Development of Problem Solving Strategies: Strategy Shift in the Tower of London Task

Richard Cooper (R.Cooper@psyc.bbk.ac.uk)
Department of Psychology, Birkbeck College, University of London, Malet St., London, WC1E 7HX

Rikke Coster Waldau (awald04@students.bbk.ac.uk)
Department of Psychology, Birkbeck College, University of London, Malet St., London, WC1E 7HX

Abstract

Problem solving performance of 3–4 year olds and 5–6 year olds was tested on two types of Tower of London problem: problems with determinate subgoal ordering and problems with ambiguous subgoal ordering. Children performed more poorly on problems with ambiguous subgoal ordering, as expected from previous research (e.g., Klahr & Robinson, 1981; Klahr, 1985). In addition, however, it was found that the difference in performance on the two types of problem was significantly greater for older children, suggesting that their problem solving strategies are more sensitive to a problem's subgoal structure. We report a series of computational models which explore strategies which the children might employ. Younger children's performance is best modelled by a hill-climbing approach with a single move look-ahead. In contrast, older children appear to employ a rudimentary form of Means-Ends Analysis.

Introduction

It has long been known that children as young as 12 months can move one object out of the way in order to gain access to a second object (Gratch, 1975). Klahr & Robinson (1981) point out that this ability contrasts with the problem solving behaviour of significantly older children when confronted with simple puzzles, such as the 2-disk version of the Tower of Hanoi (ToH). Successful solution of this problem requires moving the smallest disk out of the way in order to access the larger disk, which may then be moved directly to the goal peg. The smaller disk can then be placed on the larger disk, hence solving the problem in just three moves.

Klahr & Robinson (1981) attribute the difficulty which young children have with such tasks primarily to presentational aspects (that is, the children fail to understand the goal of the task and the rules imposed). Donaldson (1978) similarly notes that the apparent inability of young children to solve simple artificial problems could be attributed to the fact that the task environment, being outside of their normal experience, was alien to them.

In order to explore the development of problem solving strategies in preschool children Klahr & Robinson (1981) designed a version of the ToH task that was more "child-friendly". The task environment was modified so that the disks were replaced by three different sized cans placed upside-down on the pegs. This arrangement prevented a smaller can from being placed on a larger can (because the smaller can would have to contain the larger can). The revised task also included a cover story, where each individual can represented an individual member of a monkey family (a daddy monkey, a mommy monkey and a baby monkey). This story was intended to make the problem simpler for the young children to understand.

Klahr & Robinson's task eliminated several difficult aspects of the standard ToH. Children as young as 4 were able to solve some four-move problems, and development differences in ability were seen. However, the task introduced new complications. For each problem, the children were asked to verbally present complete plans that would lead to the solution of the problem. This was intended to prevent children from generating random sequences of moves. However, age-related differences in verbal abilities were not considered. It is likely that having to produce verbal plans would have disadvantaged the youngest children. Donaldson (1978) points out that developmental differences in language understanding could be one explanation of young children's limited ability to draw explicit conclusions. Asking children to give complete solution plans without moving disks also forces them to visualise how the situation changes with each successive move and construct a full plan in working memory. Developmental differences in memory capacity will thus further disadvantage the youngest children. Finally, it is questionable whether the cover-story, while engaging, was a help or hindrance to the children's problem solving. The cans were in no way similar to monkeys, and thus additional overheads relating to demands placed on the children's imaginations were likely to have affected performance.

The Klahr & Robinson task is therefore not ideal for evaluating the development of problem solving strategies. Klahr (1985) considered an alternate task: the Dog-Cat-Mouse task, in which three animals had to be moved along designated tracks to specified locations. Klahr found no clear developmental differences on performance of the task in children ranging in age from 3;9 to 5;10. (He was, nevertheless, able to provide a very plausible computational account of the children's behaviour.) Thus, although the task provides evidence for certain kinds of strategies (e.g., 2-move look-ahead, a reluctance to back-track, etc.), it provides little data on how such strategies evolve through development.

One task related to the ToH that may be better suited to children's communicative abilities is the Tower of London (ToL). The ToL (introduced by Shallice, 1982) shares many characteristics with the more familiar ToH task. However, it differs in several respects. Three coloured balls (red, green and blue) are used instead of the three different sized disks, and the three pegs are of different heights. The height of each peg limits the number of balls that may be placed on that peg. The left-most peg has space for three balls, the centre peg for two balls, and the right peg for just one ball. The constraint in ToH relating to disk size (that no larger disk be placed on a smaller disk) is replaced in ToL by a constraint enforced by the different sizes of the pegs.

