# The logic of memory search in non-human primates (*Cebus apella*)

Carlo De Lillo

PhD

The University of Edinburgh

1994

# ABSTRACT

This thesis focuses on the study of the spontaneous organization of serial behaviour as a window on primate cognitive processes. The feasibility of a research program on non-human primates focused on such issues was tested by confronting capuchin monkeys (*Cebus apella*) with a set of search tasks.

In a first group of experiments, the subjects were required to search serially an array of occluders presented with a WGTA, in order to retrieve a hidden object. Socially transmitted information allowed to infer the possible sites of the object, reducing the search space to a sub-set of all possible locations. The subjects showed a tendency towards searching in a principled way, either using the information given, or the spatial constraints afforded by the linear arrangement of the search space. However, some inadequacies were individuated in procedures and apparatus.

Successively, tasks requiring the exhaustive exploration of a set of icons, presented on touch sensitive computer monitors, were employed. These tasks implicitly demand a serial and economic search, where reiterations on sites already explored should be avoided. The structure of the search space was manipulated so that either spatial strategies or categorisation schemes could be used as a memory aid to keep track of the moves already performed.

Monkeys showed a spontaneous tendency to progressively reduce the number of redundant moves and spontaneously deployed spatial strategies when possible. This produced economic searches in sets of up to 9 locations. However, the subjects did not deploy categorisation principles strategically from the outset, but learned to use them when provided with selective negative feedback. When used in a principled way classification proved effective in sustaining highly economic searches.

Comparisons with data available from independent studies with different species allowed the conclusion that the ability to self-regulate behaviour; the qualitative analysis of search strategies and their change over time, represent promising dimensions for comparative studies. Having assessed the effectiveness of the paradigms developed here, basic modifications of the procedures are then proposed for further research.

# ACKNOWLEDGEMENTS

My heartfelt thanks to the following: The Italian Ministry of University and Scientific and Technological Research for their financial support; Dr. Brendan McGonigle from whom I have learnt so much during my time spent in Edinburgh; Dr. Jim Donnett and Dr. Benidict St Johnston for developing the software used in this project; Mrs. Rodella Purves for her genuine interest in the welfare of the monkeys; and Jennifer Marshall whose patience and efforts greatly contributed to the clarity of the final version of this thesis.

# DECLARATION

I declare that this thesis has been composed by myself, and that the work presented herein is my own.

For J.M.

# CONTENTS

# CHAPTER I

## A GENERAL PREAMBLE

### 1.1 The evolutionary perspective: the struggle to define a new scientific domain, introspection and some methodological insights.

The roots of comparative psychology can be traced back to 1871, the year of the publication of Darwin's "The descent of man". In this book human intelligence was explicitly set in an evolutionary perspective and it was suggested that precursors of human high level cognitive skills might be found in mental abilities of non-human species.

The problem faced by Darwin was to promote revolutionary ideas in times of prejudice. When applied to organic (non-mental) evolution, the theory of natural selection, had gained some credibility in restricted circles. However, in the psychological domain, even the theory's co-discoverer, Alfred Wallace, had retreated to supernatural powers (invoking a divine creation at some point of the evolution of man) in order to explain the gap between the mind of ape and man. Thus, Darwin's main concern was to show the similarities between human and non-human species, in order to make plausible the hypothesis of the continuity of mental traits.

Darwin tried to predict and challenge the difficulties raised by the application of the theory of evolution to human mental functions. He identified the following: 1) the objection that animal behaviour is guided by fixed instincts whereas humans make rational choices on the basis of past experiences; 2) the belief that nothing similar to human language can be observed in non-human species; and 3) the absence of evidence of some form of conscience and morality in animals.

Darwin addressed the first point by claiming that instinct and learning should not be regarded as mutually exclusive. A great deal of the behaviour of many species, he observed, is instinctive; nevertheless, it can be modified by experience. Moreover, ape's tool-using behaviours seemed to Darwin more of an expression of a kind of rationality than rigidly guided by instinct.

The problem of language was addressed suggesting that there was evidence for its precursors in non-human species. As examples, Darwin quoted bird songs (the development of which seemed to depend on both instinct and learning), the vocal mimicry of parrots and other birds and the ability of monkeys to communicate affective states by means of vocal calls.

The last points, conscience and morality, were seen by Darwin as the inevitable product of parental instincts, combined with the parallel evolution of intellectual skills and language.

Darwin must be accredited for having laid down the theoretical precondition for a comparative psychology of cognition, clearly defining it's goal as the comparative study of the evolution of human mind. However, his efforts were merely theoretical in their nature and he never developed a methodology for the systematic study of differences and analogies in the mental traits of species more or less taxonomically related.

Instead, George Romanes, (entrusted by Darwin himself) was to extend the methods of comparative anatomy to the study of the evolution of intelligence. In his program, the study of mental evolution featured two different phases: the first in which an affluent corpus of information on different species was collected and the second where general scientific laws were deduced from the differences and analogies observed.

In his first book, "Animal Intelligence" (1882), Romanes attempted a systematic classification of observations collected from various sources (chiefly letters sent to scientific and popular journals by amateur animal watchers). Inferences derived from these observations were then used to develop a theory of the evolution of mind, the topic of his second book: "Mental Evolution in Animals" (1884). Romanes also tried to evidence a parallel between phylogenesis and ontogenesis by comparing the competences observed in children of various ages with those of species of different zoological taxa.

The method proposed by Romanes for the investigation of animal intelligence was introspection. It was to be used to detect which mental process underpins a particular human behaviour. Once individuated, this process can then be attributed to those animals which show a similar behaviour.

Following this procedure, Romanes elaborated an additive model of the phylogeny and the ontogeny of mind (see Wasserman, 1984; 1993) where increasing complexity was seen as a function of the summation of qualitatively different competences.

An example of the effort invested in the characterization of different psychological processes is the distinction he made between sensation and perception. Sensation was considered as the direct result of a stimulation, whereas perception was assumed to be the product of inferential processes acting upon sensations. The term reflex was then confined to those behaviours which are influenced by mere sensations. Thus, behaviours which are modified by specific physical changes of a stimulus (e.g. the level of luminosity) would be classified as reflexive. By contrast, behaviours which seem to be influenced by changes in relational properties of two or more stimuli would be indexical of the action of perceptual processes.[1]

Romanes, thus, put forward the idea that mental evolution should be seen as the progressive construction of a cognitive architecture formed by different processes, whose interplay gives rise to the most complex forms of behaviour. However, his program was

---

1 An example of the latter would be the behaviour of a chick that moves towards his unseen mother, inferring her location from the spatial provenience of her calls. This ability to deal with relational properties of the environment was seen by Romanes as a necessary step towards the evolution of thought.

characterized by a lack of experimental work and most of his ideas were developed on the basis of occasional observations, often of anecdotal nature, interpreted by means of introspective methods.

Some years later, Robert Yerkes picked up the heritage of the evolutionary tradition pioneered by Romanes. In his early career, Yerkes addressed the question of whether reptiles, amphibians and invertebrates possessed the ability to learn by trial and error and demonstrated that all of them did. As a consequence, he proposed that, in order to find suitable dimensions for cross-species comparisons, more complex forms of behaviour had to be investigated.

He detected a suitable paradigm for his scope in an experimental device invented by a psychiatrist, Gilbert Hamilton. Independently from Yerkes, Hamilton (1911) had reached the conclusion that comparisons between species on the basis of quantitative measures (like speed of learning in simple problem solving situations) was of little use. What was needed instead was a method that provided the explanation of <u>why</u> some species learned particular task faster than others. With this aim in mind, Hamilton developed an apparatus in which the influence of peripheral factors (such as specific differences in perceptual or motor abilities) on the behaviour of the subjects was minimized. The apparatus featured a chamber with multiple doors. The procedure involved the presentation of a series of trials where all the doors but one (chosen in a pseudorandom fashion for every trial) were locked. The task was to find the way out from the chamber. The search patterns among the different doors in successive trials would have revealed if the behaviour of the subjects was based on some principled strategies, whereas the number of trials required in average to solve the task was taken as a measure of the relative efficiency of the different strategies.

Yerkes considered the absence of meaningful rules (specifying which choice was correct) as a weakness in this procedure. The subject was forced to rely on trial and error to find his way out. Thus, he modified apparatus and procedure to provide the subject with the opportunity to predict the correct choice on the basis of a stack of rules of increasing complexity. His aim was to compare different species on what he called "ideational behaviour" (Yerkes, 1917).

The new apparatus devised by Yerkes was a linear array of 9 compartments collocated within an open area. A subset of the compartments (ranging in number from 3 to 9), were unlocked in a particular trial. The subject, was released in front of the array and had to search for a reward hidden in one of the unlocked compartments. If an unrewarded compartment was entered, the subject was blocked inside for about a minute. In contrast with Hamilton's procedure, the correct compartment was not chosen on random basis but according to a set of rules. Typically, in successive problems, the following rules were presented: 1) chose the extreme (say) left of the subset of unlocked compartments; 2) chose always the second on the (say) left; 3) if in the current trial the rewarded location was the extreme left, chose the extreme right on the next and vice versa; 4) between the different compartments chose always the middle one.

In a first set of experiments, Yerkes tested subjects as different as human psychiatric patients, crows and pigs. Later, he stressed the importance of comparing related and increasingly removed species to establish the origins of particular human mental traits and started a research program on non-human primates.

The first experiments were conducted on two macaques and a young orang-utan (described in Boakes, 1984). The results proved interesting and unexpected. On the first problem (choice of the extreme left compartment), the two monkeys took two and three times as many trials to reach the criterion as the pigs, and the orang-utan took six times longer. Moreover, apart from the mere rate of learning, the shape of the learning curves of the orang-utan and the monkeys were significantly different. In contrast with those obtained from the monkeys, the learning curve of the orang-utan showed a very long period with no improvement, followed by a dramatic change, as the performance became suddenly almost errorless. Yerkes proposed that there were different ways of reaching the solution to the task. While the monkeys seemed to learn by means of a trial and error procedure, the orang-utan solved this problem "ideationally". The conclusion was that the number of trials to criterion were a very weak indicator of the cognitive competences of a species and as a consequence not a good dimension for comparisons.

The results obtained from the second task were taken by Yerkes as a confirmation of such a conclusion. Faced with the problem of learning to always select the second compartment on the left, the monkeys solved the task. In contrast, the orang-utan repeatedly failed, even after a long exposure to the task. Yerkes proposed that while the monkeys solved the task by means of a trial and error procedure the orang-utan failed because it was striving to reach some sort of insight. For bureaucratic reasons Yerkes never completed this set of experiments, and was not able to assess whether the orang-utan would have solved the task if given an even more protracted exposure to it.

Overall, the rationale of the program carried out by Yerkes, was to arrange different species along non-trivial dimensions of intellectual ability. These dimensions were identified in the ability of the animals to control their behaviour on the basis of complex spatial and ordinal relationships perceived between the items of a set of multiple alternatives. Yerkes put a particular emphasis on the fact that only when multiple solutions are possible, some qualitative differences between species can be detected. Moreover, protracted periods of testing were seen as necessary to highlight those differences.

In parallel and independently from Yerkes, Wolfgang Koehler stressed the importance of the study of intelligence in non-human primates to understand human cognition. His interests focused on the problem of insight in a set of studies on chimpanzee's problem solving (Koehler, 1917/1976), in order to "ascertain the degree of relationship between anthropoid apes and man in a field which seems to us particularly important" (Koehler, 1917/1976, p. 1). As Yerkes did, Koehler saw in many traditional paradigms a way of forcing the animals to find a solution by trial and error. In fact, in his view, when the contingencies between responses and rewards are artificial and established on arbitrary basis, or when the causal relationship between the components of a problem are very difficult to detect, every animal (whatever the level of its cognitive competences might be) has to rely on very basic forms of learning.

For this reason, Koehler focused the attention on tool-use skills, devising problem solving situations where it was relatively easy to perceive the instrumental value of a tool to reach a particular goal.

His experiments are well known. In one of them a banana, hanging from the ceiling of a large cage, had to be reached by climbing a wooden box (or in more complex version by assembling and climbing a structure of several boxes), after having placed it in the required position under the banana. Another problem required the use of a stick (or in analogy with the problem of the boxes, a structure of two sticks) to collect items of food otherwise out of reach.[2] Koehler identified two distinct processes at the basis of chimpanzee's problem solving: the first based on trial and error learning (used, for example to build a balanced structure of boxes) and characterized by the typical graduality of learning curves; and the second characterized by an "insight" on the reciprocal relationships of the various components of the problem and characterized by a non-gradual (all or none) path towards the solution. This "insightful" achievement of the solution was inferred, for example, from observing chimps joining at once the two short sticks together and directly reaching for the food without any gradual approximation to the goal.

---

2 According to Koehler, an important feature of these problems was that their solution required a detour in the path towards the goal (e.g. to collect the sticks or to move the boxes) and so they implied the ability to plan a sequence of actions in order to reach the goal.

However, Koehler's paradigms were designed exclusively for ape subjects. The nature of the tasks prevented their use with other species. Therefore, it was never assessed whether other organisms would have solved by trial and error the problems which apes solved "insightfully". As a consequence, an empirical comparison, in terms of relative economy (as measured, for example by solution time), between learning by insight and by trial and error was precluded. Also lacking in Koehler's experiments was an evaluation of transfer of learning from one problem to an other. A study of transfer would have helped, on the one hand, to exclude the possibility that what was interpreted by Koehler as insight was in fact the product of previous learning experience (as proposed by Harlow, 1959); and, on the other, to evaluate how insightful solutions were transferable to new problems, similar in structure to those already solved by the chimpanzees (see Wright, 1985). Furthermore, the semi-naturalistic setting of Koehler's experiments did not allow a rigourous analysis of the experimental data and they were reported in a rather descriptive fashion. From the comparative stand-point, these represent major restrictions of a paradigm which otherwise focuses on relevant dimensions (such as the ability of animals to plan a sequence of actions in order to achieve a goal).

In summary, the pioneers of the comparative study of cognition were well aware of the importance of putting human intelligence in an evolutionary context and devoted a great deal of attention to characterization of different mental processes. In fact, it was more or less explicitly assumed that an understanding of the way in which such processes were clustered in progressively more related species was essential to understand the architecture of human mind and its natural history.

The paradigms developed within this theoretical framework, were characterized by a focus on high order competences which would have been demonstrated by the achievement of non-trivial solutions to tasks otherwise solvable on the basis of simple learning processes. Moreover, a particular emphasis was posed on the importance of giving the animals the possibility to recognise meaningful relationships within the problem space. As we shall see later, these factors, i.e. the possibility to classify qualitatively the behaviour of different species, the presence of a range of possible solutions to a task, and the possibility to perceive structural relationships within the problem space (suitable to be transduced in actions) are considered important criteria for the development of the experimental paradigms used for the present study.

However, these early comparative studies lacked a standard paradigm, which was widely accepted within the scientific community. The methods and the measures were far from being refined, and much of the observations were incomplete or reported in an anecdotal fashion. In the meantime, a different tradition was born under the flag of methodological rigour. However, its development was not going to provide only benefits for the field of comparative cognition.

## 1.2 Puzzle boxes and Skinner boxes: rat, pigeon, monkey, which is which ? It doesn't matter

At the beginning of the 20th century, Edward Lee Thorndike (1911) ventured the development of a vast research project aimed at the systematic comparison of different species on the basis of what he considered a precise measure of intelligence.

The project was centred on the puzzle box, an apparatus invented by Thorndike (1911) for the comparative study of problem solving. In a typical experiment, an animal confined in wooden cage, had to select the appropriate response (e.g. pull a string, press a lever, or displace a stick) that caused the opening of the cage's door, allowing the subject to reach an item of food. In such a situation, the animal performed first a series of inappropriate behaviours, and then, accidentally produced the correct response.

The time employed to solve the task in successive trials was plotted as a learning curve. The learning curve was considered as representative of the behaviour of a particular species in a particular problem. In a typical learning curve, the latency of the solution is very high at the beginning, then gradually decreases and stabilizes after 10-20 trials. From the graduality of these curves, Thorndike deduced that the animals were unable to understand the causal relationships between the correct response and its outcome. The process responsible for the solution, he claimed, was a learning process based on the emission of random responses followed by the selection of those which lead to the food.

On the basis of this interpretation of problem solving behaviour, Thorndike formulated his famous "Law of effect", postulating that the probability of producing a response is increased when it is followed by a state of satisfaction of the organism, and reduced when followed by a state of discomfort.[3]

Thorndike tested a variety of species in his puzzle boxes and interpreted the formal similarity of the obtained learning curves as expression of a universal law, applicable to every taxa.

Thorndike's paradigm provides a first example of how the attempt to devise a uniform methodology produced, as a result, the levelling of the competence of different species on basic learning skills. As stressed by Koehler, problem boxes forced the animal to find the solution accidentally. Provided with complicated mechanisms for the opening of the door, they made it impossible for the subject to perceive the respective functions of the different parts of the apparatus (for a detailed presentation of Koehler's critical points, see Hilgard and Bower, 1975). Thus, the opportunity to evaluate how the respective level of cognitive competences of different

---

3 As pointed out by Boakes (1984), in formulating the second part of his law of effect (the inhibition of incorrect responses), Thorndike probably relied on the results obtained by Yerkes (whose experiments had just been carried out a few years before). In fact since the experiments carried out by Thorndike did not allow an analysis of the error space (which on the contrary was possible in the multiple alternatives situations devised by Yerkes) he did not have any principled way to measure the behaviours whose probability decreased.

organisms affected the way in which the problems were solved was missed with Thorndike's procedure.

Thus, Thorndike was probably the first scholar to complete a vast and systematic research program with a uniform paradigm suitable for the comparison of a wide range of species. Nevertheless, his methods were ill-suited to detect any qualitative inter-specific differences. The results were arbitrarily generalized and Thorndike ventured to claim that general problem solving skills of different species were chiefly quantitative in their nature. The consequence was that a considerable distance was implicitly taken from the evolutionary tradition that attempted to find an answer to the problem of mental evolution in the relationship between taxonomic distance and cognitive similarity.

What had been implicitly assumed by Thorndike would have been explicitly declared by John Broadus Watson (1878-1958). In his celebrated article "Psychology as the behaviourist views it" (Watson, 1913) and in the introduction to his book "Behaviour: An Introduction to Comparative Psychology" (Watson, 1914), Watson explicitly dismissed the evolutionary tradition. Animal psychology was viewed as a privileged field of study for the mere reason that it had traditionally focused on behaviour, which, according to Watson, was the only appropriate object of psychological investigation. [4] Animals were taken as adequate models of human psychology. In fact: "The behaviourist, in his effort to get a unitary scheme of animal response, recognizes no dividing line between man and brute" (Watson, 1913). Since no major distinctions were warranted between men and animals, and even less between different non-human species, the rationale itself of the comparative method started to become more and more obscure.

Concerning the problem of the mind, as we have seen, Romanes and Yerkes believed in the causal role of internal processes on animal behaviour. Thus, they tried to characterize these processes

---

4 In contrast with the use of verbal protocols, at the time so commonly adopted especially in European human psychology.

on the basis of the inferences drawn from observing the behaviour of different species. Watson, by contrast, considered the method as non-scientific.

The fact that Yerkes had proposed some forms of "ideation" to explain the performance of his subjects was bitterly attacked. Referring to Yerkes' experiments on multiple choices Watson claimed "You have made statements which are based on such flimsy and anthropomorphic evidence that for a while I seriously questioned your scientific spirit" (letter from Watson to Yerkes, 12th may 1916, quoted in Boakes, 1984 p. 199).

Watson's sceptical attitude towards the results obtained by many of his contemporaries was mainly due to the stress he put on the need for the unification of experimental procedures in psychology. Results which were not obtained from a standard paradigm of general application, had to wait further confirmation by such a methodology (when developed). Although Watson identified in the conditioning techniques of Russian reflexology a suitable paradigm, he never managed to fulfil his program of the standardization of psychological methodology.

In the following years, the development of behaviourism formulated by Watson would have taken different forms under the common label of neobehavourism. The theoretic position assumed by Watson was to become less radical and the main heritage was to be based on his general methodological directives. Methodological behaviourism might be conceived as a psychological version of neopositivism. The researchers hold that since mental processes are unobservable, the study of cognitive functions must be confined to their observable expression: the behaviour. However, the presence of internal processes here is not denied but it must be inferred from experimental procedures based exclusively on the manipulation of the input that an organism receives and the recording of its responses (see Skinner, 1977 for an account of methodological behaviourism). In this respect methodological behaviourism survives in much of contemporary comparative cognition (Wasserman, 1984; 1993).

It was Skinner who resurrected the radical behaviourism as proposed by Watson. Radical behaviourism not only denies the possibility of studying inner mental processes (if not by its behavioural ambassadors), but makes behaviour the ultimate end of any psychological investigation (Skinner, 1977; Wasserman, 1984). The shift from methodological behaviourism was warranted by the assumption that the role of any science is the prediction and not the explanation of the phenomenon it observes. The role of psychology as a science was, therefore, merely confined to the control of behaviour.

As a matter of fact, this emphasis on the control of behaviour made it necessary to divert the attention from complex problem solving situations in favour of artificial oversimplified tasks. The need to develop a simple experimental situation, where it was possible to modify the behaviour of the subjects by manipulating simple variables, led to the development of the Skinner box. This apparatus is a cage provided with a food dispenser operated by the pressure of a lever. The frequency of pressure responses of the animals is controlled according to different schedules of reinforcement.

The more different species were tested in such a simple apparatus, the more they showed similarity in their behaviour, confirming the assumption that essentially the same basic principle could have accounted for the behaviour of every organism. In line with the empiricist tradition, this single basic mechanism was identified in the association.

Since the laws of behaviour were basically the same for all of the animal realm (man included), it was worth studying one or two species only and then generalizing the results with all other species. As Skinner himself provocatively put it: "Pigeon, rat, monkey, which is which ? It doesn't matter" (Skinner, 1956, p. 230).

To explain the patent complexity of animal behaviour as observed in real life situations on the basis of the simple conditioning rules derived from Skinner box contexts, the concept of chains of responses assumed a central role in the skinnerian theory. Skinner (1938; 1953) maintained that all complex behaviours can be explained as a sequence of movements, each of which provides a feedback (internal or external) which becomes the discriminative stimulus for the one that follows it. A modified version of this theory was proposed by Skinner to account for the development of syntax in human language. As we shall see, the weakness of this simplistic explanation of grammar would have been the basis for the strongest criticism towards radical behaviourism.

Thus, radical behaviourism, created confusion among both the methodology and the objects of comparative cognition. The behaviour became the direct end of psychological investigation (rather than the indirect means of accessing mental processes) and, the association became the unique process to account for all aspects of cognition. The hierarchical and discontinuous level of complexity of different cognitive competences (see McGonigle, 1991) was conceived as being of quantitative nature, where more extensive, richer and long chains of associations accounted for increasing complexity. For the same reasons, obvious differences in the complexity of the behaviour of different species (beyond basic instinctive and reflexive behaviours with obvious innate determinants), were accounted for by a simple summation of associations more or less easily and rapidly established.

The end product of such an assumption was that the realm of mind and cognition was definitively dismissed, not only for epistemological reasons, but, mainly because they could be replaced by behaviour itself. The meaning of the comparative exercise was nullified. Different species, including humans, could have been interchanged as experimental subjects, without losing any major source of information on the determinants of their behaviour. As a consequence the evolutionary framework which originally constituted the main rationale for comparative psychology lost its original meaning.

## 1.3 WGTA: learning sets and the study of complex learning skills.

An attempt to go beyond the restriction of simple conditioning was performed by Harry Harlow. He stressed the problems arising from interpreting complex behaviours on the basis of simple behaviours observed in oversimplified tasks (Harlow, 1959).

Harlow (1949) developed a new apparatus, the Wisconsin General Testing Apparatus (WGTA) and a procedure which would have allowed the comparison of different species (only primates at the beginning) on higher order cognitive skills beyond simple associative learning.

The WGTA features a tray with two foodwells and an opaque screen interposed between the subject and the tester. While the screen is down, one of the foodwells is baited and both are covered by two objects differing multidimensionally. The screen is then lifted and the tray is pushed towards the subject to allow it to lift one of the stimuli. If it selects the stimulus covering the bait, it is allowed to take the reward, if not, the tray is withdrawn and an interval is interposed before the presentation of a new trial. Typically the subject is presented with the same pair of stimuli for a fixed number of trials. Then a new pair of stimuli is presented, and so on.

Harlow (1949) found that monkeys (trained on several hundred discriminations) eventually learned each new discrimination in one trial only. He dubbed this sort of learning to learn "learning set" and claimed to have found a higher order capability which allowed the animal to transcend the particular problem at hand and to develop strategies with a high level of generality.

In order to find even more complex forms of transfer, the paradigm has been successively modified so that, instead of simple discriminations, series of discrimination reversal were presented. In this latter paradigm, after the subject has reached the criterion on a particular discrimination, the reward contingencies are reversed, so that the stimulus previously unrewarded, now is the rewarded one, and vice versa. Typically the reversal phase features a fixed number of trials and the observed proportion of correct responses is considered as a measure of the transfer of the discrimination learning ability (acquired in the prereversal phase) to the novel situation characterized by the reversal of the reward contingencies. A series of these reversal discrimination problems, featuring different stimuli is then presented and the "transfer of transfer" is evaluated. In this way even a more abstract and general learning to learn ability is acquired by subjects who show an improvement over a series of discrimination reversal problems.

However, this higher order ability was never fully characterized (for a discussion see McGonigle, 1984). The only conceptual distinction that has been proposed is one which contraposes the development of a "win-stay lose-shift strategy" (which would allow the solution of both the reversal problem and any new discrimination in one trial only) to associative learning (Restle, 1958; Reese, 1964).

Although most of the experimental results on learning-sets support the hypothesis that monkeys rely on a "win-stay lose-shift strategy" (Restle, 1958; Levine, 1965; Mackintosh; 1974), the status of the strategy, in the context of the problems devised by Harlow, remains unclear. In fact, it can be either selected from the pre-existing behavioural repertoire of the subject when its relevance for the task at hand is recognized, or, alternatively built ex novo as a consequence of an active process of behavioural self-regulation produced by task practice.

Moreover, it can be noted that such a strategy fits only the restricted context of binary discrimination, where most of learning-sets studies have been conducted. In situations where a larger number of alternatives is presented, the mere deployment of a "win-stay lose-shift" strategy would not guarantee the solution in the minimal number of trials, unless supported by a system which keeps track of all the alternatives already tested. However, Harlow never extended his paradigm to procedures involving the presentation of multiple items and, therefore, missed the opportunity to evaluate both the generality of a "win-stay lose-shift" strategy, and, the ability of primates to deploy other complementary forms of control.

On the basis of his findings, Harlow (1959) claimed to have readdressed the question of insight as proposed by Koehler, but using rigourous experimental procedures to detect transfer of learning from one problem solving situation to others. He considered the one trial learning performance that his subjects achieved after prolonged learning set practice as a form of insight analogous to that observed by Koehler. So, he claimed: "no animal can solve problems insightfully [i.e. deploying a "win-stay lose-shift" strategy] or with maximal efficiency without a history of earlier solution of similar problems" (Harlow, 1959, p. 510).

Harlow failed to see a fundamental difference between his learning set paradigms and the problem solving situations adopted by Koehler. In fact, while in Koehler's problems the logic solution of the task is available to the subject from the outset, in the case of discrimination learning set the rewarded stimulus is selected arbitrarily by the experimenter in each successive problem. Thus, the subject is forced to find out by trial and error a consistent rule which allows the solution of successive problems. The situation faced by the subjects becomes even more chaotic in the case of discrimination reversal learning sets, and therefore it is not surprising that a protracted presentation of the task is needed to allow the subjects to detect the appropriate solution strategy.

Although misleading, if assimilated to problems where the appropriate response is not chosen arbitrarily, Harlow's paradigm tapped into interesting dimensions, such as the ability of monkeys to develop economical solutions which transcend the specific domain of a single problem (as shown by the one trial learning produced by task expertise). Moreover, the paradigm allowed the systematic manipulation of variables such as the intertrial interval and the ratio between number of instances of the same problem and number of problems presented.

For these reasons, the apparatus and the procedure proved extremely influential and, since the pioneer study by Harlow, a number of different species have been tested for their ability to form learning sets either in its original form or in various forms of discrimination reversal tasks.

However, the paradigm was never extended to problems more complex than those featuring binary discriminations. This prevented an assay of the different strategies which might be deployed by different organisms when faced with the problem of finding an economic solution to a series of more or less related tasks. The only measures allowed in the binary versions of the paradigm, such as the curves of learning set formation (based on the speed of learning and the number of errors in successive reversal phases of a learning set) provide very little information, especially when primates are compared with non-primate species. After having extensively reviewed fifteen years of research conducted on learning-sets of different species, Warren (1965) concluded as follows: "The major points in the discussion of learning-set formation by primates and other vertebrates may be summarized quite succinctly. Mammals and other birds differ from fish and reptiles in being able to learn repeated discrimination reversals in progressively fewer trials, but primates are not markedly more proficient than other mammals or birds" (Warren, 1965, p. 266).

Nevertheless, the WGTA, for its versatility was successively adapted for the implementation of several experimental procedures. The most influential proved to be the matching to sample paradigm and its clones. In its typical version a sample stimulus is presented either before (delayed matching to sample) or simultaneously (0-

delay matching to sample) with another two stimuli, only one of which is identical to the sample. A response to the correct matching is rewarded. Variants of the paradigm involve the oddity from the sample task, where the stimulus which is different from the sample must be selected; and the "Symbolic" matching to sample, where no two identical stimuli are presented and the matching is arbitrarily decided by the experimenter. It has been shown that all the versions of this paradigm can be mastered by organisms as different as pigeons (see Wasserman, 1993 for a review), monkeys (Murray and Mishkin, 1987) and apes (Premack, 1976).

Thus, Harlow must be acknowledged for having developed a standard paradigm to compare different species on abilities more complex than the simple forms of associative learning studied in Skinner boxes and simple binary discriminations. However, from the comparative point of view, even learning-sets did not prove to be a paradigm suitable to detect relevant dimensions for cross-species comparisons.

## 1.4 Ape language projects: keyboards, plastic chips, ASL and the primacy of the name

As we have seen, paradigms focused on the frequency of single responses (as in skinner boxes) or on the number of correct choices in binary discrimination tasks (as in both learning sets and matching to sample paradigms), produced a levelling effect on the performances of different species (and eventually led to a decline of the comparative method). Therefore, it is perhaps not surprising that at the end of the Sixties, the rebirth of comparative cognition focused on the study of ape's competence in one of the most structured and hierarchically organized cognitive domain: language.

Although in the twenties Yerkes had pioneered an abortive project to teach chimpanzees to articulate words, it was only after the development of cognitivism and the publication of Chomsky's (1959) famous critical dissection of Skinner's theory of Verbal Behaviour (1957) that projects aiming to the assessment of ape's linguistic skills started to blossom. Apes were taught to use the American Sign Language (ASL) (Gardner and Gardner, 1971; Terrace, 1979), and more artificial languages, based on arbitrary symbols to be selected on a keyboard (Rumbaugh, 1977), and plastic icons to be arranged on a board (Premack, 1976).

Although some of the early studies adopted basic variants of skinnerian procedures of training and were still influenced by most of the theoretical assumption of behaviourism (Gardner and Gardner, 1969; Rumbaugh, 1977), ape language projects can be considered more as a reaction to chomskyan criticism of behaviourism than as a natural development of behaviourism itself.

The role played by Chomsky in the development of ape language projects is twofold. On the one hand, the hypothesis of the discontinuity in the evolution of human language and of its independent origin from primate communication, strongly advocated by Chomsky, led to the development of experimental programs aimed to its falsification (for a review see Terrace, 1985). Beatrice and Allen Gardner, the researchers who carried out the Washoe project (Gardner and Gardner, 1969), one of the first attempts to teach the ASL to a chimpanzee, declared: "If a form of behaviour such as human language appears to be different in character from other forms of human and animal behaviour, we do not abandon the search for general laws; instead we question the adequacy of existing observations" (Gardner and Gardner, 1978, p. 37).

On the other hand, the fact that the theory of syntax was the field where Chomsky made his major contribution to linguistics and where the Skinnerian theory of language revealed most of its deficiencies (Lyons, 1991) produced a major shift in the variables to which researchers started to give attention. Instead of taking for granted the ubiquity of associative chaining in all behaviours with serial components (including the syntactic organization of linguistic production) new paradigms were developed in order to validate the hypothesis that apes possessed the ability to creatively produce long sequences of symbols, under the control of sophisticated ordering rules (Gardner and Gardner, 1974; Patterson, 1978; Terrace et al., 1979; 1980).

However, the syntactic output of language trained apes proved to be very poor. From exhaustive reviews of ape language projects (McGonigle 1980; Wallman, 1992) it emerged clearly that although apes seem to be able to acquire a relatively large "vocabulary", this does not result in a parallel expansion in the Mean Length of Utterance (MLU) (McGonigle, 1980). The ability to acquire a substantial "vocabulary" of arbitrary symbols does not represent per se either evidence of linguistic competence or of striking cognitive competences. In fact, dogs, rats, horses and other animals can learn to produce arbitrary "words" to obtain specific rewards (Terrace, 1985). By contrast, in children's linguistic development a dramatic increase in the length of MLU accompanies the enlargement of vocabulary.

According to Wallman (1992), it is impossible to find evidence for a grammatical competence in the linguistic production of apes. None of their "sentences" seem to go beyond the production of fixed sequences of "words" and those which have been claimed to be novel combinations of signs seem to be nothing more than random permutations of single "words", in an otherwise fixed list of signs.

It has also been questioned whether the constituents of apes' "sentences" carry a particular meaning even when presented in isolation (McGonigle, 1980; Terrace, 1985). When syntactic rules are not specified as permissible relations among categories of meaningful elements, the organization of serial production in language would be merely reduced to the rote learning of a number of well formed expressions. Such an approach is likely to overload the memory system of an organism which is faced with the task of producing "sentences" comparable in length to those observed in human language production. This could explain the high redundancy and recursiveness observed in the strings produced by the apes (e.g. "Give orange me give eat orange me eat orange give me you", as mentioned by McGonigle, 1989). If a subject is unable to assign a semantic role to the components of a sentence the aim of evaluating its grammatical competences might be unrealistically ambitious (Terrace, 1985).

Therefore, current research on ape language (Savage-Rumbaugh, 1986) is devoted to the assessment of whether the signs produced by the apes (when taken in isolation) share some important characteristics with the words produced by children. The field chiefly investigated is that of referential communication. The attention of the researchers is particularly directed towards the socio-pragmatic aspects of the interaction of trainer and ape, in analogy with the mother-infant interactions (such as eye-to-eye contact, pointing, joint attention etc.) which, according to recent trends in the study of language acquisition by children (Bates, 1976), support the development of intentional referential communication.

Thus, current perspectives seem to be characterized by an appraisal of the importance of studying those paralinguistic or prelinguistic skills (generally dubbed as Acquisition Support System) which might help or even be a necessary precondition for the development of language (Bruner, 1983; Bates 1976).

However, to date, the attention has been put on naming and on the social and communicative skills that might support it. Little effort has been devoted to those cognitive competences which in extra-linguistic contexts support the hierarchical and serial organization of elements (Chalmers and McGonigle, 1994).

This is a field of intrinsic interest, even setting aside the still controversial question of whether the evolution of syntactic components of language should be regarded as independent from the evolution of those cognitive skills which enable animals and humans to organize serial behaviours in other domains.

In fact, as Lashley put forward in 1951, language strikingly presents the integrative functions which are characteristic of high level cortical functions and that reach their highest development in human thought processes. However, the problem of syntax is not to be found exclusively in language: "the coordination of leg movements of insects, the songs of birds, the control of trotting and pacing in a gaited horse, the rat running the maze, the finger movements of a piano player, present a problem of sequences of actions which cannot be explained in terms of chains of responses" (Lashley, 1951).

If serial behaviours are supported by complex competences (and not merely by associative chaining), the study of the way in which they are organized might be particularly fruitful in determining different control strategies which different organisms have developed to solve the problems posed by the production of long series of responses.

Over-emphasizing the non-linguistic, communicative competences which might support the development of naming, researchers involved in ape language projects, have neglected the importance of studying the non-linguistic competences at the basis of syntactical component of behaviour. Recent research trends (some directly related with ape language projects and some independently developed from the study of seriation) which have recently produced relevant findings in this area will be reviewed in the next chapter.

## 1.5 Conclusion

In brief, this chapter aimed to show how different research projects failed to develop experimental paradigms suitable to detect relevant dimensions for cross species comparisons. Comparative psychology began as the comparative study of the evolution of human mind. Darwin and subsequent scholars within the evolutionary tradition believed in the possibility of characterizing different animal mental traits. The comparison of psychological processes of related and increasingly removed species was considered the chief method for tracing the evolution of human cognition. From the analysis of the paradigms developed within this tradition emerges the implicit assumption that to find relevant dimensions for comparison it was necessary to adopt tasks suitable to be solved either by trial and error or in a more "insightful" way. Often, the importance was stressed of providing the subjects with problem spaces where the relevant information for the solution was available to the subjects from the outset and did not require a blind, inductive process to find out the relevant characteristics of the problem. The comparison of the problem solving strategies deployed by different species would have then allowed the discovery of their respective cognitive restrictions and abilities. However, a standard paradigm which allowed the fulfilment of all the requirements of a systematic research project was never developed. In most cases the methodologies lacked in experimental rigour and this led to a strong criticism of the whole tradition.

In the name of experimental rigour the original objectives of the field have been distorted for a great part of this century. In the present interpretation it is proposed that this confusion has been the inevitable consequence of a set of circumstances. The need for methodological rigour caused a shift of attention from mental processes to behaviour and how to control it. The over-emphasis put on the prediction of behaviour led to a disregard for the problem of understanding its causes. These premises produced the design of oversimplified experimental environments, typically exemplified by skinner boxes and binary discrimination problems. The use of such experimental techniques, allowed the observation of simple forms of learning, shared by the vast majority of the species studied. Since in most of the paradigms developed in this framework the correct responses were arbitrarily defined by the experimenter, the animals were forced to rely on trial and error procedures in order to solve the tasks. The levelling effect produced by these experimental devices on the behaviour of various organisms reinforced the belief that a simple summation function of the same atomic mechanisms could explain the differences in the level of complexity of these organisms. For the same reason, concepts such as evolution and cognitive growth were trivialized as the mere expression of the number of associations and speed with which they simple formed.

Learning sets and related paradigms seemed to offer the possibility of distinguishing different species on the basis of their ability to deploy strategies of different levels of complexity beyond simple associative learning. However, no major differences were observed among species which are taxonomically very distant. Moreover, the procedures adopted in learning set studies (in contrast, for example with the problem solving situations devised by Koehler, where all the elements for the solution of the problem were at disposal of the subjects from the outset) required the subjects to achieve the solution inductively by means of a selection of those strategies which proved successful in successive problems. This led to interpretative problems regarding the origins and the status of the solution. Were the strategies part of the repertoire of the subject ? Were they developed as a consequence of task practice ? Would the subject have deployed them from the outset if given the possibility to perceive their relevance for the task at hand ? All these questions remained unanswered.

Apart from the restrictions imposed by the inductive nature of the problem posed to the subjects, a major difficulty of the paradigm was detected in the fact that it was never extended to problems beyond binary discrimination learning. Problems featuring larger sets of stimuli would have offered a much richer range of strategic possibilities and their differential use might have been a more profitable dimension for comparisons.

Focusing on competences dealing with the complex and hierarchical organization of behaviours with a strong serial component, ape language projects produced a shift of attention from the binary discrimination context. However, the assay of non-human grammatical competences in a strictly linguistic context proved to be a far too ambitious project. Language was taken as a privileged undivisable object of study, neglecting its relationship with different non-linguistic cognitive skills which might support it. In absence of independent evidence for the status of the signs produced by the apes, the "verbal" behaviour of the apes can be interpreted as the mere learning of sequence of arbitrary symbols. The failure to show any grammatical competence in non-human subjects led to the recognition of the necessity to ascertain the presence of non-linguistic prerequisite competences which apes might or might not possess or be able to acquire. However, researchers still involved in ape language training have mostly focused their attention on naming. This has led to the recent development of paradigms with strong social components and dealing with discrete responses (pointing to objects, ability to refer symbolically to a set of pictures and so on). The ability to organize series of responses by means of sophisticated strategies (a field which might evidence some continuity between pre-linguistic or proto-linguistic competences of apes and others serially organized behaviours shown by different species) has been neglected. In the next chapter, recent research trends, which focus on these latter competences will be reviewed.

# CHAPTER II

## SERIALLY MOTIVATED BEHAVIOUR

### 2.1 From Lashley to recent trends in the study of serially organized behaviours

As we have seen from the previous chapter the behaviouristic tradition postulated the ubiquity of a single process, associative chaining, in all behaviours with patent serial components. This explanation has been influential till recent years and it remains widely accepted by scholars still influenced by behaviourism.

However as pointed out by Lashley (1951), most of serially organized behaviours cannot be explained postulating a single mechanism to account for them. For example from the skilled succession of movements, which can be observed in the different gaits of an horse, such as trotting, pacing and single footing, which involves the same pattern of muscular contraction in the individual legs, but nevertheless involve a different temporal order, it can be inferred that a separate mechanism must be responsible for the serial organization of the same atomic movements. This mechanism must work quite independently from the sensory information provided by each single movement.

On the other hand, from the misplacement errors which can occasionally be observed in behaviours which otherwise maintain their general serial structure (such as slips, interference and order mistakes which occur in speech, typewriting or music playing) it can be inferred that other forces play a role in determining the serial performance quite independently from a general motor plan of the sequence. It is also necessary to postulate a mechanism which enables the coordination of temporal and spatial informations to account, for example, for reaching and grasping behaviours.

The problem of the serial organization of behaviour in real life situations has been vastly ignored, notwithstanding its relevance in foraging strategies (Krebs, 1981), navigation (McGonigle, 1991; Gallistel, 1990), explorative behaviours (Menzel, 1974), representation of social hierarchies (Cheney and Seyfart, 1992) and practically the vast majority of situations, either in natural or laboratory environments, where the basis for rational decision making is investigated (see McGonigle and Chalmers, 1992).

Ethologists have dedicated some attention to behaviours with evident serial components observed in natural environments. However, they have traditionally focused on reflexive or instinctive behaviours in lower species such as insects. Since these behavioural patterns have very strong genetic determinants and their sequence is by definition very rigid and hardly modifiable, ethologists have failed to offer any characterization of the dimensions along which different levels of complexity can be compared in various organisms to find a path toward human intelligence (see McGonigle, 1991; Terrace and McGonigle, 1994). In fact, temporally integrated actions do not reach any degree of complexity until the appearance of the cerebral cortex (Lashley, 1951).

By contrast, the understanding of the functioning of the mechanisms underpining serially organized behaviours in complex organisms might serve a twofold objective.

First, the characterization of the different sub-systems which underpin serial behaviour and their interplay might possibly offer a rich set of qualitative different possibilities in the way in which sequences of actions are organized, providing the basis for meaningful comparisons.

Second, in contrast with artificial situations featuring binary choices only, the investigation of serially organized behaviours might provide a basis for cross-species comparisons in situations which seem much more related with real life environments.

Under these conditions, it might be possible to detect, on the one hand those requirements which must have an high degree of generality among different taxa because of common problems of information management, and on the other to find those variables which account for differences in cognitive complexity between different organisms and their possible adaptive function. To use Lashley's words again:

"Analysis of the nervous mechanisms underlying order in the more primitive acts may contribute ultimately to the solution even of the physiology of logic" (1951, p. 515); since "I am coming more and more to the conviction that the rudiments of every human behaviour mechanism will be found far down in the evolutionary scale" (1951, p. 526).

More recently (as we have seen in the previous chapter) some of the serial components of behaviour have received more attention in the form of the study of syntactic competences of language trained apes and reflexive activities performed by lower organisms. However, as pointed out by Terrace and McGonigle (1994), following the arguments put forward by Lashley (1951), an important but neglected middle ground is the behaviour shown in various tasks which require the serial production of responses which are neither directly related with verbal production nor fixed and stereotyped in their nature.

In this chapter, following a review of recent studies focused on serial aspects of behaviour, it will be proposed that the study of these sorts of behaviours, tapping onto non-trivial aspects of cognition, might allow us to put back the comparative study of animal intelligence in its original evolutionary tradition. In fact it would allow the qualitative comparison of different species on the basis of a rich spectrum of organizational possibilities which might underpin such behaviours.

In the previous chapter there often emerged a contraposition between experimental paradigms characterized by a strong component of arbitrariness in the definition of the responses required to the subjects and, by contrast, tasks where the structural properties of the problem space were more or less easily detectable from the outset. Also, among recent studies on sequential behaviours, a major division can be made between those focused on serial learning of lists of items in the order arbitrarily defined at the outset by the experimenter and those, which, by contrast, focus on the way in which primates exploit some form of organization implicitly embedded in the sequence to be produced.

