		Left	Right
Not at all dominant 0..1..2..3..4..5..6..7..8	Very dominant	_____	_____
Not at all passive 0..1..2..3..4..5..6..7..8	Very passive	_____	_____
Not at all aggressive 0..1..2..3..4..5..6..7..8	Very aggressive	_____	_____
Not at all submissive 0..1..2..3..4..5..6..7..8	Very submissive	_____	_____
Not at all active 0..1..2..3..4..5..6..7..8	Very active	_____	_____
Very poorly behaved 0..1..2..3..4..5..6..7..8	Very well behaved	_____	_____

Sequence 2

Left

Right

Not at all dominant	0..1..2..3..4..5..6..7..8	Very dominant	_____	_____
Not at all passive	0..1..2..3..4..5..6..7..8	Very passive	_____	_____
Not at all aggressive	0..1..2..3..4..5..6..7..8	Very aggressive	_____	_____
Not at all submissive	0..1..2..3..4..5..6..7..8	Very submissive	_____	_____
Not at all active	0..1..2..3..4..5..6..7..8	Very active	_____	_____
Very poorly behaved	0..1..2..3..4..5..6..7..8	Very well behaved	_____	_____

Sequence 3

Not at all dominant	0..1..2..3..4..5..6..7..8	Very dominant	_____	_____
Not at all passive	0..1..2..3..4..5..6..7..8	Very passive	_____	_____
Not at all aggressive	0..1..2..3..4..5..6..7..8	Very aggressive	_____	_____
Not at all submissive	0..1..2..3..4..5..6..7..8	Very submissive	_____	_____
Not at all active	0..1..2..3..4..5..6..7..8	Very active	_____	_____
Very poorly behaved	0..1..2..3..4..5..6..7..8	Very well behaved	_____	_____

Sequence 4

Left

Right

Not at all dominant	0..1..2..3..4..5..6..7..8	Very dominant	_____	_____
Not at all passive	0..1..2..3..4..5..6..7..8	Very passive	_____	_____
Not at all aggressive	0..1..2..3..4..5..6..7..8	Very aggressive	_____	_____
Not at all submissive	0..1..2..3..4..5..6..7..8	Very submissive	_____	_____
Not at all active	0..1..2..3..4..5..6..7..8	Very active	_____	_____
Very poorly behaved	0..1..2..3..4..5..6..7..8	Very well behaved	_____	_____

Sequence 5

Not at all dominant	0..1..2..3..4..5..6..7..8	Very dominant	_____	_____
Not at all passive	0..1..2..3..4..5..6..7..8	Very passive	_____	_____
Not at all aggressive	0..1..2..3..4..5..6..7..8	Very aggressive	_____	_____
Not at all submissive	0..1..2..3..4..5..6..7..8	Very submissive	_____	_____
Not at all active	0..1..2..3..4..5..6..7..8	Very active	_____	_____
Very poorly behaved	0..1..2..3..4..5..6..7..8	Very well behaved	_____	_____

Sequence 6

Left

Right

Not at all dominant	0..1..2..3..4..5..6..7..8	Very dominant	_____	_____
Not at all passive	0..1..2..3..4..5..6..7..8	Very passive	_____	_____
Not at all aggressive	0..1..2..3..4..5..6..7..8	Very aggressive	_____	_____
Not at all submissive	0..1..2..3..4..5..6..7..8	Very submissive	_____	_____
Not at all active	0..1..2..3..4..5..6..7..8	Very active	_____	_____
Very poorly behaved	0..1..2..3..4..5..6..7..8	Very well behaved	_____	_____

Sequence 7

Not at all dominant	0..1..2..3..4..5..6..7..8	Very dominant	_____	_____
Not at all passive	0..1..2..3..4..5..6..7..8	Very passive	_____	_____
Not at all aggressive	0..1..2..3..4..5..6..7..8	Very aggressive	_____	_____
Not at all submissive	0..1..2..3..4..5..6..7..8	Very submissive	_____	_____
Not at all active	0..1..2..3..4..5..6..7..8	Very active	_____	_____
Very poorly behaved	0..1..2..3..4..5..6..7..8	Very well behaved	_____	_____

Sequence 8			Left	Right
Not at all dominant	0..1..2..3..4..5..6..7..8	Very dominant	_____	_____
Not at all passive	0..1..2..3..4..5..6..7..8	Very passive	_____	_____
Not at all aggressive	0..1..2..3..4..5..6..7..8	Very aggressive	_____	_____
Not at all submissive	0..1..2..3..4..5..6..7..8	Very submissive	_____	_____
Not at all active	0..1..2..3..4..5..6..7..8	Very active	_____	_____
Very poorly behaved	0..1..2..3..4..5..6..7..8	Very well behaved	_____	_____