Notwithstanding the similarities between the two tasks, there are significant differences. Firstly, the tasks are not isomorphs: there are 27 states in the ToH problem space, and 36 in the ToL problem space. More importantly for the current study, the principal constraint on task completion is imposed by the structure of the apparatus. Whereas, in the standard ToH, it is physically possible to place a large disk on a small disk, it is not, in the ToL, physically possible to place (for example) three balls on the centre peg. This difference, which parallels Klahr & Robinson's (1981) use of cans in place of disks, makes the task more suitable for young children. Anderson *et al.* (1996) reported ToL studies with children aged 7 and above. Pilot work revealed, however, that much younger children find the brightly coloured balls appealing, and it is relatively easy with the apparatus to engage the attention of children as young as 3 years of age. Use of the ToL in the investigation of young children's problem solving strategies therefore overcomes some of the difficulties arising the various versions of the ToH.

Shallice (1982) developed the ToL task to assess planning deficits in neurological patients. The patients were tested on a battery of 12 tasks, starting with relatively simple two move problems and progressing to rather taxing five move problems. In the analysis, Shallice (1982) equated task difficulty with the number of moves required to solve a task. The work of Klahr & Robinson (1981) on the three disk ToH suggests that this measure of task difficulty is rather coarse. Klahr & Robinson distinguished two classes of ToH problem: "tower-ending problems", in which the goal state is a tower involving all disks stacked in the correct order on a nominated peg, and "flat-ending problems", in which the goal state is one in which each peg contains a single disk. One principal finding of Klahr & Robinson was that performance was significantly better on tower-ending problems than on flat-ending problems. The difference in performance appeared to be independent of age.

A computational analysis reveals that tower-ending problems and flat-ending problems differ in terms of the ordering constraints placed on their subgoals. Both types of problem can be solved by Means-Ends Analysis (MEA), breaking the task into subgoals and then solving the subgoals in sequence. Tower-ending problems, however, have a clear ordering on their subgoals: the largest disk must be placed on the nominated peg before any other disk is dealt with. The subgoal structure of flat-ending problems is both more flexible and less obvious. On first glance it would appear that the disks could be moved to their target pegs in any order. In fact, the smallest disk should not be moved first, but either of the other disks may be moved to their target positions first.

The tower/flat distinction applies to the ToL task in much the same way as it does to the ToH task. The difference amounts simply to replacing disk size by ball colour. One might therefore anticipate that Klahr & Robinson's (1981) ToH results would extend to similar ToL tasks. In particular, it would seem to follow that children would be poorer on flat-ending ToL problems than on tower-ending ToL problems, and that young children should be poorer than older children on ToL problems of both types. Another hypothesis, however, is plausible. If the developmental progression is from semi-random trial-and-error (constrained, perhaps, by limited look-ahead and a reluctance to backtrack, as suggested by Klahr (1985)) to more goal directed strategies (such as MEA), then one would expect such a progression to interact with problem type. On this view the difference in difficulty between tower-ending problems and flat-ending problems should be sensitive to age. Young children, using strategies that don't involve subgoaling, should perform equally poorly on tower-ending problems and flat-ending problems. Older children, who have developed strategies sensitive to subgoal structure such as MEA, should do significantly better on problems where the subgoal structure is clear (towers) than on problems where the subgoal structure is ambiguous (flats).

In the remainder of this paper we report an experiment which investigates the above position. This is followed by a brief presentation of a series of computational models which attempt to capture the differences in problem solving abilities exhibited by younger (3–4 years) and older (5–6 years) children.

Experiment

Method

Seventeen 3–4 year olds and seventeen 5–6 year olds participated in the study. The children were drawn from inner-London nursery and primary schools and were of similar socio-economic status. Children were not screened for specific abilities or disabilities, but colour vision was informally assessed. All children were able to name the colours of all balls.

