A developmental study of strategies and memory in the length-transitivity task

KIMIHIKO MATSUDA

Department of Psychology, Faculty of Education, Kagoshima University, Kūrimoto, Kagoshima 890

HARUJI HIROSE and HIDETOSHI FUTAMURA

Department of Psychology, Faculty of Literature, Kyūshū University, Higashi-ku, Fukuoka 812

The purpose to examine the role of memory on the performance in length-transitivity tasks, with seven sticks as stimulus, by five year and four year old children, and examine whether solution of the transitivity task by non-transitive strategy is possible. In both Experiment I and Experiment II, the possibility was disproved that the transitivity task was solved by means of labels attached to each stick or by remembering the length of each stick. Furthermore, in Experiment I, five year olds could solve the transitivity problem if its premises were fully retained, while four year olds could not do so. Differences are suggested, between the two age groups in their ability to make correct choices on the transitivity task from the memory of its premises, and that transitivity performance cannot be described as a function of memory.

Key words: development of conceptualization, transitive inference, premise retention, non-transitive strategies, labeling strategies, Piaget (Jean).

According to Piaget (1965), and Piaget and Inhelder (1974), the children at the stage of preoperational thinking, due to the lack of logical transitivity, cannot infer $A>C$ from its two premises, $A>B$ and $B>C$. Contrary to this, Bryant and Trabasso (1971), pointing out a methodological problem that the studies in this subject have been lacking in the control of memory factor, proved that if the premises were fully retained, the children at this stage of development could make transitive inferences.

In Bryant and Trabasso’s experiment (1971), four, five and six year old children were asked to infer the relative length of composite sticks of the non-adjacent inference pairs, $AC$, $BD$ and $CE$ after having them fully retain the premise relations, $A>B$, $B>C$, $C>D$ and $D>E$ by overtraining them with visual and auditory (linguistic) feedback. The result of their experiment was replicated and confirmed by several studies such as Riley and Trabasso (1974), Harris and Bassett (1975), Riley (1976), etc., but according to Breslow (1981), Thayer and Collyer (1978), it has also invited criticisms since their assumptions and findings were drastically different from the Piagetian cognitive developmental theory or structural developmental theory.

One of the most important criticism was concerned with whether transitivity is operating in a transitivity task. With a Piagetian test of transitive inferences using three sticks, Smedslund (1966) pointed out the possibility of solving the problem by means of a non-transitive labeling strategy, since in the Piagetian transitivity task, the sticks at the ends of the stimulus series (extreme sticks) are always “longer” or “shorter” than the remaining one. In order to eliminate the possibility, Bryant and Trabasso (1971) used five sticks ($A$, $B$, $C$, $D$, $E$) and regarded the BD comparison as a critical entry of the operation of transitivity. But, De Boysson-Bardies and O’Regan (1973) proposed a non-transitive secondary labeling model to explain the performance of Bryant and Trabasso’s transitivity tasks with five sticks. Secondary labeling means that the sticks presented with the extreme sticks
also acquire the same labels as the extreme sticks and inference comparisons are made by means of the categorical labels, not of transitive inferences.

In the present study, we intended to examine the validity of secondary labeling model by giving the subjects transitivity tasks with seven sticks (A, B, C, D, E, F, G). De Boysson-Bardies and O'Regan's labeling model, when employed for the transitivity task with seven sticks, would be as follows; during learning the premises, subjects first retain each premise without combining them, they then attached the labels “long” or “short” to only the extreme sticks A and E, and finally the labels transfer to the sticks B and D which presented with the sticks A or E. Thus, at the end of training, the stimuli may be divided into three categories, “long” A and B, “short” F and G, and “nonentities” C, D and E. During the test phase, therefore, the subject make comparisons by means of these categorical labels. In the case of the AC comparison, for example, the unlabeled C is disregarded and a judgment is made on the basis of the label “long” of the stick A. According to this labeling model, therefore, no judgment can be made with the CE pair since both C and E have no labels. Consequently, it is expected that the performance on the CE comparison will be at chance level. Furthermore, in this model, as two adjacent sticks of the AB or FG pairs may belong to the same category, there will be confusion and consequently low scores with the AB or FG comparisons. In fact, the AB scores were lower than those on other adjacent pairs in the experiment of De Boysson-Bardies and O'Regan (1973).