Sequence 9			Left	Right
Not at all dominant	0..1..2..3..4..5..6..7..8	Very dominant	_____	_____
Not at all passive	0..1..2..3..4..5..6..7..8	Very passive	_____	_____
Not at all aggressive	0..1..2..3..4..5..6..7..8	Very aggressive	_____	_____
Not at all submissive	0..1..2..3..4..5..6..7..8	Very submissive	_____	_____
Not at all active	0..1..2..3..4..5..6..7..8	Very active	_____	_____
Very poorly behaved	0..1..2..3..4..5..6..7..8	Very well behaved	_____	_____

Sequence 10

			Left	Right
Not at all dominant	0..1..2..3..4..5..6..7..8	Very dominant	_____	_____
Not at all passive	0..1..2..3..4..5..6..7..8	Very passive	_____	_____
Not at all aggressive	0..1..2..3..4..5..6..7..8	Very aggressive	_____	_____
Not at all submissive	0..1..2..3..4..5..6..7..8	Very submissive	_____	_____
Not at all active	0..1..2..3..4..5..6..7..8	Very active	_____	_____
Very poorly behaved	0..1..2..3..4..5..6..7..8	Very well behaved	_____	_____

Personal Attributes Questionnaire

The items below inquire about what kind of a person you think you are. Each item consists of a pair of characteristics, with the letters A - E in between. For example:

Not at all artistic A..B..C..D..E.. Very artistic

Each pair describes contradictory characteristics--that is, you cannot be both at the same time, such as very artistic and not at all artistic.

The letters form a scale between the two extremes. You are to choose a letter which describes where you fall on the scale. For example, if you think you have no artistic ability, you would circle A. If you think you are pretty good, you might circle D. If you are only medium, you might circle C, and so forth.

- | | | | |
|-----|---|---------------|--|
| 1. | Not at all aggressive | A..B..C..D..E | Very aggressive |
| 2. | Not at all independent | A..B..C..D..E | Very independent |
| 3. | Not at all emotional | A..B..C..D..E | Very emotional |
| 4. | Very submissive | A..B..C..D..E | Very dominant |
| 5. | Not at all excitable
in a <u>major</u> crisis | A..B..C..D..E | Very excitable in
a <u>major</u> crisis |
| 6. | Very passive | A..B..C..D..E | Very active |
| 7. | Not at all able to
devote self completely
to others | A..B..C..D..E | Able to devote
self completely
to others |
| 8. | Very rough | A..B..C..D..E | Very gentle |
| 9. | Not at all helpful
to others | A..B..C..D..E | Very helpful to
others |
| 10. | Not at all competitive | A..B..C..D..E | Very competitive |
| 11. | Very home oriented | A..B..C..D..E | Very worldly |
| 12. | Not at all kind | A..B..C..D..E | Very kind |
| 13. | Indifferent to
others' approval | A..B..C..D..E | Highly needful of
others' approval |
| 14. | Feelings not easily
hurt | A..B..C..D..E | Feeling easily
hurt |

- | | | |
|--|---------------|------------------------------------|
| 15. Not at all aware of feelings of others | A..B..C..D..E | Very aware of feelings of others |
| 16. Can make decisions easily | A..B..C..D..E | Have difficulty making decisions |
| 17. Give up very easily | A..B..C..D..E | Never give up easily |
| 18. Never cry | A..B..C..D..E | Cry very easily |
| 19. Not at all self-confident | A..B..C..D..E | Very self-confident |
| 20. Feel very inferior | A..B..C..D..E | Feel very superior |
| 21. Not at all understanding of others | A..B..C..D..E | Very understanding of others |
| 22. Very cold in relations with others | A..B..C..D..E | Very warm in relations with others |
| 23. Very little need for security | A..B..C..D..E | Very strong need for security |
| 24. Go to pieces under pressure | A..B..C..D..E | Stand up well under pressure |

How much contact have you had with children? Please circle the appropriate number.

No contact 0..1..2..3..4..5 A great deal of contact

In what capacities have you had experience with children?
(check all that apply)

Am a parent/guardian
How many children do you have?

Have worked with children as a
 teacher
 camp counselor
 babysitter
 other (Specify:)

Have younger brothers or sisters
How many?

