The role of context in young children’s memory performance: Istomina revisited

Wolfgang Schneider and Hedwig Brun

According to more recent studies on memory development in young children, preschoolers and kindergarteners are able to demonstrate surprisingly good memory skills in natural as well as in laboratory-type settings. This finding is not consistent with the results of a study by Istomina (1975), conducted in 1948, leading to the conclusion that (a) preschoolers do not use voluntary remembering, and (b) children generally recall better in play situations than in typical experimental settings. In this study, two experiments were conducted to replicate Istomina’s research. In the first, it was shown that Istomina’s findings were replicable when methodological problems in the procedure were ignored. Experiment 2 improved methodologically upon Istomina’s experimental methods and did not produce results to support her conclusions. Four- and 6-year-olds showed voluntary memory in play activities as well as in laboratory-type settings, and remembered equally well in both contexts. The results did not support the assumption that memory performance in young children can be substantially facilitated by motivating contexts.

Since the early 70s, the study of the development of memory in children has been one of the most active areas in cognitive development. The majority of empirical studies has focused on memory processes in grade-school children, whereas memory abilities in preschoolers and kindergarteners were rarely investigated until recently.

There were two reasons why the study of very young children’s memory abilities was largely ignored. (1) Many studies of memory development were designed as ‘modal memory experiments’ (cf. Brown & DeLoache, 1978). Younger children were only included to secure age effects, and their (mostly poor) achievement was used as a baseline against which the progress of older subjects could be demonstrated. This approach reveals little about young children’s capabilities (cf. Gelman, 1978). (2) Even those few early studies that concentrated on preschoolers’ and kindergarteners’ memory abilities provided little evidence that young children can deliberately remember. For instance, Myers & Perlmutter (1978) reported very poor recall abilities for 2- to 5-year-old children, and Appel et al. (1972) demonstrated that children neither distinguish between perceiving and memorizing nor deliberately attempt to remember before the age of 5 or 6.

This pessimistic view of young children’s memory abilities has not been supported, however, by more recent studies. Whenever more naturalistic task settings are used instead of traditional laboratory-type tasks, young children show impressive memory. This is particularly true for memory of location (cf. DeLoache et al. 1985; Wellman, 1986). Even 2-year-olds demonstrate intentional memory behaviour when they must remember where something is hidden (e.g. a big bird doll under a pillow in the child’s living room). Thus, deliberate remembering seems possible for young children when settings are familiar and simple strategies (e.g. visually attending) can mediate a task. Given these conditions, preschoolers can coordinate memory goals with mnemonic actions deliberately (cf. Paris et al., 1985).

There is also recent evidence that preschoolers and kindergarteners carry out deliberate memory activities in traditional laboratory tasks. For example, in a methodologically improved replication of Appell et al.’s experiment, Galbraith et al. (1982) found that kindergarteners were able to distinguish between ‘look’ and ‘remember’ instructions, as
indicated by significantly better memory performance in a 'remember' condition compared to a 'look' condition. Similarly, Baker-Ward et al. (1984) and Sodian et al. (1986) demonstrated that even 4-year-olds can behave strategically when asked to remember subgroups of toys (Baker-Ward et al., 1984) or clusterable objects (Sodian et al., 1986). In both studies, some 4-year-olds and most 6-year-olds used appropriate memory activities. Nonetheless, memory behavior had a direct impact on memory performance only with the 6-year-olds.

The generally positive evidence regarding young children's memory abilities found in more recent research, particularly that reported for laboratory tasks, contrasts sharply with conceptions of memory emphasized by Soviet researchers (e.g. Smirnov & Zinchenko, 1969; Yendovitskaya, 1971). The Soviet researchers argue that remembering during the preschool years will be best if it is embedded in a meaningful activity like play.

The best-known research supporting this point of view comes from an empirical study conducted by Istomina in 1948 (first published in English in 1975). Because of the importance of this study as context for the present research, we describe Istomina's research in some detail here. The particular objectives of Istomina's study were (1) to determine the conditions under which a child can distinguish between the goals of 'remembering' and 'recalling', and (2) to find out more about the forms in which retention and recall first appear as well as about the memory operations that affect recall.