Another criticism of Bryant and Trabasso's assumption dealt with role of memory. From Bryant's (1974) and Trabasso's (1977) point of view, children at the preoperational stage can make transitive inferences if precautions are taken for the retention of the premises. In particular, according to Trabasso, who does not approve of the notion of structural change regarding cognitive development, young children can perform transitive inferences as well as adults if memory is controlled, the low scores on transitivity tasks being explained by the imperfect retention of the premises. But, studies of Glick and Wapner (1968) and Cowan (1964) provide evidence that developmental improvement in transitive performance may be found even when memory is controlled (Breslow, 1981). Furthermore, Youniss and Furth (1973) criticized the experimental procedures of Bryant and Trabasso (1971) by pointing out that in their design, memory and inference were not fully distinguished due to the lack of manipulation of memory. Bryant and Trabasso (1971) attempted to examine the effect of the memory factor on the performance in transitivity tasks by comparing the transitive scores observed with the expected values computed from the scores of the retention of the premises, but, as Youniss and Furth criticized, in such ways we cannot show clearly the role of memory or a reasoning capacity apart from memory. We classified the subjects, therefore, into subgroups and then compared the transitive scores among them. By applying this procedure in the present study, we can compare the transitive performances under the same or different conditions of premise retention, and also get clues about the way each premise is used in solving transitive tasks.

Experiment I

This experiment aimed to examine the validity of secondary labeling by using seven sticks, and to clarify the role of memory on the performance in transitivity tasks by classifying the subjects into subgroups according to the scores on the premise comparisons.
Method

Subjects. The subjects were 91 children from two kindergartens, one public, the other private, in a satellite city of Fukuoka, Japan. They were 40 children of age five class (17 boys and 23 girls with a mean age of five years and eight months) and 51 children of the age four class (33 boys and 18 girls with a mean age of four years and nine months).

Dates and place of experiment. The experiment was conducted during a one-month period covering the end of June to the end of July, 1981, in the playroom and waiting room of the kindergartens.

Stimuli. The training and testing materials used in the experiment were seven colored wooden sticks and seven containers. The seven sticks (A to G) were 16 cm, 15 cm, 14 cm, 13 cm, 12 cm, 11 cm, and 10 cm in length, and were red, white, yellow, blue, green, orange, and gray in color. The container boxes were all of the same size (5 cm long × 5 cm wide × 15 cm high) and the same color (black) but had sinking bores of different lengths so that they could be presented in such away that each stick protruded from the top of the container by 5 cm.

Procedure. The procedure of the experiment, which was basically the same as that of Bryant and Trabasso’s (1971) experiment, consisted of two phases, the training and testing of the individual subjects.

In the training phase, the sticks were presented in pairs and the subjects were asked to tell which was the longer (or shorter) stick. To avoid the possible effect of a particular combination of lengths and colors on the subject, several different combinations of them were used.

The questions used were, “Which stick is longer? ” or “Which stick is shorter? ” After each of the responses, the experimenter drew out the sticks and held them perpendicular to the table with a verbal feedback of “Yes, that’s right,” or “No, that’s not right.” This was done to facilitate the effective learning of the correct responses. The subject had to use the difference of colors as a clue in making a choice about the length, since the protruded parts of the sticks were of equal length.

The training consisted of two phases, (1) ordered presentation and (2) random presentation. In the first ordered presentation phase, the six comparisons of the adjacent pairs, AB, BC, CD, DE, EF, FG, were trained in the descending order of length for half of the subjects in each group, and, in the ascending order of length, for the other half of the children. This training was continued until each of the pairs was learned to a criterion of eight out of ten successful choices. In the second random-ordered presentation phase, which followed immediately after reaching the predetermined criterion level in the ordered presentation phase, the same training was conducted, this time with the six pairs presented in a random order, until the subjects made six successive correct responses to each pair. When the predetermined criterion was not achieved by the twentieth trial, the training was discontinued.

Upon completion of the two phases of training, a test was administered. The testing procedures were almost the same as those of the training except that there was no verbal and visual feedback provided. Each subject was tested four times on every one of eleven pairs selected out of the twenty-one possible combinations. The eleven pairs comprised the six adjacent pairs, premises for the transitive inference, and five non-adjacent inference pairs, AC, BD, CE, DF, and EG, among which BD, CE, DF were critical pairs. The pairs were ordered randomly and presented to each subject in a different order. The types of questions (“longer ” or “shorter ”) and the positions of the sticks (left or right) were randomly changed for each trial, both in the training and in the testing periods. The time
Mean percentage of correct responses on transitivity and premise retention

(five year group)