## 2.2 Serial Learning: how arbitrary sequences of responses are learnt

The main root of this type of study stems directly from ape language projects.
As we have seen in the previous chapter, on the basis of an analysis of the syntactic characteristic of the "utterances" produced by subjects of different projects, Terrace and colleagues (Terrace, 1979; Terrace et al., 1979) argued that most of the sequences of signs produced by apes could be explained more parsimoniously than assuming an underlying grammatical competence.

Although it seemed implausible to assume that the apes memorised every single sequence which they were able to produce, Terrace and colleagues proposed that, for example Sarah (Premack, 1976) and Lana (Rumbaugh, 1977) multisign utterances might have been rotely learnt sequences of symbols arranged in particular orders. There was no evidence that the apes understood all the "words" in the sequence they produced and it seemed more likely that what they learned was the meaning of several key words, which were inserted in turn in the appropriate position of an otherwise fixed sequence (e.g. "please machine give apple period" and "please machine give drink period").

To show that such a competence was not very elaborate and might have been widespread among different zoological taxa, Terrace started a program to teach pigeons to perform fixed sequences of arbitrary signs. The first positive results showed pigeons pecking four colours in particular sequences and transferring this ability to novel arrays of colours (Terrace et al., 1977; Straub and Terrace, 1981). This led to the development of a more ambitious program to study the competence of ordered responses in pigeons and rhesus monkeys.

The paradigm used by Terrace and his group consists in displaying (originally in an operant chamber, now on touch sensitive computer monitors) a set of items (icons of different colours or photographs of different objects). The spatial configuration of these items on the screen is varied from trial to trial, in order to prevent the subjects from learning a chain of spatial responses. If the subject executes the sequence of responses in the order previously decided by the experimenters, a reward is dispensed.

The subjects are trained by a "forward" procedure: a single item is presented at the beginning, and then new list items are added, one at a time, until the whole sequence is learnt. Negative feedback (omission of reward and additional intertrial interval) is provided at each point of training. For example to teach a sequence of four items (here symbolized as ABCD), item A is presented first, the subject responds to it and receives a reward, then the sequence AB is presented and the subject is required to touch the items in the AB order (a BA response would be penalized), once a given criterion (usually the 75%) of correct responses on the two items series is reached, the third item C is added to the sequence, and so on.

The main aim of this research program was to determine the sort of processes underpining serial learning. The program started when most of American experimental psychology was still heavily under the influence of radical behaviourism. Thus, as we have seen, the dominant explanation of list learning was based on the chaining theory.

The chaining theory of learning postulates that the selective reward of particular sequences of responses produces associations between temporally successive responses. Although behaviourists have proposed different versions of the chaining theory, each according a weight to associations developed between stimuli and responses or between responses to non-adjacent items of the list, the different versions of the theory share the assumption that each response represents the discriminative stimulus for the following one. Since the main associations develop between successive items, according to the theory the subjects should find it difficult to respond to subsets of the list which contain only non-adjacent items.

However, Terrace and colleagues have shown that pigeons are able to respond accurately to pairs of non-adjacent items from an already learnt list, although they do so only if the subset contains a start and/or an end item.

For example, once pigeons have learnt a five term series ABCDE, they are able to respond accurately to the subsets AB, AC, AD, AE, BE, CE and DE, but their performance falls at chance level on the subsets BC, BD and CD. Thus, Terrace and colleagues claimed that although pigeons seem unable to form an ordered representation of the series (Terrace, 1984; 1987; 1991), a simple chaining hypothesis fails to account for their behaviour. By contrast, Terrace (1987) provided some evidence that pigeons impose some forms of organization upon the sequence of items to be remembered which goes beyond a simple association of successive items.

Terrace (1987) reported that pigeons trained with colours serving as items ABC and forms as items DE seem to "chunk" the two classes of stimuli. In fact, an experimental group, trained with the two classes of stimuli, learned both four-item and five-item term series much faster than control groups for which only colours were used or for which the two forms were not presented as adjacent items but interspersed among the three colours. Moreover, in subsequent test trials where non-adjacent pairs of items were presented, the experimental group performed better than the control groups.

Thus, it seems that pigeons are able to decompose a sequence in two different "chunks", a three item one (ABC) and a two item one (DE). Furthermore, the superimposition of such a form of organization over the list to be remembered leads to a more accurate performance.

Following the first experiments conducted on pigeons by Terrace and colleagues, D'Amato and Colombo, pioneered the study of serial learning in capuchin monkeys (Cebus apella). By means of an extensive research program D'Amato and colleagues were able to demonstrate that monkeys' performance, in analogous serial learning tasks, were underpinned by processes and representations substantially different from those of pigeons.

Having trained capuchin monkeys and pigeons in an operant chamber on a five term serial learning task, in a series of successive experiments, D'Amato and Colombo tested the subjects on a variety of tasks in which the items of the learnt series were manipulated in various ways.

In one of their experiments (D'Amato and Colombo, 1988), pairs of non-adjacent items were used. It was shown that, in contrast to the pigeons, monkeys were able to perform at a high level of accuracy on all of them. Furthermore, on the basis of an analysis of the latencies of the responses to each of the items of the series, D'Amato and Colombo (1988) presented evidence for a linear internal representation of the series in the monkeys. The rationale of the time analysis was that if the monkeys in the course of training developed an internal linear representation of the series, one might expect to find an orderly relation between response latency to the first item of a test pair and the position of that item in the original series.

For example, the response latency to the first member of a pair should increase across the test pairs AE, BE, CE, DE. The reason is that in order to decide which member of a test pair to respond to first, the subject presumably would start at the beginning of its internal representation of the series and progress through the sequence until locates one of the displayed items. The more represented items there are to be consulted, the longer the response latency.

These were exactly the results obtained by D'Amato and Colombo (1988). The monkey's latency of responding to the first item of each subset of the original sequence, increased monotonically as a function of the position of the item in the original list.

The research program performed by D'Amato and Colombo progressed, and provided further evidence against a behaviouristic interpretation of serial learning in monkeys. In a successive set of experiments these authors (D'Amato and Colombo, 1989, experiment 1) introduced wild card items in the serial learning task. Monkeys already expert on the ABCDE sequence were trained with a wild card item (W) that could replace any of the items of the original sequence, thus, forming five additional sequences WBCDE, AWCDE, ABWDE, ABCDWE, ABCDW. In another experiment D'Amato and Colombo (1989, experiment 2) used two wild cards (X and Y), forming 10 different sequences (e.g. AXYDE, XBCYE, etc.). The monkeys reached high accuracy of responses in all the sequences containing wild cards. The rationale for these experiments was that, since the position of wild cards within the sequence was changed, it was unlikely that the monkey's performance was based purely on associations between adjacent items of the list. On the contrary, it seemed more plausible to attribute to the monkeys some knowledge of the ordinal position of the items within the sequence.

The evidence for some form of representation of ordinal position in monkeys has some important implications. In fact, it can be conjectured that if a monkey is able to process information about ordinal position, it might have also the competence to form an abstract template of a serial learning task. The template, once formed can then be used with novel series by filling in, at the appropriate position, each of the items of the sequence in hand. The formation of such a template would result in a positive transfer to successive series containing new items. Such a device would be much more powerful than the ability to form associations between each of the elements of a particular series because the associations developed to learn an unique sequence cannot be exported to new isomorphic problems.

Some evidence for transfer abilities on successive serial learning tasks has recently been presented for rhesus monkeys (Macaca mulatta) by Swartz, Chen and Terrace (1991). These authors trained their subjects on multiple 4 item series lists. Each sequence contained different items. While the macaques were learning successive series, Swartz et al. (1991) were able to modify the training procedure, eliminating the early phases of training. Thus, while the early series needed to be drilled using the forward procedure described above (i.e. only one item is presented first and then one more item is added to the series and so on), in acquiring successive series, the subjects were able to deal with a training procedure which featured the presentation of more than one element of the list from the outset (for example three items ABC, or the whole 4 item series ABCD).

Although the performance, under these latter training procedure was mediocre at the beginning, the accuracy steadily increased on each of the subsequent lists. On the basis of their results, the authors claim that some form of hypothesis testing and chunking of adjacent items underpinned the transfer. For example, on lists trained starting with the simultaneous presentation of three items, the performance of the subjects dramatically improved when they became able to identify the first two items (A and B). The authors argue that this result supports the idea that the subjects chunked the first two items and then responded to the item C by default. The same phenomenon was observed in lists trained starting with the simultaneous presentation of 4 items. In this latter case, the dramatic improvement appeared after the identification of the third item C, so that the fourth item D, could now be responded to by default.

These results support the idea that some changes occur in the way in which the representation of the list is organized in the course of task practice. They also indicate that the experience of learning successive lists produces both the formation of abstract templates of ordered items and the ability to assign items from novel lists to particular ordinal positions.

Overall, serial learning studies show that, in order to report an arbitrarious list of items, non-human subjects organize the items in memory in more sophisticated ways than a simple associative chaining theory would suggest. Moreover, there seem to be qualitative differences between monkeys and pigeons in how the list is represented. Therefore, these studies somehow confirm the idea that when the competences involved go beyond the conditioning of a single response and binary discrimination learning, new dimensions can be discovered, along which the cognitive complexity of different organisms and the functional value of different representational devices can be compared.

However, serial learning studies appear to be more a reaction to simplistic behaviouristic explanations of the behaviour of complex organisms then an attempt to fully characterize their cognition. As a matter of fact, the original rationale of Terrace and colleagues was in line with the behaviouristic tradition. Their aim was to show that apes' linguistic production could have been explained better in terms of associative learning than as the product of grammatical competence. Successive studies focused on the issue of whether response chaining was to be considered the only mechanism responsible for serial learning but the training techniques were still a basic modification of the standard behaviouristic procedure to teach chains of responses. The tension was mainly between the idea that associative response chaining can explain all serial behaviours of all organisms and the questioning of the ubiquity of such a mechanism. The results clearly showed the inadequacy of associative chaining and some alternative mechanisms were proposed. However, the possibility was never explored that within the class of non-associative mechanisms a variety of different strategies might be explored by different species according to their cognitive status or that a shift of strategy might take place in function of the level of expertise that a subject acquires in the course of task practice. In this latter respect an attempt was made by Terrace and colleagues with their successive list learning paradigm. The formation of a general template for list learning in the course of practice was tested but the process of formation of such a template and its features were never fully characterized.

It can be conjectured that the reason for these restrictions of the research program might lay in the nature of the paradigms used. When the sequence of the items to be learnt is chosen on an arbitrary basis and strict training procedures are employed, the subject have few opportunities to show any ability to impose some forms of organization upon the items to be reported. A different scenario emerges from studies which focus on tasks featuring items serially ordered on a non-arbitrary basis.

## 2.3 Non-arbitrary series and the role of cognitive regulation

As we have seen, the studies reviewed in the previous section featured tasks where subjects were required to reproduce a sequence of responses in the specific order established arbitrarily by the experimenter. A different line of research focuses on the problem of how the subjects organize series of items connected by implicit and non-arbitrary relationships.

This line of enquiry can be traced back to the study of discrimination learning. Every theory of learning must in some way specify what the subjects learns, when trained to a particular discrimination between a rewarded stimulus (S+) and an unrewarded one (S-). In situations where the subject faces unconnected stimuli, such as traditional discrimination tasks and Harlow's learning sets, the problem is obviously reduced to the question of characterizing the process by which the subject identifies the features constantly present in S+ and absent in S- (Hilgard and Bower, 1971). These are the relevant features of the discrimination for the very reason that they are selected before-hand by the experimenter and (by external intervention) correlated with the reward contingencies.

However, in situations where the stimuli to be discriminated are related (arranged along an ordered continuum such as size, brightness, weight, etc.) the interpretation of what is learned becomes more ambiguous. Here, in fact, the subject is allowed a certain degree of freedom. It can select, as relevant features for the discrimination, either the absolute or the relative properties of the stimuli. Therefore, appropriate paradigms become necessary to detect the properties of the stimuli which control behaviour.

Typically, the paradigms developed to clarify this issue (e.g. Koehler, 1918) use a post-training transfer test to identify the process underpinning the original discrimination learning of the training phase. For example, an animal can be presented with 2 blocks of $1cm^3$ and $2cm^3$ volume respectively, and trained to select the $2cm^3$. Then, in a transfer test the $2cm^3$ is paired with one of (say) $3cm^3$, in absence of differential reinforcement. If the animal selects the stimulus previous rewarded it can be conjectured that it responded to the absolute properties of the stimuli. If, by contrast, it selects the larger of the pair (the $3cm^3$) it is assumed that the relational properties of the two stimuli assumed control over behaviour.

Although widely employed (see Reese, 1968 for an extensive review) this method featuring <u>only one</u> training and test episode has strong limitations. In fact, different alternative explanations have been proposed in respect to the properties of the stimulus which gain control over behaviour even when subjects seem to choose on the basis of the relational properties of the stimuli in the transfer test (e.g. the classical theory of excitatory and inhibitory gradients of generalization which can produce an apparent relational response on the basis of the processing of absolute properties of the stimuli; Spence, 1942).

Moreover, as stressed by McGonigle and Jones (1978), even when evidence for relational learning emerges from the results, another question remains unanswered. This concerns the problem of whether the processing of relational properties of stimuli should be considered as based on systems which act upon primary forms of encoding which are essentially absolute in their nature (and therefore where the relational properties of a set of alternatives play a minor role in the process of decision making) or whether, by contrast:

"relational perception is itself a primary form of coding which makes fewer processing demands on the perceiver" (McGonigle and Jones, 1978, p. 636).

To answer such a question McGonigle and Jones (1978) developed a WGTA based paradigm featuring a long series training and testing episodes for the comparison of relative and absolute stimulus judgement by monkeys. In contrast with Harlow's learning set procedures,[1] the authors adopted tasks in which subjects who responded to the relational properties of pairs of stimuli (such as select always the larger or the brightest) could predict the correct choice on the first trial of novel problems featuring binary choices or triplets of alternatives.

A careful manipulation of different experimental conditions and the comparisons of groups of monkeys with different training histories allowed McGonigle and Jones (1978) to conclude that
the processing of relational properties has a primary role in discrimination learning. Moreover, it was confirmed the hypothesis that relational learning is less demanding for the memory system and

---

1 In which, as we have seen in Chapter I, transfer of learning is measured on successive problems featuring novel and unrelated stimuli and therefore the subject has to discover de novo the appropriate discrimination, on the basis of the outcome of its first choice in each novel problem.

less affected by interference produced by additional stimuli and by transformations of the context in which the task is presented (such as different lighting conditions). It also allows to make transitive choices to the largest of three stimuli following training to choose the larger in a binary discrimination.

In summary, from the results of these experiments it can be concluded that monkeys are not only able to process relational properties of a set of stimuli, when connected in an non-arbitrary way, but also that these forms of processing are very basic, robust and economic in respect to the management of cognitive resources.

These are the conclusions which can be drawn from tasks featuring a limited space of alternatives such as binary or triadic choices. However, these results suggest that, in the context of serial learning studies (such as those reviewed in the previous section), paradigms where the criterion for ordering is provided merely by the temporal succession correlated with reward might overlook fundamental properties of the mechanisms which control the serial organization of behaviour. Now a series of studies, focused on tasks with stronger serial components, which support this conclusion will be introduced.

The experiments mentioned so far featured items connected by relations which are explicitly presented and, therefore, directly perceivable by the subject. A different paradigm focuses on situations where a series of binary discriminations is connected by the fact that each pair of discriminations shares a common term. Here the relation between the items is not perceivable directly and has to be constructed at a representational level by the subject.

A version suitable for comparative studies was used by McGonigle and Chalmers (1977; 1992; Chalmers and McGonigle, 1984), in order to investigate the basis of transitive reasoning within a series of 5 items connectable on "symbolic" basis. The authors compared the role of logic as opposed to more parsimonious forms of behavioural control by means of an experimental procedure featuring a series of connected binary discriminations.

In order to develop their paradigm, McGonigle and Chalmers modified the conventional five term series task A > B > C > D > E (originally designed for children by Bryant and Trabasso, 1971)[2] so that the same procedure and apparatus could be used with human and non-human subjects.

McGonigle and Chalmer's task involved a training on couples of "premises" A+ B-, B+ C-, C+ D-, D+ E- (where the sign plus indicates reward and the sign minus nonreward), presented as coloured containers. The training was followed by a testing phase which featured the presentation of all the possible pairs obtained from the five items A B C D E, in absence of any further differential reinforcement. The authors aimed to assess whether from such a training the monkeys formed an ordered representation of the series A > B > C > D > E, which then would have led to transitive choices in the pairs of non-adjacent items of the testing phase. The subject showed a transitive bias in selecting the appropriate item in the different couples of the testing phases,

---

2 The original task as designed by Bryant and Trabasso (1971) involved a training phase where five rods, differing in length and colour were presented in pairs (the premises A > B, B > C, C > D, D > E). The rods protruded from a box, so that their actual length could not be perceived. The children were required to use the colour differences to make a choice (pointing to the selected item) between different lengths. Following a choice, either the length of the rods was shown or a verbal statement was given as feedback. After this training phase, the subjects were tested on all the possible binary combination of the five terms, without receiving any further feedback after choice.

Children as young as 4 yrs old, showed a transitive bias in their choices in the critical comparison B > D (the only non-adjacent items equally referred as "longer" and shorter during training). This finding was taken as a challenge to the Piagetian assumption that, it is only when children (at about 7 yr old) acquire a relativist notion of relations and manage to coordinate inverse relations around the same term, that they are able to make transitive inferences. It further allowed the conjecture that logic was possibly innate (see Breslow, 1981 for an extensive review of early interpretations of this finding).

including the critical pair B > D (the only non-adjacent items equally reinforced and non-reinforced during training).

The finding that monkeys, which are not well known for being logically competent, were able to solve the five term series task, pointed against the assumption that success in this task was necessarily based on logical skills (as previously interpreted, see footnote 2).

However, the key to assess unambiguously which sort of mechanism supported the transitive choices of the monkeys came from the liberalization of the binary restrictions of the task as originally conceived. In fact, McGonigle and Chalmers (1977, 1986, 1992) devised a new version of the test trials by extending the decision space. Once the subjects were proficient with the binary testing, they were presented on a triadic version of the test trials. The triads were obtained by combining the items of the original sequence in order to obtain test trials in the form, for example, of B > C > D. The rationale was that if the subjects solved the binary version of the task by means of a coordination of each of the pairs of non-adjacent items around a mental representation the absent middle term, as a logicist explanation of transitivity such as that proposed by Inhelder and Piaget (1964) would postulate, then the explicit presentation of this middle item should facilitate transitive inferences.

The results clearly showed the inadequacies of a logicist explanation of the success of the monkeys. In fact, a dramatic decrement in the transitivity of the choices in the triadic testing was observed. This originally led McGonigle and Chalmers to elaborate a stochastic sampling model (see McGonigle and Chalmers 1977 for an accurate description of the model) which allowed to predict the performance of the monkeys on both pairs and triads. That this sort of mechanism was not species specific was indicated by the fact that Chalmers and McGonigle (1984) obtained the same results in a successive study conducted on children.

These are choice based assessments of the mechanisms underlying transitive reasoning. However, other measures supporting the triadic based interpretation come from assays based on decision time analyses of the behaviour of the subjects. These analyses focus on a well known phenomenon found in reaction times of human subjects when faced with tests of ordering skills: the Symbolic Distance Effect (SDE).

In the context of the five term series task, this effect has been reported for children by Trabasso, Riley and Wilson (1975).[3] Subjects trained to criterion on items ordered along a linear dimension, such as their size (for example a series ABCDE where A is the smaller and E is the bigger), in the testing phase show a negative correlation between reaction times and the ordinal distance of the items in the series (e.g. the BC comparison takes longer than the BD comparison). In other words, the time required to perform transitive choices is shorter than the time needed to retrieve the premises taught during the training phase. Therefore, the SDE (when observed) runs against explanations which postulate that logical deductions are performed on-line at the time of testing. In fact, if the subjects memorised the premises independently and then retrieved and coordinated them during the testing phase, one should expect the reaction times to be shorter for the premises and in general to be correlated with the number of transitive inferences needed by the different comparisons (Breslow, 1981).

---

3 The Symbolic Distance Effect has also been evidenced in studies on humans, carried out within the so called mental psychophysics paradigm. In these studies the subjects are required to compare mentally the dimension of well known objects or animals. For example, in a typical experiment, a subject might be required to answer the question: "what is bigger a cat or a whale ?". On the basis of the response time, the underlying representations of the subject and the processes used to compare the different representations is derived. Since the earliest studies (Moyer, 1973; Paivio, 1975), it has repeatedly been found that the response time of adult humans varies inversely with the distance between the referents of the dimension being judged. So that, for example the comparison of a cat with a cow takes longer than one involving a fly and an elephant.

As we have seen, the results obtained by McGonigle and Chalmers (1977; Chalmers and McGonigle, 1984) from triadic testing already undermined the coordination hypothesis. However, since these studies did not incorporate an assay of decision times, the status of the SDE and the psychological mechanism that produces it remained to be explained. In fact, there remained the possibility that different psychological mechanisms were in operation when the phenomenon had been observed in other studies.

When observed by Trabasso and colleagues (Trabasso and al., 1975) the SDE was considered as evidence for the presence of a linear spatial representation of the set of items, which is scanned to find the response to a comparison. The further apart the items are along this spatial continuum, the easier it would be to discriminate them and to find a solution to the comparison. Although this hypothesis implies that no inferences are made at the time of testing, Trabasso and colleagues (Trabasso et al., 1975) argued that subjects use transitive inferences during the training phase in order to construct the linear order.[4]

For a definitive clarification of the issue, a study designed to integrate all the information provided by the presence of the SDE, an assay of the decision times shown in the binary testing and an evaluation of the performance shown by the subjects in the course of triadic testing was necessary. Such a study has recently been conducted by McGonigle and Chalmers (1992) using squirrel monkeys.[5] This study of 1992, was in part a replication of the 1977 study, but it incorporated an evaluation of decision times. Furthermore, the study featured an extensive presentation of the

---

4 Moreover, according to these authors there is a bidirectional principle in the way the premises are learned and integrated in the linear representation during the training phase. Subjects would learn the premises in a "ends-inwards" fashion. Initially they would learn the response to the premises involving the end terms of the series (i.e. AB and DE), and successively those of intermediate position (i.e. BC and CD) (Trabasso et al., 1975).

5 A set of experiments conducted on the SDE in monkey subjects, which although developed in a different context have interesting implications for the themes presented in this chapter are reviewed in appendix A.

triadic testing, in order to evaluate possible changes in performance with practice, although in absence of further selective reinforcement. Some control procedures (in the form of pseudo-triadic tests such as BCC) were also incorporated to ensure that the presentation of three items per se was not the main factor responsible for the performance decrement found between binary and triadic testing of the 1977 study.

Evidence for a Symbolic Distance Effect (SDE), was found for both group and individual subjects and, on the basis of an accurate analysis of the performance and its time correlates in the triadic testing phase, McGonigle and Chalmers were able to reinterpret the meaning of the SDE. The triadic testing, in fact, not only allowed to test a coordination hypothesis in order to explain the competences of the subjects, but also hypotheses based on a linear representation of the five term series. If the SDE is expression of the mental ordering of the set of items A > B > C > D > E, one can expect that subjects who showed the effect in their reaction times should be able to rank the items presented explicitly in a triad such as B > C > D.

As for the 1977 study, decrement in performance was observed at the beginning of the triadic testing, as compared with the level of performance achieved on the binary testing. This result suggests that it is unlikely that the performance of the subjects was either based on the coordination of non-adjacent items around the middle term or that a perfect linear representation of the set of items was formed during the binary testing.

From a further assay of the reaction times showed in the binary testing McGonigle and Chalmers (1992) obtained data which supported a novel interpretation of the SDE, not necessarily based on linear ordering. In fact, these author showed that, if plotted as a function of the end points and not of the distance of the items (i.e. AB, AC and AD vs. DE, CE, and BE), the reaction times formed two population of scores, one fast (for the pairs including the term A) and one slow (for the pairs including the term E).

Thus, the SDE observed in the binary testing can be explained by the fact that, as the distance of the separation between two items increases, the more likely it is that the two items belong to the two different populations of fast and slow responses. The results of this analysis thus supported an interpretation, as that provided for the 1977 study, where processes more parsimonious than linear ordering were proposed in order to explain the transitive choices in the binary testing.

However, an analysis of the errors observed during the triadic testing showed that the selection of incorrect items was not performed at random. Each subject showed a bias in neglecting some particular items of the set and some preferences for other items. Thus, also the stochastic model originally proposed by McGonigle and Chalmers (1977) seemed unsatisfactory in the light of these latter results. This led the authors to develop an alternative explanation based on an ordered set of conditional rules, such as if 1) E is present select it, 2) if A is present avoid it and so on. The model has also been implemented as a computer simulation based on production rules and seem to fit both group and individual data (for an accurate description of this model see Harris 1988; Harris and McGonigle, 1994).

Since it had already clearly emerged that task practice was a factor of primary importance for an accurate evaluation of the competences of the subjects, McGonigle and Chalmers (1992, experiment 3) presented the subjects with an extended phase of triadic testing without supervised learning procedures. A significant improvement of performance was observed under these conditions. In order to explain such a dramatic change in absence of explicit feedback, the authors proposed that the search for a solution which was applicable to any of the triads could have been an incentive per se. In other words, the subjects were self-regulating their behaviour in order to find a solution to the task which for the memory system was more economic than storing an ad hoc solution for each of the triads. In fact, on a set based rank, a response to only three items (C, D and E) has to be encoded in order to solve at a relatively high level of accuracy all the ten triads; every deviation from this strategy would just increase the number of rules to be remembered.

Furthermore, after having assessed that the monkeys were able to self-regulate their behaviour up to a given level of efficiency, McGonigle and Chalmers (1992, experiment 4) trained some of the monkeys on the triads, this time providing explicit feedback in the form of selective reinforcement of transitive choices. The percentage of errors of the monkeys was further reduced, falling to 15.6% in average, compared with the 36% registered in experiment 3, where triads were presented without differential feedback. In this way, it was possible to demonstrate that a perfectly transitive ranking over 3 simultaneously presented items was not beyond the cognitive competences of the monkeys. On the contrary, given enough feedback and task practice they became fully competent on the five term series problem, even presented in its triadic form. Moreover, only when they reached such a level of expertise on the task, the response times to the items C D E started to show a clear linear profile, with the fastest responses registered for the "biggest" item and the slowest for the "smallest" of the three (only reaction times for these three items were reported in the study, according to the hypothesis that learning the correct choice to a subset of the five items would have been sufficient to perform correctly all triads).

The fact that monkeys eventually showed a fully principled transitive behaviour after having practiced the task, as a product of self-regulation or further feedback, seems to show that only when the subject become expert with the task, they were able to change from a strategy effective only in the binary testing to one which was based on a some form of ranking and thus fully compatible with all the versions of the task.

Overall from this extensive study it was possible to conclude, first of all that the SDE had been over-interpreted as expression of the presence of scanning processes which operate on an ordered linear representation of a series; and, secondly, that from binary versions of transitivity tests it is much more difficult to infer principled forms of control of sequences of responses. By contrast tasks requiring the organization of multiple items presented simultaneously seem to be much better proof of the ordering abilities of the subjects.

These studies are an example of how a strategic use of monkey subjects in cognitive research can shed some light on controversial issues regarding human cognition. It allows us to evaluate whether the role of language has to be taken as essential in tasks traditionally considered as tapping on logical skills, and to assay the effects of protracted periods of testing, often prevented by the use of young children as subjects.

In summary, the paradigms developed in the context of the studies on transitivity provided the basis for interspecies comparisons in tasks where the subjects had the opportunity to detect the intrinsic "orderability" of a set of 5 items. From the results it was clear that monkeys (and young children) rank the set, that they do so quite spontaneously and by means of mechanisms more basic than logical skills. However, it was also apparent that situations where the size of the set is expanded and where items are presented simultaneously represent the best test of the status, the possible function, and the limitations of the ordering abilities of the subjects. The expansion of the decision space has in fact a twofold rationale. On the one hand, it provides a transparent window on the competences underpining the performances of the subjects. On the other hand, the problem of facing a large decision space might represent an incentive for the subject to deploy more sophisticated strategies in order to cope with the increased demands of the task.

These considerations led McGonigle and Chalmers to develop a research program aimed at the evaluation of the competences necessary to solve explicit size seriation tasks of sets of multiple connected items presented simultaneously (McGonigle and Chalmers, 1986; McGonigle, 1987; McGonigle and Chalmers, 1992; Chalmers and McGonigle, 1993).

In fact, seriation tasks usually involve the presentation of a set of objects of different sizes and a child has to reproduce a model presented by the tester, which features the items in a linear monotonic series from smallest to biggest. Since early Piagetian studies (see Inhelder and Piaget 1964), it has been shown that the ability to solve seriation tasks is one of the most robust indicators of cognitive growth. However, as originally conceived, seriation tasks do not allow an unambiguous characterization of the nature and the ontogenesis of the competences required for their solution.

Traditionally, seriation has been considered (as for the ability to perform transitive inferences) as requiring logical competences such as the understanding that an item can be the biggest of one class (the pool of the items not yet selected) and at the same time the smallest of another (the set of items already lined up in a descending monotonic series).

By contrast, McGonigle (1987) proposed the hypothesis that even logically competent subjects might solve seriation tasks using a combination of strategies (which do not necessarily require a high level of relational and ordinal competence). At the same time, failures might be accounted for by difficulties in only some sub-components of the task, such as the ability to organize sequences of responses or the ability to compute the ordinal status of each item.

Moreover, the original piagetian version of the task is very difficult to implement in a comparative context. For example, it would be very difficult to attract the attention of a monkey on an ordered construction of three-dimensional objects, and (for ergonomic constraints) to train it to select and align a set of objects scattered on a table in front of a cage. In this situation it would be very difficult to attribute possible failures to pure cognitive limitations as opposed to spurious factors deriving from motor constraints or the level of general activity of the subject.

Thus, McGonigle developed new paradigms aimed at the decomposition of seriation tasks by means of training procedures (McGonigle 1987, 1989) and suitable to be implemented with non-human subjects. One study (McGonigle and Chalmers, 1986; McGonigle, 1987) conducted on squirrel monkeys (Saimiri sciureus) was WGTA based. The animals were presented with sets of objects ordered by size and on the basis of the colour of the set had to learn to choose one belonging to a particular ordinal position (e.g. if the objects are black choose the biggest, if they are white choose the smallest, and so on). In this task monkeys reached an almost perfect performance (McGonigle and Chalmers, 1986).

Although informative regarding the ability of the monkeys to process ordinal information, this task was however based on a single response per trial, albeit performed within a relatively large set of ordered items. These sorts of tasks are not informative respect to the mechanisms which underpin a serial production within similar contexts. However, for the reasons outlined above, in order to present seriation tasks to monkey subjects new technologies such as touch sensitive computer monitors, which do not require the careful manipulation of three-dimensional objects, were required.

Therefore, an extensive project aimed to the investigation and the decomposition of seriation skills in human and non-human subjects was implemented as a set of experiments featuring the presentation of icons on touch sensitive computer monitors (McGonigle, 1987, 1989; Chalmers and McGonigle, 1993). A schema of the different paradigms used by McGonigle and Chalmers in their studies on seriation is provided in Fig. 2.1.

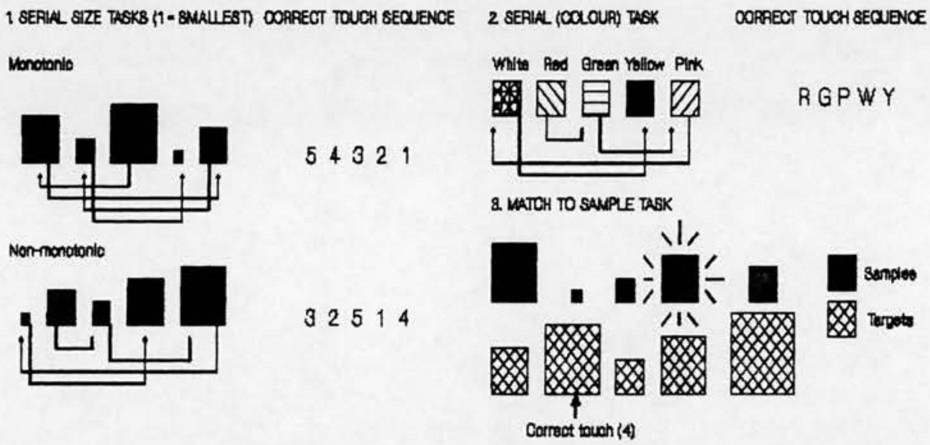



Fig 2.1. Schema of tasks employed for the decomposition of seriation skills by McGonigle and Chalmers. See text for explanation.

Children, ranging from 5 to 7 years old, were required to select one after the other icons of different sizes and received feedback for success or failure. By means of this paradigm (see McGonigle and Neapolitan, 1993; Chalmers and McGonigle, 1993) it proved possible to compare different forms of seriation, such as monotonic (e.g. from smallest to biggest) and non-monotonic (e.g. second biggest, smallest, third biggest, biggest, forth biggest) series. Moreover, two forms of seriation were also compared with serial learning, presenting the subjects with a set of icons of the same size but different colours to be touched in an arbitrary order, as in the experiments quoted in the previous section.

Even 5 year olds were able to seriate with this training procedure. However, the form of seriation proved to be an important predictor of success. In fact, 5 year olds were able to learn to seriate only monotonic sets of items. By contrast, non monotonic series proved difficult even for 7 year olds. For the 7 year old children, the speed of learning on the monotonic series was also higher than on the serial learning of arbitrarious sequences.

This finding was interpreted by McGonigle and Chalmers as suggesting that subjects were able to seriate by noticing the value of the size interval between adjacent items of the series, and thereafter selecting in turn the correct items by reiterating the same response to this size interval. The deployment of such a strategy was impossible in non-monotonic series where the size interval between adjacent items changed for each response.

This interpretation was also corroborated by the results obtained from another condition of the experiment (Chalmers and McGonigle 1993), specifically aimed to the evaluation of the role played in seriation by ordinal competences alone. The paradigm was based on a matching to sample task. Two sets of icons, the samples and the targets, were simultaneously presented on a touch screen. The size of the icons varied both within each set and between the two sets. The sample icon blinked (e.g. the third biggest in the sample set) and the subjects had to select the one which held the same ordinal position within the icons belonging to the target set (e.g. the third biggest of the target set). As compared to the seriation tasks described above, this ordinal task proved to be the most difficult for all the subjects. In sharp contrast with accounts of seriation based on high level ordinal competences, thus, there was evidence for the fact that subjects who failed in ordinal tasks were nevertheless able to reproduce a highly constrained series such as the monotonic one.

On the basis of these results, Chalmers and McGonigle (1993) proposed that seriation is to be considered as a search task, where the subjects try to detect forms of constraints afforded by the configuration of items, in order to produce long sequences of responses, without overloading the memory system. In fact, whereas, in learning a long series of arbitrarious responses (such as the non-monotonic size seriation or an arbitrary sequence of colours) the subject needs to rely on its brute memory alone, with a monotonic series (once its implicit redundancy has been detected) the reiteration of an identical response to size interval reduces almost to nothing the amount of memory resources required by the task.

This argument led McGonigle (1987) to develop a research program aimed at assessing presence of a tendency towards economy in human and non-human primates faced with different situations where the subject was given the opportunity to deploy data reduction strategies to simplify tasks otherwise demanding for the cognitive resources of the organism. This program would have featured the presentation of both seriation tasks (like those administered to the children in McGonigle and Chalmers experiments) to monkey subjects and classification tasks to monkeys and children to assess whether they would have exploited strategically and spontaneously the categorical affordability of a set of icons to be memorised. The implementation of the program started with the administration of seriation tasks to the squirrel monkeys used for previous experiments on transitivity. However, the ageing subjects died one after the other before the project could be completed. This thesis represents the first test opportunities for viable monkeys within the program.

## 2.4 Conclusions

Overall, a critical look at the literature on serial aspects of behaviour seems to indicate that the use of paradigms featuring multiple connected items represent the best policy for programs which are aimed at the characterization of the cognitive potential of primates.

First of all, it seems that when serial learning is considered on the basis of sequences arbitrarily defined by the experimenter, the subject is left with only few strategical possibilities to organize its behaviour in ways which are fitted to the task. These take the form of chunking the lists to be recapitulated on the basis of either temporal contiguity, or the belonging of the items to different classes. Moreover, there is some evidence for some forms of transfer, based on the formation of "empty" templates for a list, which can be then filled up by the particular list at hand. However, these studies failed to provide a characterization of these templates.

On the contrary, studies which have used structured material to be organized in sequences have produced results which allow some richer interpretations of the processes which underlie the performance of the subjects.

In the context of binary discrimination learning, careful studies have demonstrated that the processing of relational as opposed to absolute properties of the stimuli is a more basic, robust and economic mechanism. Therefore, it can be assumed that paradigms featuring unconnected items force the subject to rely on more demanding and less effective processes and obscure important cognitive potentialities of the organism under examination. However this observation needs further verification in contexts featuring an expanded set of stimuli.

The studies on transitivity, using paradigms series of connected binary discriminations, show that when a set of five items has an intrinsic ordered structure the subjects are able to detect it and organize the items accordingly as expressed by their transitive behaviour.

However, the mechanisms on which this ordering is based became evident only when the paradigm shifted from the binary to the triadic context. Here, strong evidence was provided for the fact that logical competences are not required for transitive reasoning.

By contrast, it seems that some forms of rational decision making, based on the ranking of rules of selection to a sub-set of the items of the ordered series, allow a satisfactory solution to the task. Interestingly, furthermore, such organization of behaviour, seems to emerge quite spontaneously, when the subject is given enough task practice, as a strategy aimed to reduce the amount of memory resources required to solve the task.

This strongly suggested the need for the use of experimental situations where the subject is faced with large sets of rankable items presented simultaneously. In fact, whereas the serial behaviour of the subjects becomes more easily interpretable when the structure of the set is presented explicitly, the enlargement of the set of stimuli might have provided an incentive for organization and, at the same time, enabled the evaluation of the length of the (non-arbitrary) series that subjects are able to cope with.

The studies of seriation helped the clarification of these issues that would have been difficult in the context of the transitivity paradigms. The use of new touch screen based procedures made it possible to set the basis for a comparative study of seriation and enabled the implementation of a set of tasks aimed to the decomposition of seriation skills.

The results showed that, when items embedded in a non-arbitrary series (of up to 7 items) have to be selected in succession, subjects benefit from constraints which the situation affords by means of the deployment of strategies which allow the reiterative use of a single response (again, economizing the amount of cognitive resources otherwise required by the task).

On the basis of these results, following McGonigle and Chalmers (1986), it can be hypothesized that there is a tendency towards economy at the basis of cognitive growth. This argument can be applied either ontogenitically and phylogenetically, so that, for example an older child or a more complex species, might be competent on tasks failed by younger children or lesser organisms, because, by strategic means, they manage to make it easier.

## 2.5 The economy/data-reduction hypothesis and the development of the experimental program presented in this thesis

The economy/data-reduction theory proposed by McGonigle and Chalmers can be characterized in greater detail as follows. First of all, a distinction must be made between different cost functions which can be applied to the solution of a serial problem.

One source of costs is provided by the environment itself. In the first instance, fitness to the task is required. The fitness of goal oriented behaviour can be considered the achievement of the goal itself. Then, within the success space, a dynamic regulatory function is provided by the external costs of different solutions. The economy of a particular solution can be evaluated in terms of the number of moves performed to reach the goal state. If one assumes that each move requires a constant amount of time to be performed, then every redundant move adds time costs to the achievement of the goal.

On the other hand, a further regulatory function is provided by the internal management of cognitive resources. A subject starts to select those strategies which sustain behavioural fitness without putting his cognitive system in danger of overloading. It is in this way that the serial organization of behaviour can be characterized as a search problem. In a situation where the subject is faced with the problem of choosing successive moves in order to achieve a goal, it will start to seek for those constraints afforded by the situation at hand which allow to decide which move to do next investing a minimum of cognitive resources. A factor of major importance in this respect is that the serial problem should be one which really challenges the cognitive system of the subject. For example the length itself of the serial production must be long enough to represent an incentive for the subject to deploy data reducing strategies.

An experimental program aimed to evaluate such issues should, thus, be characterised by the use of tasks which allow multiple solutions. Moreover, these solutions should be of a kind that allows a measurement of their respective economy defined in terms of the relative amount of external cost that each of them requires. Finally, it should be available to the subject a rich set of strategic possibilities to chose from, and they should be transparent enough to allow an interpretation of their economy, this time defined in terms of internal costs of cognitive management.

Obviously, to implement such a project a major paradigm shift is required with respect to the traditional tasks employed by comparative psychologists. Tasks based on binary choices fail to provide the dynamic spectrum of different strategical solution and their possible changes over time as a function of gaining task expertise. Serial learning tasks, seem inadequate too. The rigid training imposed on the subject does not leave enough freedom to evaluate the spontaneous interplay of different regulatory functions. The arbitrariness of the sequences, moreover, does not offer enough constraints for the subjects to impose some form of organization on their serial production.

Therefore, McGonigle (1987) proposed a series of tasks which would have allowed the subjects to express spontaneously strategic factors (such as classification and seriation along different linear dimensions) allowing a better management of the amount of data to be retained to solve the task. However, the part of the project that should have focused on non-human primates had to be suspended because of the mortality of the sample.

This thesis is a natural development of the project. In fact, at the time when the work for the present study was planned, a new colony of capuchin monkeys (Cebus apella) had just been established at the Laboratory of Cognitive Neuroscience of the University of Edinburgh. The study is an attempt to devise and implement new paradigms which provide a richer set of dimensions (such as spontaneous regulation and deployment of strategic factors) on which different organisms can eventually be compared.

The prototypical set of tasks employed here is based on an exhaustive search within a set of icons presented on touch sensitive computer monitors.

The first versions of the task feature the presentation of sets of identical icons which can be individually identified on the basis of their spatial location on the screen (see Chapter IV). Then, richer sets of organizational possibilities are offered to the subjects with the presentation of sets of icons which can be divided in different categories on the basis of distinctive features such as shape and colour (see Chapters V and VI).

The paradigm, in all its different versions, offers the subjects the opportunity to develop strategic search modes based on the constraints (either spatial or non-spatial) afforded by the search space. For the high degree of spontaneity accorded to the subject and for the rich set of organizational possibilities afforded by the sequences to be produced, the new paradigms differ radically from traditional serial learning tasks. Nevertheless, they share with them a strong serial component and the fact that multiple responses have to be produced (in contrast with tasks based on binary choices).

The main focus is on non-human primates. However, whenever analogous data on different species were collected or available from different studies, these have been described and discussed.

Despite the fact that the research project featured here, as we have seen, requires specific tests for its investigation (and therefore the development of the paradigm outlined above), a practical problem had to be solved before the core sets of experiments could be conducted. This problem rested in the fact that the experimental subjects at my disposal were wild born, and as such still completely naive in respect to any experimental environment and without any previous extended contact with humans.

Therefore one of the first objectives of this study was to familiarize the subjects with the testing environment and the tester. For this scope it seemed necessary to endure a transitional phase, offering more opportunities for the subjects and the tester to interact than touch screen based procedures would have allowed. This transitional phase will be described in the next chapter. It was implemented as a set of experiments based on the modified version of the WGTA as described by McGonigle and Chalmers (1977, 1992). The reason underlying the selection of the apparatus and the experiments will be presented in detail in Chapter III. The reader however should bear in mind that those experiments represent only a first step towards the implementation of the rest of the research program and are somehow detached from the major issues on which this thesis focuses. Therefore the next chapter should be taken as a long, albeit necessary, parenthesis in the flow of the arguments presented in this study.

# CHAPTER III

## THE FAMILIARISATION OF THE MONKEYS

### 3.1 A WGTA based search task

As mentioned in the previous chapter, the first necessary step towards the implementation of the present research project was to familiarize wild born capuchin monkeys with the testing environment and the tester himself.

In fact, the monkeys had just completed their period of quarantine and a new colony had been established at the Laboratory of Cognitive Neuroscience of the University of Edinburgh. In order to provide an optimal social environment these monkeys were kept together in a large indoor cage. Therefore, in order to test them separately, it was necessary to establish a daily routine where each monkey spontaneously entered an individual testing cage. Obviously, a precondition to this was that the monkeys had a good relationship with the tester and were faced with testing situations which they were comfortable with.

An appropriate context with which to start seemed that of search tasks where the retrieval of a hidden object within a set of three-dimensional occluders was required. The decision to begin with these particular tasks was chiefly motivated by practical needs.