In the familiarisation phase all subjects completed four ToL problems to a criterion level. Subjects were then required to solve six critical problems. Their performance was videotaped for later scoring. The familiarisation problems were all two and three move problems involving non-tower/non-flat goal configurations. The critical problems consisted of one three move, one four move, and one five move tower-ending problem, and one three move, one four move, and one five move flat-ending problem. The problems were designed such that the same start configuration was used for tower-ending and flat-ending problems requiring the same number of moves. The order of presentation of critical tasks was randomised. Each child was required to signal when she/he felt each task was complete by shaking a toy parrot (and making it squawk).[1]

The ToL apparatus was made to standard specifications (Krikorian, *et al.*, 1994). For each task, the experimenter set the problem up by placing the balls in their initial positions. Photographs of goal states were then used to convey the problem to the children.

Results

The raw data on which the analysis is based consisted of detailed transcripts of the video tape recordings of all sessions with all children. These transcripts included information about completion, problem space route taken, number of moves, violation of rules, the manifestation of these violations and any noticeable comments made by the children. The data were analysed in two ways. Firstly, a course measure of performance focussing just on successful completion was determined. A more fine-grained analyses, in which solution attempts were graded along several dimensions, was then performed. Further details of the results, including an analysis of first moves, are contained in Waldau (1999).

Task Completion Each solution attempt was scored for completion. A solution attempt was judged to be complete if the goal state was achieved without breaking any rules. This measure did not require children to minimise their moves. Table 1 shows the proportion of children who, on this measure, solved each problem.

It can be seen that the older group completed more problems than the younger group of children. It is also apparent that performance on tower-ending problems was generally better than performance on flat-ending problem. However the poor performance of the younger age group is suggestive of a floor effect. In this group, each child solved, on average, only 1.12 problems out of a possible 6. Given the apparent floor effect, the between group difference was analysed with a non-parametric test. A Mann-Whitney test on the number of problems completed by each child revealed that the effect of age was highly significant ($W = 216.5, p < 0.005$).

Overall Scores The second scoring procedure involved detailed consideration of factors that were considered to influence the children's performance. There is no standardised procedure for scoring the performance of pre-school children on the ToL task. An appropriate comprehensive scoring method was therefore developed. Several factors appeared to be relevant: configuration matched; colours matched (0, 1, 2 or 3); rule violations; minimum number of moves.

The children's performance on each problem was scored on a scale of zero to six based on these factors. One point was scored for each factor, with up to 3 points being awarded for colours matched. Any solution that was therefore complete (matched configuration and all colours, without rule breaks) and in the minimum number of moves, scored 6 points. A solution attempt that failed on all of these factors scored 0 points.

[1] Pilot work revealed that it was not always clear when children felt they had solved the tasks. The toy parrot proved to be a popular and reliable way of allowing the child to signal this.

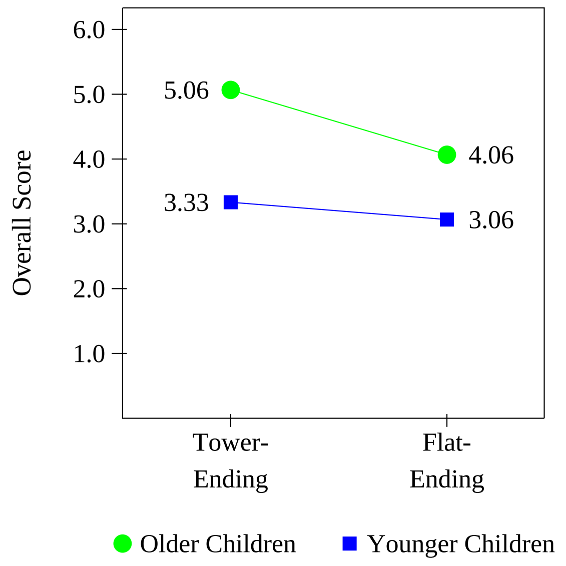

Figure 1: Mean overall score, as a function of problem type and age group

This scoring system allows performance to be graded, such that incorrect or incomplete solutions may contribute usefully to the analysis. The system also yields interval scale data that may be analysed with parametric statistical tests.

This scoring system disclosed some interesting differences. The means (table 2) indicate that differences between each group may be related both to problem difficulty (i.e., minimum number of moves) and to the configuration (i.e., tower-ending or flat-ending) of the individual problems. The younger children performed much more poorly than the older group. For both groups there appears to be little difference between level of difficulty and configuration on 3 move and 4 move problems whereas the performance of the younger children on the 5 move problem is substantially lower both on tower-ending and flat-ending problems, than the performance of the older children. It also appears that both groups are affected by problem configuration, with flat-ending problems being solved less successfully than tower-ending problems. This can be seen most clearly in figure 1, which shows the effect of problem type on mean overall score for each age group.