Two different settings were created to test the assumption that recall processes initially take place as part of a broader, more meaningful activity. In one condition, the memory task was presented as a game. Children pretended to shop for five items named by the experimenter. The game was played in two adjoining rooms. Objects like a toy stove, a doll, a toy cash register and scales were used to create a 'kindergarten' in one room and a 'store' in the adjoining room. Some children took over the roles of store employees (salesclerk, cashier), whereas others played kindergarten (teacher, cook). The two experimenters participated as a store manager and a kindergarten principal. While playing kindergarten, subjects were asked to go on an errand to buy five items named by the experimenter (i.e. the kindergarten principal). Precautions were taken that these items were not identical to food items displayed in the store and that they were also not taxonomically clusterable. The items included candy, a ball, cereal, a carrot, milk, socks, a roll and butter.*

In the second condition of the experiment, similar lists of words were given in a different context. Children were asked to participate in a 'lesson'; they were presented with lists of five items and were required to remember the words after a 60-90-second time interval. As Istomina put it, memorizing was motivated in a totally different manner in this experimental condition: here, the goal of remembering was set by the experimenter, whereas in the game condition, the children had to set these goals by themselves. Istomina's hypothesis was that the game situation would be more familiar and also more motivating, and thus, young children would remember more in the game condition compared to the lesson condition. The children in the experiment were of four ages: 3–4, 4–5, 5–6 and 6–7 years. The same children participated in both conditions.

As Istomina expected, recall improved notably with age, and children's recall in the game condition was superior to lesson recall at every age level. Particularly poor memory was observed in the 3-year-olds, who recalled an average of about one word in the play condition and 0.6 words in the lesson condition. Informal and qualitative observations of

* Given that each child was presented only five items, we suspect that the exact list content varied from child to child. It should be noted that we tried to learn more about the experimental design by directly contacting Z. M. Istomina, who is now a professor at the Pedagogical Institute, Moscow University. Unfortunately, Professor Istomina has not yet answered our request.
the children's behaviour led Istomina to conclude that the youngest age group did not understand the goal of remembering the items to buy.

The 4-year-olds remembered an average of three words in the game condition, compared to 1.5 words in the lesson condition. The major behavioural difference between 3- and 4-year-olds was that the latter group seemed to make an effort to remember. Istomina concluded that voluntary memory first emerges between the ages of four and five.

The two oldest age groups (5–6 and 6–7 years) remembered an average of 3.2 and 3.8 words, respectively, in the game condition, compared to 2.0 and 2.3 words in the lesson condition. The objectives of recall and remembering were obviously clear to the older subjects, as indicated by their memory behaviour. They made active, purposeful attempts to encode and retrieve the words. Moreover, lip movements were frequently observed in both experimental conditions and taken as an indication of verbal rehearsal. However, although these children behaved similarly in the two experimental conditions, they performed less well in the lesson condition. Istomina’s explanation for this recall difference was that the game was intrinsically motivating, and that recall as a means of shopping was more apparent in the more natural context.

Although Istomina provided a very detailed description of her study, including a number of highly interesting analysis of individual cases, it is very difficult to reconstruct the design exactly because the procedures were specified vaguely. For example, we do not know if the sequence of sessions (experimental conditions) was actually counterbalanced. The only information about this point is that the two experiments were conducted 'simultaneously'. As the need for counterbalancing the order of activities was typically not realized in Soviet studies on memory development conducted in the forties and fifties (cf. Schneider & Pressley, in preparation, for a review), there is a high probability that the sessions were carried out in a fixed order. Another methodological problem was that different word lists were used in the two conditions, and that the word lists probably also differed between subjects. Furthermore, no statistical tests were provided. Finally, a closer look at the reports of individual subjects clearly shows that the procedure was not strictly standardized. Some children, particularly in the game condition, obviously asked the experimenter to repeat the items (cf. Records No. 7, 28, 38, pp. 24ff.), which the experimenter did (cf. Record No. 31). Thus, our suspicion is that the generally superior recall found for the game condition can at least partly be explained by methodological flaws systematically biased against the lesson condition.