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>71.9***</td>
<td>61.3**</td>
<td>52.5</td>
<td>(28.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>58.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>49.4</td>
<td>58.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>61.9**</td>
<td>63.1***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td>66.3***</td>
<td></td>
<td></td>
<td></td>
<td>77.5***</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>72.5***</td>
</tr>
</tbody>
</table>

(four year group)

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>69.1***</td>
<td>69.1***</td>
<td>50.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>58.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>55.9</td>
<td>46.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>50.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td>60.8**</td>
<td></td>
<td></td>
<td></td>
<td>64.2***</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71.6***</td>
</tr>
</tbody>
</table>

*** p<.001, ** p<.005.

Notes: The figures for the adjacent pairs (e.g. A-B) and non-adjacent pairs (e.g. A-C) represent the success rates of premise retention and transitivity, respectively. The figures in parentheses show the probabilities of the Trabasso’s (1971) calculation. According to their hypothesis, the probability of making a correct inference on the BD transitive pair ($P_{BD}$) is the probability of jointly recalling the information for each of the premise pairs. Asterisks indicate the scores are significantly above chance level.

needed for each subject varied from about 35 to 40 min.

Results

Table 1 which were tabulated according to Bryant and Trabasso (1971), shows the mean percentage of the correct responses on the tests of transitivity and premise retention for the five and four year old children. As is clear from Table 1 when the premise comparisons were significantly above chance level, so were the transitive comparisons. However, only less than half of the scores were statistically significant except those on the end pairs which were all above chance level. This means that from these data, we cannot expect any further useful information as for transitive inferences or the role of memory.

For this reason, we grouped the subjects into four subgroups based on the scores of premise retention. Table 2 and Table 3 show in percentages, for the five year olds and the four year olds respectively, the correct choices on the transitive pairs for the four subgroups. The G-P subgroup for the AC comparison means, for example, that the children of this group were good (G) on the AB retention (the first premise of the AC comparison), but poor (P) on the BC retention (the second premise). Retention was evaluated as good (G) when more than two out of four performances were correct and poor (P) when
A developmental study of transitivity

Table 2
Percentage of correct responses on premise retention and transitivity for retention subgroup
(five year group)

<table>
<thead>
<tr>
<th></th>
<th>G-G</th>
<th>G-P</th>
<th>P-G</th>
<th>P-P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A C</td>
<td>86.7***</td>
<td>75.0***</td>
<td>33.3</td>
<td>15.6</td>
</tr>
<tr>
<td></td>
<td>(15)</td>
<td>(11)</td>
<td>(6)</td>
<td>(8)</td>
</tr>
<tr>
<td>B D</td>
<td>73.0†</td>
<td>53.1</td>
<td>58.3</td>
<td>37.5</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td>(16)</td>
<td>(9)</td>
<td>(10)</td>
</tr>
<tr>
<td>C E</td>
<td>95.0**</td>
<td>55.5</td>
<td>56.7</td>
<td>47.7</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td>(9)</td>
<td>(15)</td>
<td>(11)</td>
</tr>
<tr>
<td>D F</td>
<td>78.1**</td>
<td>63.6</td>
<td>51.9</td>
<td>62.5</td>
</tr>
<tr>
<td></td>
<td>(9)</td>
<td>(11)</td>
<td>(14)</td>
<td>(6)</td>
</tr>
<tr>
<td>E G</td>
<td>92.2***</td>
<td>67.0</td>
<td>77.0**</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>(17)</td>
<td>(6)</td>
<td>(11)</td>
<td>(6)</td>
</tr>
</tbody>
</table>

*** p<.001 ** p<.005 * p<.01 † p<.05
Figures in parentheses stand for the number of subjects.

less than three out of four performances were correct. In order to confirm whether the five year and four year groups with good retention are equivalent in the scores of premise retention or not, chi-square test was conducted separately for each premise pair and there were no significant differences in them between the two age groups.

The results of the analysis of the sum of squares [4(retention subgroups) x 5(pairs)] by the method of arc sine transformation (arc sine \( \sqrt{p} \)) of the scores of the five year group (Table 2), shows a significant premise retention effect on transitivity tasks. The results of the multiple comparison among the four subgroups showed that the G-G subgroup scored higher on inference comparisons than any of the other three subgroups (between G-G and G-P, \( \chi^2=37.97, df=1, p<.01 \); between G-G and P-G, \( \chi^2=30.37, df=1, p<.01 \); between G-G and P-P, \( \chi^2=77.57, df=1, p<.01 \)). It also showed that the subgroups G-P and P-G scored higher in the transitive comparisons than the P-P subgroup (between G-P and P-P, \( \chi^2=10.5, df=1, p<.01 \); between P-G and P-P, \( \chi^2=14.62, df=1, p<.01 \)), but there were no significant differences between the P-G and G-P subgroups. These facts can be confirmed by the results that the G-G subgroup showed significantly high scores on most of the transitive pairs, the P-G and G-P subgroups only on one of the inference comparisons, and the P-P subgroup on none of the inference pairs.