These practical aspects rested in the necessity to find a straightforward experimental set up. A suitable apparatus was identified in a modified version of WGTA (McGonigle and Chalmers, 1977; 1993) which allows the presentation of arrays of up to five items. The readiness to adapt to the WGTA is well documented for monkeys in general (Harlow, 1949; Warren, 1965; McGonigle and Chalmers, 1977; 1993) and capuchins in particular (De Lillo and Visalberghi, 1994). Moreover, considering the spontaneous tendency of capuchin monkeys to manipulate objects (Visalberghi, 1990), the use of an apparatus that requires the displacement of three-dimensional objects seemed particularly fitting for this species. Finally, putting the subjects directly in front of the tester, the WGTA is appropriate for socially familiarizing the monkeys with the tester. In fact, Harlow (1949) enthusiastically described how macaques implicitly develop "social learning sets" during repeated experience with different testers at the WGTA.

Moreover, with the sort of experiments presented in this chapter it seemed possible to overcome some of the difficulties which WGTA based tasks have traditionally encountered. As emerged from the review of Harlow's WGTA studies on learning-set presented in chapter I, the main limitations of WGTA based procedures can be identified in 1) the fact that correct responses were arbitrarily defined by the experimenter and therefore the subjects are forced to rely on inductive processes in order to find the solution to the task; and 2) the binary nature of the tasks which prevented an evaluation of forms of control appropriate to deal with more realistic situations featuring a multiple choice problem space.

Search tasks where an object is hidden within a set of multiple occluders in presence of the subject allow on the one hand to liberalize the binary discrimination context of traditional learning set studies. On the other hand, since the relevant information (the hiding procedure observed by the subject) about the location of the object to be retrieved is provided from the outset, the subject does not necessarily have to rely on blind trial and error learning in order to solve the task (see e.g. Spence, 1951 for a distinction between trial and error and "insightful" learning processes). This allows an evaluation of whether subjects show any tendency to search strategically, on the basis of the constraints afforded by the search space and/or of relevant information at their disposal.

Therefore, these first experiments, although somehow detached from the main theme of the rest of the thesis, allowed nevertheless the introduction of the subjects to situations encouraging the spontaneous deployment of strategic modes of search in contrast with tasks featuring a predominant role of selective reinforcement over blind trial and error learning procedures.

Search tasks of different types have a long history of exploration in research on both human and non-human species. Perhaps the oldest form is the delayed-response task pioneered by Hunter (1914), elaborated later with primates by Yerkes (1929), and culminating in the well known versions introduced by Piaget (1955) as "object permanence" tasks. A common feature of these tests is the role they give the tester who first acts to capture the subject's attention by presenting an attractive object, such as a toy, or a piece of preferred food, before hiding it within containers or behind occluding screens which remain at all times within the test field. Either immediately afterwards, or following a delay, the subject is given an opportunity to seek hidden items. Persistence of search in the absence of direct perceptual information is the first indication of object memory or event permanence. However, beyond search orientated behaviour per se, the use of strategies may indicate a great deal about the subject's ability to constrain search to the relevant occluder alone, or, in the case of direct information concerning the precise location of the bait, to infer from its absence the most likely locations which remain to be explored.

Thus, infants in an object permanence task may search the first occluder visited by the experimenter, even though the item sought has been quite explicitly removed and placed in an alternative site by the experimenter, in full view of the subject (Bower, 1974; Diamond, 1985). In tasks, furthermore, where the act of hiding could not be perceived directly, but could be inferred only from the sequence of events in each task, Haake and Somerville (1985) found a strong developmental trend from 9 month to 18 month old infants in the way they co-ordinated temporal and spatial information.

As these authors point out, the sequential nature of the hiding procedures required children: "to attend to, remember and use information about the presence and absence of the object in the context of movements among potential hiding locations. In order to search logically, events occurring at different times and places in the displacement sequences had to be linked together to determine exactly where the object had been hidden" (p. 185). Under these conditions, only the oldest children in the sample showed some response consistency searching, for example, the last place the target object was seen, following a discovery that it was now "missing".

Tests of "pure" cognitive competences are rare, if they exist at all, however, as in so many other such cases, the task used - or the conditions under which it is introduced - may itself be a factor in determining whether subjects will deploy exhaustive and relatively inefficient strategies or opt instead for controlled, efficient search. Certainly, Wellman, Somerville and Haake (1979) found considerable task-induced differences in strategies (which themselves change with age) of children aged from 2 to 6 years. Given doors in a cupboard to search versus areas of a playground, for example, children's responses indicated that "searches were more systematic and comprehensive in the cupboards than on the playground" (p. 541). The authors account for this difference by suggesting that searching "logically" makes greater demands on limited cognitive resources than other strategies, and that certain environments may be easy to search completely, as their cupboard task indicates. This factor has been given further emphasis by the findings of Somerville and Capuani-Shumaker (1984). Their study, which forms the basis of the investigation reported in this chapter found that children from 3-5 years of age were able under some circumstances at least, to constrain search on the basis of watching a tester hide or find an object within a small test field containing 4 occluders. Somerville and Capuani-Shumaker suggest, in fact, that making the children pay particular attention to the task (affording, as it does, a low cost solution even if the subject searches randomly), is a crucial factor in the subjects' success.

Whatever the role of the various factors involved, it is surely clear that the tasks described are of interest to the comparative psychologist, designed as they are both to evaluate the role of observer based and self-directed (discovery based) inference, in promoting search economy in a situation not unlike a foraging task (see McGonigle and Chalmers, 1992). Watching a conspecific visiting putative food sites, for example, may materially reduce the costs of search by an observer otherwise left to its own devices. Left to its own devices, on the other hand, the way an agent searches may also have a profound impact on the effort it expends when achieving its goal. Executing search in a random and unprincipled way, for example, could lead to costly reiterations, especially if the search space and the actual space to be searched is large (see Olton, 1982).

In summary, there would appear to be 3 main aspects to the search problem as described. The first is concerned with the extent to which an agent can improve efficiency of exploration when observing the behaviour of others. The second is concerned with the sorts of strategies which an agent, working on its own, will devise to keep search as efficient as possible. Common to both, are the cost functions which the agent must calculate in deciding if it is worth the effort of devising a strategy designed to make search efficient i.e. the "cost" of inference must be offset by the benefits of search economy.

Apart from their value as a way of familiarising the monkeys with the testing environment, the experiments reported in this chapter, aimed to determine whether socially transmitted search constraints are exploited by a non-human primate (Cebus apella). In the light of the somewhat conflicting evidence from the developmental studies just cited, however - attributed, at least in part to the specific verbal instructions and pre-task procedures adopted in the Somerville and Capuani-Shumaker study (1984) - it was decided to first test some children of ages similar to those used in the Wellman et al. (1979) and Somerville & Capuani-Shumaker (1984) investigations. In doing so the tasks were modified as necessary, making them as similar as possible to those designed for the Cebus apella - who could not benefit, of course, from linguistically based instruction. In this way, it was hoped to establish a robust template of performance in young children (when amalgamated in the research just cited), against which the non-human primates could be compared.

## 3.2 Experiment 1

### Method

#### Subjects

The subjects were 9 children (5 boys and 4 girls) with a median age of 4 years and 11 months and a range of 4 years and 8 months to 5 years. All children attended the nursery of the Department of Psychology of the University of Edinburgh.

#### Apparatus

The testing apparatus was a specially modified WGTA, designed to enable the simultaneous presentation of a maximum of 5 stimuli (McGonigle & Chalmers, 1992). For the test described here, four white plastic cups were presented in line across a 18 x 50 cm tray. The objects to be hidden were a red and a blue rubber eraser which could be enclosed in the tester's hand and placed silently under a given cup, without giving any clues that it had been secreted there.

#### Design

The experiment comprised two tasks, i.e a Hiding Task and a Finding Task. Each task involved an Absent Condition and a Present Condition.

The presentation of each task was preceded by warm-up trials of two kinds (dubbed Type 1 warm-up trials and Type 2 warm-up trials). There were 4 different types of warm-up trials i.e. a Type 1 and a Type 2, for the Hiding Task; and a Type 1 and a Type 2 for the Finding Task.

Each subject was presented with both the tasks. Five subjects, selected at random (three boys and two girls), were presented first with the Hiding Task and the remaining four (two boys and two girls) were presented first with the Finding Task.

Procedure

General procedure. Before each daily session, each child was taken from the departmental nursery to an adjacent testing room. The child sat in front of the tester, in the full view of the cups, unless occluded by a screen. Each session lasted approximately 12 minutes.

A camera positioned in front of the subject recorded his\her behaviour while looking at the hiding\finding procedure as well as while responding.

Task specific procedures. A schema of the procedure followed in the Hiding and the Finding Task is shown in Figure 3.1.

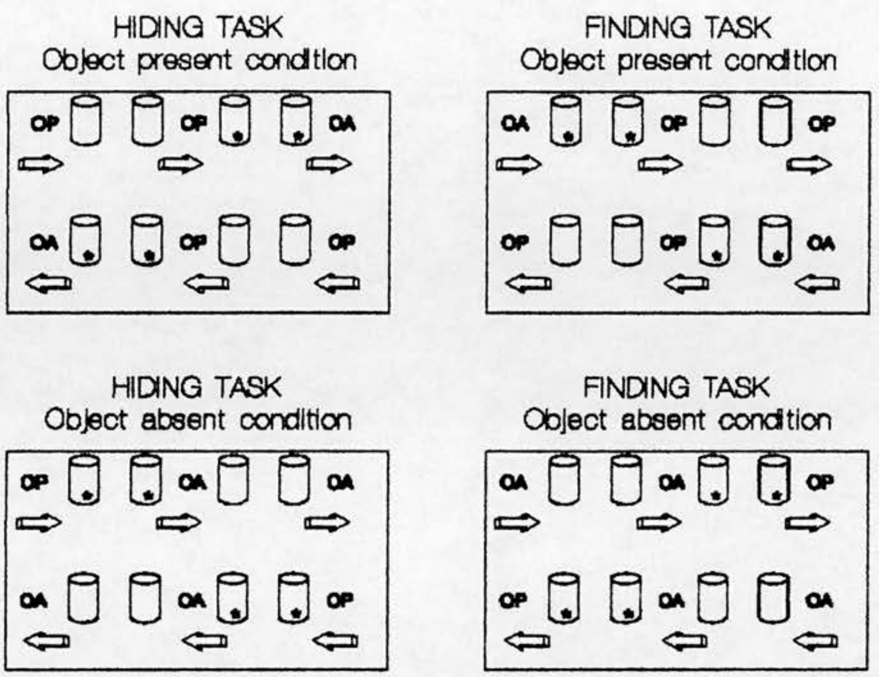



Fig 3.1 Schema of the conditions featured in the Hiding and the Finding Task. OP (Object Present) indicates that the object is shown to be present in the tester's hand at that point of the sequence of displacements. OA (Object Absent) indicates that the tester's hand is shown to be empty at that point of the sequence of displacements. Arrows indicate the direction of travel of tester's hand. Asterisks indicate the possible location of the objects after the completion of the Hiding\Finding procedure.

1) HIDING TASK

Instructions. At the beginning of each test session, each child was told that the object was going to be hidden under a cup and that their task was to find it by lifting the cups.

Each daily session comprised the following trials.

a) <u>Object-Present Condition</u>. Eight experimental trials were administered in which the object was shown in a open hand at one end of the cups array, the hand was then closed and moved under all the four cups in succession before being opened after the fourth cup to show that the object was now absent. Between the second

and the 3rd cup the hand was opened, to show that the object was still in transit.
b) Object-Absent Condition. Eight experimental trials were administered. These were identical to those of the Object-Present Condition, except for the fact that, when the hand was opened between the 2nd and 3rd cup, it was shown to be empty.

Half of the sequences in both conditions were left-to-right, half right-to-left. The four experimental trials comprised one of each of the four different sequences generated by combining the two directions of travel with the two intermediate events. The order of presentation of the four trials was selected at random for each daily session.

2) FINDING TASK
Instructions. The verbal instruction given to the children at the beginning of each daily session was: "here are two objects which always hide together under the same cup; I will find one of them and you must find the other one".
For each trial, a cardboard screen was interposed between tester and subject and the objects (now two, a red and a blue eraser) were hidden under one of the cups, out of the sight of the subject. Each daily session comprised the following trials.
Eight experimental trials analogous to those given in the Hiding Task, except that the objects were hidden, initially, out of sight of the subject. The informing event at the beginning of the sequence was now the absence of objects in the tester's hand. The intermediate event between the 2nd and the 3rd cup was either the presence of one of the objects in tester hand (Object-Present Condition) or its absence (Object-Absent Condition).

The counterbalancing of the sequences of displacements followed the same schema as described for the Hiding Task.

Warm-up.
a) Warm-up for the Hiding Task. Before the administration of the Hiding Task, each testing session featured the presentation of

warm-up trials of two different types (Type one and Type 2), as described below:

- Type 1 warm-up trials where presented until a criterion of two consecutive correct searches, each performed within a latency of 5 sec, was achieved. In each of these trials, the tester's hand was moved under one cup and then removed and opened in order to show that the object had gone. The subject was then allowed to search.
- Two type 2 warm-up trials where the object was shown in an open hand at one end of the cups array, the hand was then closed and moved under all the four cups in succession, before being opened after the fourth cup to show that the object was now absent. The subject was then allowed to search.

In both cases above, the movement of the bait alternated in direction from trial to trial.

b) Warm-up for the Finding Task. As for the Hiding Task, the administration of the Finding Task was preceded by the presentation of type 1 warm-up trials (until the achievement of the criterion) followed by two type 2 warm-up trials. These trials were similar to those used for the Hiding Task. The only difference was that the tester's hand was initially shown empty and later, after passing under the cup(s), revealed to contain one of the objects.

Data recording.

The tester recorded the location and order of occurrence of each search performed by the subject. A search was defined as the lifting of a cup.

A scrutiny of the videotape records, in slow motion mode, was performed in order to ensure that the data analysis was conducted only on those trials in which the subject watched without interruptions the whole hiding\finding procedure. In the event, none of the trials had to be eliminated.

Statistical analysis

All statistical analyses were based on the Binomial Test. In the Warm-up 1 trials the probability of occurrence of a correct search by chance was p = .25. In the Hiding and the Finding Task the probability of performing an appropriate first search by chance was p = .50, whereas that of locating the object by chance in a second search (following a correct but unsuccessful first search) was p = .33.

## Results

### Warm-up trials

Type 1 for the Hiding Task. All the children understood the hiding procedure, satisfying the criterion of two consecutive correct responses with a latency of $< 5$ sec. The group performance was of 87% correct responses ($p < .01$) and all subjects showed a proportion of correct responses above chance level ($p < .01$).

Type 1 for the Finding Task. All children but one understood the procedure, reaching the criterion of two consecutive correct responses, each performed within 5 sec. The group performance was 45% correct responses ($p < .01$). The individual scores showed that 7 children out of 9 performed a significant ($p < .01$) proportion of correct responses.

Type 2. In view of the similarity of the warm-up type 2 data from both Hiding and Finding Tasks, these have been combined in an analysis of the different modes of search adopted by the subject to explore the array. These have been divided into Systematic mode of search (SS), i.e. the subjects explored the array from one end to the other (no fixed sequences other than end to end exploration were observed) and Asystematic mode of search (AS) i.e. the search was performed at random.

As should be expected by chance ($p = .25$), in the absence of any clue about where to search, the subjects located the object on their first choice only in the 21% of the trials. Subjects adopted a Systematic mode of search in the 49% of the trials and an Asystematic one in the 30% of the trials).

### HIDING AND FINDING TASK

Group Performance

First searches. The percentages of appropriate first searches performed in the Hiding and the Finding Task are shown in Table 3.1(a) and 3.1(b), respectively. From these it can be seen that appropriate first searches were performed above chance level.

However, the most striking feature of the data is the selective effects of the sub-conditions within each task type. In the Hiding Task, the Present Condition contributes almost uniquely to the overall success within this condition; in the Finding Task, by contrast, the Absent Condition is the more successful.

Second searches. As not all searches could be correct on the first choice even when controlled by a logical strategy, it was necessary to analyse second choice behaviour following putatively appropriate if unsuccessful first choices. Thus, second searches have been included in table 3.1 and show that in the case of both tasks, second choices (following an appropriate but unsuccessful first choice), are significantly performed in the correct location.

As Somerville and Capuani-Shumaker (1984) point out, however, it is necessary also to distinguish between endpoint based second choices and mid-position ones. This is because second choices which follow on from a choice of an end location and are adjacent to the endpoint, may be simply the result of object proximity per se, and not at all a reflection of the subject's understanding of the implication of the first choice. By contrast, when the first choice is performed at an appropriate inner location, a second response performed on the basis of mere proximity would locate the object only in the 50% of the occasions. For this reason, in Table 3.1 second searches have been divided into those that followed first searches at inner and end points of the array. It can be seen that subjects were significantly correct in both tasks, even when only second searches following a first search at an inner point are considered.

Table 3.1 Distribution of children's searches in the Hiding and the Finding Tasks (see text for explanation.

**a) HIDING TASK**

| | | | | |
|---|---|---|---|---|
| Appropriate first searches | 74%** | | OP Condition | 89%** |
| | | | OA Condition | 58% |
| First search outcome | Object found 66% | Object not found 34% | | |
| Correct second searches | - | 88%** | Following Inner | 79%** |
| | | | Following End | 100%** |

**b) FINDING TASK**

| | | | | |
|---|---|---|---|---|
| Appropriate first searches | 68%* | | OP Condition | 61% |
| | | | OA Condition | 75%** |
| First search outcome | Object found 69% | Object not found 31% | | |
| Correct second searches | - | 75%** | Following Inner | 67%** |
| | | | Following End | 84%** |

* = $p < .05$
** = $P < .01$

Individual performances

Given the asymmetrical distribution of appropriate first searches between the two sub-conditions of the Hiding and the Finding Task and between the two tasks themselves, it is particularly interesting to consider how each child handled the various situations. The frequencies of appropriate first searches performed by each subject are reported in Table 3.2. From this it can be seen that only one child succeeded in both conditions of the Hiding Task and one in both conditions of the Finding Task. The behaviour of the remaining subjects is highly consistent with the group results: performing appropriate searches mostly in the easiest condition of each task. No subject succeeded on both tasks. Combining the percentages of correct first searches performed by each subject in the two conditions of each task, it emerges that three children succeeded in the Hiding Task and two in the Finding Task.

Table 3.2. Appropriate first searches (Cs) performed by each child in the Hiding Task and Finding Task.

| Subj. | Present condition | | Absent condition | | Combined | |
|---|---|---|---|---|---|---|
| | N | Cs | N | Cs | N | Cs |
| a) Hiding Task | | | | | | |
| Gi | 8 | 6 | 8 | 5 | 16 | 11 |
| Ko | 8 | 8** | 8 | 2 | 16 | 10 |
| Kr | 8 | 8** | 8 | 8** | 16 | 16** |
| Za | 10 | 10** | 10 | 6 | 20 | 16* |
| Jo | 8 | 6 | 8 | 5 | 16 | 11 |
| Br | 8 | 7* | 8 | 5 | 16 | 12 |
| Ka | 10 | 8 | 10 | 8 | 20 | 16* |
| Em | 8 | 7* | 8 | 3 | 16 | 10 |
| An | 8 | 8** | 8 | 2 | 16 | 10 |
| Tot. | 76 | 68** | 76 | 44 | 152 | 112** |
| b) Finding Task | | | | | | |
| Gi | 8 | 4 | 8 | 6 | 16 | 10 |
| Ko | 8 | 6 | 8 | 5 | 16 | 11 |
| Kr | 8 | 3 | 8 | 3 | 16 | 6 |
| Za | 8 | 2 | 8 | 8** | 16 | 10 |
| Jo | 8 | 6 | 8 | 7* | 16 | 13* |
| Br | 8 | 5 | 8 | 4 | 16 | 9 |
| Ka | 8 | 5 | 8 | 7* | 16 | 12 |
| Em | 8 | 8** | 8 | 8** | 16 | 16** |
| An | - | - | - | - | - | - |
| Tot. | 64 | 39 | 64 | 48** | 128 | 87* |

* = p < .05
** = p < .01

Identification of strategic behaviours

Up to this point, we have evaluated how successful our subjects were in selecting the two appropriate locations on the basis of the inferences that could be drawn from the observed sequences of displacements. However, it is important to take as much of the behaviour into account as possible. Some strategies (which do not lead directly to "significant" correct performance) may be at work for example in the residual error space. This would indicate that searches, even when not appropriate, are not performed at random. In order to evaluate this possibility, we subjected the choice data to a further analysis based on a taxonomy of other strategic possibilities proposed by Somerville and Capuani-Shumaker (1984) and described as follows.

For each subject, a significant difference ($p < .05$) from the value expected if first choices oscillated randomly between two possible pairs of locations (Binomial Test, two tailed) was considered evidence for conformity to a strategy.

One strategy would lead to a bias towards the selection of either the first two or the last two locations visited by the tester's hand and can be designated as temporal. Two children conformed to a temporal strategy in the Hiding Task and one in the Finding Task, selecting consistently the last two locations.

Another strategy could be a simple position bias towards the two left hand side (LHS) or the two right hand side (RHS) locations. In the Hiding Task, one subject conformed to this spatial strategy selecting consistently the right hand side locations, while in the Finding Task such a strategy was used by two subjects: one selected the right hand side locations and the other one chose the left hand side locations. Overall, only two subjects did not appear to use any strategy at all in the Hiding Task and three in the Finding Task.

Strategic consistency. Here we evaluated the relationships between the strategic behaviour of each subject on each of the Tasks. There was little evidence that strategies deployed in one task were used by the same subject in the other. For example, two of the three subjects that showed evidence for a "logical" strategy in the Hiding Task did not conform to any strategy in the Finding Task and the third, in the Finding Task, always chose the last cup visited by the tester. The two children that selected the appropriate locations in the Finding Task, turned out, respectively, to be selecting always the same location or not using any strategy at all, when presented with the Hiding Task. Ko searched according to a temporal strategy (Last 2) in the Hiding Task and according to a spatial strategy in the Finding Task (RHS); Br did not use any strategy in the Hiding Task but conformed to a spatial strategy in the Finding Task (LHS); Gi did not conform to any strategy in both the tasks; and An, who was tested in the Hiding Task only, searched according to a spatial strategy (Last 2).

**Discussion**

Children in this study are not found fully competent to use observationally based constraints on choice when both Hiding and Finding Tasks are taken as criterial. No child succeeded in both tasks. Instead, success was partial, emerging primarily in the Absent Condition of the Hiding Task, and the Present Condition of the Finding Task. A similar trend has been found by Somerville and Capuani-Shumaker (1984). This result, combined with our evidence that the behaviour, even when unsuccessful, was essentially non-random indicates that the children's failure was not merely the result of boredom, or of inappropriate testing procedures.

Overall the picture which emerges from the child data appears a heterogeneous one. Some children seem to perform observationally constrained searches in one or the other of the two tasks. However, no one subject conforms fully to the criteria set to determine "logical" search in both tasks and for both conditions of each task. There are, nevertheless, many consistent features of performance which suggest that data are not idiosyncratically generated by each subject, nor the product of poor test conditions or procedures. Instead, in my opinion, this suggests that children of the age we tested are only partially competent at dealing with some of the implications of events which they perceive directly, even in a situation as (ostensibly) simple as the one we describe. There is evidence of weakness in the control of the negative search space, motivated by the experience of absence of object or event. In addition, the requirement to couple simple background knowledge (conveyed linguistically) with the interpretation of directly perceived events is far from optimal.

These lacunae aside, however, there is also evidence of a gradient of constraint on search and object choice which these subjects may exploit on the basis of the observations of the behaviour of a third party. Would this also be true of the behaviour of a non-human primate, the Cebus apella? Informed by a behavioural base for 5 year old humans, this question was addressed in experiment 2.

## 3.3 Experiment 2

The experiment comprised of two different phases, phase A and phase B.

A precondition for the administration of the Hiding and Finding Tasks which feature complex sequences of displacements is that the subject will search at all, under conditions where the size of the set to be searched is four items and sometimes under delays of at least 3sec.

Phase A was essentially an attempt to give the monkeys experience of searching under these conditions.

Phase B featured the presentation of the Hiding and the Finding Task to the monkeys.

### PHASE A

#### Method

#### Subjects

The subjects were two adult males (Al and Ch) and three adult females (Lu, Ki and Ol) wild born tufted capuchin monkeys (Cebus apella). They were housed in a colony compound within the

Laboratory for Cognitive Neuroscience of the University of Edinburgh. At the time of the experiments the colony was composed of two adult males and five adult females. The enclosure was equipped with perches, water sources, tree branches and hangers suspended from the ceiling in order to provide locomotor opportunities. A layer of wood shavings covered the floor. To encourage foraging behaviour, a mixture of seeds was dispersed into the wood shavings on daily basis. Water was available ad libitum. The monkeys were transferred in individual cages for the testing sessions that took place in the morning. Reward was based on highly preferred food (grapes). All the monkeys were experimentally naive and had just terminated a period of quarantine of 6 months.

Apparatus

The testing apparatus was a modified version of WGTA, similar to that used for the children in experiment 1. The tray was especially wide, designed to take 5 discriminanda simultaneously (McGonigle and Chalmers, 1986, 1992). The occluders were the same sort of white cups used for the children in experiment 1. The bait was a white grape that could be enclosed in the tester's hand and placed silently under the cups without giving the subject any auditory or visual clue.

Stimuli

Up to 5 polystyrene white cups were used as occluders.

Design

Essentially the design was motivated to enable the monkeys to cope with up to 5 occluders per trial and a delay interposed between hiding and retrieval. To achieve this, 3 conditions were presented to the monkeys in the following order: a visible baiting condition, comprising 5 phases featuring the presentation of 1, 2, 3, 4 and 5 occluders, respectively; a control condition; and a delay condition.

Procedure

The following conditions pertained:

1) Visible baiting condition. A cup was baited while the subject was looking. Starting from trials in which only one cup was presented on the tray, the number of cups was increased until a linear array of five cups was presented. For each trial, the cup to be baited was randomly chosen within the array. When the subject reached a criterion of five consecutive correct responses, each performed within a latency of five seconds, one more cup was added to the previous array.

On reaching this criterion for the five cups array, subjects were overtrained for several sessions to ensure a stable performance before the administration of the next stage.

2) Control condition. An array of 5 cups was presented. The procedure consisted in moving a second cup simultaneously with the displacement of the bait. Thus, the mere movement of a cup could not be taken as a sign indicating which cup was being baited. The second cup to be moved was randomly selected for each trial.

3) Delay condition. The task featured the presentation of a 5 cups array. First, a 3 sec. delay was introduced between the displacement and the retrieval of the bait, subsequently increased to 5 sec. for those subjects which did not show a noticeable performance decrement.

Data recording

The data recording was exactly as described for Experiment 1.

Statistical analysis

A Binomial Test was performed on data obtained from the extensive testing with the 5 cups array, the Visible Baiting Condition, the Control Condition and the Delay Condition. The probability of occurrence of a successful search by chance was $p = .20$.

Results

1). Visible baiting condition. All subjects immediately searched for the bait when it was hidden under the only cup presented. In the phases featuring the presentation of 2, 3, 4 and 5 occluders, the averaged numbers of trials to criterion were 24.6 (sd = 20.9), 8.6 (sd = 4.1), 8.8 (sd = 4.1), 6.8 (sd = 2.2), respectively. The highest number of trials to criterion was found (for four subjects out of five) in the condition were two cups were presented. For one subject (Lu) the highest number of trials to criterion was found when it was presented with 4 cups.

The averaged percentage of correct choices in the extensive testing with the five cups array was 92.8% (sd = 4.4, $p < .01$). The overall percentage of correct choices made by each subject was also highly significant (Al = 97%, $p < .01$; Ch = 91% $p < .01$; KI = 94% $p < .01$; Lu = 96% $p < .01$; Ol = 86% $p < .01$).

2. Control condition. The mean percentage of correct choice in the control condition was 90% (sd = 9.5), $p < .01$). As for the visible baiting condition with five cups, all the subjects showed a highly significant percentage of correct responses. Only one subject (Ki) showed, in the control condition, a percentage of correct responses lower than that shown in the visible baiting condition with five cups.

3. Delay condition. The interposition of a 3 sec. delay between hiding and retrieval did not disrupt the performance of 4 subjects out of 5. The percentages of correct choices made by Ol (90%), Ch (94%), Lu (80%) and Al (92%) were all highly significant, while that performed by Ki (32%) did not reach statistical significance. When the delay was increased from 3 to 5 sec. most subjects became frustrated and distressed. We were thus obliged to terminate the administration of this condition. However, the two subjects (Ol and Ch) that received enough trials to compare their percentages of correct responses (Ol = 68% and Ch = 73%) to chance, searched correctly ($p < .01$) even when a 5 sec. delay was interposed between hiding and retrieval.

### Discussion

Visible baiting condition. Results obtained from this first set of tests show that all subjects were committed to search for an object they had seen disappear under a cup. The incentive to search for an object, now out of sight, was apparent even from the first condition in which only one cup was presented. However, in a single cup condition, the action of lifting it might be expression only of a manipulative disposition on the part of the subject, and not necessarily motivated by the bait per se. The fact that the subjects chose the baited cup only, under multiple cup conditions, indicates the contrary. Nevertheless, this selective response in the presence of multiple occluders was acquired in the course of testing and was not expressed spontaneously. It would seem, therefore, that there is already a disposition to use a self-directed mode of search. When one cup is presented, this mode is sufficient for immediate success; where there are alternatives, however, it is not. Informed by failure, however, the monkey's search, once constrained through observation of the tester, appears unaffected by the addition of further distractors (up to five cups presented in a linear array).

Control condition. Results obtained from the control condition show that subjects were not using cup movement alone as a unique clue to location of reward. Instead, the whole displacement procedure was taken as the informing event.

Delay condition. An evaluation of the delay that subjects were able to tolerate was necessary before the Phase B of Experiment 2; that required an attentional phase of approximatively 2-3 sec., if the tasks were to be administered successfully. As four of the subjects proved able to cope with an interval of at least 3 sec, a necessary precondition for Phase B was satisfied.

PHASE B

Method

Subjects

The four monkeys (Al, Ch, Lu, and Ol) that proved able to tolerate at least a 3 sec. delay in the preceding stage.

Apparatus

The testing apparatus was the WGTA used for Phase A. Four white plastic cups, identical to those used in Experiment 1 and in Phase A of experiment 2, served as occluders. The baits were black and white grapes.

Design

The experimental design was as described for children in Experiment 1. Two monkeys (Ch and Lu) were first presented with the Hiding Task and the remaining two (Al and Ol) began with the Finding Task.

Procedure

The procedure adopted with the monkeys in the Hiding and the Finding Task followed the same schema featured in Figure 3.1, with the following minor modifications:

1) HIDING TASK

At the beginning of each daily session subjects were motivated with five trials in which the bait was hidden under a cup and the subject had to retrieve it. This was followed by 4 experimental trials.

2) FINDING TASK

Before the administration of the Finding Task, it was necessary to convey, non linguistically to the subjects the crucial information that two baits were always hidden together under the same cup. This was attempted by administering a task identical to the visible baiting condition described above, except for the fact that two baits (a black and a white grape) were hidden together under one of the cups. This was followed by 4 experimental trials per session using the procedures as described for children in Experiment 1.

Warm up

The warm up trials presented to the monkeys, both for the Hiding and the Finding Task, were identical to those presented to the children and described for Experiment 1.

Statistical analysis

The statistical data analysis followed the schema described for Experiment 1.

Results

Warm-up trials

Type 1 for the Hiding Task. The group performance averaged 84% correct responses (p <.01). Three of the monkeys (Ch, Lu and Ol) were individually correct above chance level (p < .01, Binomial test, one tailed)) indicating that the hiding procedure was understood. The fourth subject (Al), that was presented first with the Finding Task, became so stressed during the presentation of the Finding Task that the experiment had to be terminated before the presentation of the warm up trials for the Hiding Task.

Type 1 for the Finding Task. The monkeys were given on average 25 visible baiting trials with two baits (range = 21-30) before the administration of two Warm-up 1 trials. The averaged percentage of correct searches was 95.2% (range = 90%-100%). The percentage of correct searches of all the subjects was above chance level (p < .01, Binomial test with a chance probability of occurrence of a correct search = .25). The combined percentage of correct searches for the two monkeys (Al and Ol) presented with the Warm-up 1 trials was 73% (p < .01) and also significantly above chance level (p < .01) considered as individuals. The other two monkeys stopped searching after a few failures and the administration of this task was terminated.

Type 2. As for the child sample, preliminary data analyses showed no major differences between the results obtained from the warm-up 2 trials for the Hiding Task and the Warm-up 2 trials for the Finding Task. Therefore, the results from the two tasks have been combined.

As with children, monkeys were at chance when locating the object, on their first attempt (22% of the trials). They adopted a Systematic mode of search more often than an Asystematic one, although the percentage of occasions in which the subjects were Asystematic is considerable. In fact, 36 searches (50% of the total number of searches) were performed in a Systematic way, while 20 (28%) were Asystematic.

HIDING AND FINDING TASK

Group Performance

<u>First searches</u>. Table 3.3 shows the percentages of appropriate first searches performed by the monkeys the Hiding (section (a) of the Table) and the Finding Task (section (b) of the Table). From Table 3.3, it can be seen that the percentage of appropriate first searches was above chance level in the Hiding Task but was not significant in the Finding Task. This taken together with a selective effect of sub-conditions within each task type, analogous to that found for the children, leads to a major difference between tasks.

<u>Second searches</u>. Table 3.3 shows also the number of correct second searches performed after an appropriate but unsuccessful first search.

In the Hiding Task the monkeys identified the correct location after a first appropriate choice. Correct second searches were above chance level either when following a first appropriate choice at an end point or at an inner point of the array.

In the Finding Task, the combined percentage of correct second searches for the two conditions is again above chance level; when analyzed according to first search location, second correct searches following an end point first choice are highly significant; those that followed a first choice to an inner location, whilst correct in all cases (3) are too few to yield to statistical test.

Table 3.3 Distribution of monkey's searches in the Hiding and the Finding Tasks (see text for explanation.

a) HIDING TASK

| | | | | |
|---|---|---|---|---|
| Appropriate first searches | 87%** | | OP Condition | 93%** |
| | | | OA Condition | 81%** |
| First search outcome | Object found 54% | Object not found 46% | | |
| Correct second searches | - | 94%** | Following Inner | 88%** |
| | | | Following End | 100%** |

b) FINDING TASK

| | | | | |
|---|---|---|---|---|
| Appropriate first searches | 53% | | OP Condition | 41% |
| | | | OA Condition | 65% |
| First search outcome | Object found 73% | Object not found 27% | | |
| Correct second searches | - | 100%** | Following Inner | 100% |
| | | | Following End | 100%** |

* = $p < .05$
** = $p < .01$

**Individual Performances.**

As for children, given the asymmetrical distribution of first choices across tasks and conditions, the data were analyzed on an individual basis. The results are presented in Table 3.4.

Table 3.4 Appropriate first searches (Cs) performed by each monkey in the Hiding Task and Finding Task.

| Subj. | Present condition | | Absent condition | | Combined | |
|---|---|---|---|---|---|---|
| | N | Cs | N | Cs | N | Cs |
| a) Hiding Task | | | | | | |
| Ch | 20 | 18** | 22 | 21** | 42 | 39** |
| Lu | 17 | 16** | 14 | 13* | 31 | 29** |
| Ol | 19 | 18** | 21 | 12** | 40 | 30** |
| Tot. | 56 | 52** | 57 | 46 | 113 | 98** |
| b) Finding Task | | | | | | |
| Al | 7 | 2 | 7 | 4 | 14 | 6 |
| Ol | 20 | 9 | 22 | 15 | 16 | 11 |
| Tot. | 27 | 11 | 29 | 19 | 56 | 30 |

* = p < .05
** = p < .01

In the Hiding Task, three subjects searched appropriately above chance level in the Object-Present Condition and two in the Object-Absent Condition. Neither of the subjects presented with the Finding Task performed above chance in either the Object-Present nor the Object-Absent Condition. Combining the results obtained from the two conditions of each of the tasks secures the conclusion that all three monkeys tested in the Hiding Task correctly performed first searches above chance level; in contrast, the two that were presented with the Finding Task both failed it. The main differences are, therefore, between tasks, not conditions. What seems clear is that the finding procedure is itself difficult to understand. In fact, the subject that was presented with the Finding Task first never solved it, and became so frustrated and distressed that we were forced to discontinue testing. Moreover, the two subjects that proved successful on the Hiding Task, failed the warm-up trials of the Finding Task; after a few unsuccessful attempts they eventually ceased search for the hidden bait.

Identification of strategic behaviours.

We have already seen that monkeys behaviour was "logical"[1] while performing the Hiding Task. However, when presented with the Finding Task they used a spatial strategy. In particular each subject in the Finding Task always chose the same end location on the same side (one subject chose left, the other right).

Discussion

Monkeys in this experiment appear able to use the behaviour of a third party in the Hiding but not the Finding Task. This indicates that even in this small task domain, the cost functions are appropriate to the induction of an observationally based strategy. Special, desirable food may well be a strong factor, energising the subject in circumstances where children may need strong social facilitation to maintain attention and devise more complex solutions perhaps than the task may otherwise warrant.

However, where the tester acts as finder, the monkeys fail. While this may indeed be a direct result of a failure to appreciate a "finding" role, as described by Fischer & Jennings (1981) and Berthental & Fischer (1983), the fact that the performance of children in this study was also relatively poor (as indeed was the performance of subjects in the Somerville and Capuani-Schumaker study) indicates that the task is difficult to comprehend. One index of this is given by the number of verbal prompts required in the Somerville and Capuani-Shumaker study (1984). Task communication apart, however, the subject must link some background knowledge with the perception of object displacement i.e. it is crucial that the

1 For an extended analysis and discussion of inferential behaviour in monkeys see McGonigle & Chalmers, 1992.

subject interprets the object event at the end of the finding sequence in the light of the background knowledge that the objects are always together. The most likely reason for the failures in the Finding Task, therefore, would thus appear to be based on the extra demand it makes on the subject.

## 3.4 General discussion

Apart from the pratical aim to familiarize the subjects with the testing environemnt, the tasks reported in the experiments described in this chapter have been designed to evaluate the strategies subjects may use when searching for unseen objects. Ostensibly they are about the ways in which the agent constrains search both on the basis of information received and (as in the Finding Task) on the basis of prior information, needed to interpret the events under the subject's interrogation. The results vary both across and within species. In the former case, the variation seems to have something to do with the type of task used, the effort of searching and the cost functions attaching to that search (see Wellman et al., 1979). In addition, there may well be social factors at work. The (social) costs of mistake in situations where the adult tested has carefully coached the child in the rubric of the test may well contribute to the performance recorded by Somerville and Capuani-Schumaker (1984). Whilst these are unlikely to apply in the case of the monkey, failure to retrieve highly preferred food, albeit in situations where search is otherwise un-costly may dispose the Cebus in these experiments to pay particularly good attention to the behaviour of the tester.

A further social factor is the sort of role assigned the tester, as Hider or as Finder, as collaborator or as deceiver. These latter factors may all play a part, particularly in encounters with conspecifics. For these reasons, it may be best to specialise and develop experimental paradigms which target a cohort of closely related issues. For example, the social aspects of the encounters are already represented in experiments on social inference and imitation. In non-human primates, recent studies Visalberghi & Fragaszy (1990) and Povinelli et al. (1991) are examples of controlled assessments of the use of socially derived information by non-human subjects.

A complementary but separable line of inquiry is one concerned with self-directed search under conditions which do not presume socially based observational competences as a precondition for its operation. Here the tasks used in the study described are flawed, when considered from this perspective alone. One major reason for this is that each manual interrogation of an object displaces it from its test position, thus leaving a visible trace of a visit. Under these circumstances, it would be a very foolish subject who attempted to reiterate visits to previously interrogated locations. Yet a measure of reiteration is essential if we are properly to evaluate the extent to which subjects can keep track of choices made serially over time.

For these reasons, as outlined in the previous chapter, for the rest of the research program, serial-order search tasks for human and non-human primates were developed using touch screen based paradigms and procedures designed to evaluate size seriation skills (McGonigle, 1987b; McGonigle, 1989; Chalmers and McGonigle, 1993). These, intended primarily to evaluate self directed search strategies without reference to a third party, will be described in the following chapters. The aim of the following experiments was to assay whether a non-human primate, Cebus apella could devise its own strategies in an exhaustive search task, becoming more economical by paying attention to the spatial organization of items in the search space (McGonigle et al., 1992).

As a consequence of the experiments described in chapters IV-VI used a new paradigm (McGonigle, 1987b) based on computer-interactive touch screen technology which enables the experimenter to display a wide range of items through which the subject must search. However, unlike our present procedures, each touch leaves no lasting trace of a touch, leaves the subject to discover its own best (most efficient) route through the search space, and is sufficiently motivating to keep the subject working for protracted periods, thus enabling a comprehensive in-depth analysis of each case.

In short, once the object of familiarizing the subjects with the tester and the testing environment had been achieved, it was believed that restricting the research to issues concerning self regulatory factors in cognition and leaving social regulatory ones to other studies, offered the best prospect for the study of information organisation and management by primates.

# CHAPTER IV

## TOUCH-SCREEN BASED SEARCH TASKS

### 4.1 Introduction

Now that the monkeys to be used as experimental subject had been familiarized with the testing environment and the tester, it was possible to proceed along the experimental project as announced and outlined at the end of chapter II.

In this chapter two experiments will be reported. They are the first implementation of the new technologies mentioned already. Following the experimental procedures adopted by McGonigle and Chalmers with children and monkeys using touch sensitive computer monitors (see Chapter II for a detailed review), the new apparatus was employed in an attempt to overcome the limitations of WGTA based experiments.[1]

---

1 For example, the WGTA used for the experiments described in the previous chapter did not allow protracted periods of testing, the social nature of the cues provided to the subjects prevented an unambiguous interpretation of the difficulties which they encountered in solving the task, and each move left a permanent trace of the choice, preventing an evaluation of the ability of the subjects to keep track of their moves throughout the search space. These were major difficulties for a research program aimed at the investigation of issues such as economy and data reduction in serially organized behaviours. First of all, when subjects do not tolerate protracted testing, it is impossible to observe those strategy shifts, which, as we have seen in Chapter II, might accompany the gradual acquisition of task expertise (McGonigle and Chalmers, 1992). Secondly, the fact that the displacement of an object leaves a permanent mark of each choice made, deprives the subject of the incentive to behave strategically provided by the need to keep track of long series of action. It can be expected that the larger the space to be searched, the more the system is at risk of overloading if strategic data-reducing factors are not deployed. Obviously, no incentive to use data-reduction strategies is present when the subjects leave involuntary external clues of each of the items to be remembered (i.e. the locations

The new paradigms

As already outlined at the end of Chapter II, the general paradigm is a search task where subjects are required to respond to each icon of a set, presented on a touch sensitive computer monitor. The subjects are provided with a variable number of icons to search. So long as search is exhaustive, the task requirements are achieved, a termination signal is given, and a peanut is dispensed as reward.

The paradigm allows the manipulation of the length of the production as an independent variable to assess its effects as an incentive for organization. In fact, since the requirement of the task is an exhaustive search, the increase of the size of the search space, would result in a longer series of responses being performed. Moreover (in case the subjects do not spontaneously perform economic searches), the procedure and the apparatus were designed to provide the subjects with explicit feedback (penalization of surplus redundant moves by omission of reward).

The tactic followed for the implementation of the set of experiments consisted in starting with versions of this task where the search of the subjects is completely non-supervised. In fact only when the role of training is made minimal in the first stages of the study, does it become possible to understand the potentiality of an autonomous cognitive system.

Thus, here subjects are required to perform an exhaustive search of the set but are left free to select the items in any order and are allowed to reiterate on items already selected. In this chapter, experiments based on a spatial version of the task will be reported, i.e. the subject

---

already explored). Thus in such a circumstance the interplay between length of the production and strategic control over serial behaviour becomes impossible to assay.

is presented with a set of identical icons and each choice can be identified only on the basis of the spatial location of the icon selected. The following chapters will focus on experiments featuring search spaces with richer structural properties.

The general procedure summarized above, was designed according to the criteria which, as stressed in Chapter II, tasks aimed to investigate the issues of economy and data reduction must satisfy, namely: 1) the subject must be allowed a high degree of **spontaneity** in the selection of his own path toward solution; 2) there must be a **variety of solutions** within the context of behavioural mastery to allow the subject to find the most effective one, and; 3) the task **must be repeated** sufficiently to enable the subject to express genuine strategic shifts as a function of practice.

With the new paradigms featured here, spontaneity is granted by the absence of selective feedback on the best trajectories through the search space. As long as the search is exhaustive, the trajectories selected by the subjects represent different solutions to the task. Finally, given the minimal requirements of the task, it was reasonable to predict a high tolerance to task exposure by monkey subjects.