In view of these obvious methodological problems and the influence of Istomina's research project (cf., Neisser, 1982; Rogoff & Mistry, 1985; Weissberg & Paris, 1986) it is surprising that no one has tried to replicate the Istomina procedure. A few experiments, however, by Wippich (1981, 1985) came close to replicating Istomina's study with German children. Wippich (1981) reported that a game task failed to produce better recall in preschool children than did a lesson situation. The only finding in favour of Istomina's position was that, when asked to predict their own recall, preschoolers were more realistic in the game situation than in the lesson situation. However, Wippich's study was not a close replication of Istomina's experiments, because (1) the dependent measure (serial recall) did not correspond to the free recall measure used by Istomina and (2) a between-subjects design was used (i.e. different children served as subjects in the game and lesson conditions). In addition, Wippich (1985) contrasted laboratory with game conditions for various memory paradigms (e.g. sort recall, recall readiness). In general, older children always performed better than younger children in these studies, but most critically, memory context never affected performance.

We decided to carry out research designed to replicate Istomina's own experiment and to improve upon her experimental methods. In particular, the first experiment reported here
was aimed at reconstructing the experimental conditions presumably used by Istomina including the shortcomings as we perceived them. The second experiment was designed to correct shortcomings of the original procedure. Here, the order of play and lesson conditions was counterbalanced, lists of items were counterbalanced across the two experimental conditions, and precautions were taken to provide identical, standardized instructional settings in the game and the lesson activities. Thus, the second experiment was thought to represent a more appropriate test of the context hypothesis.

Experiment 1
To see whether Istomina's results could be replicated, Expt 1 compared children's memory performance in an experimental setting that we think came close to Istomina's original design.

Method

Subjects. Sixty children from eight kindergartens participated in the experiment. There were two groups of children, with 15 boys and 15 girls in each. The mean ages of the younger and older groups were 4 years and 2 months (range 3'9-4'7) and 6 years and 1 month (range 5'6-6'6), respectively.

Materials. In the game ('grocery store') condition, we used a table with a toy cash register as a 'store counter'. A big store sign in front of the desk showed drawings of various food items that were different from those actually used in the experiment. Two word lists were used in the 'grocery store' as well as in the 'lesson' condition. List 1 was apples, pretzels, cheese, salt, tooth paste, pencil, handkerchief and candy. List 2 included carrots, chocolate, bread, sausage, sugar, eraser, laundry soap and cotton balls. Children were given toy money, a purse and a basket to do the shopping.

Procedure. All subjects were tested individually in the kindergarten. Children were told that they would be playing some games with the experimenter and her friend. In the game condition, the experimenter asked the children to go on an errand to the store to buy a list of items. They were told that the items would be read once to them, and that they had to listen carefully. Then the experimenter named the items slowly and carefully at three-second intervals. Next, children were given the purse with toy money and the basket to do the shopping. In the store, the experimenter's assistant waited with the children and chatted with them to make sure that 60 to 90 seconds elapsed between presentation of the items and recall. Then he asked them what they were sent to buy. When it appeared that the children could not remember more items, the store manager prompted them by asking 'Is there anything else?' Whenever a subject decided to get back to the experimenter (who was still in the room) and asked her the same question (What else was there?), the experimenter repeated the entire word list. After this, the child usually returned to the store manager who again chatted with the children before he asked them which items they wanted to buy. The items recalled and any signs of memory behaviour (i.e. rehearsal) were recorded. The children then paid for the purchased items and went back to the experimenter who waited for them.