Other (Specify:)

Attitudes toward Women

The statements listed below describe attitudes toward the roles of women in society which different people have. There are no right or wrong answers, only opinions. You are asked to express your feeling about each statement by indicating whether you, (A) agree strongly, (B) agree mildly, (C) disagree mildly, or (D) disagree strongly.

1. Swearing and obscenity are more repulsive in the speech of a woman than a man.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

2. Under modern economic conditions with women being active outside the home, men should share in household tasks such as washing dishes and doing laundry.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

3. It is insulting to women to have the "obey" clause remain in the marriage service.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

4. A woman should be as free as a man to propose marriage.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

5. Women should worry less about their rights and more about becoming good wives and mothers.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

6. Women should assume their rightful place in business and all the professions along with men.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

7. A woman should not expect to go to exactly the same places or to have quite the same freedom of action as a man.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

8. It is ridiculous for a woman to run a locomotive and for a man to darn socks.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

9. The intellectual leadership of a community should be largely in the hands of men.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

10. Women should be given equal opportunity with men for apprenticeship in the various trades.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

11. Women earning as much as their date should bear equally the expense when they go out together.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

12. Sons in a family should be given more encouragement to go to college than daughters.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

13. In general, the father should have the greater authority than the mother in the bringing up of the children.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

14. Economic and social freedom is worth far more to women than acceptance of the ideal of femininity, which has been set up by men.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

15. There are many jobs in which men should be given preference over women in being hired or promoted.

A	B	C	D
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly

Appendix D

Statistical Tables for Part A

Table A

Median Number of Times Each Behavioral
Category was Reported for Girls and for Boys

Behavioral category	Sex of observer			
	Male		Female	
	Boys	Girls	Boys	Girls
Aggression	1.5	.9	1.5	.5
Relaxation	.1	.2	.0	.0
Nurturance	.7	.7	.7	.8
Concentration	.9	.7	.3	.8
Playful physical contact	1.1	.4	.8	.5
Group play	1.6	1.5	1.8	1.8
Solitary play	1.0	1.1	1.0	.8
Cooperation	1.8	1.6	1.7	1.9
Vigorous play	.4	.5	.4	.4
Dependency	.2	.1	.1	.1

Note. Subjects had two opportunities to report each category for boys and two opportunities to report each category for girls.

Table B

Wilcoxon T Values for the Difference
Between the Number of Times Each Behavior
was Recorded for Boys and for Girls

Behavioral category	<u>T</u> value
Aggression	0 (11)**a
Relaxation	0 (6)*
Nurturance	30 (11)
Concentration	35 (13)
Playful physical contact	39 (14)
Group play	46 (14)
Solitary play	63 (17)
Cooperation	70 (14)
Vigorous play	73 (16)
Dependency	(6) ^b

Note: T values are for the combined data from male and female observers, since a minimum of six non-zero differences between pairs is required to calculate T and on many categories there were fewer than six such differences within each sex of observer.

^aNumbers in parentheses indicate the number of pairs having non-zero differences.

^bA meaningful T value could not be calculated because there were equal numbers of positive and negative ranks.

* $p < .05$.

** $p < .01$.

Table C

Point Biserial Correlation Between Presence/Absence of

Bias in Recording Aggression and Questionnaire Responses

Sex of observer	Personal Attributes Questionnaire scale		Attitudes toward Women, Scale
	M	M-F F	
Male	-.20	-.08 +.37	-.11
Female	-.11	-.22 +.18	-.23

Note: No correlation reached significance at the .05 level.

Appendix E

Statistical Tables for Part B

Table A
 Source Table for Analysis of Variance Examining Effect of
 Presentation Order on Degree of Bias in Rating Aggression

Source	df	SS	MS	F
Presentation order	3	84.60	28.20	9.951*
Within subjects	36	1067.40	29.65	
Total	39	1152.00		

*n.s.



Table B

T-tests Comparing Ratings Given to Boys
with Ratings Given to Girls on Each Scale

Scale	Average total rating		t
	Boys	Girls	
Male observers			
Aggression	45.60	42.70	.245*
Dominance	45.35	45.40	-.05
Goodness	36.10	37.30	-.78
Activity level	47.80	47.40	.28
Passivity	36.55	37.05	-.37
Submission	33.65	33.45	.16
Female observers			
Aggression	43.95	43.85	.08
Dominance	46.95	45.75	1.31
Goodness	35.40	35.90	-.55
Activity level	46.20	45.00	.91
Passivity	35.50	36.10	-.38
Submission	34.10	34.25	-.16

Note. Each t-test has 19 df.

* $p < .05$.