The same type of analysis [4(retention subgroups) x 5(pairs)] was conducted for the performance of the four year olds (Table 3). There were significant effects of both the subgroups and the pairs, but no interaction between them. The results of the multiple comparison among the transitive performances of the four subgroups showed that the G-G group scored significantly higher on the transitive comparisons than any of the other three groups (all at the .01 level) but there were no significant differences among the rest of the three groups. The results of the multiple comparison among the five inference pairs showed significant differences between each of the two extreme pairs and each of the other three critical pairs at the .05 level except between the AC and DF pairs.

Moreover, an analysis of the sum of the squares [method of arc sine transformation, 2(age) x 5(pairs)] was conducted for the G-G subgroups of the four and five year olds, since there seemed to be differ-
ence in the effects of the pairs between the G-G subgroups of the five year and four year olds. The results showed that the effects of the two factors were significant and there was a significant interaction between them (Table 4). Therefore, the inference scores were compared among the pairs separately for each age group, the results of which presented no significant differences between any inference pairs for the five year group, but for the four year group, significant differences between AC and BD, AC and CE, and CE and EG (all at the .05 level). When the scores on the five inference pairs were compared between the two age groups, there was a significant differences only on the CE comparison ($\chi^2=9.17$, $df=1$, $p>.005$).

Table 4
Analysis of sum of squares of premise retention scores for G-G subgroup (arc sine transformation)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>$df$</th>
<th>$\chi^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairs (A)</td>
<td>368.72</td>
<td>4</td>
<td>14.04</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Ages (B)</td>
<td>166.38</td>
<td>1</td>
<td>6.33</td>
<td>&lt;.02</td>
</tr>
<tr>
<td>Interaction (A×B)</td>
<td>301.62</td>
<td>4</td>
<td>11.48</td>
<td>&lt;.025</td>
</tr>
</tbody>
</table>

Experiment II
In Experiment I, which was intended to exclude the possibility of labeling strategy by increasing the number of sticks, there still existed a possibility of non-transitive strategy due to the flaws of the experimental procedure. It was a possibility that subjects might remember the absolute length (size) of the sticks and, based on this, make a direct comparison without using an intermediating stick, since the full lengths of the sticks were shown to them at the time of visual feedback. Except for Bryant and Trabasso (1971), there have been few experiments of transitivity task reported which examined the subjects' memory of the individual sticks primarily. With this understanding as a background, Experiment II was designed to examine this point by asking subjects to recognize the absolute lengths of the individual sticks out of the series of seven sticks.

Method
Subjects. The subjects were 25 children of the age five class of a private kindergarten in a satellite city of Fukuoka, Japan (15 boys and 10 girls with the mean age of five years and nine months). The kindergarten was the same as in Experiment I.

Dates and place of experiment. The experiment was conducted in the waiting room of the kindergarten, in the beginning of August, 1981.

Stimulus and procedure. The stimulus materials were identical to those used in Experiment I except the series of seven sticks used in the testing phase. The experiment consisted of training and testing as Experiment I did so. The procedures were the same as those in Experiment I in the training period, but different in the testing period in that the subjects were shown one stick (protruded from the container), instead of being presented the two sticks, and were asked to choose the stick which they thought to be the same length from the series of seven sticks (revealed full length and arranged in order), each of which is the same stick used in Experiment I except that they were not painted different colors. In the testing period, the sticks were presented six times. The presentations were randomly ordered throughout the 42 test trials.

Results
The responses of the subjects to each of the sticks were described in terms of the length (cm) of the stick chosen from the seven unpainted testing sticks. The results of the responses to each of the seven sticks were as follow (average scores of 25 subjects): A=14.55 cm, B=13.30 cm, C=13.50 cm, D=13.54 cm, E=13.54 cm, F=13.25 cm, G=12.74 cm. An analysis of variance showed that there were statisti-
A developmental study of transitivity

131

cally significant differences among the seven sticks \([F(6,168)=4.37, p<.005]\). The results of the multiple comparison among the seven sticks showed that there were significant differences between only A and B, A and C, A and F, and A and G at the .05 level, and, there were no significant differences between the two sticks of the inference pairs.