In the lesson condition, children were seated on a small table next to the experimenter, who told them that they had to listen to a list of words, and that the task would be to remember as many as possible. As in the game condition, they were asked to listen carefully because the list would only be read once to them. After presentation of the item list, the experimenter chatted with the child to make sure that about 60 to 90 seconds elapsed between list presentation and recall. When it became obvious that children could not remember any more, the experimenter prompted them by asking 'is there anything else?' Again, words recalled and signs of memory behaviour were recorded. At the end of the experiment, the child was praised and thanked.

Results and discussion

Unless otherwise stated, all reported statistics were significant at the 0·01 level. The data was initially examined for sex and list effects. There were none, and thus, the data were collapsed across these variables.

A 2(age) x 2(order of activities) x 2(activities) analysis of variance on the recall data yielded main effects of age \((F=23·98, \text{d.f.}=1,56, \text{P}<0·01)\), and activity \((F=76·97, \text{d.f.}=1,56, \text{P}<0·01)\). On average, 6-year-olds recalled more items than 4-year-olds \((4·27 \text{ vs.} 2·55)\), and memory performance in the game condition was significantly better than performance in the lesson condition for both age groups \((4·51 \text{ vs.} 2·35)\). No further main effects or interactions proved statistically significant \((\text{all }\text{Ps}>0·20)\).

In short, we were able to replicate Istomina's major findings. Even the proportional
differences between play and lesson activities were remarkably similar: whereas Istomina found that recall was about 1.8 times better in the game condition, the ratios found in the present study varied between 1.5 and 2.2, depending on age and order of activities. According to our assumptions, order of activities and the not strictly controlled instructional setting could have contributed to the consistently better recall in the game condition. Although recall always tended to be better for the task presented first, the main effect and interactions obtained for the order of activities factor did not prove statistically significant.

However, the liberal instructional setting, i.e. the possibility to get back to the experimenter and ask her to repeat the items, appeared to have different impact in the two experimental conditions. In the lesson condition, subjects did not find it necessary to ask for a repetition, whereas 14 4-year-olds and 24 6-year-olds in the game condition did so after the store manager asked them if there was anything left. Apparently, in the more motivating play condition, children felt a greater need to remember all items and therefore tried harder than in the lesson condition. Not surprisingly, those who asked the experimenter to repeat the items also recalled more than those who listened only once to the stimulus list. While the difference was significant for the older children (t = 4.49, d.f. = 28, P < 0.01; means 6.00 vs. 3.32), it approached significance for the younger subjects (t = 1.90, d.f. = 28, P < 0.10; means: 4.13 vs. 2.93). Further evidence for the impact of repetition on recall was derived from a separate 2(age) x 2(order of activities) x 2(activities) analysis of variance on children's initial recall (i.e. prior to any repetition by the experimenter). While there was a significant main effect of age (F = 10.62, d.f. = 1,56, P < 0.01), no further main effects or interactions proved statistically significant. Most importantly, the effect of activity was negligible (F = 0.61, d.f. = 1,56, P > 0.20). Thus initial recall was comparable for the lesson and game conditions. We concluded from this that a major weakness of Istomina's experiment concerned the instructional setting. Better recall in the game condition may have been related to additional repetitions of the stimulus list that subjects requested in the play compared to the lesson condition.

In order to explore this hypothesis more thoroughly, a second experiment was carried out.

**Experiment 2**

The only difference between Expt 1 and Expt 2 concerned the instructions given to the subject. In both experimental conditions of Expt 2, the children were told that they had to remember a list of items which the experimenter would read to them only once. In addition to the instruction given in Expt 1, the experimenter explicitly told them that they should listen very carefully because the list would not be repeated. In the game condition, they were told that the rules of the games were that they could only return to the experimenter after the shopping was completed.

Another 60 children participated in this experiment. Half were about 4 years old (mean age = 4.3, range 3.10 to 4.7), and half of them about 6 years of age (mean age = 6.2, range 5.9 to 6.7). There were equal numbers of boys and girls in each group.