Moreover, in such a context, it is possible to identify unambiguously the different factors postulated by an economy/data reduction hypothesis. These are the following. First of all, since success is defined unambiguously as the ability to perform exhaustive searches, an objective measure of fitness to the task is granted. Then, within the success space, there is a range of more or less efficient searches, again suitable to be measured objectively. In fact (in a task where a set of locations has to be visited exhaustively and only one visit per location is required), a behaviour best fitted to the task would produce only one visit per each location. On the other hand, a less fitted behaviour would produce some reiterations on locations already visited. Thus, the ratio between the number of icons presented to the subjects and the number of moves performed to exhaust the set provide an index of the relative efficiency of behaviour within the success space. The costs attached to unnecessary "surplus" moves, in relation to the task requirement are extra-cognitive in their nature. The task implicitly provides the currency in terms of time (and/or energy) spent in searching before obtaining a reward and, as a consequence, feedback information to evaluate the relative efficiency of successive searches.

A second type of costs is related to the amount of cognitive resources which are employed in order to maintain a particular level of fitness to the task. These sort of costs are intra-cognitive and might vary according to different ways of solving the task.

It can be conjectured that a system striving to achieve a high level of behaviour fitness at a low cognitive cost will start to self-regulate. In other words, given the opportunity to practice with a task, it will try to select those constraints afforded by a particular problem space which allow the subject to satisfy the task requirements with an optimal management of internal resources.

With tasks of this sort, it becomes possible to assess, whether subjects spontaneously deploy strategic factors, and, if they do so, whether they shift to more economic strategies in the course of practice. The observation of (if any) the sort strategies used by the subject and their change over time, can then become a window on the dynamic of the management of intra-cognitive resources.

Empirical templates of "fit" and "unfit" behaviours, in a spatial version of the paradigm, similar to that used for the experiments reported here, was provided by a parallel study conducted on children (McGonigle, De Lillo and Dickinson, 1992; Dickinson, in preparation). While 4 and half year olds, showed to be able to search sets up to nine items, showing maximal efficiency from the outset, the behaviour of 3 year old children was characterised by the presence of a high percentage of redundant moves.

Moreover, the searches of the older children were based on a number of spatial strategies which included the use of fixed starting points (in a particular configuration, always start touching the same icon), an adjacency principle (when possible, touch the item spatially next to the one just touched), and the use of preferred directional vectors (as your search path, always follow a vertical trajectory). In contrast with the older subjects, younger children deployed only some of these strategic factors and never combined them all to produce a very principled trajectory through the search space.

Thus, children provide an illustration in principle of the relationship between principled searches and success. The strategic use of a principled trajectory through the search space, which at any point of the search tells the subjects the portion of the search space which has already been explored and which icons remain to be touched, can be considered an effective device to reduce the demands which would have been put on the memory system by storing all the icons already touched.

However, children (according to their age group) show either immediate success or failure. Young children who are not economic at first have poor tolerance of task repetition in this context. Nevertheless, it would be of great interest to assess whether increase in economy of search can be obtained in the course of protracted task practice. It is here that the monkey is an ideal experimental subject, since its tolerance to prolonged task exposure is well documented (Harlow, 1949; McGonigle and Chalmers, 1977; 1993; Shwartz et. al; 1991).

The experiments which will be soon reported were, thus, an attempt to evaluate if these criteria of rationality could have been applied to capuchin monkeys. Now for a longer assay of strategic factors in mature subjects capable of task repetition tolerance.

## 4.2 Experiment 1

The focus of this experiment is on spontaneity. It aimed to evaluate whether monkeys would have spontaneously searched exhaustively a set of icons presented on a computer monitor; at the assessment of whether they would have deployed strategic factors in doing so; and at the characterization of the relationship between the size of the search space and the economy of search.

### Materials and Methods

#### Subjects

The subjects were 6 (2 males, Al and Ch; and 4 females, Ki Lu, Mi and Ol) adult, wild born tufted capuchin monkeys (Cebus apella).[2] None of the subjects had previous experience with touch screen based procedures. The monkeys were housed in a colony compound within the Laboratory for Cognitive Neuroscience of the University of Edinburgh.

#### Apparatus

The apparatus consisted of a IBM clone computer equipped with a Microvitec Touchtec 501 frame. This touch sensitive apparatus is based on a infra-red scanning technology which

---

2 One of the subjects used in this study (Mi) has not been mentioned in the previous chapter. This is because, at that time, Mi was pregnant and successively lactating. Therefore it was impossible for her to be tested on regular basis and to undertake all the different phases of the WGTA based study. However, when available she was extensively exposed to WGTA procedures and therefore undertook, as well as the other subjects the necessary familiarisation phase.

enables it to detect a finger or object placed within the plane of the monitor screen and communicate the coordinates of this position to the computer. The monitor could hold up to 9 stimuli simultaneously in a 3 x 3 (3 rows and 3 columns of icons) matrix of possible positions. The monitor was fitted with a transparent plexiglass screen with hand-holes corresponding to each of the location in which a stimulus could appear on the screen. A touch of any of the stimuli displayed resulted in a blink of the corresponding stimulus, a 0.5-seconds flash around the stimulus, and a simultaneous beep.

Figure 4.1 depicts the apparatus and the experimental set-up used for this study.


Fig. 4.1 The experimental set up used for this and the following experiments. a) The indoor cage where the colony of capuchins is housed. b) The monkey spontaneously enters the individual testing cage. c) The testing station provided with touch sensitive computer monitor and video camera. d) The cage with the monkey is hooked in front of the monitor (on the side of the cage, the food dispenser can be observed).

The system's software recorded the location of the stimuli touched as well as the length of the time interval between two consecutive touches and allowed to automatically dispense peanuts as rewards.

As shown in Fig. 4.1, a camera was positioned over the experimental apparatus to allow an accurate recording of the actions performed by the subjects during the testing sessions.

Design

The experiment was divided into a pre-training phase, where subjects were familiarized with the apparatus, and 3 testing phases.

Each of the 3 testing phases was designed and implemented on the basis of the results obtained from the previous one. For this reason what is described here is an a posteriori schema of the experiment as it was carried out.

For the same reason, the features of the method shared by all the testing phases will be described first in a section called General Procedure. The specific rationale and procedure of each phase will then be described in details, only after the presentation and the discussion of the results obtained from the previous phase.

Phase I consisted of 4 conditions each featuring the presentation of a different number of icons. In particular, conditions featuring the presentation of configurations composed of 2, 3, 4, and 5 items were presented. The conditions were presented successively, i.e. after the completion the 2 stimuli condition, the 3 stimuli condition was presented and so on.

Phase II was carried out in the course of 10 experimental sessions. In each of sessions 1-5, subjects were presented with 5 different conditions, administered in random order. The 5 conditions featured the presentation of 4, 5, 6, 7, and 8 icons, respectively. As for sessions 1-5, each of sessions 6-10 featured the presentation of 5 conditions presented in random order. This time the 5 conditions involved the presentation of 5, 6, 7, 8, and 9 icons, respectively.

Phase III featured a protracted presentation of sets of 9 icons, until a stable state of performance was observed.

General Procedure

The testing environment comprised two working stations, each provided with a computer, a touch screen and a video camera. The subjects were tested in pairs, each on one of the two working stations. Before each daily testing session, the subjects were transferred to individual cages. The cages were then hooked in front of the monitors.

Each session consisted of 50 trials. In each trial a set of stimuli, all identical in shape (squares), size (3cm x 3cm), and colour (green), was displayed on the screen. For each trial, the stimuli were randomly distributed among 9 locations defined by a 3 x 3 matrix (3 rows and 3 columns).

The subject was required to touch all the stimuli presented on the monitor in order to clear the screen, produce a series of tones, and obtain a peanut as reward. A new trial was presented after an interval of 15 sec. Both repetitions (i.e. temporally contiguous touches of the same stimulus) and reiterations (redundant moves to stimuli previously touched in the same trial), were allowed. The number and the spatial relationship of the stimuli varied according to the requirements of each of the conditions, as described below.

Pre-training

The subjects were faced with a monitor displaying one stimulus. For each trial, the location of the stimulus on the screen was selected at random.

A number of peanuts were dispensed at regular intervals in order to familiarize the subject with the use of the feeder and the noise produced by the dispenser. Then, a peanut was dispensed selectively when the subject approached the screen until, eventually, the stimulus was touched. At this point, a peanut was automatically dispensed whenever the subject touched the stimulus.

Phase I: do monkeys spontaneously perform exhaustive searches ?

This first phase was an exploratory one. In fact, given the novelty of the paradigm, the assessment of how the monkeys behaved when faced with the task was mandatory. Therefore, a primary aim of this phase was to check whether they would have spontaneously started to produce a series of responses, leading eventually to exhaustive search and reward.

The only information about the memory span of monkeys in a related task, comes from serial learning studies where the ability of macaques to acquire 6 term series has been reported (Swartz et al., 1991) and of capuchin monkeys to learn 5 term series (D'Amato and Colombo, 1988). The assessment of the length of the series of responses which capuchin monkeys would have been able to cope with in the new paradigm featured here was therefore necessary. For the present study, it was particularly relevant to find the number of items that would begin to challenge the memory system of the subject, since an important assumption of the hypothesis under scrutiny was that an incentive to the deployment of strategic devices is constituted by the length of the serial production itself.

Thus, the first aim of Phase I was to assess whether the subjects would spontaneously perform exhaustive searches within sets comprising from 2 to 5 identical stimuli and, if they did so, to evaluate their search behaviours.

This evaluation was based on a number of natural measures of efficiency that the paradigm afforded. First, having a well defined search space, it was possible to define unambiguously the minimal number of moves necessary to exhaust it. This number, obviously, corresponded to the number of icons itself. Every redundant move, as a reiteration on an item already touched, would increase the disparity between actual and optimal performance. Thus, the "fitness" of the behaviour of the subjects to the task, was first evaluated according to two different measures based on the number of responses: the percentage of minimal path searches (i.e. trials where no redundant moves were observed) and, the percentage of non-redundant moves in the course of successive trials. It can be observed that although the two measures are related (a subject registering 100% minimal paths, will necessarily register 100% non-redundant moves, and vice versa) they do not always converge. For example, in exploring 10 configurations of 5 items each, a subject could register an overall percentage of non-redundant touches equal to 83%, by making only one "surplus" move per trial, yet no minimal paths would be observed.

Apart from a criterion of "fitness" based on the number of response, the time spent by the subject in each search was recorded as another natural measure of fitness to the task. Again, measures based on the number of responses and measures based on search time are obviously related, but not always converging. In fact, a very fast subject can make more redundant touches than another and still register the same search time.

A further major aim of the study was to evaluate the effect of task repetition on search efficiency. Thus, in case the subject did not show a maximal efficiency of search from the outset, the task would have been presented repeatedly, in order to observe eventual changes in efficiency (in terms of a spontaneous reduction of redundant moves) over time, without giving the subject any differential feedback.

Finally, by means of an overall inspection of video recordings, it was aimed to gain further qualitative information about subjects' search behaviour.

Procedure

Sets of stimuli comprising from 2 to 5 monochromatic icons were displayed on the monitor. On each trial, the location of the stimuli varied at random.

Fig. 4.2 depicts some examples of possible configurations for each of the different conditions.

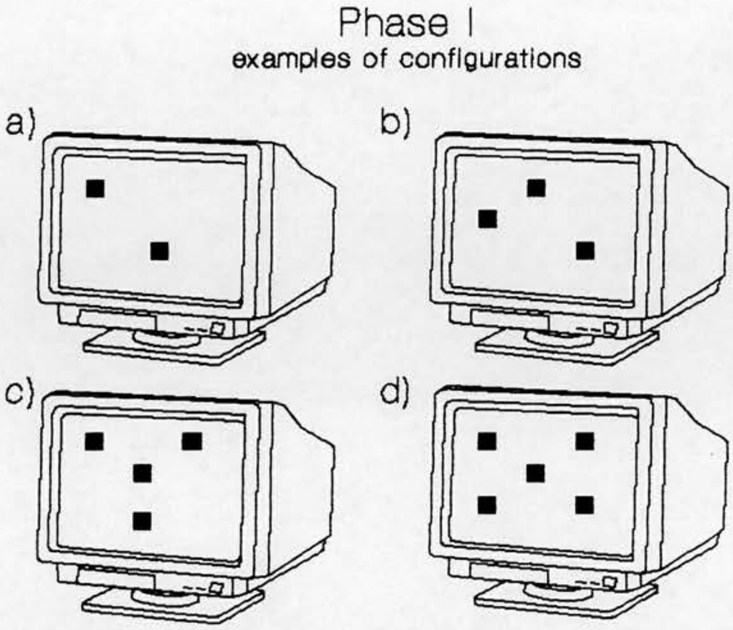



Fig 4.2. a) Condition 1, an example of configuration with 2 stimuli. b) Condition 2, an example of configuration with 3 stimuli. c) Condition 3. An example of configuration with 4 stimuli. d) Condition 4. An example of configuration with 5 stimuli.

Each subject was first presented with 2 stimuli, then the set size was gradually incremented up to five stimuli. Before a further stimulus was added to the set, a criterion of 75% non-redundant touches had to be achieved within an experimental session. The phase continued until sets of 5 icons were displayed. After the administration of the 2 stimuli condition (where the only possible redundant touches were repetitions), in evaluating the achievement of the criterion, repetitions were deleted from the analysis and only reiterations were considered as redundant moves. The removal of repetitions from the analysis appeared necessary for two different reasons. First, while reiterations can reasonably be accounted for by a memory failure in keeping track of the icons already explored, the status of the repetitions is more ambiguous. Being performed on the icon just touched, they can hardly be caused by a memory fault[3]. The second reason was that repetitions are ignored by authors who have used related tasks and a similar apparatus (see Terrace, 1987; 1991; Swartz et al., 1991; D'Amato and Colombo, 1988); and, thus, it seemed reasonable to adopt the same procedure in order to provide data comparable with those obtained from these studies.

After the achievement of the criterion, and before the set size was increased, the subjects were presented with the same condition, until they showed no further increment in the percentage of non-redundant touches, in at least 2 successive testing sessions.

## Results

### Comparison with chance

To compare the distributions of scores obtained in the different conditions with those expected by chance, the Kolmogorov-Smirnov one sample test was used (see Siegel and Castellan, 1988). This test is based on the comparison of a theoretical and an observed cumulative distribution.

The theoretical distribution was obtained from a mathematical model of random moves within an analogous search space. The model was developed by St Johnston (1993) and provided the probability of each search length, assuming a random walk across the search space. The mathematical details of the random model are provided in appendix B.

The observed percentages of trials in which the search was completed in a particular number of moves, and those

---

3 On the contrary, it can even be conjectured that they might be expression of the emphasis posed by the subject on his choice (like in situations where the subject is not sure that the machine has registered the touch).

predicted by the random model were respectively transformed in an observed and in an expected cumulative distribution. The distributions were built according to the following procedure.

The observed and the theoretical distribution were both constructed on the basis of 50 classes, defined by trial length (i.e. class 1 contained the number of trials where the set of icons was exhausted in the minimal number of moves; class 2 incorporated class 1 plus the number of trials which contained one redundant move;...class 50 incorporated class 49 plus the number of trials containing 50 or more redundant moves). Then the percentages of trials in each class were calculated. Finally, with the Kolmogorov-Smirnov test the probability of obtaining the observed sample of scores from a population distributed as the theoretical distribution was determined.

All the subjects spontaneously performed exhaustive searches and reached criterion (75% non-redundant touches in 2 successive sessions) in all the different conditions. The comparison of their performance with chance (for each of the single conditions), will be reported in the following paragraphs.

3 stimuli condition.

The distribution obtained in the 3 stimuli condition is depicted in Fig. 4.3.

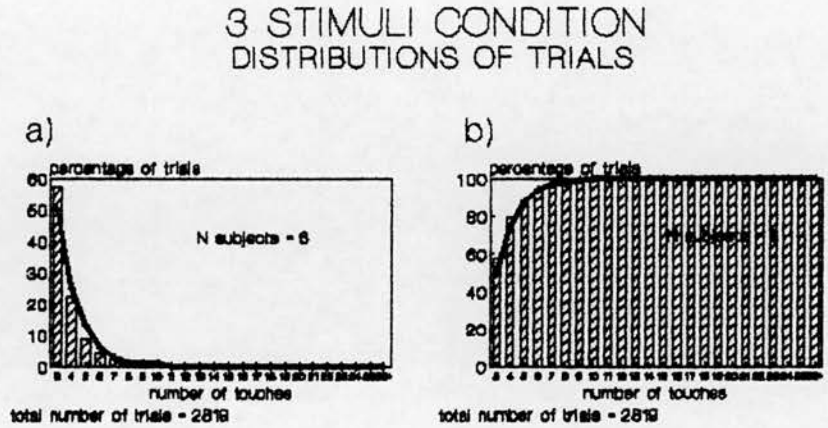



Fig. 4.3. 3 stimuli condition a) Observed (bars) and expected (line) percentage of trials of a particular length. b) The observed (bars) and expected (line) cumulative distribution of trials on which the Kolmogorov-Smirnov test was based.

In this condition, the distribution of scores of none of the subjects differed significantly from the expected. This result is not surprising, considering that with such a small number of icons to be explored, even a random trajectory throughout the set would lead to a high number of exhaustive searches and several even in the minimal number of moves. In fact, on the basis of chance, a minimal path search of a set of 3 icons is expected in the 50% of the trials (the probabilities being equal to 1 for the first and the second response, and 0.5 for the third. Thus $1 \times 1 \times 0.5 = 0.5$, would be the probability of a minimal path search).

4 stimuli condition.

Fig. 4.4 depicts the distribution observed in the 4 stimuli condition.

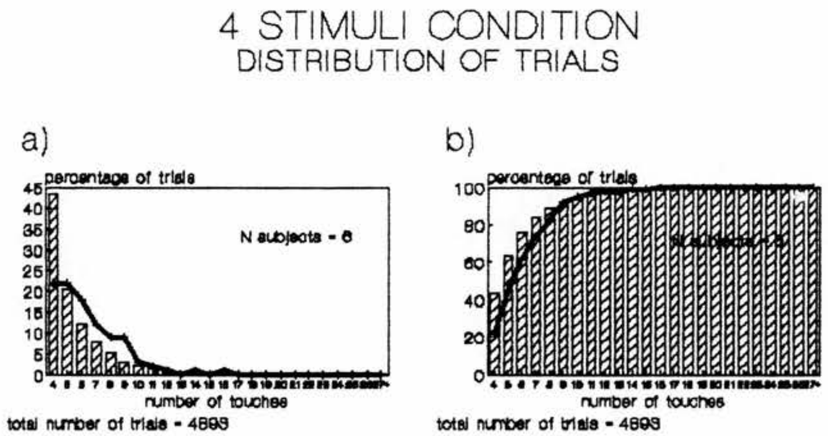



Fig 4.4. 4 stimuli condition a) Observed (bars) and expected (line) percentage of trials of a particular length. b) The observed (bars) and expected (line) cumulative distribution of trials on which the Kolmogorov-Smirnov test was based.

The distribution obtained combining the scores of the six subjects was significantly different from the expected (D = .2633 p <.005). Results obtained for each of the individuals showed that the distribution of scores of five of the six monkeys was different from the expected (Al, D = .2691, p <.005; Ch, D = .2987, p <.005; Ki, D = .3835, p <.005; Mi, D = .2524, p <.005; Ol, D = .1896, p <.05).

5 stimuli condition.

The distribution obtained in the 5 stimuli condition is shown in Fig. 4.5.

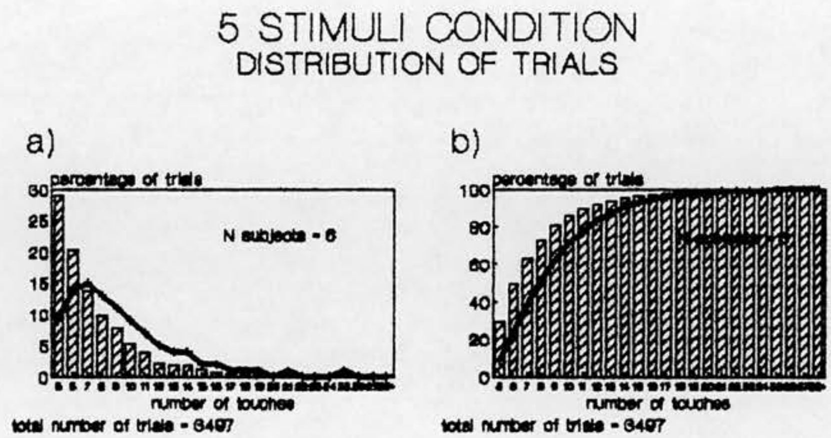



**Fig 4.5. 5 stimuli condition a) Observed (bars) and expected (line) percentage of trials of a particular length. b) The observed (bars) and expected (line) cumulative distribution of trials on which the Kolmogorov-Smirnov test was based.**

**The overall distribution obtained for the six subjects was significantly different from the expected (D = .3671 p <.005). Individual results showed that the distribution obtained from five of the monkeys was significantly different from the expected (Al, D = .3671, p <.005; Ch, D = .3176, p <.005; Ki, D = .4307, p <.005; Mi, D = .2121, p <.005; Ol, D = .2516, p <.005).**

**Overall, these results show that (with the only possible exception of the 3 stimuli condition), the subjects not only spontaneously performed exhaustive searches within sets containing up to 5 stimuli, but they did so in an economic way. Evidently, for each of the stimuli condition the costs associated with a random search provided enough incentive for the subjects to perform at a level significantly above chance. A detailed evaluation of how these constrained**

searches developed in the course of practice will be reported in the following section, where the trends observed in the level of performance of the subjects at different stages of testing (and the strategy that possibly sustained it) will be analyzed.

Evidence for improvement with practice.

This trend analysis aimed to evaluate the tendency of the subjects to progressively reduce the number of redundant moves while practising the task. The global set of data obtained for each subject in each of the different conditions was divided in vincent sixths. The percentage of non-redundant moves obtained for each of the sixths was considered representative of the economy of search at a particular time (the first sixth representing the initial phase of testing and the last sixth its end phase). A Page's trend test was performed for the six subjects. A positive trend was considered as evidence for improvement with practice.

Trends (Page's L trend test) were significant for all conditions at $p < .05$ (see Fig. 3.6-3.8).

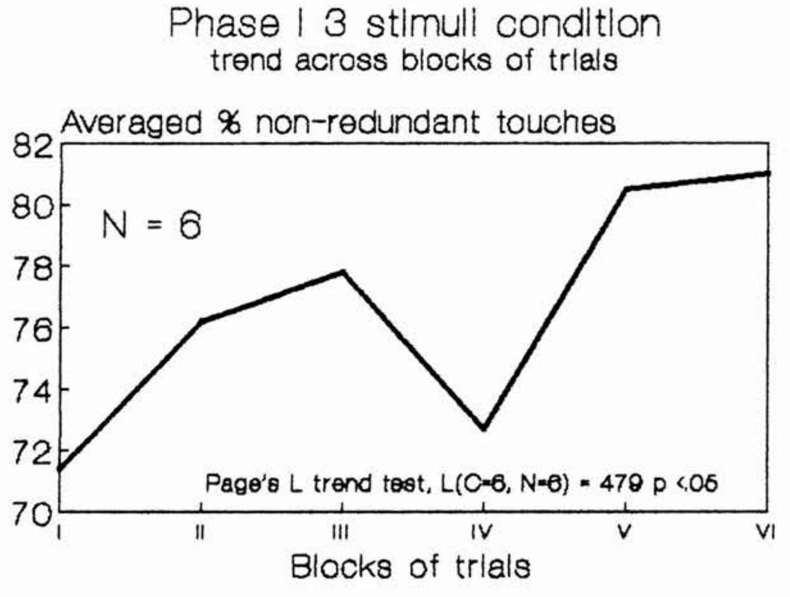



Fig. 4.6. 3 stimuli condition. Trend across blocks of trials (vincent sixths). On the horizontal axis the different vincent sixths are reported in which the global distribution was divided. On the vertical axis the percentage of non-redundant touches observed in each of blocks of trials is reported.

Figure 4.6 depicts the trend observed in the condition featuring 3 stimuli. This was the condition where the overall distribution of scores of the subjects was found to be non-significant. From this figure it can be seen that the behaviour of the subjects was not uneconomic as the previous analysis seemed to suggest. On the contrary a significant trend in the percentage of non-redundant moves across time can hardly be accounted for by chance alone. From the figure it can also be observed that on the IV block of trials the performance of the subjects dropped dramatically. Although difficult to interpret, this phenomenon can account for the failure of the previous analysis to show a significant result.

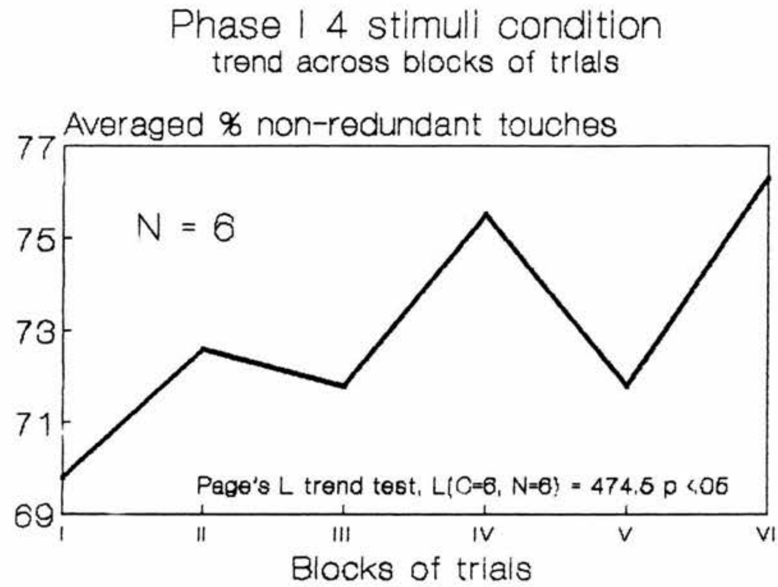



Fig. 4.7. 4 stimuli condition. Trend across blocks of trials (vincent sixths). On the horizontal axis the different vincent sixths are reported in which the global distribution was divided. On the vertical axis the percentage of non-redundant touches observed in each of the blocks of trials is reported.

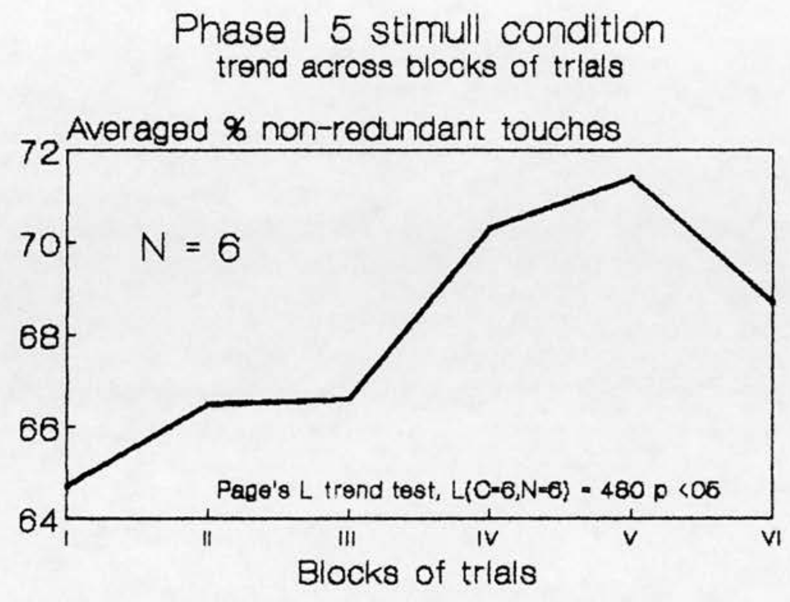



**Fig. 4.8. 5 stimuli condition. Trend across blocks of trials (vincent sixths). On the horizontal axis the different vincent sixths are reported in which the global distribution was divided. On the vertical axis the percentage of non-redundant touches observed in each of the blocks of trials is reported.**

**From figures 4.6-4.8 it emerges clearly that the subjects were benefiting from task practice.**

**Since there was clear evidence that the behaviour of the subjects was fitted to the task and improved spontaneously with practice, an investigation of the strategies on which such improvement was based would have been interesting. In other words, the evaluation of whether the subjects were exploiting some of the spatial constraints afforded by the different configurations would have provided useful information. However, the fact that the configurations changed on each trial and the huge number of different possible configurations, prevented a strictly numerical**

analysis of the paths followed by the subjects in searching the sets.

Nevertheless, from a visual inspection of the videotapes (counterchecked daily by two experimenters), it seemed clear that the subjects were deploying spatial organizational strategies such as using an adjacency rule (i.e. "Always select the closest item to one just touched").[4]

An analysis of such sorts of strategies will be reported for Phase III, where the presentation of a single condition, featuring a set of icons which completely filled up the 3 x 3 matrix, made it possible to collect a more affluent data base.

Relationship between performance and number of stimuli.

Comparing the percentage of non-redundant touches observed in the different stimuli conditions, the number of stimuli was not correlated with performance (Friedman test, $Xr^2(N = 6\ C = 4) = 9.6$ n.s.). Thus, given the opportunity to practice the subjects achieved a similar performance, irrespective of the size of the search space.

The averaged number of trials to criterion for the group was of 413.33 for the 3 stimuli condition, 785.83 for the 4 stimuli condition, and 490 for the 5 stimuli condition. Analogous profiles were obtained for individual subjects. This result supports the conclusion that with such a small number of stimuli, there was not a linear relationship between the number of items presented and the difficulty of the tasks. The 4 stimuli condition appeared more difficult to master than the 3 stimuli condition, but it seems as if a transfer of expertise occurred from the 4 stimuli condition to the 5

4 It can be noted that with sets ranging from 2 to 5 items, as those featured so far, the costs (in terms of length of the serial production required to search exhaustively) associated with searches were not particularly high. Therefore, the incentive to deploy strategic behaviours might not have been particularly strong in this phase of the study.

stimuli condition. In fact the number of trials required to master the 5 stimuli condition was only slightly above the number of trials required to reach criterion in the 3 stimuli condition.

In summary the results obtained from Phase I show that the subjects spontaneously performed exhaustive searches, and reduced redundant moves in the course of practice. This improvement resulted in a similar profile of performance in sets ranging from 3 to 5 stimuli.

The following Phase was designed to evaluate the effect of set size in conditions where the subjects did not have extensive exposure to each search space.

**Phase II: assessment of the effects of set size.**

The results obtained from the previous phase show that monkeys spontaneously performed exhaustive searches within sets containing up to 5 icons. Moreover, they showed a spontaneous tendency towards economy of search and their performance seemed to remain relatively unaffected by the increase of set size. Overall, it was clear that even the largest set (5 icons) was well within monkey's competence.

However, in the previous phase the set size was increased progressively and thus it was difficult to separate the effect of set size from that of transfer of expertise from one condition to the next.

For these reasons an independent assay of the effect of set size on performance was necessary. Such an assay had a twofold rationale. First of all, the effect itself of length of the serial production over the spontaneous deployment of strategic devices constitutes an important assumption of the economy/data reduction hypothesis. In fact, the hypothesis states that only when memory span risks overloading by the amount of information that it has to keep track of, a cognitive system will strive to find data reducing strategies. So, it was necessary to validate the assumption that the increase of set size was accompanied by an increase in task difficulty.

Secondly, it was necessary to find the set size that corresponded to an optimal level of difficulty such as one where the subjects had enough incentive to behave strategically but did not result in a high level of frustration that could have affected monkey's tolerance to protracted task exposure.

In order to isolate the effect of set size from the effect of task practice, a number of conditions, featuring different number of items, were presented in blocks of trials, randomly interspersed within each testing session.

As for the previous condition, the percentage of non-redundant touches was taken as an expression of the level of performance. On the basis of this measure it was evaluated the effects produced by the set size.

<u>Procedure</u>

10 experimental sessions were administered to each subject. Each session involved the presentation of 5 blocks of ten trials each. The set's size was hold constant within each block. The first 5 sessions administered to the subjects comprised blocks featuring 4, 5, 6, 7 and 8 stimuli, whereas the last 5 sessions comprised blocks featuring 5, 6, 7, 8 and 9 stimuli.

The stimuli were all identical in shape and colour. Their location on the monitor varied randomly for each trial (except for the blocks in which 9 stimuli were presented, where the 3 x 3 matrix was completely filled up). Fig. 4.9 depicts some examples of configurations featuring 6, 7, 8, and 9 stimuli respectively (conditions featuring 4 and five stimuli were analogous to those exemplified in Fig. 4.2, to which the reader can refer).

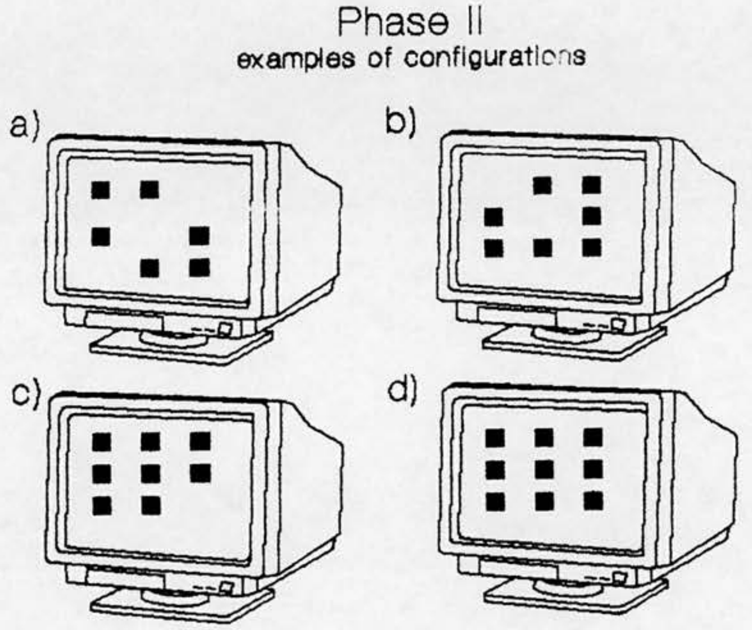



Fig 4.9. Some examples of configurations pertaining Phase II a) An example of configuration with 6 stimuli. b) An example of configuration with 7 stimuli. c) An example of configuration with 8 stimuli. d) An example of configuration with 5 stimuli.

Within each session, the order of presentation of the 5 blocks was based on a 5 (sessions) x 5 (blocks) Latin Square design. The randomization of the Latin Square was performed according to Winer (1971, p. 689).

<u>Results</u>

Relationship between performance and number of stimuli

In contrast with the results obtained from Phase I, in Phase II, a significant difference between performance in the various conditions was found (Freedman test, $Xr^2(N = 6\ C = 6) = 18.2\ p < .01$), now that the number of icons had increased beyond a low level. Moreover, a negative correlation between performance and the number of stimuli presented was observed (Page's L trend test, L(N = 6, C = 6) = 520 p <.001). Fig. 4.10 shows the averaged percentages of non-redundant touches and minimal path trials obtained by the group of subjects in the different conditions of phase II.

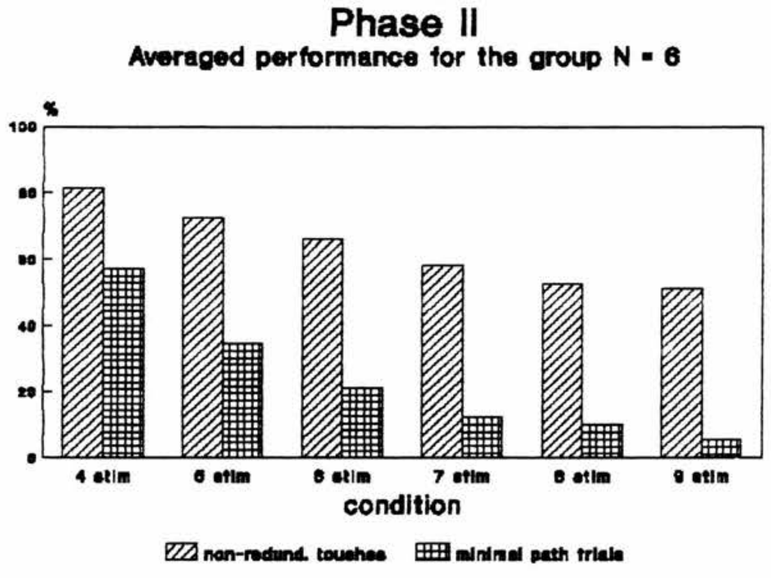



Fig 4.10. Percentage of non-redundant touches and minimal path trials obtained in the different conditions of Phase II.

In Fig. 4.11 individual data are reported, from which it can be observed that the same relationship observed between performance and number of stimuli for the group holds for each of the subjects too.

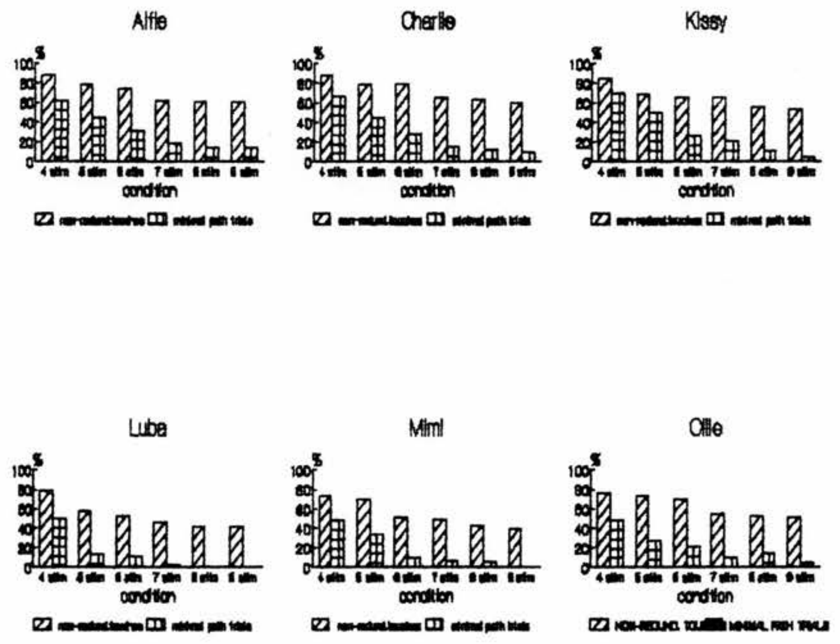



Fig. 4.11. Individual percentages of non-redundant touches and minimal path trials observed in the different conditions of Phase II.

Comparison with chance

After the global evaluation of the effect of set size on performance, an analysis was conducted to evaluate how economic the subjects were in each single condition. For the comparison of the distribution of scores obtained in the different conditions of Phase II with chance, the Kolmogorov-Smirnov test described for the previous phase was used. The results were the following.

4 stimuli condition.

The distribution obtained combining the scores of the six subjects was significantly different from the expected ($D = .3564$ $p < .005$). Likewise, on individual basis, the distribution of all the six monkeys was significantly different from the expected (Al, $D = .3778$, $p < .005$; Ch, $D = .4378$, $p < .005$; Ki, $D = .5066$, $p < .005$; Lu, $D = .2778$, $P < .005$; Mi, $D = .2578$, $p < .005$; Ol, $D = .2578$, $p < .05$).

5 stimuli condition.

The overall distribution obtained grouping the scores of the six subjects was different from the expected ($D = .3332$ $p < .005$). On individual basis, the test was significant for five monkeys (Al, $D = .4656$, $p < .005$; Ch, $D = .4091$, $p < .005$; Ki, $D = .4663$, $p < .005$; Mi, $D = .3591$, $p < .005$; Ol, $D = .3591$, $p < .005$).

6 stimuli condition.

The distribution obtained combining the scores of the six subjects was significantly different from the expected ($D = .3204$ $p < .005$), as well as the distributions obtained for each of the monkeys (Al, $D = .4507$, $p < .005$; Ch, $D = .4150$, $p < .005$; Ki, $D = .3948$, $p < .005$; Lu, $D = .1875$, $P < .005$; Mi, $D = .2374$, $p < .005$; Ol, $D = .3570$, $p < .05$).

7 stimuli condition.

The distribution obtained combining the scores of the six subjects was significantly different from the expected ($D = .2936$, $p < .005$). On individual basis, the distribution obtained for five of the monkeys differed significantly from the expected (Al, $D = .3282$, $p < .005$; Ch, $D = .4738$, $p < .005$; Ki, $D = .4060$, $p < .005$; Mi, $D = .2361$, $p < .005$; Ol, $D = .3282$, $p < .005$).

8 stimuli condition.

The distribution obtained combining the scores of the six subjects was significantly different from the expected ($D = .2698$ $p < .005$). On individual basis, the distribution of scores of four monkeys was significantly different from the expected (Al, $D = .4271$, $p < .005$; Ch, $D = .4769$, $p < .005$; Ki, $D = .3427$, $p < .005$; Ol, $D = .2615$, $p < .005$).

9 stimuli condition.

The distribution obtained combining the scores of the six subjects was significantly different from the expected (D = .2690 p <.005). On individual basis, the distribution of scores of four monkeys was significantly different from the expected (Al, D = .4807, p <.005; Ch, D = .5118, p <.005; Ki, D = .3519, p <.005; Ol, D = .3532, p <.005).

Overall, the results obtained from Phase II show that the performance of the subjects was above chance level in all conditions. This finding raised the question of whether performance was underpinned by the deployment of some strategic factors. Moreover, the assessment of whether the subjects (when faced with larger sets of icons) tried to compensate the inevitable increase in the delay of reward (caused by the necessity to perform a longer search) by speeding up their response latency was considered interesting. The following section features a time analysis performed in order to clarify these issues.

Analysis of temporal aspects of performance

The time analysis had two overall rationales.

a) Speed-accuracy trade-off

Since a negative correlation was found between the number of icons presented and the performance of the subjects, it was important to check whether the decrease of accuracy could have been accounted by the subjects speeding up their responses at expense of carefulness.

b) Evidence for organization.

A number of experiments conducted on adult humans (for a review see Wright, 1990), have put evidence for motor programming in relation with various time analyses in tasks which share some similarities with those featured in this thesis. In these experiments (Sternberg et al. 1982; Wright,

1990) the subjects are required to report, either by speaking or typewriting, a list of items (ranging from, single letters or numbers to words of different lengths) previously presented on a computer monitor. The results showed a relationship between the length of the list to be reported and the averaged inter-item interval during retrieval. This finding has been interpreted as evidence for the presence of motor programming, as opposed to simple chaining of responses as a behaviourist account of the process underling serial retrieval would postulate. In fact, a chaining hypothesis would suggest that the production of each component of the list to be reported is elicited by the stimulus provided by the response to the previous one. This latter process being local and always the same for different list lengths, it cannot explain a difference in latencies as a function of global features of the list to be reported (such as its length). On the contrary, more sophisticated processes used to organize the items in memory, to detect, retrieve, and report them, might well be sensible to global features of the list such as its length or its structural properties.

Of course, the sort of tasks used with human adults differ in many aspects from those featured in this study and the specific model proposed to account for the results (Sternberg et al., 1982) do not allow to draw useful predictions in the present study.

However, the tasks featured here require the production of sequences of actions, in a context that provides both the incentive and the opportunity to use some form of organization of the series to be produced, as those devised by Sternberg and colleagues (1982).

Thus, of particular interest was considered an assay of the relationship between the length of the serial production and the averaged inter-response time in monkey subjects, to see if it was possible to suggest some form of organization and planning in their serial production.

The time analysis aimed to the evaluation of the relationship between inter-touch intervals and the number stimuli presented.

The specific rationale behind the present analysis was the assumption that the larger the search space the more organizational factors are recruited to control the search behaviour (in order to reduce the memory load required by keeping track of all the moves). If the behaviour of the subjects is controlled by some forms of motor planning, according to Sternberg and colleagues (Sternberg et al., 1982), the length of the serial response should be correlated with the averaged intertrial interval. Similarly, in our case, a positive correlation between set size and intertrial interval should be found.

The analysis was made as similar as possible to those used by Sternberg and colleagues. In their experiments the time analysis was conducted only on correct trials, and all trials which contained pauses or where the subjects were not fluent (in speaking) were eliminated from the analysis (Wright, 1990). Thus, the present analysis was conducted only on minimal path trials, and all trials containing at least one intertouch interval of 2 seconds or more were excluded. A Page's L test was performed on the mean intertrial intervals for the different sub-conditions.

Results

The test was highly significant (L, C=6, N=6 = 506 p <.001), showing that there was a regular increase of the intertouch interval moving from the 4 stimuli sub-condition up to the 9 stimuli sub-condition (see Fig. 4.12).