The data were analysed using a 3-way analysis of variance for mixed designs ($2{\times}3{\times}2$) to examine the effect of age, difficulty and configuration. Significant effects were found for all three variables. There was a significant effect of the between group variable age ($F(1, 32) = 14.66, p < 0.001$), substantiating the effect found in the overall completion scores. The

Configuration	Tower-Ending			Flat-Ending		
Difficulty	3 move	4 move	5 move	3 move	4 move	5 move
Older children	0.71	0.71	0.65	0.47	0.53	0.47
Younger children	0.24	0.29	0.18	0.12	0.24	0.00

Table 1: The proportion of children completing each problem by age, difficulty and configuration

Configuration	Tower-Ending			Flat-Ending		
Difficulty	3 move	4 move	5 move	3 move	4 move	5 move
Older children	4.94 (0.75)	5.18 (1.38)	5.06 (1.20)	3.71 (1.70)	4.47 (1.28)	4.00 (1.73)
Younger children	3.47 (1.28)	3.53 (1.74)	3.00 (1.62)	3.24 (1.30)	3.65 (1.70)	2.30 (1.57)

Table 2: Means (and standard deviations) of overall scores by age, difficulty and configuration

effect of the within group variable difficulty was also significant ($F(2,64) = 4.10, p < 0.025$), as was the effect of the within group variable configuration ($F(1,32) = 14.70, p < 0.001$).

No significant interaction was found between age and difficulty ($F(2,64) = 2.34$), suggesting that the effect of difficulty is the same in younger and older groups. Similarly, no interaction was found between difficulty and configuration ($F(2,64) = 1.79$) or between age, difficulty and configuration ($F(2,64) = 0.54$). There was, however, a significant interaction between configuration and age ($F(1,32) = 4.76, p < 0.05$). This interaction (see figure 1) indicates that the effect of problem type is dependent on age. In particular, children's performance on tower-ending problems appears to increase with age, but this increase is not matched by a similar increase in performance on flat-ending problems. Post-hoc analysis of the configuration/age interaction revealed that the largest contributing factor was number of moves: older children tended to produce minimal move solutions for the tower-ending, but not the flat-ending, problems; younger children rarely produced such solutions. Colour also contributes substantially. Rule breaks and configuration were not contributing factors.

Discussion of results

The effect of age, seen in both the completion scores and the overall scores, demonstrates that the task is sensitive to age. In this sense it provides a better developmental measure than Klahr's (1985) Dog-Cat-Mouse task, were no age-related effects were found.

The effect of difficulty (i.e., number of moves required) seen in the overall scores provides some support for Shallice's (1982) categorisation of problem difficulty in terms of number of moves. However, the effect appears not to be monotonic: children in both groups performed better on 4 move towers and flats than on the equivalent 3 move problems. These effects are likely to be due to confounding factors relating to the specific start positions used in the 4 move problems. The starting configurations in both 4 move problems allowed only two possible first moves. Four first moves were possible in the other problems.

The effect of configuration, i.e., tower-ending or flat-ending, supports our contention that ToL problems provide similar ambiguity of subgoal ordering to that found in ToH problems.

Given previous research, none of the above effects are particularly surprising. However, the interaction between age and configuration is of some import. It suggests that older children employ strategies which rely upon subgoal ordering, such as MEA, and the utility of such strategies is significantly impaired when subgoal ordering is ambiguous. Younger children, who do not use such strategies, perform at a generally lower level, but are not sensitive to the tower/flat distinction.

The factors contributing to the overall scores, and hence to the interaction, deserve further consideration. As noted above, older children tended to produce minimal move solutions for tower problems but not flat problems. Younger children generally failed to produce minimal move solutions, irrespective of goal configuration. This further supports the view that older children have goal-directed strategies such as MEA in their problem solving repertoire. By applying such strategies they are able to solve tower-ending problems efficiently — significantly more efficiently than younger children. The strategies are of little assistance when solving flat-ending problems, however, and both young *and* old children exceed the minimum number of moves required on these problems.

A further feature of the behaviour of all children was a tendency to arrange the balls in the correct configuration (i.e., tower or flat), whilst disregarding their colour. This tendency also contributed to the interaction. The number of colours correctly matched by the younger children was not dependent upon problem type, but older children were more likely to get three colours (and hence configuration) right in tower-ending problems than in flat-ending problems. We suggest that when a problem is perceived as difficult both groups of children disregard colour and take the easier option of aiming to get the configuration right. For the younger group this pattern occurs both in tower-ending and in flat-ending problems. For the older group, the pattern only occurs in flat-ending problems. Again, this is consistent with older children using subgoal-sensitive strategies such as MEA, which break down when applied to problems lacking a clear subgoal ordering (such as flat-ending problems).