**Results and discussion**

Again, preliminary analyses revealed no significant effects of sex and lists and the data were collapsed across these variables. A 2(age) x 2(order of activities) x 2(activities) analysis of variance on the recall data yielded significant main effects of age (F = 7.56, d.f. = 1,56, P < 0.01), but no effect of activity (F = 3.56, d.f. = 1,56, P > 0.05). On average, 6-year-olds recalled more than 4-year-olds (2.61 vs. 1.85). Although performance tended to be better in the game condition, regardless of age, the overall difference (2.45 vs. 2.11) was no longer
significant. A significant interaction between order of activities and the amount recalled in the two experimental condition \( (F=11\cdot12, \text{d.f.}=1,56, P<0\cdot01) \) indicated that, regardless of age, better recall was obtained for the experimental task that was presented first.

The data from Expts 1 and 2 as well as Istomina's data are shown in Table 1. The effect of the strictly standardized instruction used in Expt 2 is obvious: explicitly mentioning the fact that the word lists would not be repeated affected memory performance particularly in the play condition. In both age groups, average recall in this condition was significantly lower than the corresponding scores obtained in Expt 1. On the other hand, no significant differences were found when recall in the lesson condition of the two experiments were directly compared. As indicated in Table 1, memory performance was still slightly superior in the game condition. However, the play recall/lesson recall ratios were substantially lower than those obtained in Istomina’s study and in Expt 1.

<table>
<thead>
<tr>
<th>Age</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lesson</td>
<td>Game</td>
</tr>
<tr>
<td>4</td>
<td>1·05</td>
<td>1·63</td>
</tr>
<tr>
<td>6</td>
<td>2·15</td>
<td>3·06</td>
</tr>
</tbody>
</table>

*Eight items per set were used in our experiments while Istomina used 5.

Data were collapsed across the 3–4- and 4–5-year-olds on the one hand and the 5–6- and 6–7-year-olds on the other hand.

Note. MSE (56) = 2·33 (Expt. 1); MSE (56) = 3·48 (Expt.2).

Thus, the major conclusion to be drawn from this experiment is that we failed to repeat Istomina’s findings when we introduced stricter experimental controls than those used in her study. This confirms our hypothesis that Istomina’s findings were due in part to methodological shortcomings.

**General discussion**

There were two major reasons for replicating Istomina’s widely cited study on voluntary memory in young children. Since it was conducted with Soviet children about 40 years ago, it was far from certain that the same results would be obtained with German kindergarten children raised in 1980s. In addition to possible differences due to history and culture, several methodological problems were identified in Istomina’s experimental design that could have biased her findings. She did not counterbalance the order of activities, and word lists were only repeated to those children who explicitly asked for repetition.

In our first experiment, we tried to provide a close replication of Istomina’s procedure, and we were successful in repeating Istomina’s findings. However, this was probably due in part to the fact that the play and lesson conditions differed in more than one respect. More children in the play than in the lesson condition asked the experimenter to repeat the word list and consequently recalled more items than the other subjects did. The effect of this methodological flaw was assessed in Expt 2 where the procedures were more directly
comparable across the two experimental conditions. On average, recall dropped in the play condition of Expt 2 in comparison to the play condition of Expt 1, regardless of age. Thus, there were non-significant differences between the game and lesson conditions in Expt 2. In short, there was a failure to replicate Istomina’s research after methodological problems of her study had been removed. Although the majority of our children seemed to be more interested during the play than during the lesson activities, differences in context and motivation did not substantially influence recall.

It should be noted that there is additional confirmation for the validity of our findings that was brought to our attention after our study had been completed. Weissberg & Paris (1986) did a similar replication study of Istomina’s research with 3–7-year-old American children. In that study, the methodological shortcomings of Istomina’s experiments were corrected by counterbalancing the order of activities and by using the same words for all subjects. Further, in addition to the shopping situation, a ‘party’ situation was created and set up as a game versus a deliberate lesson activity. The main result of the Weissberg & Paris experiment was a failure to replicate Istomina’s findings. In contrast to our results, however, children given the lesson condition showed better recall than those given the game condition, regardless of age. Weissberg & Paris also found generally higher levels of performance than either Istomina or we did. For example, 3–4-year-olds remembered an average of 2.8 words in the lesson condition, compared to 0.6 words reported by Istomina. Similarly, the 6–7-year-olds in the Weissberg & Paris study recalled almost twice as much in the lesson condition than did their Soviet counterparts in Istomina’s study (4.5 vs. 2.3).