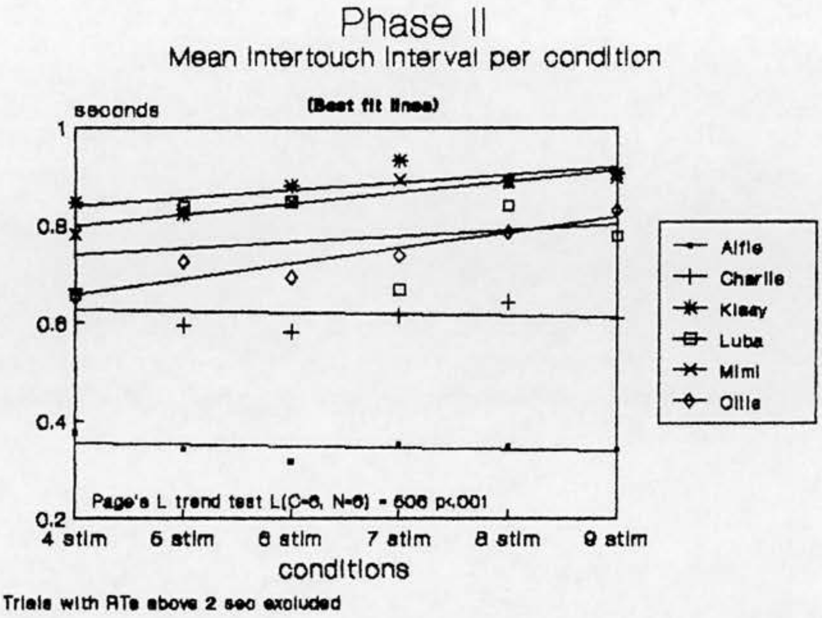



Fig. 4.12. Averaged inter-touch interval for the different conditions featured in Phase II.

From Fig. 4.12, it can be observed that only 4 subjects out of six showed a positive trend. However, this was enough to account for an overall positive trend.

The finding reported in this section, thus, can be taken as evidence that 4 subjects were principled in their searches and organized their responses according to global features of the search space, such as its size. This finding rules out the possibility that the observed negative relationship between set size and economy of search was due to a speed accuracy trade-off. On the contrary it seems better explained by an increase of task difficulty produced by the increase of the length of the serial production necessary to search exhaustively a larger set of icons.

Conclusions

Overall, the results obtained from Phase II show that the subjects were principled in all conditions. Moreover, from the time analysis some evidence of the use of organizational constraints emerged. In fact, capuchins seemed to control their behaviour according to some global features of the search space such as its size. On the basis of these results, therefore, it appeared worthwhile to evaluate the behaviour of the monkeys during protracted testing with a configuration which affords strong spatial constraints, such as the 9 stimuli one. This was done in Phase III.

Phase III: assay of strategic behaviour

By presenting the subjects with the same configuration of 9 items until a stable state of performance was achieved, phase III aimed at detecting possible changes over time in the level of economy showed by the subjects, and the spectrum of strategies which might support search efficiency.

From the results obtained from the time analysis performed for phase II some evidence emerged supporting the fact that larger set sizes produce incentive for organization. The present phase, therefore, aimed at determining whether a long exposure to the task would have changed search behaviour in a large set with spatial affordability.

Filling up the 3 x 3 matrix, the 9 stimuli configuration affords strong linear constraints. Moreover, it offers the opportunity of using a fixed location from which the search can begin, as well as the possibility of moving along fixed paths through the search space, and of conforming to an adjacency principle in exploring the array.

As mentioned above, a parallel study conducted on children (McGonigle et al., 1992; Dickinson, in preparation) showed that four and half year olds, when faced with an analogous set, exploited all the spatial constraints afforded by the configuration by means of very principled trajectories through the search space. This led to an almost perfect performance (about 100% non-redundant touches). The fact that the use of such strategic factors sustained performance, was confirmed by the observation that younger children (3 years old) were never as principled as their older companions, and, as a consequence, their performance was defective. The relationship between age and minimal path trials is shown in Fig. 4.13a. Examples of typical trajectory throughout the search space used by subjects of the two age groups are shown in Fig. 4.13b and c.

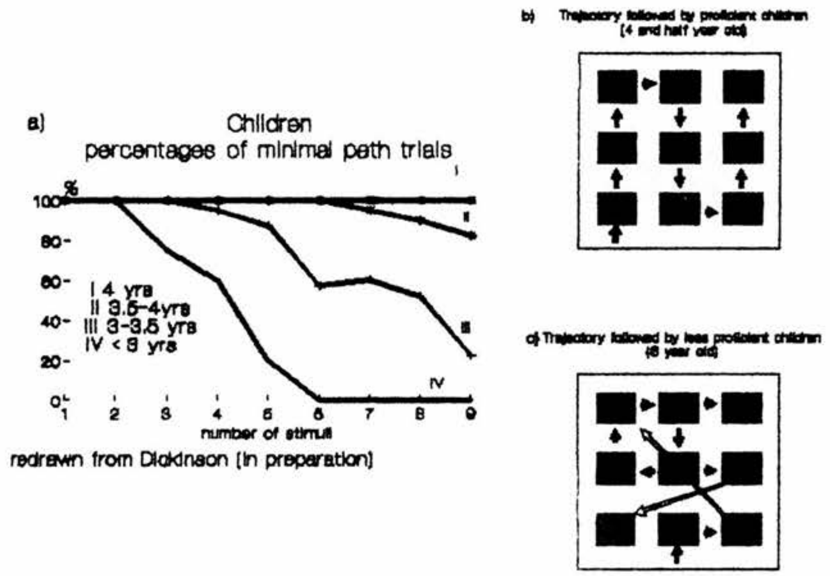



Fig. 4.13. a) Percentage of minimal path trials performed by children of different ages, redrawn by permission from Dickinson (in preparation). b) A trajectory frequently observed by the most proficient children (whose performance is represented by line I of the graph). c) an example of a less principled trajectory performed by the youngest children (whose performance is represented by line IV of the graph). The unfilled arrows represent non adjacent moves towards the pointed item.

Although it was clear that older children were able to use the perceptive information about the spatial constraints of the set to control their search behaviour, while the younger children failed to do so in such a principled way, the experiment on children did not bring results about the regulatory mechanisms, which, according to the working hypothesis for the present experiment, might be deployed in the course of task practice.

Nevertheless, the behaviour of older children provided an empirical template for maximal fitness to the task. On the other hand, the mathematical model of the probabilities associated with searches of a particular length, obtained assuming a random walk through the search space (St Johnston, 1993), was taken as representative of the lower end of the distribution of possible search patterns. Within this scale, ranging from unprincipled to highly strategic behaviour, an assay was attempted of the spatial organization imposed by monkey subjects on the search space and its change over time.

Procedure

All the 50 trials presented in each daily session featured the presentation of 9 monochromatic stimuli. The administration of this condition continued if a positive trend in the percentage of non-redundant touches was observed.

Data analysis

The data analysis featured an evaluation of the overall performance shown by the subjects in phase III and a trend analysis to assay changes in economy of search over time. Both were based on the percentage of non-redundant touches performed by the subjects.

The analysis of the overall performance of the subjects in phase III (as for the previous phases) was a comparison of the obtained distributions of scores with those expected by chance, performed using the Kolmogorov-Smirnov test.

The trend analysis was based on the Page L trend test (Siegel and Castellan, 1988).

Results

OVERALL ANALYSIS OF PERFORMANCE

a) group performance

The combined distribution for the six subjects was significantly different from the expected ($D = .3281$, $p < .005$). The distributions of trials for this condition are shown in Fig. 4.14.

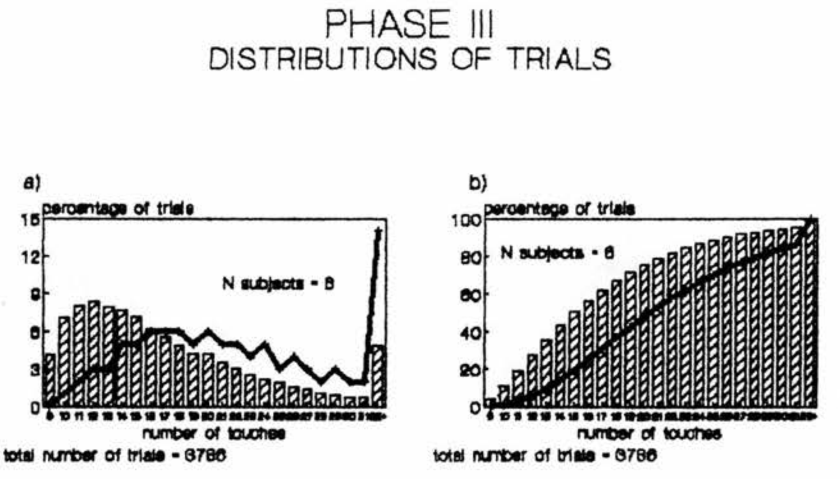



Fig. 4.14. a) Observed (bars) and expected (line) percentage of trials of a particular length observed in Phase III. b) The observed (bars) and expected (line) cumulative distributions of trials on which the statistical test was based on.

From Fig. 4.14 it can be observed that the obtained distribution differed from the expected mainly for its modal tendency toward economic trials.

b) individual performance

On individual basis, the distribution of scores of four monkeys was significantly different from the expected (Al, D = .2844, p <.005; Ch, D = .5427, p <.005; Mi, D = .2887, p <.005; Ol, D = .4183, p <.005). Fig. 4.15 shows the distributions of trials for each subject.

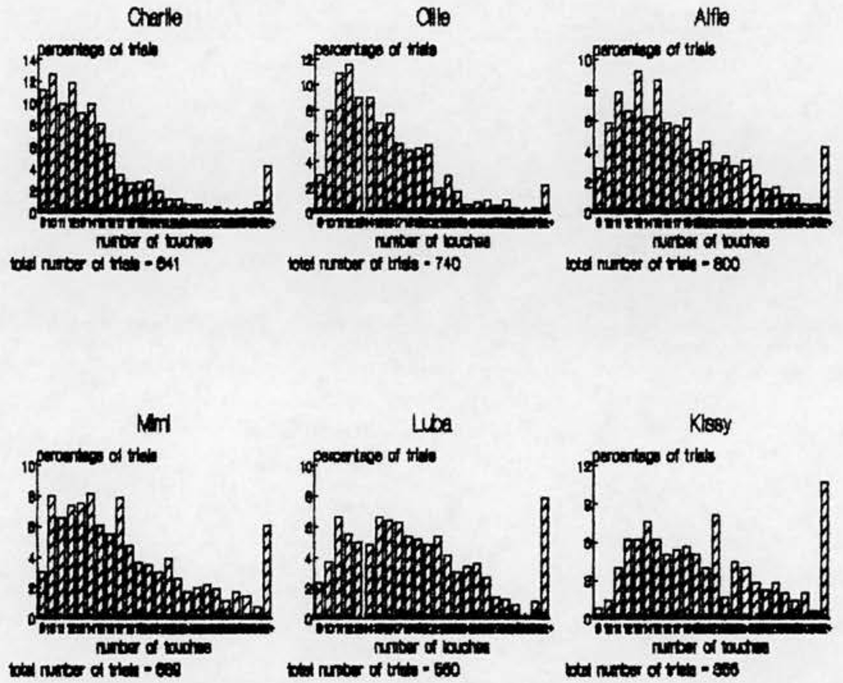



Fig. 4.15. Distribution of trial lengths obtained in Phase III for each of the monkeys.

Thus, 4 monkeys performed exhaustive searches in a 9 stimuli set above chance level. The shape of their distribution of scores was so different from the expected to lead to a significant result for the group.

In the next section, an analysis of the development over time of such a proficiency will be presented.

EVIDENCE FOR IMPROVEMENT WITH PRACTICE

As for Phase I, a Page L trend test was performed on the observed percentages of non-redundant touches obtained for each of the vincent sixths in which the overall distribution was divided. A positive trend in the percentage of non-redundant touches over these successive different blocks of trials was considered as a suggestion that the subjects were self-regulating their behaviour in order to achieve a better economy of search.

The Page L trend test performed on the combined data for all the subject was significant (p <.05), as shown in Fig. 4.16.

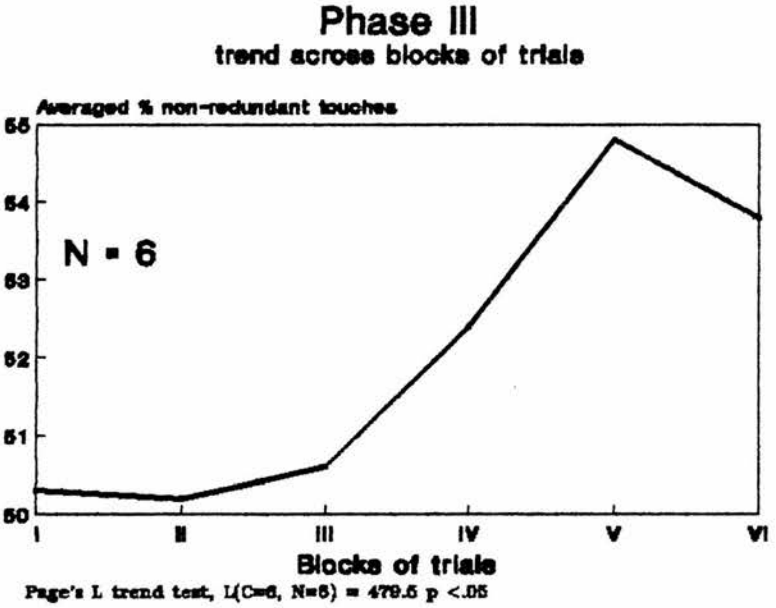



Fig. 4.16. Percentage of non-redundant touches. On the horizontal axis the different vincent sixths, in which the global distribution was divided, are reported. On the vertical axis the percentage of non-redundant touches observed in each block of trials is reported.

Individual trends for the 9 stimuli sub-condition are shown in Fig. 4.17.

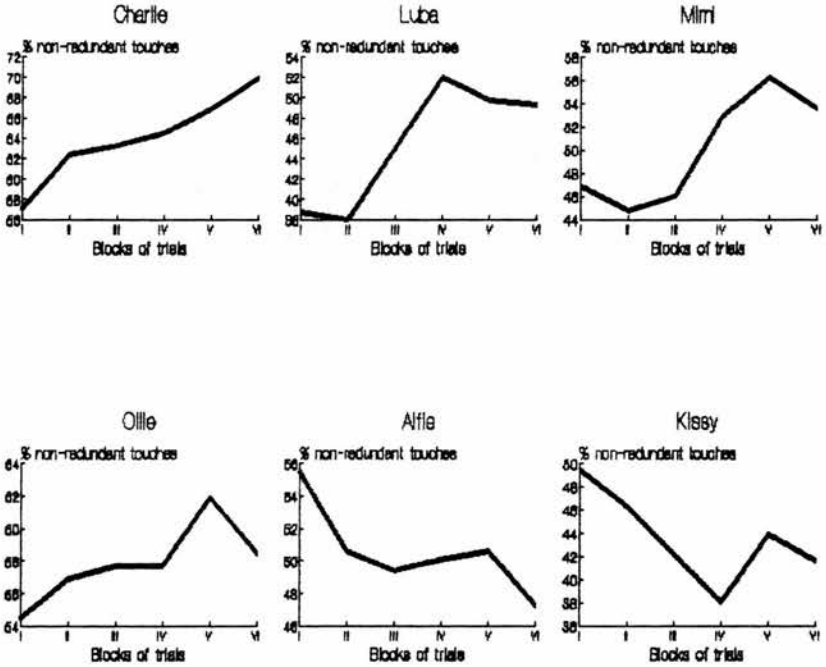



Fig. 4.17. Individual trends observed in Phase III. See Caption of Fig. 4.16 for explanations.

From a visual inspection of Fig. 4.17, it can be noted that the performance of 4 monkeys (Mi, Ol, Ch and Lu) accounts for the overall trend, while the other 2 monkeys (Al and Ki) did not show a positive trend.

Fig. 4.18 shows how the distribution of trial length changed in the course of practice. From it, the fact that the improvement of performance was mostly due to shift of the mode towards searches progressively closer to minimal paths can be observed.

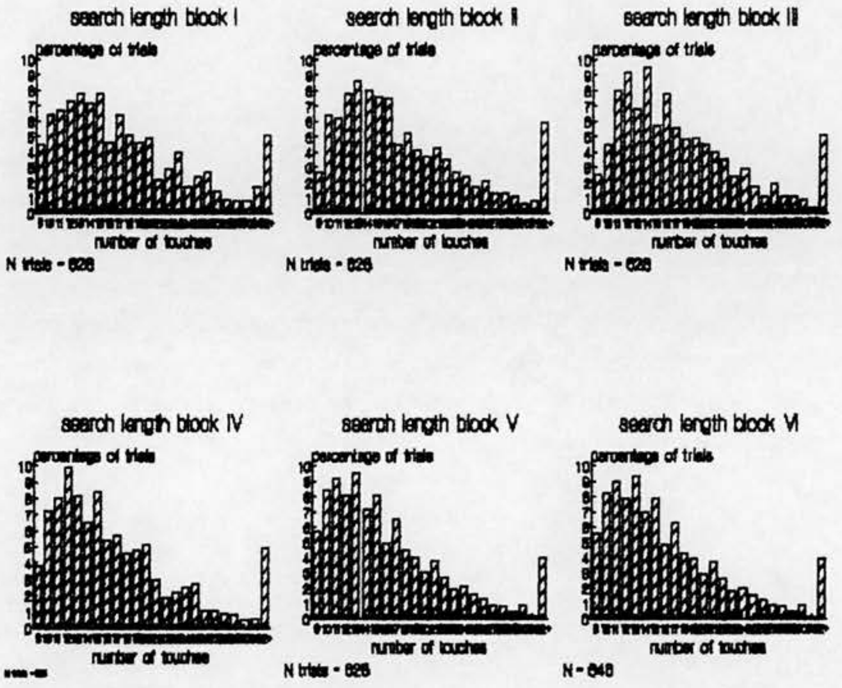



Fig. 4.18. Change in the distribution of trial lengths with practice. Blocks I-VI, respectively from top left to bottom right.

Overall, these results showed the ability of the subjects to achieve great economy of search and self-regulate their actions over time. In the next section the use of some forms of spatial constraints available to the subjects in this 3 x 3 matrix, will be scrutinised, in order to evaluate which sort of strategic factors (if any) their performance was based on.

## EVIDENCE FOR THE USE OF CONSTRAINTS

### Starting points

In order to assess if subjects were beginning their searches from preferred positions, a chi squared was performed on the location selected to start the search in each trial.

The results showed that starting points were not equally distributed among the 9 possible locations (Chi squared (df = 8) p <.001 for all the monkeys.

This result was considered to be the first indication of the fact that the subjects were deploying strategic factors in order to reduce the memory effort required by a high level of fitness to the task. In fact, it can be assumed that a subject which knows the location (or set of locations) from which its searches always start, is less likely to reiterate on this location in the terminal phase of the search. It will know that the starting location is likely to be already visited. Thus, the number of items to remember is reduced of one element.

Adjacent moves

In order to assess if the trajectory followed by the monkeys in the search space was based on successive moves on adjacent locations, chi squared were performed on the four classes of possible transitions defined by the distance of two successive touches in the set of stimuli. The analysis was based on the frequency of each transition type.

The chi squared performed on the combined frequencies for the group of six monkeys was highly significant (chi squared = 17854.8, df = 3, p <.001). Fig. 4.19 shows the obtained and the expected percentages for each transition distance (between successive moves).

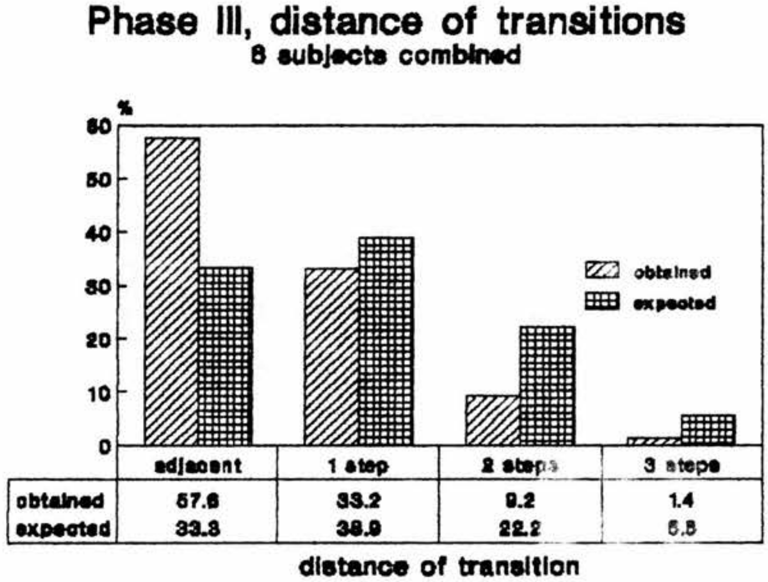



Fig. 4.19. Obtained and expected percentages of transitions for each distance (calculated in terms of items interposed between two locations on the screen).

From Fig. 4.19, the fact that only adjacent moves were more than expected can be observed. In contrast, all other types of transitions were below the expected value. This, together with the high total number of transitions observed, (n = 61027) explains the very high value of the obtained chi squared.

The same results were obtained for individual subjects (Al, chi squared = 4263.6, df = 3, p <.001; Ch, chi squared = 3715.4, df = 3, p <.001; Ki, chi squared = 4748.2, df = 3, p <.001; Lu, chi squared = 1894, df = 3, p <.001; Mi, chi squared = 3556, df = 3, p <.001; Ol, chi squared = 4568.3, df = 3, p <.001). The obtained and expected percentages of each transitions type are shown in Fig. 4.20, for each subject.

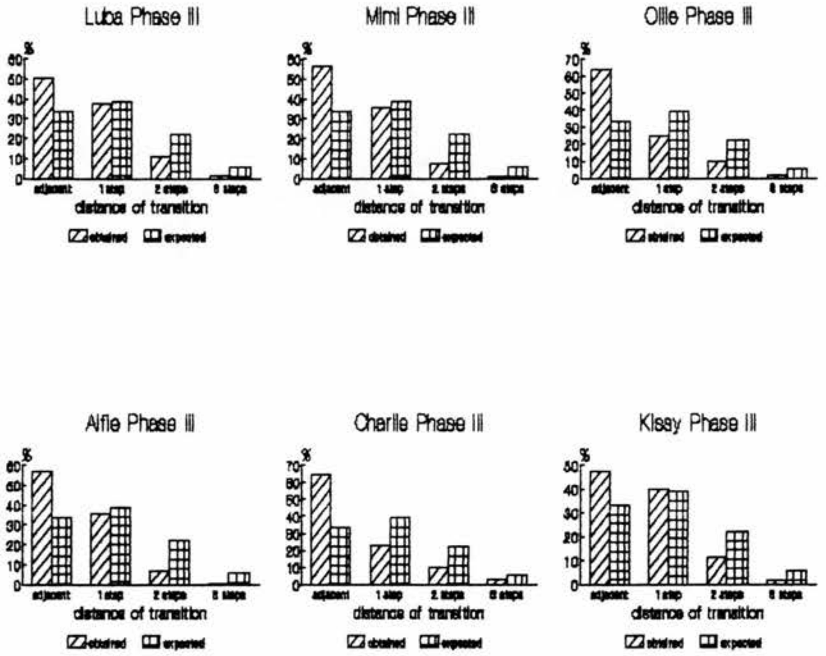



Fig. 4.20. Individual monkeys. Obtained and expected percentages of transitions for each distance (calculated in terms of items interposed between two locations on the screen).

Fig. 4.20 shows that results based on individual subjects support the conclusions drawn for the group as a whole.

Thus, monkeys were using adjacent moves more than expected. However, their percentages of adjacent moves never reached the level shown by the most proficient children of the study quoted earlier. In fact, a trajectory such as the one showed in Fig. 4.13b, consists of a 100% of adjacent moves. The percentages of the monkeys never approximated this value.

In addition, a visual inspection of the videotapes of the testing sessions, showed a high variability among different trials in the paths followed by the subjects during their searches. For this reason, it was impossible to evaluate on quantitative basis if the monkeys were using some form of vectorial organization of their search paths like children, even if just on probabilistic basis, and perhaps only on some sub-sequences of responses. However, from ocular inspection of the videotapes the fact that the subjects were using some privileged directions of travel seemed quite clear.

Examples of common directions of transitions (taken from the behaviour of one the most proficient subjects, Ch) are shown in Fig. 4.21. There, the fact that not all the possible directions of transitions were represented in the search patterns of the monkey (and that some transitions were highly preferred) can be observed.

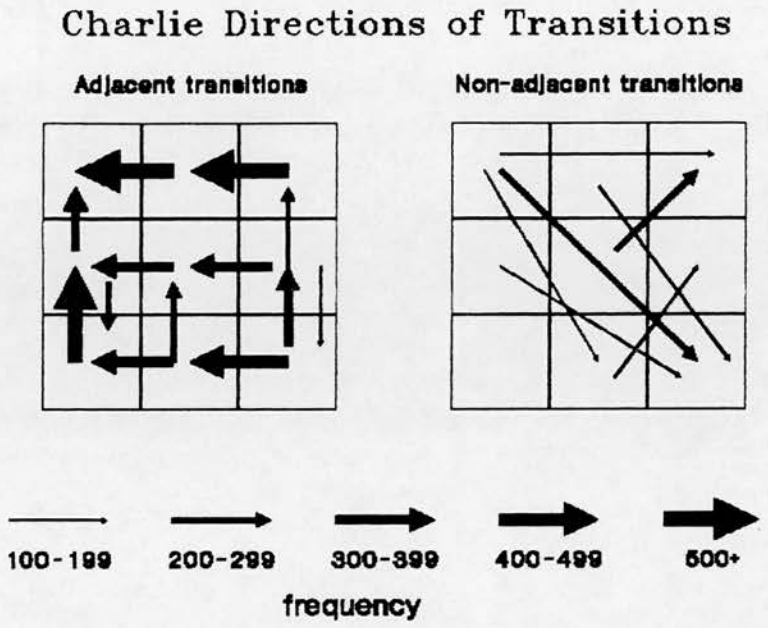



Fig. 4.21. Subject Charlie, frequencies of transitions in particular directions. The transitions are calculated on the first 9 moves (8 transitions). Only frequencies above chance level are reported in the figure. The thickness of the arrows is proportional to the frequency of that particular transition.

Overall, the results presented so far point to a spontaneous tendency to economically search a set of 9 stimuli. Moreover, the proficiency of the monkeys improved with practice and their performance seemed to be sustained by the strategic use of some constraints afforded by the spatial configuration of the search space. Faced with a fixed and spatially constrained configuration, they tended to start from preferred locations, move predominantly on adjacent items and, probably, they followed some preferred directions of travel. This behaviour stands in marked contrast with the "performance" of the random trajectory probability model mentioned earlier.

However, the searches of the monkeys were not as principled and economic as those of 4 and half year old children. On the contrary, it seems that the behaviour of the monkeys partially overlapped with that of younger children, falling midway between the completely principled and economic search behaviour of the older children and the completely unprincipled and mostly uneconomical "behaviour" of the random model.

At this point a question still to be answered is why monkeys did not self-regulate even more, approximating the very principled behaviour shown by successful children. Two alternative explanatory hypotheses can be formulated. The first is that, in the course of task practice the subjects tended to speed up their responses to detriment of accuracy. The results presented for Phase II already ruled out the fact that the subjects were increasing the speed of response when more items were present. However, in the present case, it was still possible that they were increasing the speed of response in the course of task practice. The second hypothesis is that within their search trajectory, subjects were, at some point, loosing track of where they were and thus found it difficult to remember the items already touched.

In the following section these hypothesis will be evaluated, on the basis of a time analysis of the behaviour of the monkeys.

TIME ANALYSIS

The first hypothesis formulated to explain why the subjects economy of search reached a plateau before becoming optimal, was based on a speed-accuracy trade-off argument. If the currency used by the subjects to evaluate their economy was the time spent in each search, they might have tried to increase the speed of response instead of trying to search in a more principled fashion. Fig. 4.21 shows the averaged duration of searches completed in different number of touches for each of the successive vincent sixths. The figure also shows the trial's duration, which would have been expected if the intertouch interval remained constant (calculated on the basis of the averaged inter-trial interval registered on minimal paths).

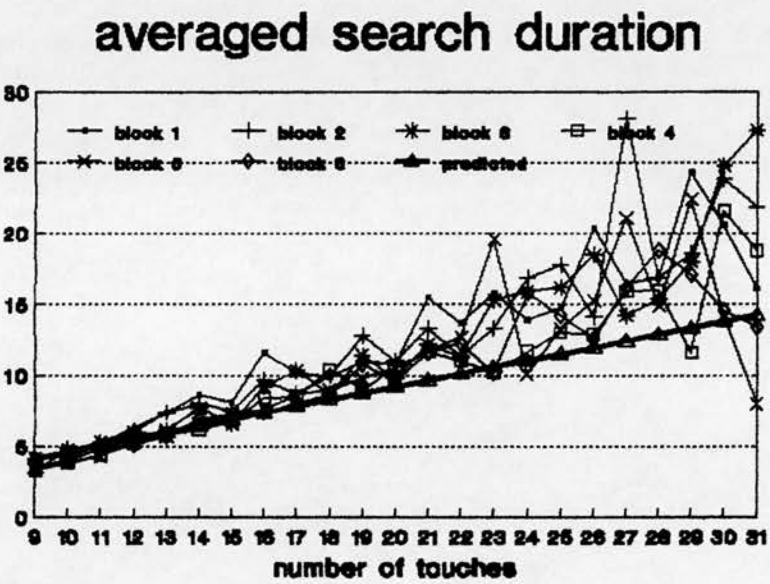



Fig. 4.22 Observed trial duration for searches of different lengths (on the horizontal axis are the number of touches reported) in the different blocks of trials, and duration predicted on the basis of the averaged inter-touch interval recorded on minimal paths (thick line).

From the figure it can be observed that the curves for the different blocks of trials practically overlap and the monkeys never became faster than expected.

From these results some interesting observations can be made. The reader will remember that in Phase II, a positive trend was found when the intertouch intervals were plotted according to the size of the search space. Following Sternberg and colleagues (Sternberg et al., 1982; Wright, 1990), these results were interpreted as evidence for the presence of some form of motor programming and as a suggestion that the larger the search space, the more organizational factors were recruited to explore it. Now, from the current results, it is apparent that even in the less economic trials the subjects were not reducing their intertouch interval. Thus, the Sternberg argument, if applied to the current data, would allow the conclusion that the same organizational constraints that the subjects were using when maximally economic (on minimal paths) were still deployed in the less economic trials. Therefore, the latter should not be the expression of occasional relaxations of the subject to unprincipled searches. On the contrary, it seems that because the strategies were never maximally principled, they (although applied in every trial) led to some long and non-economical searches. It can also be conjectured that the discrepancy between the predicted and the observed trial time in the very long searches might be an expression of the fact that when the subjects were losing track of where they were (along the search trajectory), some extra-time was spent in the attempt of identifying those icons that still needed to be visited.

Thus, apparently, a speed-accuracy trade-off account cannot explain the failure of the monkeys in reducing the search costs below a given point. A more plausible hypothesis can be based on the relative frequencies of long trials recorded in the different vincent sixths. In fact, the distribution shown in Fig. 4.14, 4.15, and 4.17, shows that the frequency of the most uneconomical trials sharply decreased in the course of practice. Thus, the hypothesis can be formulated that in the late trials the subjects were lacking in part of an important source of negative feedback, and for this reason were not able to self-regulate their behaviour above a certain level. In other words, they might have become victims of their own success.

In order to test the hypothesis that, when provided with some additional negative feedback for very long searches, the subjects would have continued to regulate their actions, a second experiment was devised. It will be presented in the next section.

## 4.3 Experiment 2

### Rationale

This experiment focused on the effects of explicit negative feedback on task fitness, and on the self-regulatory processes towards economy, which, as we have seen, accompany practice.

As outlined in the introduction, an economy/data reduction hypothesis would postulate that an organism, faced with the problem of organising long serial productions of acts, will tend to find an optimal balance between external costs (associated, for example, with the time spent to get a reward) and internal costs (associated with the management of cognitive resources[5]).

From the distribution of trial lengths obtained from experiment 1, we have seen that, in the course of task practice, the frequency of very long trials (i.e. characterised by a large number of surplus moves over the minimum required) was progressively reduced (see figure 4.18). However, Fig. 4.18, shows also that after a long task exposure, the (right) tail of the distribution was still characterized by the occasional presence of very long trials.

On the basis of these results, the hypothesis was formulated that the subjects, after having self-regulated their behaviour up to a certain standard of economy, started to lack a valuable source of negative feedback (which at the beginning of testing was provided by a high frequency of very long searches). As a consequence, the conjecture was made that monkeys began to encounter difficulties in the evaluation of their own behaviour (by comparing the costs of

---

5 Aimed, for example, at the attainment of a satisfactory trade-off between the cognitive strain required by the development of strategies (which allow the reduction of the memory load) and the advantage deriving from the reduction of memory load itself.

"good" and "bad" runs) and that this led to a stable state of their performance.

Thus, in this experiment much more explicit feedback of behaviour fitness (albeit at the end of each sequence) was given to the subjects. The mostly unfitted sequences of responses were penalized by omitting the reward and prolonging the intertrial interval when the number of redundant touches exceeded a given criterion. This was done in order to "prune" the distribution of trials by "cutting-off" the tail of occasional extremely long searches. The rationale was to assess whether such a perturbation of the equilibrium reached by the subjects between internal and external costs could have pushed them towards further self-regulation, in order to find a new satisfactory balance between the two types of costs.

Materials and Methods

The same subjects and apparatus used in Experiment 1 were employed in Experiment 2.

Procedure

The procedure followed in Experiment 2 was similar to the one adopted in phases I and III of Experiment 1. The subject was presented with different conditions featuring different numbers of stimuli. In particular, conditions featuring 3, 4, 5 and 9 stimuli were presented.

As in Experiment 1, the subject was required to touch all the stimuli on the monitor, in order to clear the screen, produce a series of tones and obtain a peanut as a reward. The stimuli were exactly the same as those presented in Experiment 1.

However, in Experiment 2, the subject was penalized when the number of reiterations in a trial exceeded a criterion (repetitions were still allowed and were not penalized). The penalty consisted in the omission of the reward and in the addition of further 5 sec delay to the 15 sec intertrial interval.

During the extra 5 sec delay a white screen was presented and a tone (different from that one presented before the administration of a reward) signalled that the trial was to be penalized. The criterion for penalty was established, for each subject and for each stimulus number condition, on the basis of the distribution of trial lengths obtained from Experiment 1. As criterion, the number of reiterations that allowed the subject to obtain the reward in the 85% of the trials was chosen. If the subject improved, the criterion was modified accordingly.

## Data Analysis and Results

### ANALYSIS OF THE DISTRIBUTION OF TRIALS OBTAINED IN EXPERIMENT 2

#### Data Analysis.

As for Experiment 1, to compare the distributions of scores obtained in the different conditions with those expected by chance for the same number of stimuli, the Kolmogorov-Smirnov one sample test was used (see data analysis of Experiment 1). A Wilcoxon test for related samples was used to compare the performance shown by the subjects in the present experiment with the data obtained from experiment one. A trend analysis was performed to evaluate the improvement with practice. Finally, the chi square test was used for a microanalysis of the spatial strategies adopted by the subjects.

#### Results

3 stimuli condition. The combined distribution of scores for the six subjects was not statistically significant, however, when the data were analyzed for each individual

subject, the distribution of four of the six subjects was significantly different from the expected (Al, D = .4600, p <.01; Ch, D = .0833, p <.05; Mi D = .2071, p <.05; Ol, D = .2400, p <.01).

Fig. 4.23 shows the distribution of trials obtained in the 3 stimuli condition.

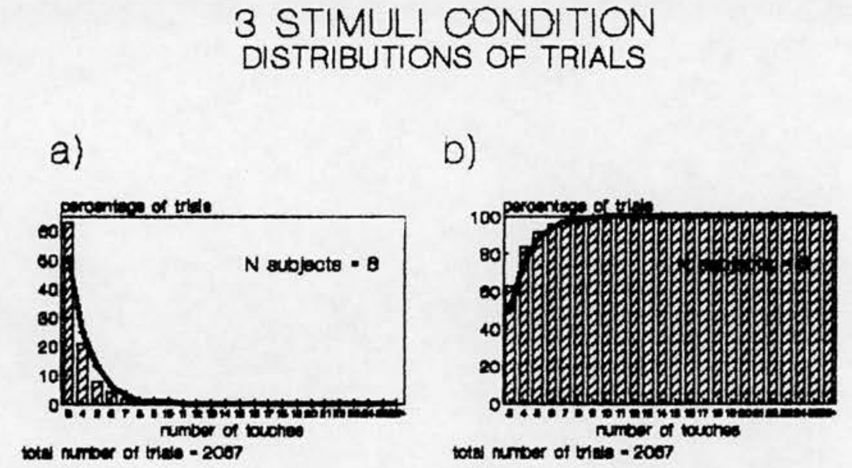



Fig. 4.23. 3 stimuli condition a) Observed (bars) and expected (line) percentage of trials of a particular length. b) The observed (bars) and expected (line) cumulative distribution of trials on which the Kolmogorov-Smirnov test was based.

Comparing Fig. 4.23, with Fig. 4.3, which shows the results obtained from the 3 stimuli condition of experiment 1, it can be observed that the two distributions look very similar. This confirms the hypothesis that, with such a small search space, even a random exploration of the configuration would produce a satisfactory percentage of minimal path trials, thus producing no incentive to self-regulate.

4 stimuli condition. The distribution obtained combining the scores of the six subjects was significantly different from the expected (D = .3697 p <.005), as well as the distributions obtained for each of the subjects (Al, D = .2984, p <.005; Ch, D = .3702, p <.005; Ki, D = .2704, p <.005; Lu, D = 3844, p <.005; Mi, D = .4540, p <.005; Ol, D = .4183, p <.005).

The distributions obtained from this condition are shown in Fig. 4.24.

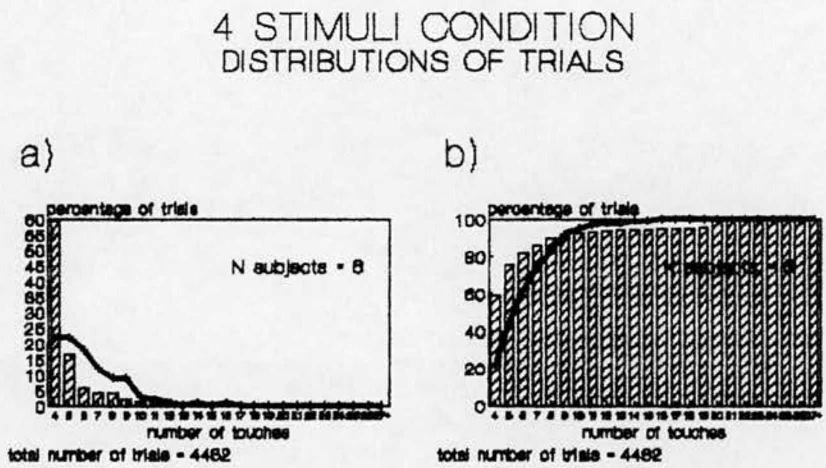



Fig. 4.24. 4 stimuli condition a) Observed (bars) and expected (line) percentage of trials of a particular length. b) The observed (bars) and expected (line) cumulative distribution of trials on which the Kolmogorov-Smirnov test was based.

Comparing the distribution shown in Fig. 4.24 with that obtained for the 4 stimuli condition of experiment 1 (see Fig. 4.4), it can be observed that there was an effect of selective feedback, and it found expression especially as an increase of the percentage of minimal path trials.

5 stimuli condition. The distribution obtained combining the scores of the six subjects was different from the expected (D = .4629 p <.005), as well as the distribution obtained from each of the monkeys (Al, D = .4208, p <.005; Ch, D = .5350, p <.005; Ki, D = .4876, p <.005; Lu, D = .4261, p <.005; Mi, D = .4270, p <.005; Ol, D = .5224, p <.005). Distributions for the 5 stimuli condition are presented in Fig. 4.25.

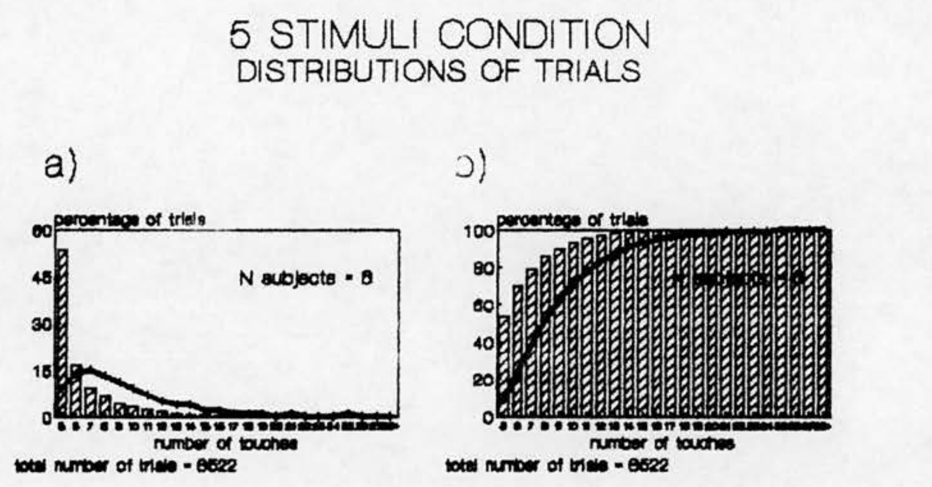



Fig. 4.25. 5 stimuli condition a) Observed (bars) and expected (line) percentage of trials of a particular length. b) The observed (bars) and expected (line) cumulative distribution of trials on which the Kolmogorov-Smirnov test was based.

Comparing Fig. 4.25 with Fig. 4.5 which shows the distribution obtained from the 5 stimuli condition of experiment 1, it can be observed that the distribution obtained from experiment 2 is characterized by a sharp reduction of long trials and by an increase in the percentage of both minimal paths and trials characterized by only few surplus moves.

9 stimuli condition. The distribution for the group of the subjects was different from the expected (D = .5529 p <.005) as the distribution obtained for each of the individuals (Al, D = .5153, p <.005; Ch, D = .6683, p <.005; Ki, D = .6044, p <.005; Lu, D = .4318, p <.005; Mi, D = .4148, p <.005; Ol, D = .5723, p <.005). The distribution of trials obtained from this condition is depicted in Fig. 4.26.

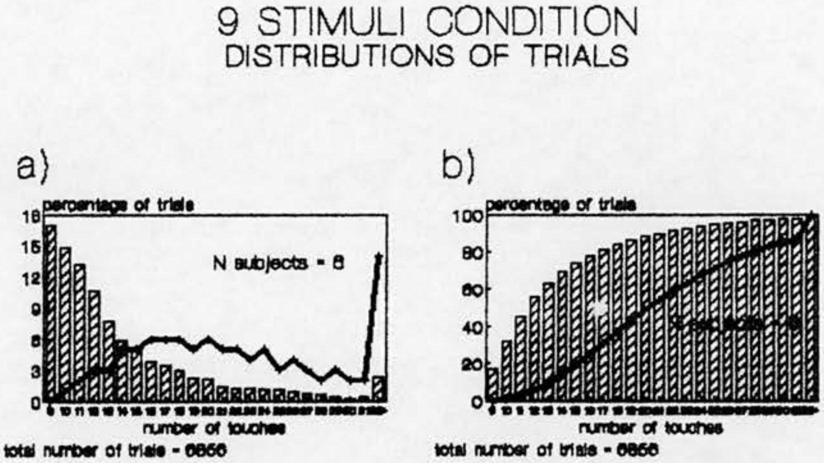



Fig. 4.26. 6 stimuli condition a) Observed (bars) and expected (line) percentage of trials of a particular length. b) The observed (bars) and expected (line) cumulative distribution of trials on which the Kolmogorov-Smirnov test was based.

The 9 stimuli condition is the condition where the effect of selective feedback was most conspicuous. The distribution of trials differs sharply from that obtained in experiment 1 (compare with Fig. 4.14). Thus, it seems clear that the feedback provided in this experiment had its most powerful effect in this condition which featured a large number of items and strong spatial constraints to be exploited.

In order to check quantitatively the overall reduction of redundant moves produced by the selective feedback in conditions featuring 9 stimuli, a Wilcoxon test was performed to compare the percentages of non-redundant touches observed in the equivalent conditions of experiment 1 and 2.

The mean percentage of non-redundant touches, performed by all the subjects in experiment 2 was higher than in experiment 1 (Wilcoxon's test $T(N = 6) = 0$ $p < .05$). Thus, as shown in Fig. 4.27, the performance of all subjects benefited from the explicit feedback provided in experiment 2.