The tendency to disregard colour, by younger children in all problems and by older children in flat-ending problems, suggests that children perceive colour and configuration to be separate aspects of the ToL task. The two features dissociate, and it appears that children either attach less importance to colour than to configuration, or perceive colour to be more difficult to match than configuration. Consequently, in difficult problems, they are willing to settle for achieving a configuration match. This is consistent with Klahr & Robinson's (1981) suggestion that, when faced with a problem that is too difficult, children will focus on that aspect of the problem which is simpler.

Rule breaks were recorded because earlier research had suggested that younger children were more likely to resort to breaking the rules if faced with difficulty. However the incidence of rule breaks does not appear to contribute to the interaction. In fact, the older children tended to break the rules more frequently than the younger children. Thus, breaking of the task rules is not indicative of development differences. This finding vindicates our use of the ToL task with very young children: the task rules do not appear to discriminate against the younger children.

Models

In order to better understand the children's behaviour we have developed a series of models of the ToL situation employing a variety of heuristics. The simplest models use a random move selection strategy. The most complex model employs a primitive version of MEA. In this section we describe the features and behaviours of these models which, in the context of the above empirical results, reveal most about the possible computational mechanisms underlying children's behaviours.

Simulation of children's performance on the ToL task raises two immediate difficulties: simulation of rule breaks and determination of when the task is finished. As noted above, rule breaks occurred at similar levels across problem types and across age groups. Although there appeared to be some identifiable conditions triggering rule breaks, we do not attempt their simulation. Determining task completion is more difficult. Children frequently signalled task completion when all three balls were not in the desired positions. Again though, we do not attempt simulation of incorrect assessment of task completion. Instead, we focus primarily on the ability to actually solve each task, and the number of moves required.

Model 1: Move at Random

Children's performance on the ToL is clearly non-random. A model in which moves are selected at random, even when backtracking is prohibited, performs poorly in comparison to either group of children. The principal behavioural difference with such a model and the children concerns number of moves to solution: the random model may make several hundred moves before stumbling upon the goal state. In addition, the random model frequently approaches a solution, getting one or two balls in place, before moving away again. The model confirms that even the youngest children are far from random in their approach to the ToL.

Model 2: Hill-Climb at Random

One obvious refinement to the move at random model involves the inclusion of a naïve hill-climbing constraint: if a ball is in the correct position, don't move it away (unless a ball underneath is not in the correct position). This model does not deliberately move up-hill. Rather, it moves at random with the constraint of not moving down-hill. Thus, the model moves at random until one ball is in place. Once a ball is in place, it is not moved again.

This model fares better than the move at random model. Children are reluctant to move balls out of position (suggesting a tendency towards hill-climbing), but their initial moves appear less guided. This is apparent within the model from the number of moves that are frequently made before a single ball is in its correct position.

A second positive feature of this model concerns the occurrence of "dead-end states" — states in which the solution requires moving away from the goal (i.e., moving balls that are in their correct positions out of position). This model is unable to escape from such states. Observational evidence suggests that the occurrence of such states in children's problem solving (at all ages) can trigger rule breaks.

Model 3: Hill-Climb with Look-Ahead

Addition of minimal look-ahead to the hill-climbing model, in terms of a comparison of the states resulting from each possible move, radically improves the model's performance. This type of model, with a look-ahead of just one move, outperforms most children, even in the older age group. Its performance could be reduced to that of the older children with the introduction of, for example, random errors due to attentional lapses, but the model fails to discriminate between tower-ending and flat-ending problems. Unlike the older children, its performance on both types of problems is equivalent.

The superior performance of the single-move look-ahead model contrasts with Klahr's (1985) findings relating to young children's performance on the Dog-Cat-Mouse puzzle. Klahr found that two-move look-ahead was required in order to adequately model children's performance. Notwithstanding this difference, the hill-climb with look-ahead model, supplemented with occasional random moves, is capable of providing a good account of the younger children's performance. Of particularly significance is that fact that, like the hill-climb at random model, it falls prey to the problem of dead-end states — a problem which, as noted above, seems to characterise children's performance at all ages.