Although the reasons for these pronounced recall differences are not entirely clear, there are at least two possible reasons for the differences observed: (1) Weissberg & Paris presented the word list twice, whereas the lists were only given once in Istomina’s research and in our study. Results from our pilot work indicated that young children indeed did much better when the word lists were repeated to them. (2) Weissberg & Paris reported that a large proportion of their subjects exhibited some form of verbal rehearsal. Interestingly, nearly half of the 3-year-olds and even more of the older children rehearsed words spontaneously, particularly in the lesson conditions. As it could be shown that use of rehearsal obviously corresponded with better recall, it is not surprising that memory performance was generally superior in the lesson conditions.

The analysis of verbal rehearsal in the study by Weissberg & Paris provides some support for the assumption that even 3–4-year-olds are capable of voluntary memory. This finding clearly contradicts Istomina’s assumption that young children do not understand the goal of remembering per se, and is also in accord with the more optimistic view of preschoolers’ memory abilities given by Baker-Ward et al. (1984) and Sodian et al. (1986). However, given our results, one might question the generalizability of these findings. For example, Weissberg & Paris argue that contemporary American preschoolers are more familiar with deliberate memory tasks and the request to remember than Soviet children were 40 years ago, mainly because of the influence of television, preschool education, and parental behaviour. However, although these influences seem to operate on German children as well, verbal rehearsal was only rarely observed in our sample (in total, only three children rehearsed the items). Again, we can only speculate about the underlying causes for the differences found between the American and German samples. Of course, there is the possibility that Weissberg & Paris recruited a sample of very bright children, whereas our subjects were of only average intelligence, but IQ was not assessed in either study. There might be also cross-national differences in preschool education that could have contributed to the differences in memory behaviour. Finally it could also be that the repeated presentation of the word lists in the Weissberg & Paris study elicited rehearsal activities in their subjects.
In spite of these differences, however, both replication studies clearly indicate that the conclusions Istomina drew from her research are no longer valid. While the context of a play activity may be more interesting and motivating for most subjects, there is no evidence that the more favourable context also leads to better memory performance, compared to a laboratory-type memory situation. The results of Weissberg & Paris as well as those obtained in our experiments clearly show that even 3–4-year-olds can do at least equally well in a laboratory-type lesson condition. Hence, differences in internal conditions like active involvement in the task, which were also observable in our study, did not lead to differences in memory performance, as Istomina hypothesized.

Compared to the findings of Baker-Ward et al. (1984) and Sodian et al. (1986) where the stimulus material was presented visually and for a longer period of time, overall memory performance was relatively poor in the present study, where learning materials were presented acoustically and only once. Apparently, the experimental procedures used in our study and that of Istomina are not very appropriate for demonstrating the memory skills of preschool children. To take full advantage of their memory abilities, young children obviously need pictorial cues and more time or trials to encode the incoming information. Thus, we suspect that preschoolers can show considerably higher levels of recall in this task whenever the experimental design allows for acoustical and visual encoding of the learning material, and also gives sufficient time to encode the information. From the empirical evidence presented in this study, however, there is no reason to assume that presenting the items in the context of a play or a lesson activity would make a difference in the direction that Istomina expected.

Acknowledgements

We are grateful to Beth Kurtz and Michael Pressley for helpful comments and suggestions.

References


Received 9 December 1986; revised version received 10 February 1987

Requests for reprints should be addressed to Wolfgang Schneider, Max-Planck-Institute for Psychological Research, Leopoldstrasse 24, D-8000 München 40, West Germany.

Hedwig Brun is also at the above address.