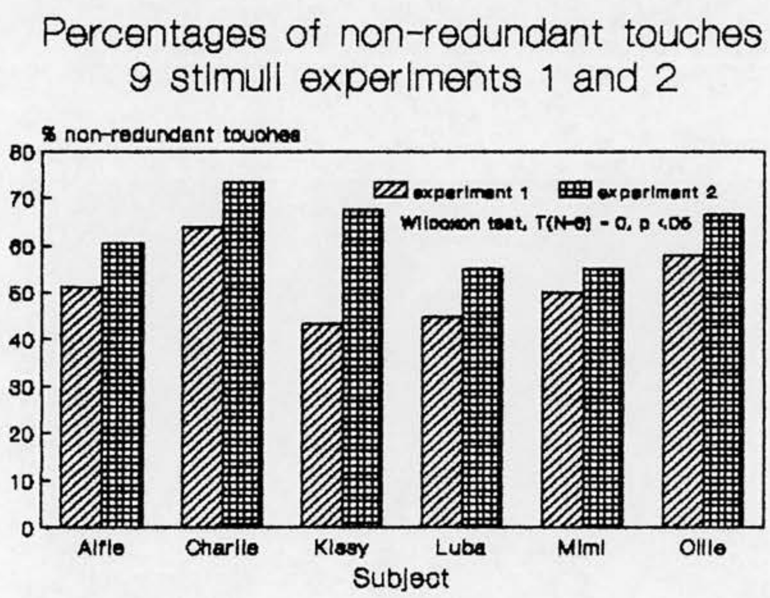



Fig. 4.27. Percentage of non-redundant touches obtained in the 9 stimuli conditions of experiment 1 (Phase III) and 2.

From Fig. 4.27, a positive correlation can also be observed (Rho = .895 $p < .05$) between performance of each subject in the 2 experiments (i.e. the most proficient subjects in the non-feedback condition are the best subjects in the feedback condition as well).

The effect of explicit feedback is even more evident when the portion of the improvement space gained in experiment 2 is considered, as shown in Fig. 4.28.

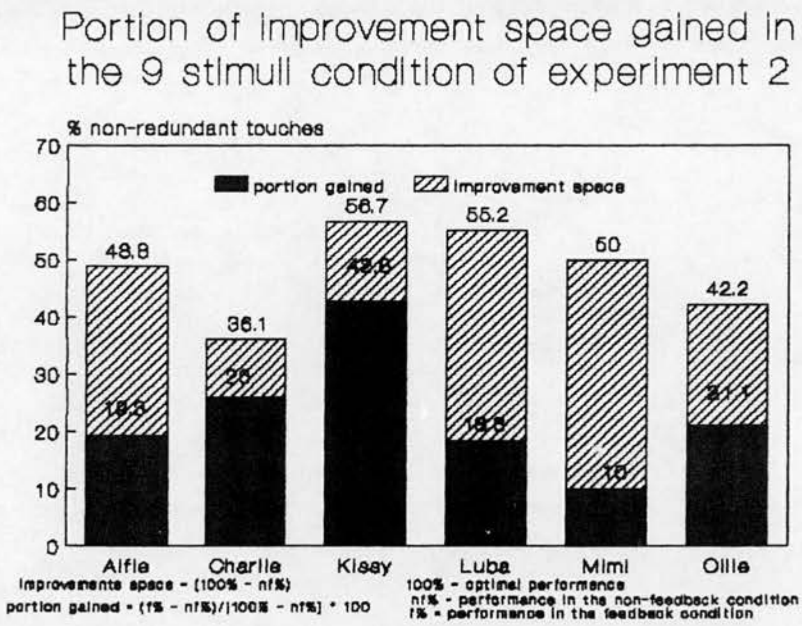



Fig. 4.28. Portion of improvement space gained in the 9 condition stimuli of experiment 2.

EVIDENCE FOR THE USE OF CONSTRAINTS

Since, in experiment 1 some evidence was found of the fact that performance was sustained by the deployment of spatial strategies, an analogous analysis was performed for experiment 2, where a better overall performance was found.

The analysis of the search strategies deployed by the subjects was conducted (as for experiment 1) on the condition featuring 9 stimuli, where the complete filling-up of the 3 x 3 grid, afforded stronger spatial constraints than the other conditions.

Starting points

As, in experiment 1, it was found that starting points were not equally distributed among the 9 possible locations ($Chi^2$, df = 8, p <.001 for all the monkeys).

However, since each subject did not chose an unique starting location and there was a variability between the starting points selected by different subjects, it proved impossible to compare quantitatively the differential use of this strategy in the two experiments.

Nevertheless, a quantitative comparison between the two experiments was performed for adjacent moves, which, allowing a measure of their overall percentage, made it possible to compare directly the results obtained from experiment 1 and 2.

Adjacent moves

The rationale and the procedure for the analysis of adjacent moves was the same as delineated for experiment 1.

As for experiment 1, the $chi^2$ performed on the combined frequencies for the group of six monkeys was highly significant ($chi^2$ = 405999.3, df = 3, p <.001).

Fig. 4.29 shows the relationship between the obtained and the expected percentages of transitions according to the distance between two successive moves.

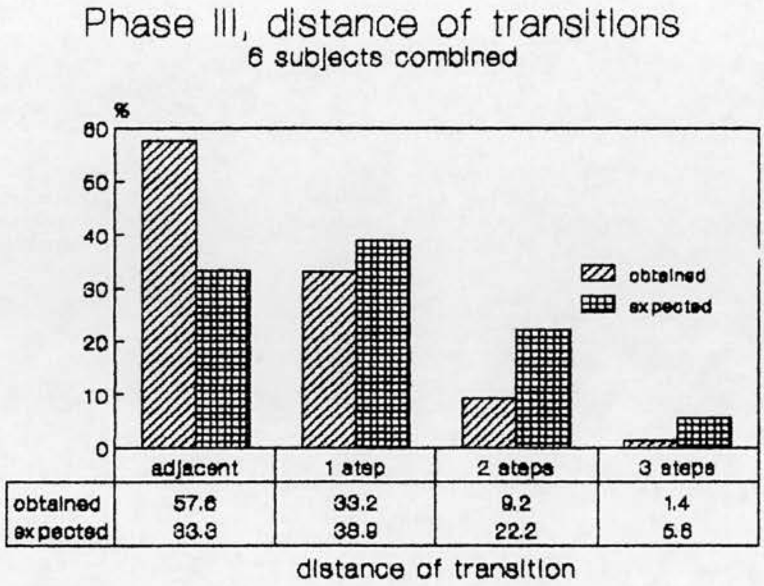



Fig. 4.29. Obtained and expected percentages of transitions for each distance (calculated in terms of items interposed between two locations on the screen.

From Fig. 4.29 it can be observed that only adjacent moves were more than expected. In contrast, all other types of transitions were less than expected. The sample on which the analysis was conducted was very affluent (n = 89159), as for experiment 1, and again this explains the very high value of the obtained $chi^2$.

Moreover, comparing the percentage of adjacent moves observed in this experiment and those reported for the 9 stimuli condition of experiment 1 (see Fig. 4.19), it can be observed that in experiment 2 the subjects performed a 10% more adjacent moves than in the previous experiment.

As shown in Fig. 4.30, when individual subjects were considered, analogous results were found (Al, $chi^2$ = 6067.4, df = 3, p <.001; Ch, $chi^2$ 7708.8, df = 3, p <.001; Ki, $chi^2$

= 5402.6, df = 3, p <.001; Lu, $chi^2$ = 2521.2, df = 3, p <.001; Mi, $chi^2$ = 8184.4, df = 3, p <.001; Ol, $chi^2$ = 22507, df = 3, p <.001).

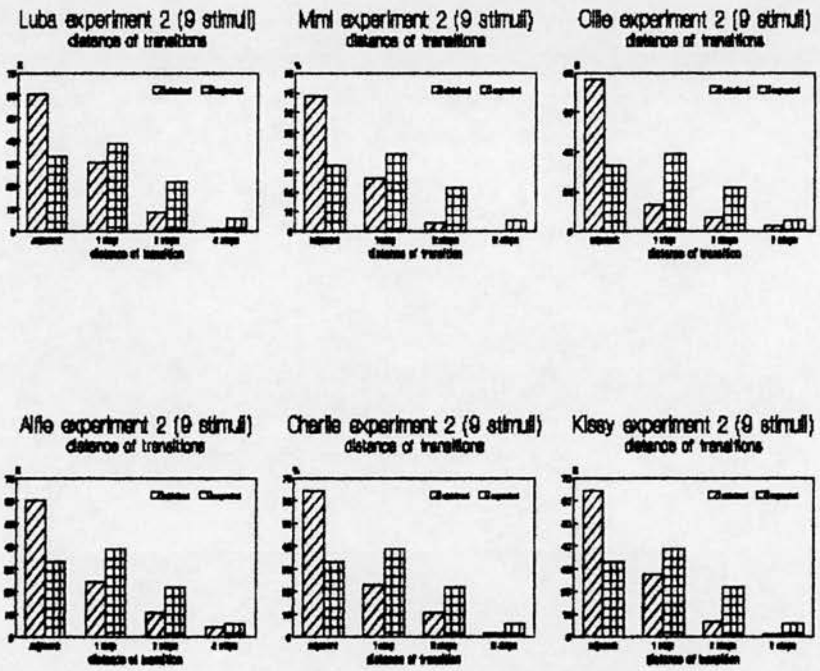



Fig. 4.30. Individual monkeys. Obtained and expected percentages of transitions for each distance (calculated in terms of items interposed between two locations on the screen).

EVIDENCE FOR IMPROVEMENT WITH PRACTICE

As for Experiment 1, a Page's trend test was performed for each subject, in each condition, on data divided in vincent sixths, to assess if performance improved across successive periods of testing.

In all the sub-conditions none of the trends were significant. To exemplify the type of curves obtained from the analysis, Fig. 4.31 shows the combined performance of all the monkeys in the different vincent sixth of the 9 stimuli sub-condition.

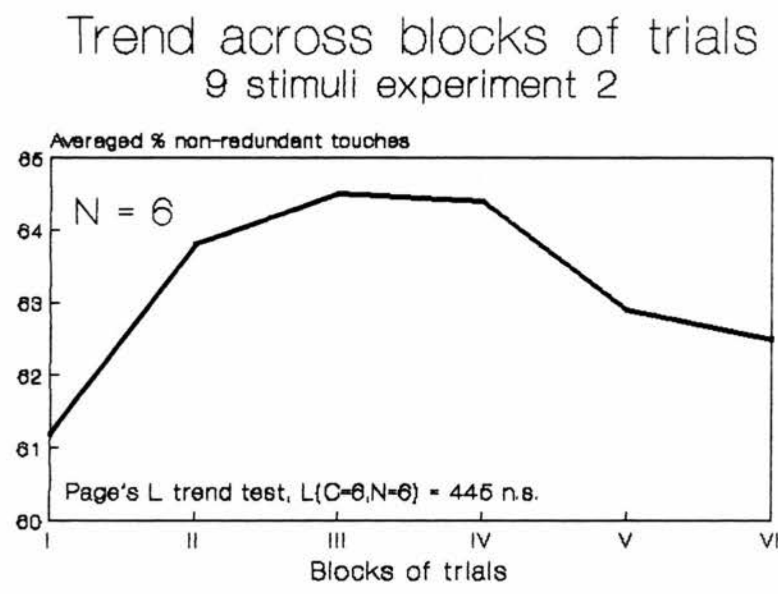



Fig. 4.31. Percentage of non-redundant touches. On the horizontal axis the different vincent sixth, in which the global distribution was divided, are reported. On the vertical axis the percentage of non-redundant touches observed in each block of trials is reported.

From individual performances (shown in Fig. 10) it can be noted that only one monkey (Ch) showed a positive trend across blocks of trials.

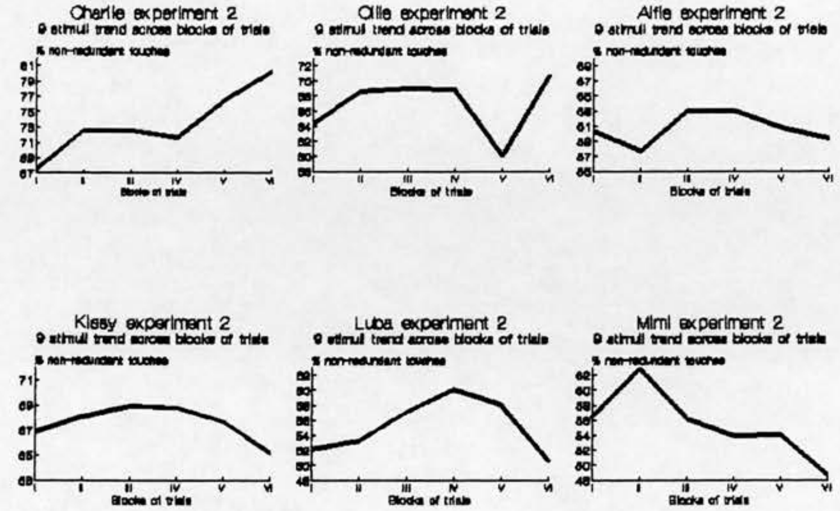



Fig. 4.32 Individual performances observed in the 9 stimuli condition. See Caption of Fig. 4.31 for explanations.

Thus, it seems as if the explicit feedback, had a strong effect on the performance as soon as it was introduced but, further task practice, even, when accompanied by a progressively more severe criterion for reward, did not find a parallel in the performance of the subjects.

This conclusion was supported by an analysis of the relationship between percentages of non-redundant touches and the criterional mastery level required at different points of testing. The results are shown in Fig. 4.33.

Since the criterion was established for each monkey on the basis of its performance, the criterional levels were divided into three groups for the analysis. The first group included the lower group of criterional levels, class two the intermediate, and class 3 the most severe ones (for example, for a subject whose criterion was reduced from 12 to 3 redundant-touches allowed for reward, group 1 would include the performance obtained when the criterion ranged from 12 to 10 redundant touches, group 2 the performance when the criterion was comprised between 9 and 7, and group 3 when the criterion ranged from 6 to 4 redundant moves).

Performance in the feedback condition followed the trend of the cut-off point for reward in only two subjects (overall results non-significant), as shown in Fig. 4.33.

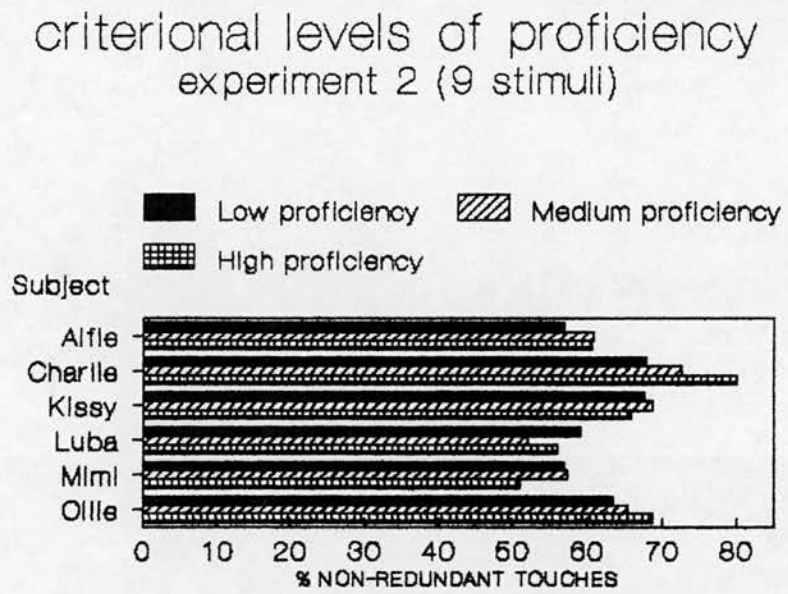
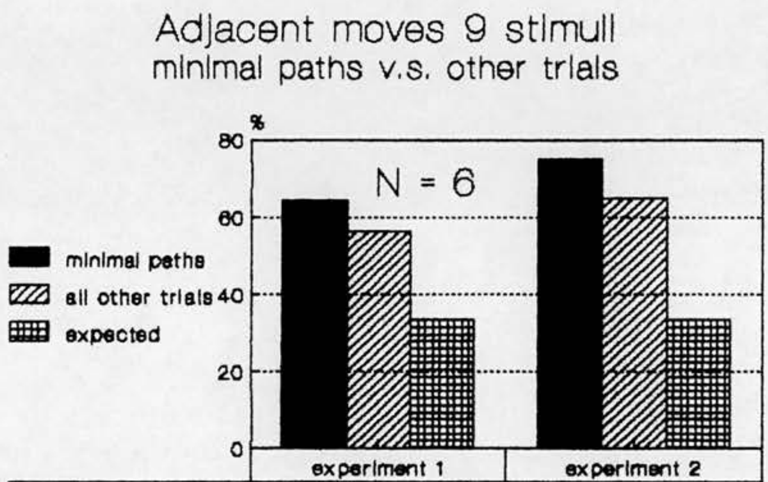



Fig. 4.33. The three groups of criterional mastery levels against which the percentages of non-redundant touches were plotted, are described as low, medium and high proficiency (see text for explanations).

RELATIONSHIP BETWEEN USE OF ADJACENT MOVES AND PERFORMANCE

The only evidence for a correlation between percentages of adjacent moves and performance was obtained comparing the difference between the adjacent moves observed in minimal path trials and those observed in all the other trials containing at least one surplus move. From Fig. 4.34, it can be observed that the percentage of adjacent moves was larger when minimal paths were compared with the other trials, in both experiment 1 and 2.




| | experiment 1 | experiment 2 |
|---|---|---|
| minimal paths | 64.27 | 74.87 |
| all other trials | 56.39 | 64.93 |
| expected | 33.33 | 33.33 |

**Fig. 4.34. Observed and expected percentages of adjacent moves in minimal and non-minimal path trials.**

Only for Alfie were adjacent moves correlated with performance in Phase III of experiment 1 (N.B. Alfie shows a negative trend in this condition).

## Discussion

Explicit negative feedback proved effective in increasing the economy of search of the subjects in all the conditions featuring 4 icons or more. Hence, there is evidence in support of the hypothesis that the ceiling effect observed in experiment one was due to a difficulty in differentiating between good and bad runs once the subjects had spontaneously reduced the frequency of inefficient searches in the course of practice.

Further support in favour of the economy data/reduction hypothesis was given by the finding that economic searches were accompanied by an increase in the use of spatial strategies such as moving to adjacent icons in successive moves. However, even with explicit feedback, monkeys do not spontaneously develop such principled searches which allow 4 and half year old children to achieve a maximally economic search efficiency.

## 4.4 Conclusions

In summary, with the experiments described in this chapter, the following things seem to have been achieved.

Touch screen based procedures have been tested with capuchin monkeys. The monkeys readily adapted to the tasks and the apparatus. The system proved effective in allowing a precise measure of sequential patterns of responses and associated time variables.

Using touch screen based tasks, it proved possible to overcome the limitations of the WGTA. The paradigms and the apparatus afforded protracted testing, giving the opportunity to collect an affluent data base and to evaluate changes in expertise over time.

It was found that, practicing the task, the subjects spontaneously reduced the number of surplus moves, advancing towards the maximal level of behaviour fitness, objectively defined as the ratio between number of items to be explored and total number of touches observed.

As a comparative note, pigeons that are currently being tested with the same apparatus and procedures (Dickinson, in preparation) show a very different behavioural profile. The fact itself that the versatility of the paradigms permits the collection of data on birds too, in my opinion supports the idea that the method is a comparative tool of extreme value. Pigeons tested on search spaces up to 5 icons show only little spontaneous improvement on sets with 3 icons. In sets of 4 and 5 icons their economy of search remains constant even when they are given protracted exposure to the task and selective feedback for economy (omission of reward for trials exceeding a particular number of surplus touches). Thus, the ability to self-regulate seems to be an important dimension for cross-species comparison.

Practising the task, the fitness of the behaviour of the monkeys reached similar levels for arrays comprising from 3 to 5 stimuli, albeit more trials were required before a plateau was observed in conditions featuring more stimuli.

Nevertheless, when the possibility to improve with practice was prevented, the length of the serial production required for the subjects, (objectively defined as the number of icons presented on the screen), proved to be an important predictor of the performance of the subjects. In fact, when different conditions (featuring different number of stimuli) were embedded in the same testing session, an inverse relationship between performance and number of stimuli was found.

A complementary time analysis indicated that, when more icons were presented, the subjects where slower in their responses. This suggested that more sophisticated organizational devices were deployed when a longer serial production was required. A measurement of latencies shown by species (such as pigeons) which do not improve their economy of search with practice would be of great interest. An absence of the time related phenomena observed in monkeys would give indirectly support to the hypothesis that these organizational factors sustained self-regulation and the increase in the economy of search.

In fact, principled sequential patterns (strongly contrasting with a random walk through the configuration to be explored) were detected in the behaviour of the monkeys. The trajectory through the icons was constrained by the use of preferred starting locations, the high probability of moving to a spatially adjacent item after each choice, and possibly the use of preferred directions of travel.

Since a parallel study with children showed a cross correlation between constrained sequential patterns and task fitness, the strategic factors detected in the behaviour of the monkeys was interpreted as a tendency towards economy. Consistently with a data/reduction hypothesis, I proposed that the subjects were using search strategies (aimed to reduce the number of items to be tracked), in order to avoid an excessive number of redundant moves. The fact that the subjects failed to achieve the maximal fitness shown by 4 and half year old children, was taken as a validation of the idea that, when the production of long sequences of responses is required, performance cannot be based exclusively on brute memory. On the contrary, very strong organizational devices are needed for high levels of economy. If these devices are not used in a fully principled way, the system does not reach maximal economy.

From the results of experiment 2, it was seen that by giving the subjects explicit feedback at the end of each trial, their behaviour shifted further towards maximal fitness, before showing a definite ceiling effect.

The improvement produced by the explicit feedback was accompanied by an increase in the use of some forms of constraints. In particular, it was shown that the subjects were spontaneously using more adjacent moves in situations where an explicit feedback of behaviour fitness was provided. Moreover, it was shown that adjacent moves were used more in the most fit trials (the minimal paths) as compared with those trials which contained at least one redundant move). This phenomenon gave strong support for the idea that the level of fitness to the task was effectively correlated with the extent to which constraints were exploited in a principled fashion.

Nevertheless, it proved impossible to find a clear correlation between the change in fitness to the task and the use of particular spatial constraints. This was due to two main reasons. Firstly, some of the spatial constraints afforded by the configurations featured in these experiments did not allow a precise measure of their use. Especially considering that they were used on probabilistic basis and not in such a fully principled way as observed for 4 and half year old children. Secondly, only the use of spatial (vectorised) strategies is transparent in these sorts of tasks and almost exclusively in the 9 stimuli condition, where the configuration was highly constrained and remained constant across different trials. The conjecture can be made that when the spatial affordability of the set (such as its linear arrangement) is not fully perceived at the outset and transduced in search trajectories (as for the 4 and half year old children) the combinatorial space of alternative paths (within the search space) is to high to induct the potential response templates, except at the local level illustrated by adjacent moves.

However, it was clear that the procedures and the apparatus were promising and that the subjects showed incentive towards self-regulation, (in terms of a search for sources of constraints through the search space) to achieve economy. The general strategy of moving from weaker to stronger forms of external intervention (i.e. selective feedback) proved extremely appropriate given that it allowed the observation of a tendency towards a completely spontaneous form of regulation. Obviously, if selective feedback was used from the outset, the detection of such a tendency would have been impossible.

Thus, given that some basic parameters had been set, the way was open for new tasks, which afforded more strategic opportunities. The specific question posed was whether or not the subjects would have spontaneously shown a tendency to use forms of constraints other than spatial. One obvious candidate (as proposed by McGonigle, 1987) seemed to be the superimposition of classificatory schemes over a search space which allowed the subdivision into different chunks according to different features shared by groups of icons to be explored.

Since the strategy of implementing first untutored versions of the task (no selective feedback) proved effective here, it was adopted for the rest of the program. The experiment on spontaneous classification, will be described in the next chapter.

# CHAPTER V

## AN EXPERIMENT ON SPONTANEOUS CLASSIFICATION

### 5.1 Introduction

In the light of the conclusions mentioned at the end of the previous chapter, the experiment that will be described here aimed to assess if monkeys would have spontaneously imposed categorisation schemes over a set of icons varying according to multidimensional features (see McGonigle, 1987).

Very few researches have focused on spontaneous classification in capuchin monkeys. Some data derive from a developmental and comparative study conducted by Antinucci and colleagues (Antinucci, 1989). Inspired by the Piagetian tradition, their general methodology involved leaving the subject free to interact with sets of randomly scattered objects. The features of the objects were varied according to a 4 x 4 classification matrix (4 forms x 4 materials).

In one first experiment (Antinucci, 1989; Natale, 1989) the differential application of four manipulatory schemata (mouthing, handling, visual exploration and secondary schemata: actions that involved both the object acted upon and another object or surface) to the different kinds of objects was evaluated.

According to the authors, capuchin monkeys distinguished the four types of objects from one another and this was taken as evidence for classification through action schemata. However, most of the difference was accounted for by the comparison of visual exploration versus secondary schemata. Only two of the four different groups of actions were taken into consideration. The objects maximally differentiated were cups (high frequency of visual exploration but a very low frequency of secondary schemata) and sticks (high frequency of secondary schemata and a very low level of visual exploration).

Although some evidence of spontaneous classification in capuchin monkey emerged from this study, several methodological weakness make the results far from being conclusive.

Firstly, some of the comparisons between the patterns of action schemata on different objects were merely based on the visual inspection of the frequencies and not supported by any statistical analysis. Secondly, one of the authors claims that: "a certain amount of arbitrariness in identifying, equalling as instances of the same class, and, especially, segmenting manipulative events cannot be avoided" (Natale, 1989, p. 148).

Thirdly, the design, claimed to be partially longitudinal and partially cross-sectional, was based on a very small sample of subjects: one cebus was tested at 16, 36 and 48 months, a second was tested two times at 36 and 48 months and a third only at 16 months (however, the results were obtained from data collapsed across ages, since no differences were observed between the age subgroups).

Apart from the methodological shortcomings, it should be noted that classification, as defined in that study, is at its most elementary level i.e. how and to what extent types of objects are assimilated by different action schemata depending on their properties. That is what Inhelder and Piaget (1964) claimed to be a sort of practical classification, somewhat reminiscent of the later definition by use, but not yet comparable to functional classification.

A somewhat more sophisticated form of classification was detected in their second study (Spinozzi and Natale, 1989). This was based on an independent analysis of the videotaped session obtained from the first study. Videotapes were scanned to detect all cases in which 2 or more objects had been grouped together (compositions) and to find out the class properties of the successions of objects manipulated to form each composition (object selecting).

Some evidence for classification was found even here, where the analysis was not based exclusively on the weak criterion of classification by differential use.

However, in this case too, the criteria for establishing instances of classificatory behaviour were far from being strict. In fact, both UNMIXED (objects belonging to the same class) and MIXED (objects belonging to different classes) compositions (as well as selections) were considered instances of classificatory activity, because both were meant to imply that the subjects were taking into account the classes into which the set of objects could be divided.

Furthermore, as far as I understand, to be considered MIXED the compositions and selections did not require that all the objects belonged to different classes. In addition, it is very difficult to find out what the evidence for statistical significance was, given that it is not stated what test was used to evaluate the deviation from a random distribution.

Moreover, the fact that most of the compositions were of two objects only, and that those involving the grouping of more that 3 objects were extremely rare makes the evaluation of the significance of the results even more difficult. Furthermore, most of the compositions were generated by manipulating one object only and those involving the manipulation of two objects never exceeded 50%.

Thus, an independent evaluation of capuchin's spontaneous deployment of classificatory skills (based on richer patterns of responses from a larger sample of subjects) would have provided important information still lacking in the literature.

Apart from that, from the perspective adopted in this thesis, a major deficiency was detected in the studies of Antinucci and colleagues. This is the lack of any task requirement and therefore of any incentive to use classification as a data reducing device.

Classification can easily be seen as a skill which enables the chunking of otherwise heterogeneous material, difficult to perceive or remember (Miller, 1956; McGonigle, 1987).

Developmental psychologists have stressed the role of classification as a data reducing device in memory tasks (for a review see Flavell and Wellman, 1977) and a developmental trend has been observed in the ability to recognize the data reduction properties of a set of items which can be chunked in classes (as opposed to a comparable set of conceptually unrelated items). For example, Moynahan (1973) found that 7 year old children seem able to recognize that serial recall of a linear sequence of coloured blocks is likely to be easier when blocks of the same colours are adjacent rather than randomly placed in the series (a sequence red-red-blue-bue-yellow-yellow would be easier to reproduce from memory than a sequence red-red-blue-yellow-blue-yellow). However, 9 and 10 year old children are significantly more likely than 7 year olds to predict that the set of categorized items would be easier to recall (Moynahan, 1973; Flavell and Wellman, 1977) and 6 year olds encounter difficulties in recognizing the mnemonic advantages of categorical organization, even when explicitly made aware of its presence (Salatas and Flavell, 1976).

Although the developmental studies mentioned above have concentrated on retrieval of stored information from memory (internal search), the adaptive value that similar strategical factors might have when the search space is external is obvious. As stressed by Drozdal and Flavell (1975; see also Flavell and Wellman, 1977) what is usually referred to as <u>knowledge about memory</u> may itself be too narrow a designation, since some of the knowledge one might wish to talk about in this connection may not be about memory as it is conventionally understood. It might, for example consist of <u>knowledge about how to search the external world intelligently</u>, a form of knowledge that also undergoes a marked development with age (Drozdal & Flavell, 1975; Flavell and Wellman, 1977).

It is in this latter sense that the role of categorisation as a data reducing device in the sort of tasks introduced in the previous chapter becomes transparent. In searching exhaustively a set of icons on a touch screen monitor the subjects might find the possibility of chunking them useful. In doing so the search of items identifiable only by their (individual) spatial location would be restricted to those belonging to one class. Once the class is exhausted the subject would know that reiteration on items of that sort is to be avoided. If search behaviour is controlled by this sort of chunking, the number of individual items to be remembered would then be reduced to the number of classes in which the search space is divided.

The experiment that follows aimed to assay whether the search behaviour of cebus monkeys would show some evidence of spontaneous classification, and (if so) whether the monkeys made a strategic use of such a competence by exploring the set in a principled way i.e. exhausting one class before moving to the next.

## 5.2 Material and Methods

Subjects and Apparatus. The same subjects and apparatus used for the experiments described in Chapter IV were employed for this experiment.

Design. The experiment featured three experimental phases and a control phase. Phase 1 aimed to assess whether subjects would have spontaneously imposed categorisation schemes over a set of icons divided in two disjunctive classes. In phase 2 the multiple features of the icons were permutated to avoid possible bias due to the salience of a particular combination of features defining a class. In phase 3 the trade-off between class size number of classes was manipulated to assess its effects on classification. Finally, with a series of control trials featuring identical icons, the possibility of classifying was prevented in order to compare the level of economy of search shown in "chunkable" and "non-chunkable" sets of icons.

Procedure

General procedure.

As in the set of experiments presented in the previous chapter, the task was to search exhaustively a set of icons presented on a touch screen.

The subject was left free to choose his own trajectory through the search space. However, as in Experiment 2 of chapter IV, in each trial a criterion of non-redundant touches had to be satisfied in order to obtain a peanut as reward. When the subject failed to meet the criterion a white screen was presented for 5sec, followed by an intertrial interval of 15 seconds.

The criterional level of economy was set, for each monkey, on the basis of its performance on the final phase of the previous experiment, so that the 20% of the trials remained unrewarded. Then, the criterion was progressively tightened according to the level of performance of the subject to keep the 20% of the trials unrewarded.

In order to reduce the spatial constraints afforded by the search space, the location of the icons on the screen was selected at random for each trial between all the possible locations defined by a 4 x 4 matrix (instead of 3 x 3 as in the previous experiments).

Each experimental session featured the presentation of 50 trials.

Phase 1

The set of icons presented on the screen was divided in two groups differing multidimensionally and disjunctively. In details, a set of 8 icons was divided into a sub-set of 4 Small Red Circles (SRC) and a sub-set of 4 Large White open figures which, for their shape resembling the letter I, will be referred as Is (LWIs). A schematic example of such a configuration is provided in Fig. 5.1

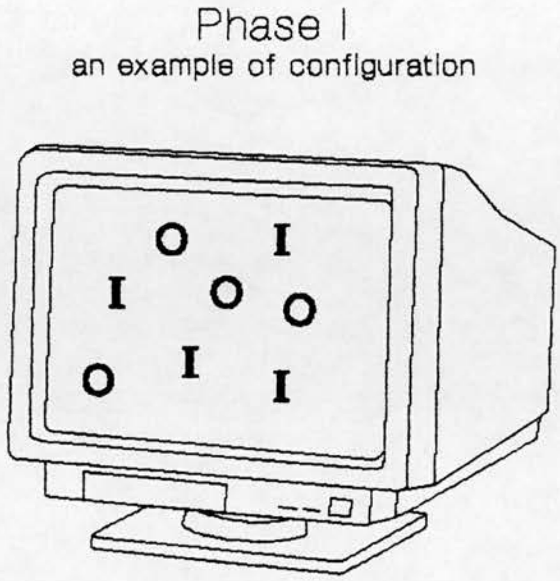



Fig. 5.1

Phase 2

Each daily testing session featured the presentation of one of the four different conditions obtained permutating the features of the icons of phase 1, in ways that preserved the disjunctiveness of the two classes of 4 items each. In greater detail, condition I featured the presentation of SRCs and LWIs; condition II Large Red Circles (LRCs) and Small White "Is" (SWIs); condition III Small White Circles (SWCs) and Large Red "Is" (LRIs); and condition IV Large White Circles (LWCs) and Small Red "Is" (SRIs).
The presentation of the conditions was randomized according to a Latin square design.

Phase 3

In this last Phase the trade-off between number of classes and class size was modified. Thus, the 8 items set featured 4 classes (referred as A, B, C and D) of two stimuli each. In greater detail, class A was composed of two empty brown triangles; class B of two solid yellow crosses; class C of two solid magenta egg-timers; and class D of two patterned cyan hexagons. An example of such a configuration is presented in Fig. 5.2.

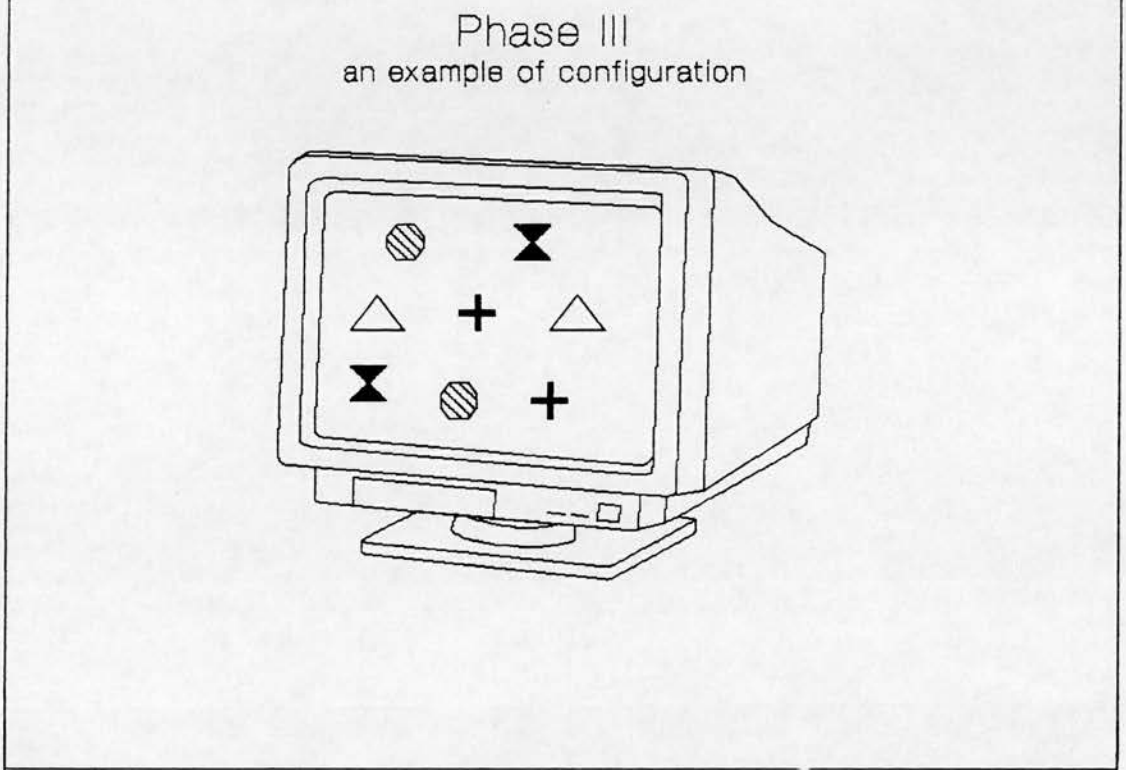



Fig. 5.2

Control

The procedure being the same as that featured for the previous phase, a set of 8 identical icons (green squares) was presented to the subjects.

## 5.3 Data Analysis and Results

### Evidence for classification

Data Analysis.

This analysis was based on the assumption that a subject moving at random throughout the search space (or conforming to any sort of spatial strategy), would score a number of pairs of successive moves within the same class and between different classes equal to the probability expected by a random trajectory across the search space.[1] Thus, for each monkey, a $chi^2$ was performed on the transitions from and to each class (e.g. when the search space was divided in two classes, A and B, two $chi^2$ were performed: one comparing the obtained number of transitions from A to A and from A to B with the expected values and one comparing the obtained number of transitions from B to B and from B to A with the expected values). This analysis made it possible to detect any systematic use of the non-spatial constraints afforded by the set of icons. In fact, either a tendency to stay within the same class in successive touches or to shift to a different class at every touch would lead to statistical significance.

Results.

A summary of the results obtained in the different phases is given in Table 5.1.

1 It must be kept in mind that the spatial configuration of the icons was randomly changed for each trial.

Table 5.1 Probabilities of $chi^2$ performed on transitions from each of the classes presented in the different phases. Repeats excluded from the analysis. In brackets the results are reported of a visual inspection of the frequencies contained in the cells of each significant $chi^2$ i.e. (=) = transitions within the same class; (#) = transitions between different classes (see text for further explanation).

| Subj | Phase1 | Phase2 | | | | Phase3 | |
|---|---|---|---|---|---|---|---|
| | | CondI | CondII | CondIII | CondIV | | |
| Al | p<.001(#) | ns | p<.01(#) | ns | ns | p<.02 | p<.01 |
| | p<.001(=) | ns | p<.001(=) | ns | ns | ns | p<.001 |
| Ch | ns | p<.001(#) | ns | ns | ns | p<.05 | ns |
| | ns | ns | ns | ns | ns | ns | ns |
| Ki | ns | ns | ns | ns | ns | ns | ns |
| | ns | ns | ns | ns | ns | ns | ns |
| Lu | - | ns | ns | p<.01(#) | ns | ns | ns |
| | - | ns | ns | p<.02(=) | ns | ns | ns |
| Mi | ns | p<.01(#) | p<.001(=) | p<.01(#) | p<.01(=) | ns | ns |
| | p<.001(=) | ns | ns | ns | p<.02(#) | ns | ns |
| Ol | p<.001(#) | ns | p<.001(=) | p<.001(#) | p<.001(=) | ns | p<.02 |
| | p<.001(=) | ns | p<.001(#) | p<.001(=) | p<.001(#) | p<.05 | ns |

Phase 1 (one condition, two disjunctive classes: SRC and LWI)
Significant results were found for three subjects (Al, Ol, Mi) out of five (Lu was not tested on this phase). Whilst all the $chi^2$ performed for Al and Ol were significant ($p <.001$), for Mi only the $chi^2$ performed on transitions from LWI was significant ($p <.001$).

A visual inspection of the relative frequencies of transitions showed that Al selected LWIs after having touched either a SRC or a LWI; the opposite tendency was found for Ol that moved to SRCs either from another SRC or a LWI; Mi showed a tendency to make successive moves within the class LWI.

Phase 2 (four conditions, each featuring two disjunctive classes, i.e. condition I: SRC-LWI; condition II: LRC-SWI; condition III: SWC-LRI; condition IV: LWC-SRI)
For all the subjects but one (Ki), at least one significant $chi^2$ was found in each condition. The individual results showed:
2 significant $chi^2$ ($p <.001$) for Al in condition II; 1 significant $chi^2$ ($p <.001$) for Ch in condition I; 2 significant $chi^2$ ($p <.01$ and $P <.02$, respectively) for Lu in condition III; 2 significant $chi^2$ ($p <.01$ and $p <.02$ respectively) in condition IV and 1 significant $chi^2$ in all other conditions ($p <.001$ in condition II and $p <.01$ in conditions I, III, and IV) for Mi; and 2 significant $chi^2$ for Ol ($p <.001$) in conditions II, III, and IV.

A visual inspection of the frequencies contained in the different cells of significant $chi^2$ showed that Al (condition II) selected LRCs, either after having touched a SWI or another LRC;
Ch (condition I) shifted from SRCs to LWIs; Lu (condition III) moved on SRIs either after having touched another SRI or a LWC; Mi shifted from SRCs to LWIs (condition I), remained within LRCs (condition II), shifted from LRIs to SWCs (condition III), and selected LWCs either after having touched another LWC or a SRI (condition IV); Ol selected LRCs either after having touched another LRC or a SWI (condition II), selected LRIs either after SWCs other LRIs (condition III), and chose LWCs, either after SRIs or other LWCs (condition IV);

Phase 3 (4 classes)

The analysis of this Phase involved four $chi^2$ for each subject. For all the subjects but one (Mi) at least one $chi^2$ was significant.

In details, three $chi^2$ were significant for Al ($p <.02$, $p <.01$, and $p <.001$ respectively). Two $chi^2$ were significant for Ol ($p <.02$ and $p <.05$, respectively). Only one $chi^2$ was significant for Ch ($p <.05$), Ki ($p <.05$) and Lu ($p <.01$).[2]

Evidence for strategic use of classification

Data Analysis

To be an effective tool in exhaustive search, classification should be used to reduce the number of icons to the number of chunks represented by the classes. Therefore, a proficient searcher should be expected to stay within the same class until it is exhausted, and only then, to shift to the next one, keeping track of those already explored (instead of keeping track of each of the single icons touched). Thus, a z test (Binomial for large samples, one tailed, corrected for continuity) was used to compare the expected chance probability ($p = .0286$) with the number of trials where one of the classes was exhausted before shifting to the other class. All redundant touches were excluded from the analysis. Data for all the conditions comprising two disjunctive classes were combined.

Results

Three subjects tended to exhaust one of the two classes within the first four non-redundant touches: Ol ($z = 3.53$, $p < .001$); Ki ($z = 3.12$, $p <.005$); Mi ($z = 1.65$, $p <.05$).

Evidence for stimulus preference

Data Analysis

---

2 In this phase (for the large number of possible transitions) an assay of the relative contribution of particular moves on the basis of a visual inspection of the single cells of the $chi^2$ was impossible.

The analysis of transitions, and especially the visual inspection of the data, showed that all the subjects had a tendency to choose stimuli belonging to a particular class irrespective of the kind of stimulus touched immediately before.

Such a tendency could have biased both the analysis of transitions and the likelihood of exhausting one of the classes before moving to the other.

In order to statistically test the hypothesis that the results reported above could be accounted by stimulus preference, a z test (Binomial test for large samples) was performed on the frequencies of reiterations on each stimulus type. The analysis was conducted on individual data for those conditions where evidence of classification or strategic classification was found.

Results

A summary of the results obtained from this analysis is provided in Table 5.2.

**Table 5.2 Preferred stimuli in conditions where evidence for classification or strategic classification was found. The table does not include Phase 3 (see text for explanation).**

| Subj | Phase1 | Phase2 CondI | CondII | CondIII | CondIV |
|---|---|---|---|---|---|
| Al | LWI (z=8.5, p<.001) | no classification | LRC (z=6.8, p<.001) | no classification | no classification |
| Ch | No classification | no classification | no classification | no classification | no classification |
| Ki | No preference | LWI (z=3.67, p<.001) | no preference | LRI (z=2.12, p<.05) | no preference |
| Lu | - | no classification | no classification | LRI (z=6.6, p<.001) | no classification |
| Mi | LWI (z=5.5, p<.001) | LWI (z=5.5, p<.001) | LRC (z=4.5, p<.001) | SWC (z=4.1, p<.001) | LWC (z=4.7, p<.001) |
| Ol | SRC (z=7.7, p<.001) | LWI (z=10.5, p<.001) | LRC (z=11.5, p<.001) | LRI (z=7.8, p<.001) | SRI (z=7.2, p<.001) |

Phase 1 (one condition, two disjunctive classes: SRC and LWI)
The three subjects which showed evidence for classification in this phase were Al, Ki, Mi and Ol.
Al's frequencies of reiterations on SRC and LWI were 3769 and 4544 respectively ($z = 8.5$ $p < .001$); Ki performed 1773 reiterations on SRC and 1827 on LWI and the z test was not significant; Mi reiterated 1313 times on SRC and 1612 times on LWI ($z = 5.5$ $p < .001$); and Ol showed 2010 reiterations on SRC and 1577 on LWI ($z = 7.2$ $p < .001$). Thus, stimulus preference seem to be likely to explain the behaviour of Al, Mi and Ol but not that of Ki.