Model 4: Rudimentary Means-Ends Analysis

None of the above models are sensitive to the tower-ending/flat-ending distinction. The inclusion of rudimentary MEA leads to a model in which this difference does have behavioural consequences.

The rudimentary MEA model solves ToL tasks by comparing the current and desired states, selecting a difference between those states, and then eliminating the difference. The first differences to be eliminated are those relating to balls whose desired position is at the foot of any peg. Once these differences have been eliminated, any ball whose desired position is one above the foot of a peg becomes a candidate difference to be eliminated. Finally, any ball whose desired position is two above the foot of a peg becomes a candidate difference to be eliminated. Flat-ending problems have multiple initial differences to be reduced. Tower-ending problems have at most one difference to be reduced at each stage.

This rudimentary form of MEA leads to the successful solution, in minimal moves, for each tower-ending problem on which the children were tested. The strategy does not lead to minimal move solutions for the flat-ending problems, thus capturing one of the key between-task differences exhibited by the older children. The strategy is also plausible in terms of working memory demands. The subject needs maintain at most one goal — to eliminate a nominated difference by moving a specific ball to a specific position — in working memory at any time. If that goal cannot be immediately realised, it is either because the ball to be moved is not clear, or because another ball occupies the destination position. In either case, for the problems considered in this study, it is possible to move the blocking ball(s) out of the way without further problem solving.

General Discussion

We have suggested that children aged 5–6 years rely on problem solving strategies that are qualitatively different from those of their younger peers. In the ToL problems investigated here, this qualitative difference leads to an interaction between age and problem type on an aggregated measure of problem solving success. We now highlight three significant issues arising from this work.

First, from a practical perspective, the Tower of London task proved to be an appropriate task for investigating young children's problem solving strategies. Children as young as 3 years of age were engaged by the task, and were able to solve some of the simpler problems. They also did not break the task rules more frequently than older children, indicating that any differences in performance could not be attributed to differences in understanding the constraints imposed by the task. Furthermore, although the more complex problems were beyond the capabilities of our youngest subject, it was still possible to score solution attempts in a meaningful way. The procedure employed also allowed for the children's own assessment of task completion to be considered.

Second, and as a consequence of our scoring procedure, two primary factors contributing to the interaction between age and problem type were identified. Older children tended to employ optimal solution plans in solving tower-ending problems. Their solution plans for flat-ending problems, and the solution plans of younger children's attempts at both types of problem, were non-optimal. In addition, children generally tended to concentrate more on configural aspects of the solution than on matching the colours of balls. The one exception, where both colour and configuration were systematically matched, occurred in older children's solution plans to tower-ending problems. Taken together, these factors lead to significantly superior performance of the older children on tower-ending problems. This contrasts with the performance of younger children, which was found not to differ between tower-ending and flat-ending problems.

Third, the two age groups may be characterised by different problem solving strategies. On the basis of a series of computational models, it appears that the older children deploy a rudimentary form of Means-Ends Analysis. It is this strategy which leads to their superior performance on tower-ending problems. Younger children, on the other hand, appear to employ more limited hill-climbing strategies. These strategies are not sensitive to the differences in problem type.

Acknowledgements

We are grateful to the staff and children of Marquess and Willow Under-Fives Education Centres and Newington Green Primary School for their involvement with this work.

References

Anderson, P., Anderson, V., & Lajoie, G. (1996). The Tower of London task: Validation and standardization for pediatric populations. *The Clinical Neuropsychologist*, *10*, 54–65.

Donaldson, M. (1978). *Children's Minds*. Fontana Press, London, UK.

Gratch, G. (1975). Recent studies of Piaget's view of object concept development. In Cohen, L. B., & Salapatek, P. (Eds.), *Infant perception: From sensation to cognition*, pp. 51–102. Academic Press, New York, NY.

Klahr, D. (1985). Solving problems with ambiguous subgoal order: preschoolers' performance. *Child Development*, *56*, 940–952.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology*, *13*, 113–148.

Krikorian, R., Bartok, J., & Gay, N. (1994). The Tower of London procedure: A standard method and developmental data. *Journal of Clinical and Experimental Neuropsychology*, *16*, 840–850.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London*, *B298*, 199–209.

Waldau, R. (1999). A developmental study of problem solving in children aged 3–6: Development of planning strategies. Undergraduate dissertation. Department of Psychology, Birkbeck College, University of London.