Discussion

With regard to the strategy model, the results of the Experiment I were inconsistent with the non-transitive secondary labeling model. If the hypothesis of De Boysson-Bardies and O'Regan (1973) is true, the scores on the extreme adjacent pairs must be low. But, as is shown in Table 1, the AB and FG comparisons got very high scores. The irrelevancy of this hypothesis as an explanatory model of the children's performance on the transitivity task can be supported also by the fact that the G-G subgroup of the five year olds scored significantly high on the critical CE pair (Table 2). On the other hand, the result that the AC scores of G-P subgroup and EG scores of P-G subgroup tended to be well above chance level, is consistent with the labeling model in which the inference comparisons are made on basis of the labels attached to the end sticks. From these facts, we can conclude that the labels "long" or "short" attached to the end sticks A and G but didn't transfer to the sticks B and F, and therefore, the comparisons on the critical pairs were solved without resorting to the labeling strategies. Furthermore, the results of Experiment II, which was designed to examine the possibility of direct comparison based on the memory of the absolute lengths of the sticks, proved that children have better memory on the longest end stick than on the rest of the sticks. However, this memory may not be about the absolute size of the sticks, since children's estimation of the stick A averaged 14.55 cm the length between B and C. As for the rest of the sticks, it was found that the children did not remember the lengths clearly enough to distinguish between them. From this we can say that there were few possibilities that the visual feedback gave a clue for direct comparison on the inferential pairs.

Relating to the role of memory, the results shown in Table 1 was in accord with those of Bryant and Trabasso (1971), De Boysson-Bardies and O'Regan (1973), etc. in that the high scores of premise retention were followed by high inference scores, although there were differences between them in the number of significantly high scores of retention and inference. That is, in Bryant and Trabasso and De Boysson-Bardies and O'Regan, all the entries scored positively high above chance level on both the premise retention and the transitive inference, while in the present study only half of the entries scored high above chance level, which, it was considered, was caused by the fact that the seven stimulus sticks increased the load of retention. Nevertheless, it would be unwise to conclude from this that the performance on a transitivity task is a function of the retention of its premises. In order to explain the transitive performance as a function of the retention of its premises, it is also necessary to prove that those who were good at the retention of its premises also scored high on transitive inferences and those who were not scored low. It was with this aim that in the present study the subjects were classified into subgroups according to the scores on the premise comparisons.

A comparison of the transitive performances between the five and four year groups, based on Table 2 and Table 3, reveals the fact that in the five year olds, both the G-P and P-G subgroups scored higher than the P-P subgroup while in the four year olds, there were no significant differences among them. This means that the performance on the transitive pairs cannot be explained only by the quantity
of premise retention but affected by the age factor. However, it is not enough to base this claim on these evidences, as almost all of the scores shown in Table 2 and Table 3 were at the chance level except for the G-G subgroups. Perhaps, a good evidence, showing that the performance on transitivity tasks may not be always function of premise retention, would be a comparison between the G-G subgroups of the five and four year olds. Although the results of $\chi^2$-test showed that two age groups were equivalent in regard to premise retention, the comparison between the two G-G groups revealed clearcut difference in transitive performance, as is expected from the existence of interaction between the pair and age factors. That is, all the scores of the five year olds on the critical pairs (BD, CE and DF) were well above chance level, while the four year old children scored low on the critical pairs. These facts can be taken to disprove Bryant and Trabasso's claim that both the five and four year old children can make transitive inferences through the combination of two separate comparisons.

In conclusion, the evidences of Experiment I and Experiment II, relating to the non-transitive strategy, proved that there are few possibilities of solving the transitivity task by means of the absolute labels or lengths of each stick. If we conclude, from this, that the transitive comparisons were solved by means of transitive inferences (although there still remains the possibility of such a non-transitive strategy as was pointed out in the sequential-contiguity model), the evidences, that five year old children could solve the transitivity task when memory is controlled while four year olds could not, suggest that the ability to make transitive inference develops several years later from birth. Provided that the criticism concerned with age of appearance of transitive reasoning do not challenge the basic assumption of Piaget's account (Breslow, 1981), the results of the present study will be compatible with the Piaget's structural developmental theory rather than the Trabasso's hypothesis that there exist no developmental change in transitivity.

References


Trabasso, T. 1977 The role of memory as a system in making transitive inferences. In R. V.


(Received June 14, 1982; accepted July 14, 1984)