Phase 2 (four conditions, each featuring two disjunctive classes
In this Phase all the subjects but one (Ki) showed evidence of classification. The results of the z tests can be summarized as follows:
- Al (two significant $chi^2$ in condition II). The frequencies of reiterations were 1889 on SWI and 2331 on LRC ($z = 6.8$ $p < .001$) and this preference for LRCs can explain the significance of both $chi^2$.
- Ch (one significant $chi^2$ in condition I) reiterated 1003 times on SRC and 1278 times on LWI ($z = 5.7$ $p < .001$) and this preference for SRCs can account for the significance of the $chi^2$.
- Ki (significant z test for trials where a class was exhausted before shifting to the other) reiterated 537 times on SRC and 665 times on LWI ($z = 3.67$ $p < .001$) in condition I; 594 times on LRC and 572 times on SWI in condition II ($z = .614$ ns); 554 times on SWC and 629 times on LRI ($z = 2.12$ $p < .05$) in condition III; 858 times on LWC and 827 times on SRI ($z = .756$ ns). Although the preference for the stimuli belonging to one of the classes in some conditions can have facilitated the fact that Ki exhausted one set before shifting to the other, it is likely that in some occasions she made a strategic use of classification.
- Lu (two significant $chi^2$ in condition III) showed 2184 reiterations on SWC and 2654 on LRI ($z = 6.6$ $p < .001$). This preference for the LRIs can explain the significance of both the $chi^2$ .
- Mi (one significant $chi^2$ in condition I, II and III; and two significant $chi^2$ in condition IV) reiterated 1681 times on SRC and

1837 on LWI ($z = 5.5$ $p < .001$) in condition I; 2016 times on LRC and 1738 times on SWI ($z = 4.5$ $p < .001$) in condition II; 1927 times on LRI and 2190 times on SWC ($z = 4.1$ $p < .001$) in condition III; 1727 times on LWC and 1458 times on SRI ($z = 4.7$ $p < .001$) in condition IV. The preference shown for one of the stimulus type in each of the conditions can explain all the $chi^2$ found significant for this subject.

- Ol (two significant $chi^2$ in conditions II, III, and IV; and a significant z test for trials where a class was exhausted before shifting to the other). Ol showed 1114 reiterations SRC and 1671 reiterations on LWI ($z = 10.5$, $p < .001$) in condition I; 1604 reiterations on LRC and 1015 on SWI ($z = 11.5$ $p < .001$) in condition II; 1157 reiterations on SWC and 1567 on LRI ($z = 7.8$ $p < .001$) in condition III; 1577 reiterations on LWC and 2010 reiterations on SRI ($z = 7.2$ $p < .001$) in condition IV. For this subject it is evident that a stimulus preference can account for all the statistical significance found in all previous analyses.

Phase 3 (4 classes)

In this phase 4 different stimulus types were presented. For this reason, instead of a binomial test, a $chi^2$ goodness-of-fit was performed on the reiterations on each stimulus type. All the $chi^2$ were found significant: Al ($chi^2 = 58.7$, df = 3 p .001); Ch ($chi^2 = 24.0$, df = 3 p .001); Ki ($chi^2 = 9.7$ df = 3 p .05); Lu ($chi^2 = 25.6$, df = 3 p .001); Ol ($chi^2 = 49.7$, df = 3 p .001). This result suggests that classification was based on stimulus preference in this phase too.

Comparison with control condition (set of identical icons).

The results presented above show that for most of the subjects the presence of non-spatial constraints lead to reiterations on icons belonging to a particular class. None of the subjects (with the possible exception of Ki) made a strategic use of classification.

In this situation a decrement might be expected in the number of reiterations if the non-spatial constraints are withdrawn from the search space. Table 5.3 shows the overall percentage of non-redundant touches made in the classification phases and in the control condition. There it can be observed that the percentages of non-redundant touches of five subjects was higher in the control condition than in any of the classification phases.

Table 5.3 Total number of touches (N) and percentages of non-redundant touches obtained in Phase1, Phase2 (all conditions combined) and in the control condition.

| Subj | Phase1 | Phase2 | Phase3 | Control |
|---|---|---|---|---|
| Al | N=954<br>48.1% | N=2407<br>54.2% | N=495<br>53.1% | N=181<br>63.3% |
| Ch | N=591<br>62.8% | N=2389<br>68.0% | N=575<br>68.5% | N=229<br>79.9% |
| Ki | N=795<br>63.9% | N=1732<br>72.6% | N=564<br>80.8% | N=135<br>78.1% |
| Lu | -<br>- | N=2316<br>51.5% | N=965<br>61.5% | N=150<br>72.9% |
| Mi | N=317<br>46.4% | N=1968<br>51.9% | N=398<br>67.2% | N=97<br>73.8% |
| Ol | N=564<br>55.7% | N=2355<br>63.7% | N=742<br>68.1% | N=207<br>74.3% |

The only exception is again Ki, whose percentage of non-redundant touches was higher in Phase 3. As mentioned above, in Phase 3 Ki showed one significant $chi^2$ out of four. However, a visual inspection of the cells of this $chi^2$ shows that the significance can be accounted to by the systematic shifting from stimulus type B to stimulus type D, and this tendency per se can hardly account for an increase in performance.

## 5.4 Conclusions

Overall, the results point to a lack of strategic use of classification as a spontaneous data reducing strategy in search. In other words, monkeys failed to spontaneously develop the strategy of comparing different classes and exhaust each of them at specified points of the serial production. However, there is evidence for a spontaneous tendency to take into account the non-spatial constraints afforded by the search space. This tendency was expressed as a non-random exploration of the search-space, where icons belonging to a particular class were selected more frequently than those belonging to another class.

Thus, there is evidence for a spontaneous tendency to pick up potentially strategical information, without however using it efficiently. In the literature on human cognitive development, a related phenomenon is described by Sophian and Wellman (1987) who found that 3 year old children search for potentially strategic information before they are able to implement it according to the task requirements. However, at later stages of development, children become able to use strategic information according to task requirements. Thus, there remains an open question as to why adult monkeys do not make a strategic use of classification skills.

It is possible that classification in monkeys is present only as a rudimentary skill and lacks those forms of control which would allow its strategical and flexible use. This would not necessarily mean that sophisticated forms of control are precluded to the subjects by their taxonomical status. A number of alternative interpretations of the results of the present study must be ruled out before such a conclusion can be drawn.

First, an aspect to be taken into account is the novelty of the task faced by the monkeys. The subjects entered the present experiments after a long practice with search-spaces which did not afford the use of strategies other than spatial ones.
As pointed out by Brown and Deloache (1978) in the context of human development, and by Luger and Stern (1990) in Artificial Intelligence, the novelty of a situation plays an important role in the failure to recognize critical aspects of the task at hand. Novices, not knowing much about either their competences on a new task or the strategies that can help to perform it in a more efficient way, lack the information which would allow them to self-regulate

their actions. Moreover, an important aspect underpining the ability to self-regulate is the possibility of evaluating the efficiency of a particular strategy by predicting it (for example realising that a strategy that proved effective in a different context might be of some use in the situation at hand). It is thus important to have previous experience with a task that shares some commonality with the current one.

A second important process on which self-regulation is dependent, is the monitoring (cost evaluation) of the effects of naturally occurring tendencies in order to develop more control over them. In this case, it is important to have an effective differential feedback on "conjectures" generated spontaneously. In the present experiment, (and as stressed before in order to follow the strategy of avoiding the use of rigid training schedules from the outset) the procedure was not designed to provide a strong incentive to classify strategically. Thus, the subjects might have found a satisfactory trade-off between the cognitive strain that strategical classification would have demanded and the cost attached to a surplus of redundant moves, before having the opportunity to test the efficiency of strategical classification.

On the other hand, it is still possible that the ability to compare different classes and exhaust each of them at specified points of the serial production is beyond cognitive skills of monkeys. This ability, in turn, can be decomposed in two complementary components. One is the ability to make an exhaustive and economic search within a given class. It has been documented that young children find difficulties in doing so. In a study by Kobasigawa (1977), first graders who spontaneously used an available category cue to retrieve items stored in memory, still recalled fewer items per category than third graders did. In other words even when they though about using the retrieval cues, the younger children failed to conduct an exhaustive search for the items associated with each cue. The other component is the ability to keep track of the classes already exhausted and so avoid backwards errors in the sequence of classes to be explored.[3]

Finally a further alternative, is that the apparent failure of the subjects to classify strategically can be due to a deliberate rejection of such a strategy, intra-cognitively costly, in favour of spatial strategies still affordable. In fact, also in the context of developmental psychology have been identified a number of factors which may attenuate the empirical evidence of an ideal relationship between judgements about the usefulness and a particular strategy and its use. The first is obviously the fact that the subject might think that something else might be better in the situation at hand (Flavell and Wellman, 1977). In our case it would not be at all surprising that subjects previously over-exposed to a search space affording only spatial constraints would continue to use (to a certain degree) the strategies which paid off in he past. A second factor is the possibility that a child may have enough knowledge to judge that categorisation would be a good strategy if asked about it, but not enough to think to use it spontaneously (Flavell and Wellman, 1977). Similarly, it was still open the possibilities that monkeys would have started to classification strategically only if prompted to do so.

For all these reasons, it was necessary to devise a set of conclusive experiments to assess the relative contribution of these different factors to the apparent failure of the subjects in using classification in a principled and strategic fashion. These experiments will be presented in the next chapter.

---

3 Note that what has been referred to here as a stimulus preference can be an expression of these sorts of errors. For example, in a situation where 4 classes A, B, C, and D are presented, a subject exhausting class A and class B, and then going back to A before exhausting C, and so on, would (apparently) show a strong preference for stimuli of type A, although his behaviour could be the expression of a deficiency in ordering A, B, C and D. The same argument applies for a situation where the subject fails to recognize when a particular class has been exhausted and just repeats its search again within the same class.

# CHAPTER VI

## SERIALLY ORDERED CLASSIFICATION

### 6.1 Introduction

From the results of the experiment described in Chapter V, it emerged that monkeys did not spontaneously deploy categorisation skills as a data reducing device in search. Nevertheless, when the search space afforded categorical organization, their searches were not based exclusively on spatial strategies. The subjects were able to detect the non-spatial affordability of the search space (as shown by a non-random distribution of choices among the different classes into which the search space was divided) even if they did not use them in a principled way to control search behaviour. Thus, the monkeys apparently lack the ability to adapt classification skills to the particular task at hand.

In order to be used as a data reducing device in the search tasks that we are dealing with, classification must be deployed according to two independent although related principles. The first is the prescription to perform exhaustive and non-redundant searches within classes. In other words, the subject should confine the search to the icons belonging to one particular class and keep track of the items touched (identifiable by their spatial location). Doing so it can avoid both to reiterate and to exit the class before all the similar items have been touched. The second is the prescription to avoid reiterations between classes. In other words the subject must keep track of the classes already explored and avoid reiterations on items of particular kinds (i.e. identifiable by their features other than spatial position).

A failure to conform to either (or both) of these two principles would produce a classificatory behaviour not fitted for the task and a profile similar to the one observed in the experiment on spontaneous classification.

A number of possible interpretations of the performance of the monkeys were proposed at the end of the previous chapter. They can be summarized as follows: 1) monkeys possess only rudimentary classification skills which do not enable the ordering of multiple classes in a serial production; 2) the fact that principled strategies based on categorisation skills were not deployed systematically prevented the subjects from evaluating their benefits (in term of reduction of both intra- and extra-cognitive costs associated with search); 3) having had a protracted exposure to search spaces which only afforded spatial constraints, subjects continued to rely (to a certain degree) on strategies which had proved effective in the past. Being incompatible with a classificatory principle, these might have contaminated the potentially strategical search behaviour of the subjects.

However, the experiments presented in the previous chapter were a necessary step to be undertaken in order to implement the strategy followed in this entire study, namely to start with unsupervised versions of a task and only afterwards (on the basis of the results obtained) provide explicit instructions with selective feedback procedures.

The experiment described here was aimed to disambiguate the alternative interpretations proposed for the unsupervised version of the classification task. Once it had been assessed that the subjects were spontaneously differentiating the different icons presented (preferring one type over another), it was a matter of empirical investigation to assess whether they possessed the competence to learn how to order the different classes in a sequence of responses.

Thus, in order to clarify the issues mentioned above, a paradigm proposed by McGonigle (1994) was implemented as a hybridization of procedures based on drilled serial learning and free classification. The new paradigm, was designed to provide from the outset the information that it was necessary to impose a sequential order to the different icons presented on the screen. This ordering requirement is conveyed in a training phase, during which the subject learns a particular sequence containing a single icon for each ordinal step. This training phase is similar to the classical experiments on serial learning (e.g. McGonigle, 1987; D'Amato and Colombo, 1988; Terrace, 1987).

After the training phase, the subject is faced with different versions of the learned sequence, this time containing multiple identical icons for each ordinal step. If the subject shows a high level of economy in searching these sets, the search space can be enlarged by adding new members to the different categories.

This procedure gives the subjects the opportunity to fully appreciate the data-reduction benefits (if any) deriving from the use of classification in search. Then, the following phases (featuring the enlargement of the search space), would allow an assay of whether they recognize the usefulness of the strategy (learned in the previous phase) and would export it to a search space structurally similar to the previous one (because organized on the basis of the same categories).

Each successive phase, in fact, can be considered as a transfer test because it makes it possible to assess to what extent the subject generalizes (or, in other words, spontaneously classifies) the ordinal position of a given icon (or class of icons) contained in the training set to all its duplicates included in the test sequence.

Moreover, in order to learn a search strategy based on the serial ordering of items according to their class membership, the subject must refrain from using spatial constraints such as adjacency or linear organization. In fact, spatial and categorical strategies would be fully incompatible (unless spatial strategies are confined to searches within a particular class). In order to prevent a potentially satisfactory trade-off between intra-cognitive and extra-cognitive costs, a principle of maximal efficiency (no reiterations allowed) is required to gain the reward, both in training and in the generalization phase.

In addition to the disambiguation of the issues which emerged from the experiment described in the previous chapter, the new paradigm allows the collection of further interesting data. In fact, the procedure permits the manipulation of the generalization conditions for a given sequence length in order to evaluate the effects of the composition of the transfer sequence on generalization (such as the trade-off between class size and number of classes or the relationship between class size and ordinal position of the class). The search space can also be organized in hierarchically nested sets of classes. Some of these further procedures were also tested in the present experiment.

An experiment conducted on children, using an analogous procedure (McGonigle and Jaswall, 1993) provides template data on which the behaviour of the monkeys can be mapped. Five year old children showed evidence for spontaneous classification starting from series of four items grouped in two classes up to sets of fifteen items organized in five classes of three items each. The fact that most of the children were able to retain lists of such an impressive length (as compared with the magical number 7 +/- 2 described by Miller, 1956 as the working memory span for human adults), searching almost errorless from the outset, strongly supports the hypothesis of the important role of classification as a data reducing device.

Moreover, the pattern of the few errors observed in children showed strong regularities. On the basis of this principled error space it was possible to advance some interpretative hypothesis of the psychological mechanisms underpining performance. For example, errors typically tended to occur at class boundaries and reiteration errors within a class were extremely rare. This strongly suggested that the subjects had chunked the list in classes and that if difficulties arose, they were chiefly due to the demands posed by retaining the order in which the classes had to be explored. When the class ordering was well represented the demands of keeping track of the items <u>within</u> a class were minimal.

The paradigm, thus, seems to provide a data set of qualitatively different behaviours, affording rich interpretative possibilities (in terms of either success or failure) of potential comparative relevance. In other words (as pointed out by McGonigle, 1994) it was hoped that this hybrid paradigm would have combined the best of the arbitrary and highly supervised serial learning procedures with the free choice procedures that have characterized this study so far. This experiment conducted on monkey subjects was therefore also intended as an evaluation of the potentialities of the paradigm.

## 6.2 Design

The experiment involved a pretraining stage and 7 following phases. Their order of administration proceeded from 1 to 7. Phase 1 and 3, were training phases in which ordered sequences of 2 and 3 stimuli respectively were taught to the subjects. Phase 2 and 4 were the corresponding generalization phases for phases 1 and 3, i.e. the sequences of the training phases were multiplexed (adding a new item at each ordinal node). Phase 5 and phase 6 aimed to evaluate the role of class size per ordinal position and featured the presentation of conditions with a variable number of icons at each ordinal node of the sequence. Phase 7 was a generalization phase which involved the multiplexing of a three ordinal step sequence to form a 9 item search space. The design of each phase was informed by the results obtained from the previous phases. Thus, in order to make the overall plan of the experiment intelligible, the details of the procedure for each phase will be presented together with a discussion of the results.

## 6.3 Materials

Subjects. The six monkeys used for the previous experiments entered this study. However, when the analyses of the data reported in this thesis were conducted, three years of testing had already been carried out. Therefore, the results reported here concern only those subjects who had completed the research program up to a particular phase. A brief update of further results obtained in the time in which I am writing will be given in a post-script.

Apparatus. The same apparatus used for the experiments described in Chapter V was employed for the present experiment.

Stimuli. The stimuli for each phase and each monkey were selected from a set of 8 icons differing multidimensionally and disjunctively from each other: 1) blue hourglass; 2) green square; 3) red star; 4) white triangle; 5) yellow plus; 6) brown circle; 7) open hexagon; and 8) pink I-shape.

## 6.4 Procedure and results

I) Pretraining

Rationale. The aim of the pretraining stage was to familiarize the monkeys with the serial order requirements of the new task.

Procedure. For each trial, the subject was presented with an icon appearing at random in one of the 16 possible locations defined by a 4 x 4 grid. Once touched the icon disappeared and a second icon was presented, without delay, in a random location on the screen. A touch on this second icon terminated the trial and a peanut was dispensed as reward.
Each subject received two pretraining sessions of 60 trials each.

Results. All the monkeys adapted to the task within two sessions, readily touching the icons as soon as they appeared on the screen.

II) Phase 1: teaching a 2-item-list

Rationale. In Phase 1 the subjects were trained to respond in a particular order to two items simultaneously presented on the screen. The phase was a precondition for phase 2 where the sequence would have been increased at each ordinal node.

Procedure. The particular icons received by a subject in the pretraining were used as stimuli for the same subject in Phase 1. The two icons were presented simultaneously on the screen and the subject had to touch them in a given order (the same in which the icons appeared in the pretraining). Each subject received a different sequence of icons (the sequence will be referred in abstract terms for all the subjects as AB). This phase was presented until a criterion of 75% correct was reached over the last 20 trials.

Results and Discussion. None of the subjects failed to reach criterion. The probability of obtaining by chance the required level of performance specified by the criterion (75% correct in the last 20 trials) is $p < .05$ (cumulative binomial distribution: probability of a single success equal to .5). Thus, at the end of training, all the subjects were able to order two icons and their performance was above chance level. The averaged number of trials to criterion was 276.2. On individual basis: 386 for Al; 167 for Ch; 235 for Ki; 169 for Lu; 246 for Mi; and 315 for Ol.

III) Phase 2: generalization of a 2-item-list to a 2-"chunk"-sequence

Rationale. Phase 2 was the first condition where a sequence was multiplexed by adding items at each of its ordinal nodes. Since in the experiment presented in chapter V no evidence for spontaneous classification was found, spontaneous chunking was not expected here either. However, by giving enough exposure to the task and providing strong incentive, it was attempted to teach the monkeys to chunk identical icons at each ordinal step. Moreover, by manipulating the sequence length and the presence of "chunkable" items in either first or second ordinal node, it was aimed to characterize the sort of expertise which was transferred from the previous phase to the present one.

Procedure. This phase featured the presentation of four different conditions: 1) a control condition identical to the training of Phase 1 (condition AB); 2) a condition featuring three icons, two identical to the first element of the sequence AB and one identical to the second (condition AAB); 3) a condition featuring three stimuli, one identical to the first element of the sequence AB and two identical to the second (condition ABB); and 4) a condition featuring four stimuli, two of which were identical to the first element of the sequence AB and the other two were identical to the second (condition AABB). The last three conditions will be dubbed generalization conditions. In all the conditions, the subject was required to exhaust the set of icons in the minimal number of moves and in the particular order practiced in the training phase. In other words, all stimuli of type A had to be touched before stimuli of type B. However, as long as this requirement was satisfied, the order in which two icons of the same type were touched was irrelevant. All trials where the subject satisfied the two requirements of maximal economy and of serial order were rewarded with a peanut.

The presentation of each single condition was terminated when a percentage of 75% correct trials over the last 20 was achieved. In the meantime, the subject had to maintain a performance of 3 correct responses out of 4 trials of the control condition. If it failed to retain this level of performance in the control trials, it was presented with additional consecutive control trials, until the criterion was met.

Results and Discussion. Five subjects (Al, Ch, Ki, Lu, and Ol) reached criterion in all the conditions. In the control condition (AB) the overall percentage of correct responses for all the five subjects was well above the required 75% (Al, N = 141, CR = 122 = 86.5%;; Ki, N = 160, CR = 147 = 91.9%; Lu, N = 297, CR = 288 = 97.0%; Ol, N = 301, CR = 283 = 94.0%).

The number of trials required by each subject to reach criterion in the different transfer conditions are shown in Table 6.1(a). Table 6.1(b) shows the overall percentages of correct trials performed by each subject in the transfer conditions.

Table 6.1 a) number of trials to criterion required by each subject in the transfer conditions featured in Phase 2; b) overall percentage of correct responses shown by each subject in transfer conditions featured in Phase 2;

| a) SUBJ | CONDITIONS | | |
|---|---|---|---|
| | AAB | ABB | AABB |
| Al | 137 | 24 | 139 |
| Ch | 118 | 38 | 238 |
| Ki | 64 | 30 | 161 |
| Lu | 102 | 21 | 300 |
| Ol | 113 | 93 | 376 |
| Mean | 107 | 41 | 243 |

| b) SUBJ | CONDITIONS | | |
|---|---|---|---|
| | AAB | ABB | AABB |
| Al | 33.6%** | 66.7%** | 16.5%** |
| Ch | 56.8%** | 65.8%** | 49.2%** |
| Ki | 59.4%** | 73.3%** | 49.7%** |
| Lu | 44.1%** | 71.4%** | 42.0%** |
| Ol | 54.9%** | 63.4%** | 42.3%** |
| Mean | 49.8%** | 68.1%** | 39.9%** |

** binomial test p <.001. The test was based on the cumulative binomial distribution of obtaining n or more successes, considering that the probability of a single success in the conditions AAB, ABB and AABB was of .165, .165 and .036, respectively.

These results show that five monkeys, after having learnt to order two different icons were able to chunk their multiple copies at each ordinal position.

Comparing the number of trials to criterion of the different generalization conditions it seems clear that the composition of the set of icons plays an important role independently from its size. The AAB condition, far from being equivalent to the ABB condition with which it shares the total number of items, requires more trials to be mastered. This can be explained by the fact that subjects transfer the already learnt sequence AB to the generalization conditions. Thus, while success in the ABB condition can be achieved by touching AB first and then the remaining B by default, the same strategy would lead to failure when deployed in the AAB condition.

The analysis of the errors observed in this phase is perfectly consistent with this interpretation and is summarized in Table 6.2. In fact, errors of type AB are far more abundant than both errors of type AAA and B.

Table 6.2 Frequencies of types of errors observed in the different conditions of Phase 2.

a) Condition ABB

| Subj | Type of error | | |
|---|---|---|---|
| | ABA | B | Tot |
| Al | 5 | 3 | 8 |
| KI | 4 | 4 | 8 |
| Lu | 5 | 1 | 6 |
| Ol | 19 | 15 | 34 |
| Tot | 33 | 23 | 56 |

b) Condition AAB

| Subj | Type of error | | | |
|---|---|---|---|---|
| | AAA | AB | B | Tot |
| Al | 27 | 53 | 11 | 91 |
| Ki | 7 | 15 | 4 | 26 |
| Lu | 13 | 40 | 4 | 57 |
| Ol | 14 | 35 | 2 | 51 |
| Tot | 61 | 143 | 21 | 225 |

c) Condition AABB

| Subj | Type of error | | | | |
|---|---|---|---|---|---|
| | AAA | AB | B | AABA | Tot |
| Al | 5 | 79 | 13 | 19 | 116 |
| Ki | 5 | 51 | 9 | 16 | 81 |
| Lu | 7 | 121 | 13 | 33 | 174 |
| Ol | 25 | 156 | 11 | 25 | 217 |
| Tot | 42 | 407 | 46 | 93 | 588 |

From the comparative point of view it is interesting to note that one of the 5 year old children tested by McGonigle and Jaswall (1993), who failed to show spontaneous classification in similar testing conditions, showed an identical pattern of errors. Not once did this subject respond correctly to the sequences AAB or AABB in 39 and 36 trials respectively. Instead, she omitted the second A and incorrectly executed the sequence AB. Only after explicit instruction she responded correctly to an AAB trial. On the following trials the subject executed both AAB and AABB correctly. By contrast from table 6.1a, it can be observed that monkeys, once they have learned to respond correctly to the sequence ABB, still needed further trials to learn the sequences AAB and AABB.

It seems therefore that when not 100% errorless, the patterns of errors shown by children is similar to those observed in cebus. However, a major difference can be identified in the fact that when the child acquired a strategy based on classification in the AAB condition she immediately exported it to condition AAB and even AABB irrespective of the fact that the sequence had been increased by one item.

Of course no strong claims can be made on the basis of the observation of an individual case, but it seems that further study would be of great interest especially if younger children (who might better overlap with the monkeys) are used as subjects.

A second important factor which contributes to the difficulty of the task for the monkeys seems to be the length of the series to be produced. In fact, the highest number of trial to criterion was found for the condition with 4 items (AABB). This condition, as the AAB condition, is incompatible with an AB sequence of responses and this type of response accounts for most of the errors observed in this condition. However, both the number of trials to criterion and the total number of errors testify an increase in difficulty compared with the AAB condition. Thus, the surplus demand must be an expression of the extension of the serial production required by the AABB condition which increases the probability of perpetrating errors of different types.

IV) Phase 3: teaching a 3-item-list

Rationale. In this phase the training of the monkeys was extended to a 3-item-list. This was a precondition for the following phases featuring the multiplexing of a set of icons composed of three ordinal steps.

Procedure. A third novel stimulus C was added to the sequence AB. Thus, three stimuli were simultaneously presented in random positions on the screen. The subject was required to touch all of them, in the required order ABC. This phase was presented until a criterion of 75% correct trials was reached over 120 trials.

Results and Discussion. Four of the five subjects (Al, KI, Lu, and Ol) reached criterion in this condition. The probability of obtaining by chance the criterional level of mastery in this phase (75% correct in the last 20 trials) is less than .001 (cumulative binomial distribution: probability of a single success equal .33 x .5 x .5 = .08). Thus, the performance of the four subjects was significantly above chance at the end of training.

The averaged number of trials to criterion for the four subjects was 262.5. In detail, the number of trials performed by each subject was 199 for Al; 356 for Ki; 300 for Lu; and 195 for Ol.
Table 6.3 shows the frequencies of the different types of errors observed in Phase 3.

**Table 6.3 Frequencies of types of errors observed in Phase 3.**

| Subj | Type of error | | | | |
|---|---|---|---|---|---|
| | ABA | B | C | AC | Tot |
| Al | 51 | 4 | 0 | 34 | 89 |
| Ki | 33 | 13 | 13 | 129 | 188 |
| Lu | 32 | 9 | 6 | 62 | 109 |
| Ol | 25 | 6 | 4 | 40 | 75 |
| Tot | 141 | 32 | 23 | 265 | 461 |

A comparison of the mean number of trials required to reach criterion in Phase 3 and Phase 1 suggests the presence of a positive transfer from the early phases to the first encounter with a sequence composed of three ordinal steps. This result contrasts with comparable studies in capuchin monkeys (D'Amato and Colombo, 1988) and pigeons (Straub & Terrace 1981; Terrace and Chen, 1991) multiple arbitrary list learning, where subjects require more time to complete each phase as a new item is added at the end of a list. By contrast the monkeys in the present experiment show some similarity with the profile obtained from 5 year old children (McGonigle and Jaswall, 1994) who seem to export what is learnt in one phase to subsequent phases. In other words an already learned sequence is recognized in the new one and executed accordingly, while any novel item is responded last by default.

However, on individual basis, it is evident that only two subjects (Al and Ol) required less trials to reach criterion on a three items sequence after practicing with different conditions involving two ordinal steps only. For the other two subjects (Ki and Lu) the addition of a further ordinal step resulted in a relative increase in the difficulty of the task. These subjects seem to fail to recognize the similarity of the learned and the novel sequence and possibly have to relearn the entire sequence in each phase by trial and error.

However, the learning process seemed to be more sophisticated than associative chaining. The most common error observed in Phase 3 was the AC type. This points against a classical skinnerian interpretation of how the series is learnt by the subjects. In fact, an associative chaining hypothesis would predict a strong association between the adjacent items A and B and cannot account for errors produced by skipping an item of the series. On the contrary, there seems to be evidence for a strong representation of the end items A and C. When errors occur, they are caused by neglect of the middle item B. Only one subject (Al) produced ABA errors more often than AC type of errors. Al's result is open to two different interpretations: on the one hand it is possible that the subject tended to neglect the novel item C because of memory constraints, on the other it is equally possible that he misunderstood the requirements of the task and persevered on what proved successful in the previous phases i.e. the completion of the sequence AB, reiterating it instead of completing the sequence by touching the item C.

V) Phase 4: a transfer test. From a 3-item-list to a 3-"chunk"-sequence.

Rationale. After the subjects had learnt to master a 3-item-sequence in Phase 3, Phase 4 was administered as a transfer test of three sessions, to assay whether the subjects would have spontaneously chunked 2 identical items at each of the three ordinal steps. By now the subjects had also experienced phase 2, where they learnt to chunk two items at each ordinal step of a 4-item-sequence. It was therefore interesting to observe if chunking had by now become part of their repertoire of strategies. If so they could have spontaneously chunked the new sequence from the out-set recognizing the structural similarity of the novel search space with the sets presented in Phase 2.

Procedure. The sequence ABC was doubled in order to form a six items set (AABBCC). The subject was required to exhaust the set respecting the order ABC and to avoid redundant moves within each of the classes of identical stimuli. Trials of ABC type (identical to those of training Phase 2) were interspersed with trials of AABBCC type at a ratio of 1 to 1 for the first session and 1 to 4 for the second and the third session. The second and the third sessions featured the presentation of a warm-up stage comprising of ten trials of type ABC. A criterion of 7 correct warm-up trials out of 10 had to be reached. In case the subject failed to meet this criterion, additional blocks of ten warm-up trials were presented, until the required level of mastery was reached. Each daily session featured a total of sixty trials, including the warm-ups. The four monkeys which had completed all the conditions of Phase 3 were tested for transfer in Phase 4.

Results and Discussion. Three subjects (Al, KI and Ol) required only one block of warm-up trials to reach the criterion of 7 correct responses out of 10, before each daily testing session. The number of warm-up trials required by the fourth subject (Lu) ranged from 10 to 30.

The level of performance retained by all the subjects in the control trials interspersed with the generalization condition was well above the specified criterion (Al, N = 41, CR = 39 = 95.1%; Ki, N = 32, CR = 31 = 96.9%; Lu, N = 29, CR = 25 = 86.2%; Ol, N = 35, CR = 31 = 88.6%).

The percentages of correct responses are shown in Table 6.4.

Table 6.4 Frequencies and percentages (in brackets) of correct responses obtained in Phase 4.

| Subj | N. trials | correct |
|---|---|---|
| Al | 176 | 18 (10.2%)** |
| KI | 163 | 17 (10.4%)** |
| Lu | 54 | 7 (13.0%)** |
| Ol | 135 | 19 (14.1%)** |
| Mean | 132 | 61 (11.9%)** |

** binomial test p <.001. The test was based on the cumulative binomial distribution of obtaining n or more successes, considering that the probability of a single success was p = .0004.

The results seem to indicate some degree of transfer from the early phases to this one in which three classes of two items were presented. In fact, although the percentages of correct responses is always below the 15%, the very low probability of performing an AABBCC sequence by chance (p = .33 x .2 x .4 x .2 x .4 x .2 = .0004) makes the results obtained from all the subjects highly significant, even when the number of trials completed was very small (see subject Lu in Table 6.4).

However, from the error space it seems that none of the subjects showed a strong spontaneous tendency to chunk the two items at each ordinal step. The frequencies of the different types of generalization errors are reported in Table 6.5.

Table 6.5 Frequencies of different types of errors observed in Phase 4.

| | Subject | | | | |
|---|---|---|---|---|---|
| Error type | Al | Ki | Lu | Ol | Tot. |
| B | 6 | 12 | 0 | 1 | 19 |
| C | 1 | 0 | 1 | 0 | 2 |
| AB | 65 | 51 | 0 | 60 | 176 |
| AC | 10 | 7 | 8 | 6 | 31 |
| AAA | 5 | 1 | 0 | 4 | 10 |
| AAC | 13 | 19 | 8 | 6 | 46 |
| AABA | 0 | 5 | 2 | 0 | 7 |
| AABC | 47 | 34 | 12 | 28 | 121 |
| AABBB | 2 | 0 | 1 | 1 | 4 |
| AABBA | 0 | 5 | 0 | 3 | 8 |
| AABBCA | 0 | 8 | 0 | 1 | 9 |
| AABBCB | 9 | 4 | 2 | 6 | 21 |
| Tot. | 158 | 146 | 34 | 116 | 454 |

From the table it can be observed that the most common errors were those of type AB and AABC. Thus, it seems that the original sequence was retained well by the subjects. Given the very low number of errors due to reiterations on items within a class or to items belonging to classes already exhausted, it seems unlikely that the errors were caused by memory constraints. On the contrary, it seems (again) that they can be explained by a lack of spontaneous strategic use of classification, or, in alternative, to a misunderstanding of the task requirements (the subjects simply tried out the learnt sequence ABC, neglecting the second item to be touched at each step).

VI) Phase 5: diagnostics

Rationale. A lack of generalization was observed in the previous phase and, as a consequence, the monkeys for three sessions had been rewarded only in a small percentage of trials. It was necessary to give the monkeys an easier task, to allow them to get more rewards per session and avoid possible frustration. At the same time an assessment of the following points was necessary: 1) that the sequence ABC was well retained by the subjects; and 2) that the monkeys remembered the "chunking" requirements of the task. Furthermore, an evaluation of how well the items were represented at each different ordinal position would have been interesting. It was therefore decided with Phase 5 to present sets of 4 items in three conditions featuring the three possible permutations of the three chunks AA, BB and CC.

Procedure. Phase 5 featured the presentation of three conditions obtained decomposing the sequence AABBCC in its constituents AABB, BBCC, and AACC. The session lasted until a criterion of 75% correct over the last 20 trials was reached in all conditions. The order of presentation of the three conditions followed the pseudorandomization procedure adopted for the previous phases. Moreover, control trials ABC were interspersed at a ratio 1:4.

Results and Discussion. Two subjects (Al and Ol) always reached criterion in the first block of ten warm-up trials. The other two subjects required from 10 to 20 (Ki) and from 10 to 30 trials (Lu) to reach criterion in the warm-ups.

The performance of 3 subjects in the control trials interspersed with the generalization conditions was well above the specified criterion (Al, N = 10, CR = 10 = 100%; Ki, N = 13, CR = 12 = 86.7%; Ol, N = 8, CR = 8 = 100%); Lu occasionally failed to hold the specified criterional level of mastery and, for this reason, received more control trials to recover her performance (Lu, N = 29, CR = 21 = 72.4%).

The number of trials to criterion and the percentages of correct responses obtained from the different generalization phases are reported in Table 6.6.

Table 6.6 a) Trials to criterion required to achieve criterion (75% correct responses in 20 trials) and overall percentage of correct responses (in brackets) obtained in the generalization conditions of Phase 5.

| Subj | Condition | | |
|---|---|---|---|
| | AABB | BBCC | AACC |
| Al | 42 (69.0%)** | 23 (65.2%)** | 52 (57.7%)** |
| KI | 20 (75.0%)** | 49 (59.2%)** | 77 (59.7%)** |
| Lu | 102 (43.1%)** | 97 (54.6%)** | 53 (64.6%)** |
| Ol | 39 (61.5%)** | 54 (51.9%)** | 18 (83.3%)** |
| Mean | 50.7 (62.1%)** | 55.7 (57.7%)** | 50.0 (66.3%)** |

** binomial test P <.001. The test was based on the cumulative binomial distribution of obtaining the observed percentage of correct responses or higher, considering that the probability of a single success was p = .5 x .33 x .66 x .33 = .036.

The results reported in the table show that, given enough exposure to a series requiring two chunks and two ordinal steps the subjects were able to classify the items at the appropriate point of the series. The fact that the ABC series was well represented as a whole is shown by the highest performance obtained in the condition AACC. This again demonstrated that the subjects had a strong representation of the starting and ending points of the series, while the representation of the intermediate item B was weaker.
Table 6.7 shows the frequencies of the different types of errors which occurred in each generalization condition.

Table 6.7 Frequencies of different types of errors made in the different generalization conditions of Phase 5.

a) Condition AABB

| Error type | Subject Al | Ki | Lu | Ol | Tot. |
|---|---|---|---|---|---|
| B | 0 | 5 | 11 | 0 | 16 |
| AB | 12 | 24 | 30 | 15 | 81 |
| AAA | 1 | 0 | 0 | 0 | 1 |
| AABA | 0 | 2 | 3 | 0 | 5 |
| Tot. | 13 | 31 | 44 | 15 | 103 |

b) Condition BBCC

| Error type | Subject Al | Ki | Lu | Ol | Tot. |
|---|---|---|---|---|---|
| C | 0 | 1 | 6 | 1 | 8 |
| BC | 5 | 12 | 33 | 15 | 65 |
| BBB | 0 | 0 | 4 | 2 | 6 |
| BBCB | 3 | 7 | 15 | 8 | 33 |
| Tot. | 8 | 20 | 58 | 26 | 111 |

c) Condition AACC

| Error type | Subject Al | Ki | Lu | Ol | Tot. |
|---|---|---|---|---|---|
| C | 2 | 5 | 0 | 0 | 7 |
| AC | 6 | 24 | 16 | 2 | 48 |
| AAA | 12 | 0 | 3 | 0 | 15 |
| AACA | 2 | 2 | 10 | 1 | 15 |
| Tot. | 22 | 31 | 29 | 3 | 85 |

The distribution of the errors shows again that the majority of them (AB in condition AABB, BC in condition BBCC, and AC in condition AACC) occurred because the subjects tended to neglect the second item to be touched at each ordinal step. Once again, it seems that they were trying to cope with the task performing the learnt sequence ABC without spontaneously chunking the two items at each step.

VII) Phase 6: training "chunkable"-6-item-sequences.

Rationale. The results obtained from Phase 4 showed only a weak tendency to classify spontaneously a sequence of three chunks. In Phase 6 it was attempted to teach the monkeys to do so by means of extensive task exposure. Moreover, the possible effects of the composition of the set were taken into account. The relative speed of acquisition of sets organized in three classes of two stimuli each and two classes of three stimuli each was evaluated.

Procedure. Four different conditions were presented i.e. a condition featuring trials analogous to those of the generalization Phase 2 (AABBCC) and three conditions featuring trials similar to those presented in Phase 5 but containing three items instead of two for each class of icons (AAABBB, BBBCCC and AAACCC). Trials of these four types were interspersed according to the standard pseudorandomization procedure described above.

The presentation of each single condition was terminated when a percentage of 75% correct trials over the last 20 was achieved.
Each daily testing session was preceded by the administration of warm up trials identical to those presented in conditions 4 and 5. As for Phase 5, the warm up trials were presented in blocks of ten until a criterion of 7 correct out of 10 was reached.

Results and Discussion. In the warm-up trials, all the subjects usually reached the criterion of 7 correct responses out of 10 in the first 10 responses, only in two occasions three subjects (Ki, Lu and Ol) required up to 30 trials to reach criterion.

Two subjects (Al and Lu) learned all the conditions of this phase, Ol learnt the AAACCC condition only and Ki none. The number of trials to criterion in the four generalization conditions are shown in Table 6.8. Table 6.8 also depicts the overall percentages of correct responses obtained in these conditions and the correspondent probability respect to the binomial distribution.

Table 6.8 a) number of trials to criterion required by each subject in the transfer conditions featured in Phase 6; b) overall percentage of correct responses shown by each subject in transfer conditions featured in Phase 6.

| a) SUBJ | CONDITIONS | | | |
|---|---|---|---|---|
| | AAABBB | BBBCCC | AAACCC | AABBCC |
| Al | 214 | 449 | 121 | 537 |
| Ki | - | - | - | - |
| Lu | 182 | 60 | 109 | 272 |
| Ol | - | - | 20 | - |
| Mean | 198 | 254.5 | 83.3 | 404.5 |

| b) SUBJ | CONDITIONS | | | |
|---|---|---|---|---|
| | AAABBB | BBBCCC | AAACCC | AABBCC |
| Al | 34.6** | 42.8%** | 57.0** | 35.4%** |
| Ki | - | - | - | |
| Lu | 52.2%** | 60.0%** | 59.6%** | 27.6%** |
| Ol | - | - | 75.0%** | |
| Mean | | | | |

** binomial test p <.001. The test was based on the cumulative binomial distribution of obtaining n or more successes, considering that the probability of a single success in conditions AAABBB, BBBCCC and AAACCC was of p = .5 x .4 x .4 x .6 x .4 x .4 = .008; and .0004 in condition AABBCC.

Results obtained from Phase 6 show that, given enough trials 2 monkeys were able to classify in conditions involving three ordinal steps composed of two items each and two ordinal steps composed of three items each. Taken together the number of trials to criterion and the overall percentage of correct sequences observed for the different phases seem to show that the composition of the set influences the level of difficulty of the task. In general, monkeys found sets of two chunks of three items each easier than sets of three chunks of two items. The overall series was well retained as (consistently with the results obtained from Phase 4) the AAACCC sequence seem to be the easiest to master. Ol, in this phase showed a perfect transfer, and perhaps for the first time evidence for a spontaneous tendency to classify, since reached criterion in the minimal number of trials. By contrast she failed to learn the other sequences.

Table 6.9 depicts the frequencies of the different types of errors observed in the different conditions of this phase.

Table 6.9 Frequencies of different types of errors made in the different generalization conditions of Phase 6.

a) Condition AAABBB

| Error type | Subject Al | Ki | Lu | Ol | Tot. |
|---|---|---|---|---|---|
| B | 4 | - | 1 | - | 5 |
| AB | 31 | - | 14 | - | 45 |
| AAB | 84 | - | 43 | - | 127 |
| AAAA | 18 | - | 26 | - | 44 |
| AAABA | 2 | - | 3 | - | 5 |
| AAABBA | 1 | - | 0 | - | 1 |
| Tot. | 140 | - | 87 | - | 227 |

b) Condition BBBCCC

| Error type | Subject Al | Ki | Lu | Ol | Tot. |
|---|---|---|---|---|---|
| C | 1 | - | 0 | - | 1 |
| BC | 5 | - | 3 | - | 8 |
| BBC | 30 | - | 16 | - | 46 |
| BBBB | 16 | - | 1 | - | 17 |
| BBBCB | 0 | - | 3 | - | 3 |
| BBBCCB | 0 | - | 1 | - | 1 |
| Tot. | 52 | - | 24 | - | 76 |

c) Condition AAACCC

| Error type | Subject Al | Ki | Lu | Ol | Tot. |
|---|---|---|---|---|---|
| C | 1 | - | 2 | 0 | 3 |
| AC | 5 | - | 3 | 0 | 8 |
| AAC | 30 | - | 32 | 3 | 65 |
| AAAA | 16 | - | 7 | 2 | 25 |
| AAACA | 0 | - | 0 | 0 | 0 |
| AAACCA | 0 | - | 0 | 0 | 0 |
| Tot. | 52 | - | 44 | 5 | 101 |

d) Condition AABBCC

| Error type | Subject Al | Ki | Lu | Ol | Tot. |
|---|---|---|---|---|---|
| B | 3 | - | 7 | - | 10 |
| AB | 104 | - | 49 | - | 153 |
| AAA | 35 | - | 23 | - | 58 |
| AABA | 11 | - | 2 | - | 13 |
| AABBA | 2 | - | 1 | - | 3 |
| AABBB | 0 | - | 5 | - | 5 |
| C | 2 | - | 2 | - | 4 |
| AC | 8 | - | 5 | - | 13 |
| AAC | 19 | - | 25 | - | 44 |
| AABC | 132 | - | 75 | - | 207 |
| AABBCA | 3 | - | 0 | - | 3 |
| AABBCB | 15 | - | 3 | - | 18 |
| Tot. | 334 | - | 197 | - | 531 |

Consistently with the results obtained from previous phases the most common error observed in all the different conditions was the "exit" from a particular ordinal step before all the items of the chunk were touched. By contrast the skipping of an entire class or reiteration errors were extremely rare.