Table C
 Source. Table for Analysis of Variance Examining Effect of
 Sex of Observer on Degree of Bias in Rating Aggression

Source	<u>df</u>	<u>SS</u>	<u>MS</u>	<u>F</u>
Sex of child	1,38	45.00	45.00	3.19*
Sex of child x sex of observer	1,38	39.20	39.20	2.77
Sex of child x within subjects	38	536.80	14.13	
Sex of observer	1,38	1.25	1.25	.03
Within subjects	38	1833.70	48.26	

* $p < .10$.

Table D

Source Table for Analysis of Variance Examining Effect of Sex of Observer on Degree of Bias in Rating Dominance

Source	<u>df</u>	<u>SS</u>	<u>MS</u>	<u>F</u>
Sex of child	1,38	6.61	6.61	.75
Sex of child x sex of observer	1,38	7.81	7.81	.89
Sex of child x within-subjects	38	333.08	8.76	
Sex of observer	1,38	19.01	19.01	.66
Within subjects	38	3754.97	98.82	

Note. All effects are nonsignificant at the .05 level.

Table E
 Source Table for Analysis of Variance Examining Effect of Sex
 of Observer on Degree of Bias in Rating Goodness of Behavior

Source	<u>df</u>	<u>SS</u>	<u>MS</u>	<u>F</u>
Sex of child	1,38	14.45	14.45	.90
Sex of child x sex of observer	1,38	2.45	2.45	.15
Sex of child x within subjects	38	610.10	16.06	
Sex of observer	1,38	22.05	22.05	.24
Within subjects	38	3512.50	92.43	

Note: All effects are nonsignificant at the .05 level.

Table F
 Source Table for Analysis of Variance Examining Effect of
 Sex of Observer on Degree of Bias in Rating Activity Level

Source	df	SS	MS	F
Sex of child	1,38	12.80	12.80	.68
Sex of child x sex of observer	1,38	3.20	3.20	.17
Sex of child x within subjects	38	712.00	18.74	
Sex of observer	1,38	80.00	80.00	.86
Within subjects	38	3543.20	93.24	

Note. All effects are nonsignificant at the .05 level.

Table G
 Source Table for Analysis of Variance, Examining Effect of
 Sex of Observer on Degree of Bias in Rating Passivity

Source	df	SS	MS	F
Sex of child	1,38	6.05	6.05	.28
Sex of child x sex of observer	1,38	.05	.05	.00
Sex of child x within subjects	38	824.90	21.71	
Sex of observer	1,38	20.00	20.00	.23
Within subjects	38	3295.80	86.73	

Note. All effects are nonsignificant at the .05 level.

Table H
 Source Table for Analysis of Variance Examining Effect of
 Sex of Observer on Degree of Bias in Rating Submission

Source	df	SS	MS	F
Sex of child	1,38	.01	.01	.00
Sex of child x sex of observer	1,38	.61	.61	.05
Sex of child x within subjects	38	457.88	12.05	
Sex of observer	1,38	7.81	7.81	.13
Within subjects	38	2291.17	60.29	

Note. All effects are nonsignificant at the .05 level.

Table I
 Step-wise Multiple Regression of
 Degree of Males' Bias in Rating Aggression on Attitudes
 toward Women Scale and Personal Attributes Questionnaire

Step	Variable entered	Simple R	Multiple R	R square	R square change
1	PAQ F scale	-.497*	.497*	.247	.247
2	PAQ M scale	-.405	.608*	.370	.123
3	AWS	-.142	.612	.374	.005
4	PAQ M-F scale	-.193	.615	.378	.004

*p < .05.

Table J /
 Step-wise Multiple Regression of
 Degree of Females' Bias in Rating Aggression on Attitudes
 toward Women Scale and Personal Attributes Questionnaire

Step	Variable entered	Simple R	Multiple R	R square	R square change
1	PAQ F scale	.312	.312	.097	.097
2	PAQ M-F scale	.073	.339	.115	.018
3	PAQ M scale	-.083	.391	.153	.038
4	AWS	-.012	.410	.168	.015

Note. All regression coefficients are nonsignificant at the .05 level.

Table K
Pearson Correlations among Questionnaire Scales

	AWS	PAQ F	PAQ M-F
Male observers			
PAQ M	.05	.12	.52*
PAQ M-F	-.09	-.10	
PAQ F	.41		
↑ Female observers			
PAQ M	.32	-.11	.70**
PAQ M-F	.11	-.18	
PAQ F	-.22		

* $p < .05$.

** $p < .01$.