VIII) Phase 7. A final transfer test.

Rationale. Since two monkeys learned to classify a 6-items-sequence in the previous phase. It was of major interest to evaluate whether now they would have spontaneously transferred this competence as a data reducing strategy in a new set featuring 9 items (divided into three sub-sets of three). As we have seen in chapter IV, the length of the serial production required to search a set of 9 items was an incentive for the subjects to deploy spatial strategies. Now that spatial strategies could not be deployed, an assessment was necessary of whether the subjects would have spontaneously used classification as a strategy. Moreover, on the basis of the performance of the subjects, the evaluation of the relative efficiency of strategies based on classification as compared with spatial ones would have been feasible.

Procedure. This phase was analogous to Phase 2 but at each node of the sequence AABBCC an item was added, in order to form a nine items set (AAABBBCCC). As for Phase 2, the subject was required to exhaust the set according to the order ABC and to avoid redundant moves within each of the classes of identical stimuli. A further difference in respect to Phase 2 was that in Phase 7 no trials of the ABC type were interspersed with trials of AAABBBCCC, (since the subject had already had a long practice with this sort of trials). However, in order to be sure that the performance on the series ABC was not disrupted, five warm-up trials of ABC type were presented before each daily session. In case the subject failed to perform correctly at least 4 of the 5 warm-up trials, other blocks of 5 warm-up were presented until the criterion was reached. In addition to the warm-up trials, each daily session featured fifty trials of AAABBBCCC type. Two sessions were administered as a transfer test to the monkeys.

Results and Discussion. In the warm-ups, both Al and Lu always reached criterion in the minimal number of trials (4 out of 5). The overall performance of the 100 trials presented was of 13% and 18% correct for Al and Lu, respectively.

Since the number of trials presented was relatively small and the probability of performing a correct sequence extremely low (being p = 3/9 x 2/8 x 1/8 x 3/8 x 2/8 x 1/8 x 3/8 x 2/8 x 1/8 = .000001416) it seems that some transfer from previous phases had occurred.

Once acquired, the ability to impose classificatory schemes over the search space seems to be effective as a data reduction strategy. Some evidence for this can be evinced from a comparison of Lu's current performance with the performance shown in those conditions of the experiments reported in chapter IV, which featured the presentation of nine items (as in the last condition of the present experiment), but afforded spatial constraints only. The percentage of minimal paths performed by Lu when faced with 9 stimuli in the experiments presented in chapter IV was of 2.32% and 6.49% in the non-feedback and in the feedback condition, respectively. The 18% of correct trials performed by Lu in the last condition of this experiment sharply contrasts even with the peak performance she registered in the previous experiments (the 4th block of the feedback condition) which was only of 10.81% minimal path trials.

Similar conclusions can be drawn from the behaviour of subject Al, whose percentage of minimal path trials was respectively of 2.87% and 8.29% in the phases featuring 9 items of the non-feedback and feedback conditions of the experiments presented in chapter IV. Thus, also Al with his 13% of minimal path trials shown in Phase 7 of the present experiment, provides evidence for the effectiveness of the superimposition of classification schemas over the search space. However, the peak performance registered by Al in the experiments presented in chapter IV (3rd block of the feedback condition) was of 16.42%, slightly higher than that shown in Phase 7 of the present experiment.

The frequencies of the different types of errors observed in this phase are shown in Table 6.10.

Table 6.10 Frequencies of different types of errors observed in Phase 7.

| Error type | Subject Al | Lu | Tot. |
|---|---|---|---|
| B | 1 | 0 | 1 |
| AB | 9 | 7 | 16 |
| AAB | 51 | 32 | 83 |
| AAAA | 1 | 5 | 6 |
| AAABA | 0 | 2 | 2 |
| AAABBA | 0 | 1 | 1 |
| AAABBBA | 0 | 0 | 0 |
| AAABBBB | 1 | 4 | 5 |
| C | 1 | 4 | 5 |
| AC | 1 | 0 | 1 |
| AAC | 0 | 1 | 1 |
| AAAC | 1 | 2 | 3 |
| AAABC | 11 | 5 | 16 |
| AAABBC | 10 | 19 | 29 |
| AAABBBCA | 0 | 0 | 0 |
| AAABBBCB | 0 | 0 | 0 |
| AAABBBCCA | 0 | 0 | 0 |
| AAABBBCCB | 0 | 0 | 0 |
| Tot. | 87 | 82 | 169 |

The table shows that the most common errors were the AAB type, followed by the AAABBC type. This pattern seems consistent with those observed for the other generalization phases in which the subjects tended to export lists already acquired to the new multiplexed ones, without considering the increase of the number of icons at each ordinal position. Similarly, in the present phase the subjects (which learned an AABBCC sequence in Phase 6) tended to neglect the third item at each ordinal node.

This pattern of errors was observed in some of the 5 year old children tested by McGonigle and Jaswall (1993) on an AAABBBCCC.

sequence. Six out of 8 subjects committed errors of the AAB type as the monkeys did, albeit with a much lower frequency. However, another common error observed in children was of the AAAC type, i.e. where a whole class of icons was neglected. This type of error was taken by McGonigle and Jaswall as evidence for the fact that the subjects recalled the lists as classes of items and not as individual units. Monkeys very rarely performed this type of error.

## 6.5 Conclusions

Overall, the results obtained from this experiment help to clarify the issues raised from the previous chapter. If monkeys do not spontaneously classify (or do so only at a certain extent) they are able to learn to do so. The ability to take simultaneously into consideration multiple classes of items and to put them in a principled order seem to be within their competence.

However, the subjects do not seem to readily deploy classification as a data reducing strategy, even after they have learnt to classify similar sequences featuring a smaller number of items. By contrast, five year old children seem to do so in comparable tasks (McGonigle and Jaswall, 1993). Therefore, the spontaneous use of classification (in these sorts of tasks) seem to be a promising dimension for species comparisons.

The patterns of errors, moreover, offer a rich set of interpretative possibilities which can help to characterize the psychological processes underpinning performance in search tasks. The systematic study of differences and analogies between children of different ages and different species of primates could provide valuable information about the relationship between the ontogeny and the phylogeny of complex cognitive skills.

## 6.6 Post-script

During the writing phase of this Phd thesis, further testing has been carried out on the paradigm featured in this chapter by a colleague (Dickinson, Phd in preparation). The additional data he collected using the same apparatus, subjects and procedures featured in this chapter seem to provide a further indication of the potentialities of the paradigm.

In fact, after an extended period of training all the six monkeys reached the criterion of 75% correct on the six item series AABBCC sequence. Four monkeys reached the same criterion on the 9 item sequence AAABBBCCC and one even on a 12 item sequence AAAABBBBCCC. For this latter subject (Ch) there is also evidence for some degree of transfer from the 9 to the 12 items condition. This is expressed by both the number of trials to criterion (1686 for the 9 items condition and 569 for the 12 items condition) and the overall percentage of correct trials (37% in the 9 items condition and 42% in the 12 items condition).

In the light of these results the effectiveness of classification in the organization of serial behaviour is striking. Especially comparing the maximum sequence length obtained in serial learning studies with the performance of the monkeys observed here. As mentioned in Chapter II, capuchin monkeys can learn to produce a maximum of 5 items in a specified order (D'Amato and Colombo, 1988).

A preliminary time analysis undertaken with some of the data available at the time of writing provides further evidence of the psychological reality of the classificatory competences of the monkeys. In fact, if classification has to be considered a data reducing strategy which allows the subjects to reduce the total number of items to be remembered to the number of chunks, one should expect a longer latency at chunk boundaries (see McGonigle and Jaswall, 1993). This would be because once the subject has identified the first member of a class, execution of subsequent items should require relatively little effort.

This is what has been observed by analysing data obtained from Phase 4. For example, in a sequence AABBCC (suitable for the analysis since it is formed by three classes of two items each), the typical pattern of latencies would show peaks in correspondence of the first A touched, the first B, and the first C (Fig. 6.1).[1]

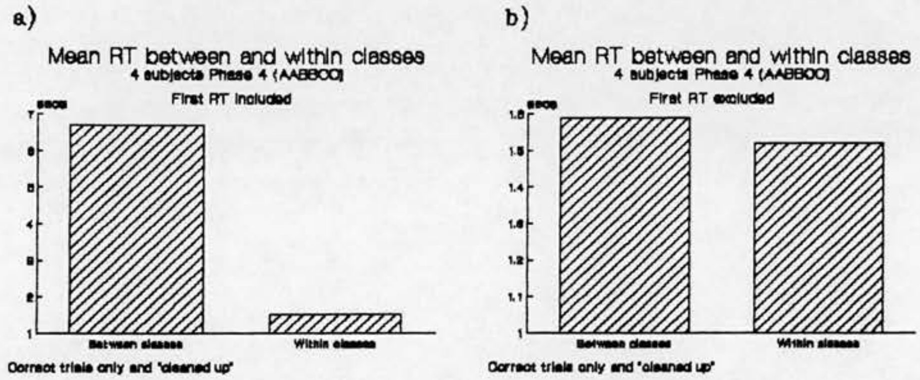



Fig. 6.1 Mean RTs obtained for the first (between classes) and the second item (within classes) touched for each class. Only correct trials were included in the analysis. In a) the analysis included the first reaction time, usually very long. In b) the first reaction time has been excluded from the analysis to show that the effect (longer latencies at class boundaries) cannot be accounted merely by the long latencies recorded at the beginning of each trial.

The same effect was observed for each of the individuals. Even when the first reaction time (the length of which might sometimes be biased by the fact that the subject does not immediately start to respond when the stimuli appear on the screen) is excluded from the analysis the effect emerges for three out of four subjects.

---

1 An analysis currently in progress (Dickinson, in preparation) on further data collected by the time of writing seems to support the results when a more extensive data base is taken into consideration. There is also evidence for a similar pattern on longer series such as AAABBBCCC.

Subject Ollie showed an inverse relationship (with shorter decision times at class boundaries). However this subject was very fast in responding compared to the others and this might have produced a difficulty in differentiating between the two types of responses.

The same pattern of latencies has been observed in children by McGonigle and Jaswall (1993) who also found a rich set of other time related phenomena. For example, these authors found a correlation between latencies of the responses to the first item of each chunk of a list and the length of the list itself. This demonstrates that children do not merely pause between chunks; rather, the first-item latency truly represents search time and increases when more classes have to be taken into account.

It goes without saying that a comparison of the latency pattern shown by the monkeys when faced with lists of different length would provide comparative data of great value.

# CHAPTER VII

## CONCLUSIONS

### 7.1 A critical assessment of the paradigms developed in this thesis

The rationale for the study of primate cognition can be considered threefold. First of all, only through comparative exercises, aimed at determining differences and analogies between related species might it be possible to trace the course of the evolution of human intelligence. Thus, the investigation of primate's cognitive skills serves what has traditionally been considered the pure objective of comparative psychology. Secondly, the characterization of primate cognition might help the mapping of human higher order cognitive functions onto animal models more plausible than those based on rodents for neuropsychological research. This can ultimately lead to bio-medical applications of the discipline. Finally, as suggested by McGonigle (1982; McGonigle and Chalmers, 1984), the use of monkey subjects, in relevant experiments, might prove an effective way of unpacking some high order cognitive abilities, which are difficult to disentangle working exclusively on adult humans and children.

Moreover, monkeys are more resistant to prolonged exposure to the same task than both human adults and children (see, for example, McGonigle and Chalmers, 1992). The study of non-human primates, therefore, allows a better investigation of changes in expertise over time. In addition, the study of experimentally naive and non-verbal subjects permits the observation of how expertise is gained from the very beginning and without the mediation of language (see Terrace and McGonigle, 1994).

However, following the arguments put forward by McGonigle (1984), in the first two chapters of this thesis, it was argued that, to date, most of the research conducted in comparative cognition has failed in finding those dimensions along which species might meaningfully be compared.

At first it was argued that conventional paradigms like discrimination learning, reversal learning, learning sets, delayed responses, and matching to sample, in their binary forms, although sometimes successful in showing quantitative and qualitative differences between broad taxonomic groupings (like the distinction between mammals and non-mammals), constantly failed to indicate differences between more complex organisms, particularly between primate species. It was put forward that the main limitation of those paradigms lay in their inadequacy to determine what is learned and the structure and content of what is remembered, as contrasted with the evaluation of mere rates of learning or simple memory span over time (see Warren, 1965; McGonigle, 1984). From a critical assessment of the traditional studies, it emerged that the serial aspects of behaviour have been crucially under-represented in these studies. Aspects which from many different perspectives have been recognized as fundamental for the characterization of cognition, (Lashley, 1951; McGonigle 1986; Terrace and McGonigle, 1994; Miyashita, 1988).

When studies on serial learning and related issues were examined, it emerged that serial learning paradigms overlook the strategies that subjects might spontaneously develop, when faced with the problem of organizing long serial productions of actions. In fact, when a subject is rigidly trained to perform an arbitrary sequence of responses (or, in general, to solve any particular problem), it becomes impossible to know whether it would have recruited competences (either built in or transferred from previous practice with similar problems), in order to maximize its proficiency in the task in hand. It was suggested that this issue might be particularly relevant for the understanding of key questions in comparative cognition. Namely, what is the generality of a particular competence ? How does it develop ? How can general epistemological procedures be characterized ? I argued that all these questions deal with the issue of self-regulation, in other words with the problem of which resources an organism might be able to recruit in order to compensate for its own cognitive limitations.

In Chapter V, those studies which have concentrated on spontaneous manipulations in capuchin monkeys were reviewed. Two main restrictions of these studies emerged. On the one hand it was suggested that spontaneity per se, in absence of task requirements, cannot show the emergence of any strategical use of competences. On the other, as a practical point, the sort of paradigms based on spontaneous play do not allow sufficient exposure to the situation from which the data are derived to provide an analysis of change of expertise over time. Yet, an analysis of what has been dubbed "microdevelopment" (Karmiloff-Smith, 1992) is of primary importance for the stance of self-regulation. Only by allowing the subject to become expert with a task, we give it the chance to recognize the relevant aspects of the situation at hand and the opportunity to self-monitor its own behaviour in order to recognize the possible limitations of its own cognitive resources (for a discussion of a similar point of view see also Brown and Deloache, 1978; and Luger and Stern, 1990).

Thus, it was emphasized that a major aim of the present study was to devise new paradigms, to overcome the major difficulties of the traditional ones. Keeping in mind the major sources of criticism of the conventional paradigms, I will now proceed with an overall evaluation of the research program implemented in this study.

The first set of experiments of WGTA based experiments, was derived from paradigms already developed in the context of developmental psychology. In this thesis, they were conceived as a pilot study to familiarize the wild born and experimentally naive subjects with the testing situation itself and to introduce them to a set of problems based on the exploration of multiple alternatives. In fact, the familiarisation of the subjects with search tasks, where multiple locations had to be explored serially, was considered a necessary precondition for the rest of the research program. Moreover, the tasks seemed appropriate because they also afforded the possibility of using strategic devices in order to reduce the search space.

Finally, as for all the experiments which would have followed, the opportunity to detect relevant information and deploy the appropriate strategy was given to the subjects from the outset. This approach differs from the majority of tasks implemented with WGTA, where the solution had to be found inductively by means of what Spence (1951) called blind trial and error procedures and, on the contrary, is in some way reminiscent of the early experiments on insight reviewed in the first chapter (e.g. Koehler, 1917).

For all these aspects, these first experiments were a necessary preamble to the ones which followed. Only if subjects were familiar with testing, and showed incentive to perform search tasks within a set of multiple alternatives, as well as expressing some evidence for strategic use of the information available, it would have been possible to proceed along the experimental program. The results showed that the enterprise was viable. The subjects proved keen to search and prone to deploy strategical devices as economic, rational, data-reducing strategies.

Apart from helping the implementation of the rest of the research program, this first set of experiments, never conducted before with non-human subjects, provided the first comparative data on the relationship between search for a hidden object and inference.

Moreover, an accurate analysis of the sort of information needed to be coordinated by the subjects, in order to solve the different conditions of the experiments, led to the detection of an implicit hierarchy of complexity within the tasks, previously overlooked by the original authors (Somerville and Capuani-Shumaker 1984).

However, the tasks themselves shared some of the difficulties outlined for the traditional experiments based on spontaneous play. In fact, they did not allow protracted periods of testing and the subjects, especially in the most difficult conditions of the task, showed evidence of distress and demotivation. Moreover, given the social nature of the cues, on the basis of which the subjects had to solve the task, it was difficult to attribute failures unambiguously to restrictions in the competences of the subjects, as opposed to spurious problems in dealing with information transmitted by another agent (and, for the monkeys, not even a conspecific). Last but not least, it was clearly confirmed that in order to study search and self-regulation, it was necessary to move on to an experimental apparatus, which allowed a measure of reiteration to evaluate the extent to which subjects could keep track of moves made serially over time. In this respect, the WGTA appeared to be inadequate, because each manual interrogation of an object displaces it from its test position, thus living a visible trace of a visit.[1]

The second set of experiments was, thus, conducted using touch-screen based procedures, which had already proved efficient for the administration of size seriation tasks in children (McGonigle, 1987a, b; McGonigle, 1989; McGonigle et al. 1992; Chalmers and McGonigle 1993). The first two experiments based on such procedures, were an inaugural attempt to test such technologies with capuchin monkeys. In fact, although they proved effective with macaques (Swartz et al., 1991), touch screens had never been tried out with capuchin monkeys. The only related apparatus that has been used successfully with this species is the Lehigh Valley operant chamber, a device based on projectors fitted with transparent plastic keys which serve as response mechanisms. Such a device differs in many features from a touch screen. First, there is a difference in the resolution of images produced by projectors and those generated on computer monitors. Secondly, the operant keys move slightly backwards when pressed, providing a kinesthetic feedback to the subject for each stimulus touched. Such a feedback is not provided

---

1 Actually, this sort of external indicators of previously explored alternatives, can be used as part of a notational system that can help the subject to overcome the limitations of its own memory, and, as shown for children, the change in the strategical use of notational systems is a privileged window on cognitive growth (Karmiloff-Smith, 1992). However, in order to study the ability of the subjects to self-regulate, its use must be optional and not inevitable as with the WGTA.

with touch screens. Moreover, it was needed to assess the sensitivity of the touch screens to the touch style, the size and the moisture of the fingers of capuchins.

The finding that capuchins readily adapted to this sort of apparatus, thus represented a first result of the study. It opened the way to novel researches with this species, previously precluded by the use of the operant chamber alone. In fact, touch screens allow the presentation of stimuli much more flexibly than the operant chamber, both in terms of possible locations on the screen and speed of presentation. The results that followed clearly demonstrated the potentiality of this new apparatus.

The gradual increment of the set size proved an effective procedure to instantiate search tasks with large numbers of alternatives. In fact, the subjects showed to be capable, not only to perform spontaneously exhaustive searches within large sets of icons, but also the length of their sequences of responses has no precedents in the literature on primate cognition. However, the major innovation of the paradigm was the freedom that it accorded to the subject in finding its own path through the search space.

The evidence for positive trends in the percentages of non-redundant moves, in absence of any arbitrarily imposed penalization for reiterations, was the first direct demonstration in laboratory studies that, without any explicit feedback, monkeys are nevertheless able to self-regulate their serial production by monitoring the delay (for a discussion on the relative role of time/muscular effort in foraging strategies see Gallistel, 1992) inevitably associated with the production of reiterative moves.

This phenomenon corroborated the hypothesis formulated by McGonigle and Chalmers (1992) on the basis of a related finding in the context of transitive inference in squirrel monkeys, which was a key background for the present research program. Consistently with the hypothesis of McGonigle and Chalmers, the results of this experiment showed a natural tendency towards economy. Such a finding provides an alternative hypothesis to traditional interpretations of learning, phylogenetic trends and ontogenesis.

This tendency towards economy might, in fact, be a motivating factor for evolving systems (i.e. undergoing transformations either in terms of learning, speciation, or development) to invest in complexity and cognitive growth, in order to gain in terms of better management of resources when faced with non-trivial tasks. The assessment of the presence of such a spontaneous tendency towards economy was a precondition for the exploration of the sorts of constraints monkeys might be able to exploit when dealing with complex serial tasks which heavily challenge their memory capacity.

However, the finding that subjects tried to maximize their economy of search only up to a given level of performance, supported the idea that we need to distinguish between two sorts of cost factors which the self-regulating system might take into account while monitoring its own behaviour. One type of cost was designated as intra-cognitive and was related to the amount of internal resources which the deployment of a particular strategy demands. The other was designated as extra-cognitive and was related to the time spent to gain the reward. It was hypothesized that self-regulation was active only until an optimal (or subjectively judged so) trade-off between these two cost factors was achieved.

From a comparison of the performance of monkeys and children (McGonigle et al., 1992), it emerged that the former were never as principled in their searches as the latter. This observation led to the instantiation of a procedure then used throughout the rest of the experimental program. The subjects, having had the opportunity to self-regulate (on the basis of the feedback implicit in the situation at hand, as in experiment 1 of Chapter IV), were then (in the second experiment presented in Chapter IV), exposed to a selective feedback, in order to check whether the standstill of their improvement was the expression of a true "resistance point" for the species.

Results obtained from the second experiment presented in Chapter IV, gave some support to the hypothesis of the trade-off between intra- and inter-cognitive costs. Adding extra-cognitive costs to the search task (by means of a penalization of the most inefficient trials), produced further self-regulation and a dramatic improvement in the economy of search. Moreover, as the selective feedback was based on the penalization of redundant moves and not search time, it compelled the subjects to self-monitor the number of their touches as well as the time spent in searching.

The fact that efficiency was sustained by the deployment of strategic factors, and not just by brute memory, was evinced from the larger number of adjacent moves observed in minimal path trials compared to trials where some redundant moves were performed. It was also expressed in the finding that the strategy of moving on adjacent items was employed more in the second experiment of Chapter IV (where a better overall performance was observed) than in the first experiment.

In any case, even with selective feedback monkeys were never as principled as children. Thus, it seems that with the second experiment the upper level performance for this species was reached. This was an important point to assess before some conclusion of comparative relevance could have been made about the "resistance points" for capuchins in this task. As mentioned above, children, presented with an identical task, showed an almost perfect performance. Nevertheless, when the forms of constraints used by the two species were compared a similarity emerged in the data reducing strategies adopted by both species. The most interesting finding in this respect was that, by a qualitative analysis of the strategies adopted by the monkeys, it was possible to decompose the very principled trajectory through the search space shown by the children into different additive sub-components. In fact, facing a configuration of 9 items, a principled path through it could be obtained by using particular starting points, an adjacency principle, and moving along privileged vectors. Although, all these sub-strategies were employed by the monkeys only to a certain degree (i.e. on statistical basis), the most evident source of difference between their behaviour and that of children was in the lack of a principled use of the third one. Moreover, and interestingly enough, younger children approximated monkey's behaviour better than older ones.

With these set of experiments the common problem rising from floor or ceiling effects in performance was overcome (effects which can completely obscure the rich set of intermediate possibility which ultimately are the only way of characterizing cognitive change in the course of either ontogenesis or phylogenesis).

This has been made possible by both the selection of the appropriate paradigms and the appropriate species. If the nature of the paradigms allowed a clear specification of the strategies adopted by the subjects, the use of monkey subjects allowed an in-depth investigation of how the knowledge of the task emerged gradually. As we have seen from comparable data reported for other species, the perfectly strategic behaviour shown by the majority of children and the completely flat performance of pigeons would have prevented such an analysis of microdevelopment.

These results themselves prove the validity of the paradigm in respect to the criteria outlined above. In fact, there is enough similarity between the forms that the organization takes in the two species to promise the possible use of monkeys as a model of some aspects of human cognition. On the other hand, the task elicits the emergence of inter-species differences, both quantitative and qualitative, supporting the hypothesis that these sorts of paradigms would be better comparative tools than those traditionally employed.

Furthermore, by means of these tasks it was possible to observe how complex forms of organization can be assembled from simpler sub-components. It seems that with this research it has been possible to work at an appropriate level of analysis as contrasted with radical forms of reductionism in psychology. Here we have the possibility to observe how basic elements (which nevertheless have got their own idiosyncratic properties if taken in isolation) give rise to complex behaviours, when embedded in a set of meaningful relationships with each other. As pointed out by Luria (1987), to avoid becoming futile, the reduction of complex psychological phenomena should be operated only up to certain limits. This is the level where it is possible to characterize "units" which "can preserve all the richness of the behaviour while at the same time pointing out models for which it can be the subject of an accurate study" (Luria, 1987, p. 675). It seems that in the context of this study it has proved possible to identify these "models" in the different sub-strategies and the nature of their interaction.

Finally, the procedure and the apparatus proved suitable for a prolonged presentation of the tasks. Only allowing subjects to practice enough with the task and presenting them with different feedback conditions was it possible to see which spatial constraints afforded by the stimuli configuration the subjects were able to use. This favoured the advancement of the experimental program to evaluate (as proposed by McGonigle, 1987) whether the subjects would have been able to profit from forms of constraints less restrictive than the spatial ones alone.

The experiments that followed explored the extent to which capuchins showed some propensity to focus their attention on the non-spatial features of the configuration of items to be searched. Furthermore, the study tried to asses whether, in order to provide an external aid for their memory system, they would have shown a tendency to structure the search space on the basis of classification schemas and organize their behaviour accordingly. These experiments can be considered as a novel approach to the comparative study of classification from different points of view. First of all, it was the first time that a classification task was presented to monkeys on touch sensitive computer monitors. Moreover, it was the first time that capuchin monkeys were faced with a task where the superimpression of classification schemas was completely optional, albeit in the presence of an overall incentive to perform an economic search. From this perspective, the paradigm radically differed from the study of spontaneous classification (Antinucci, 1989) mentioned in Chapter V, where the subject is left free to manipulate a set of three-dimensional objects in absence of any task requirement. The study provided the first opportunity to evaluate whether, when the opportunity to classify is offered more as an useful accessory to the economic solution of the task than a requirement of the task itself, it would have been exploited as a strategic, data reducing factor.

The results showed that the subjects were perceptively able to distinguish between the different icons presented on the monitor and that they used the opportunity to classify offered by the configuration. However, this did not lead to an increase in performance and a number of possible explanations were offered.

Firstly, it was conjectured that, as it has been observed in 3 year old children (Sophian and Wellman, 1987), monkeys failed to recognize the potential data reducing function of organizing the search space in a principled way. This could have been due to a lack of those skills necessary to evaluate their own memory constraints and to find an aid in the features afforded by the search space.

Secondly, the possibility was considered that monkeys lacked those categorisation skills that allow the simultaneous comparison of different classes.[2] However, it was possible, as observed in the previous experiments, that the subject found a satisfactory trade-off between the cognitive strain that such a form of classification would have required and the extra-cognitive costs attached to a search not maximally economic.

Thus, although not conclusive for the stance of self-regulation and strategical classification, the results of these experiments

2 For a discussion of the difference between selecting similar items belonging to only one class, the typical behaviour shown in spontaneous play by children below 2 years, and the awareness of the presence of more than one class, shown by older children see Sugarman (1983).

provided the indication that the subjects were able to search the set of icons in a non-random fashion according to non-spatial features. Therefore, the same approach which had proved effective to disambiguate the results of the experiments presented in Chapter IV, was used for the last set of experiments. Namely, the strategy of moving from weak to strong forms of training was adopted again.

In this case the strongest forms of constraints were implemented as the hybrid paradigm proposed by McGonigle (1994) which incorporates tutored serial learning and free classification procedures. A penalty was imposed over unprincipled searches, giving the subjects more incentive to classify in a principled way.

The possibility to observe the strategy adopted by proficient subjects (children) provided optimal information for use in training monkeys. In this way it was possible to asses whether they were able to learn to use classification as a memory aid and if so to observe the effects on performance.

This approach contrasts with most of research conducted in developmental psychology where once a particular strategy is inculcated in subjects who did not use it spontaneously the experiment is considered complete.[3] In fact, here it was possible to go beyond the mere assessment of monkey's capability to acquire a strategy and evaluate the extent to which the strategy could be exported to similar tasks. Thus, the paradigm was designed to satisfy not only the behavioural criteria of observing that a particular performance was achieved but also to observe whether a real cognitive change had taken place. In other words the paradigm allowed the evaluation of whether the subject's acquisition of the strategy would change its future understanding of similar tasks.

---

3 Usually, following the distinction proposed by Flavell (1970), this approach has been adopted to try to answer the question of whether the lack of strategies is to be considered as a production or as a mediational deficit.

Moreover, as a complementary set of information, the possibility of analysing both time related aspects of performance and the quality of the errors would have enabled the characterization and assessment of the psychological status of the phenomena observed.

The last set of experiments proved effective in clarifying most of the issues which had remained ambiguous in the previous ones. The results showed that capuchins are able to compare different classes and to order them according to a particular sequence. The usefulness of classification as a data reducing device emerges from a comparison of the results obtained in the present study with data reported in the literature on serial learning. To date, the maximum sequence length observed for capuchins has been of five items (D'Amato and Colombo, 1988), while macaques have been trained to perform multiple series of six items each (Swartz et al., 1991). A serial production of up to nine items above chance, as shown by two subjects in this experiment, thus seems to have no precedent in the literature on serial learning. However, for the differences in the serial requirements of the tasks used by D'Amato and Colombo (1988) and Swartz et al. (1991) on the one hand, and those used in the present study on the other, this comparison with their data must be considered only as an indication of the promise of serially ordered classification in the study of serial aspects of behaviour.

Some more indications of the usefulness of classification in sustaining long serial productions were manifested in the increased performance of the subject, when compared with the economy shown in searching sets of icons that did not afford the superimposition of classificatory schemas (although featuring the same number of icons).

However, strong evidence for spontaneous classification was not found. Whilst the fact that subjects classified above chance level (even in the first trials of each of the multiplexed versions of the core series) can be considered as evidence that some form of transfer had occurred, the phenomenon was not strong enough to warrant the claim that the subjects spontaneously generalized the expertise acquired in the core series to the multiplexed versions of it. On the contrary, from the qualitative analysis of the errors, it emerged quite clearly that the subjects were simply extending what they had learned in the previous version of the task to the conditions where more items were presented at each ordinal node, thus neglecting some of the items and performing forward errors along the sequence.

From preliminary analyses of reaction times (and from the first results obtained from the analysis currently in progress, mentioned in the post-script in chapter VI), it seems evident that the subjects were chunking the set of icons on the basis of the classes in which it could be organized, as was observed for children (McGonigle and Jaswall, 1994). Thus, in this last experiment emerged again a similar profile for monkeys and children (see McGonigle and Jaswall, 1994) in both error types and temporal aspects of performance.

However, when the spontaneity of classification was compared, major differences were found between the two species. While children do not need to learn to chunk items together, monkeys seem to do so only if given enough task practice and incentive.

These findings strongly support the hypothesis put forward by McGonigle (1984) that one of the most important cognitive dimensions (and thus along which different species might meaningfully be compared) is the extent to which a particular organism is able to individuate, on the one hand the limitations of its cognitive resources, and, on the other, the relevance of their strategic use in different tasks. In other words, the ability to deploy strategic skills and to self-regulate, a competence that, as put forward in the introduction, has recently received much attention as an index of cognitive growth in the context of human cognitive development (Flavell and Wellman, 1977; Brown and Deloache; 1978; McGonigle and Chalmers, 1992).

In fact, as for the previous paradigms, this last experiment proved the adequacy of the present research program. Once again, there seem to be enough similarities in the behaviour of monkeys and children to make the matching of some human cognitive skills onto animal models plausible, while the emergence of both quantitative and qualitative differences between species (human and non-human in this case) shows that the tasks are tapping into non-trivial aspects of cognition which might help its characterization.

To summarize, this thesis has shown that it is possible to study self-regulation in non-human subjects. Forms of organization of external cues and changes in expertise emerge in capuchin monkeys in absence of any artificial feedback imposed by the experimenter.

The focus on search proved fruitful as a general framework, allowing at the same time: 1) the use of well defined measures of efficiency for quantitative analyses; 2) the observation of those strategies which the subjects selected from a rich set of qualitatively different possibilities; 3) the way in which they can be combined together and their relation with efficiency; 4) how they are achieved; 5) and how much they are exportable to similar problems.

Moreover, the paradigms implemented in this thesis allow the manipulation of external cost factors in order to see to what extent an organism is able to rearrange adaptively its competences. On the basis of this sort of information it becomes possible to asses on which grounds different organisms converge and, in contrast, where they find different solutions to a particular problem, and how effective (in terms of the trade-off between efficiency in a particular task and exportability to different tasks) these solutions are.

Detached from the criterion of comparative relevance, and at a more general and conceptual level of analysis, one hopes that this research has set the basis for a novel theory of cognitive change and might ultimately answer ambitious questions concerning what cognition should be considered about. This latter statement should not be taken as mere rhetoric since it is far from being clear what the term cognition stands for in animal psychology.

Sometimes we have witnessed a simple restatement in cognitive terms of old concepts mutuated from traditional associative learning theories. For example, in a most recent review of current researches, with the ambitious title of "Comparative Cognition: Beginning the Second Century of the Study of Animal Intelligence", Wasserman (1993) speaking in praise of the discipline, declared that recent developments in human cognitive psychology have again made the study of animal intelligence central to a traditional goal of psychology. This goal, in his opinion (borrowed from Gluck and Bower, 1988) is to view complex human abilities as emerging from a configuration of elementary associative processes that can be studied in simple organisms.

Sometimes the effort of those who advocate cognitive approaches has been dedicated to demonstrate the ability of animals to form internal representations, without however having the possibility to go further and characterize the nature of those representations and the processes which act upon them.

When a differentiation between different levels of representative skills has been attempted, often the research has been biased by what Brown and Deloache (1878) have individuated as a fatal error in developmental psychology too, namely the selection of unrelated tasks each designed to assess the presence of particular competences. Results from animal studies for example have answered questions such as whether a particular organism possessed the capacity to form concepts at different levels of abstractions (e.g. Hernstein, 1984; Premack, 1976); whether it forms internal representations of lists of items in a serial learning task (D'Amato and Colombo, 1988); or whether it possesses counting abilities and uses numbers in a representational way (Boysen and Bernston, 1989); or in a Piagetian perspective, what stage it might reach on an object permanence scale (Antinucci, 1989). In all these cases the underlying assumption was that an animal either possessed or did not possess a particular competence or, going a little further, that it was possible to assess the degree of development of that particular competence and the sort of representations that supports it. As pointed out by McGonigle (1984, 1991) the question of what the function of a particular competence is, when the cognitive system is considered as a whole, has been very rarely posed.

The present research, by contrast, tapped onto the understanding of how different capacities (inferential skills, visual scanning and detection of spatial constraints, categorisation skills and motor planning) might converge on the goal of making a potentially demanding task (serial, non-reiterative exploration of a large set of alternatives) relatively easy to solve. In other words on the issue of strategical intervention over a problem space.

As pointed out by Brown and Deloache (1978), even experiments which focused on similar topics have traditionally been confined to the crude assessment of the presence or absence of strategic intervention. Such an approach would inevitably lead to a dead end, once this basic question is answered. The research that has been presented here seems to go much further because it brought to light a rich set of intermediate stages of strategic control over behaviour.

On these grounds, following McGonigle's theory as outlined above, I would propose that a profitable way of characterizing cognitive change might be to consider it as the product of gaining increasing control over different competences. This would ultimately lead to a flexible management of internal resources whenever the organism is presented with problematic situations to solve.

In the following section the way in which the natural progression of the present research program is currently being implemented and possible future developments of it will be outlined.

## 7.2 Indications for further research

As mentioned at different points of this thesis the work described here is part of a vast project stemmed from the research conducted by McGonigle and colleagues on the mechanisms underlying serially ordered behaviours in human and non-human primates. The project is still in progress and here some current implementations of it will be delineated, which represent the natural progression of the program inaugurated with the present study.

The classification studies presented in this thesis have dealt with disjunctive classes composed each of identical icons. It has now been assessed that the subjects are able to classify and order classes according to particular sequences. Moreover, there is a sufficient number of subjects that is able to deal with sequences of-up to nine items (see post-script at the end of chapter VI). Therefore, with such a long serial production, supported by some form of chunking, it is now possible to manipulate the composition of the set in various ways and proceed to the second step of the program as proposed by McGonigle (1987; 1994). This consists in dividing the classes in different ways (for example icons might be non identical within each class, sharing only one common feature on the basis of which they can be distinguished from the members of the other classes). In this way the classes can then be organized in a hierarchical fashion, to see the extent to which the subjects would also be able to exploit also the constraints afforded within each class to organize their serial production.

Within each class, moreover, it is possible to transform the icons according different linear dimension in order to provide the subjects with the opportunity to organize their search behaviour according to seriation principles. As specified by McGonigle (1994) these transformations can be performed along different types of dimensions such as <u>rotation</u> (e.g. one of the icons within a class is presented in frontal position, another rotated of 45 degrees, another of 90 degrees and so on), <u>size</u> (transformations based on percentage of area change), <u>outlines</u> (each level of the transformation is based on the thickness of the outline) and many others. Then, it would be possible to present linear transformations between classes so that, for example class A (say squares) has items larger then class B (say circles) and class B has items larger than class C (say triangles). Within each of the classes the opportunity to seriate can be offered by manipulations of other dimensions (such as rotation, thickness etc.). This would provide the possibility to organize a directional search based on categorisation and seriation which demands very little memory effort to explore exhaustively a large set of icons.

The implementation of this procedure is already in progress, and started with the subject who reached criterion on a twelve items set. In order to construct a set of hierarchically nested classes the multiple disjunctive features currently defining the classes are progressively being eliminated within the categories. Presently one subject has reached criterion on a set of nine icons where colour has been eliminated and the classes are now distinguishable on the basis of their shape and texture alone.

For the other subjects, the results obtained from the last experiment presented here are currently being integrated with the collection of novel data and additional analyses. Although, in fact, from the results reported in this study, a lack of spontaneous generalization from "core" lists to multiplexed versions of it emerged, the question concerning what would have happened giving the monkeys even more practice with multiplexed versions of the list, (before confronting them with sets including more items at each node) was still an open one. It was still possible that, at the end of the study, the subjects had not yet reached the sort of expert solution of the task that allows its generalization at different levels of complexity. Indeed, the data currently being collected, as mentioned in the post script to Chapter VI, support this hypothesis.

From this consideration stems another interesting variation of the paradigm. This would be concerned with a different type of transfer which has not been explored in this study. What in this study has been evaluated, is the extent to which the subjects spontaneously generalize the expertise learned on core lists of items to conditions where each ordinal node is multiplexed. The exploration of the ability to transfer the expertise acquired in a particular series of ordered classes to other series, featuring the same structure and size of the search space but different icons (as in the classical learning set, but without the limitations of its binary version featuring unrelated items) would be worth undertaking.

It would be of particular interest to compare these two forms of transfer. In fact, the relationship between the level and the modality of acquisition of a particular strategy (expressed for example by the degree of transfer between "core" sequences and their multiplexed versions) and the transfer of knowledge from one task context to another (for example sequences featuring novel elements) would be of outstanding value for the understanding of cognitive change. Far beyond the mere assessment of behavioural mastery of a particular task, it would give the opportunity to evaluate the cognitive objectives which a system satisfies when it is taught to use a particular strategy and (by contrast) when the strategy is acquired by means of self-discovery.

Moreover, this modification of the paradigm would make it possible, in a successive experimental phase, to teach the subjects to join together different lists (for example just teaching them to order in succession the last item/class of a sequence and the first one of another). This would open the way to the evaluation of their capability to deal with very long serial productions (for a similar proposal, however based on core lists only, see Swartz et al., 1991) once they have been meaningfully organized in different types of chunks (each of the sequence per se, according to Terrace[4] can be considered a type of chunk, and the classes, as shown in this study, another).

---

4 See Terrace and McGonigle, 1994.

Furthermore, having subjects proficient in dealing with different sequences, organized in classes, it would be possible to mix the different series, preserving only the ordinal position of each class. This procedure has been used by Swartz, Chen and Terrace, albeit with series composed of just one item for each ordinal step. The paradigms featured in this thesis would give the opportunity to work with subjects dealing with classes of items, to which is attached an ordinal value. This might open the way to the study of competences which share some analogies with human language (where classes linguistic units, defined at different levels of abstraction are ordered according to grammatical rules). The approach might ultimately prove more profitable than those mentioned in chapter one for the understanding of the actual syntactic competences of non-human primates.

Apart from the natural extensions of the present paradigm, the apparatus and procedures featured in this study can be used for the implementation of several related ones. An example would be a paradigm proposed by McGonigle (see McGonigle and Chalmers, 1986) based on matching to sample procedures, featuring sets of multiple icons as opposed to the classical binary version of the task. The subjects would be presented with a template consisting of variable numbers of items. Following a delay, they would then be required to identify the template in a larger configuration containing it as a sub-set. Given that the template is composed of multiple items, the subject would have to select an order of report. The spontaneous ordering profile selected by the subject would then reveal the way in which the stimuli were organized in memory. Different ways of organizing the set can be then put in relation with the accuracy of retention.

Quite detached from the possible implementation of new paradigms, an interesting enterprise would be to test different samples of subjects on exactly the same tasks employed for the present study.

Since the ability to self-regulate is subject to developmental trends (Flavell and Wellman, 1977), and from the results of all the experiments described in this thesis monkeys and 4 year old children overlapped in some aspects or others of their performance (McGonigle et. al, 1992; McGonigle and Jaswall, 1994), an obvious continuation of the present study, would be to present the same tasks to younger children to see whether the gap between human and non-human competences can be reduced even more. This would provide valuable information about the relationship between ontogenesis and phylogenesis, a central issue in both comparative psychology and evolutionary biology (Gould, 1977). For the same reason, it would be interesting to collect data with apes (i.e. species even more related with humans) using the same paradigms.

In summary, the present study aimed at developing and implementing novel experimental apparatus, paradigms and procedures, to intrude in a new territory of comparative cognition and survey its potential resources. The mission apparently was accomplished and the province seems to promise rewards for a cohort of novel experimental enterprises.

# REFERENCES


Antinucci, F. (1989) (Ed.). Cognitive Structure and Development in Non-Human Primates. Hillsdale, N.J.: LEA.

Bates, E. (1976). Language and Context: The Acquisition of Pragmatics. New York: Academic

Berthental, B.I. & Fischer, K.W. (1983). The development of representation in search: A social-cognitive analysis. Child Development, 54, 846-857.

Boakes, R. (1984). From Darwin to behaviourism. Cambridge: Cambridge University Press.

Boysen, S.T. & Berntson, G.G. (1989). Numerical competence in a chimpanzee (Pan troglodytes). Journal of Comparative Psychology, 103, 23-31.

Bower, T.G.R. (1974). Development in infancy. San Francisco: W.H. Freeman and Company.

Bruner, J. (1983). Child's Talk. New York: Norton.

Breslow, L. (1981). A Re-evaluation of the literature on the development of transitive inference. Psychological Bulletin, 89, 2, 325-351.

Brown, A. L. & DeLoache, J.S. (1978). Skills, plans and self-regulation. In: R.S. Siegler (Ed.), Children's Thinking: What Develops? Hillsdale, N.J.: LEA

Bryant, P.E. & Trabasso, T. (1971). Transitive inference and memory in young children. Nature, 232, 456-458.

Chalmers, M. & McGonigle, B.O. (1984). Are children any more logical than monkeys on the five term series problem. Journal of Experimental Child Psychology, 37, 355-377.

Chalmers, M. & McGonigle, B.O. (1993). An experimental analysis of size seriation skills in children. Quarterly Journal of Experimental Psychology (in press).

Cheney, D.L. & Seyfart, R.M. (1992). The representation of social relations by monkeys. In: C.R. Gallistel (Ed.), Animal Cognition. Cambridge, Massachussetts, pp. 167-196.

Chomsky, N. (1959). Review of B.F. Skinner, Verbal Behavior. In Language, 35, 26-58.

D'Amato, M.R. & Colombo, M. (1988) Representation of serial order in monkeys (Cebus apella). Journal of Experimental Psychology: Animal Behaviour Processes, 14, 2, 131-139.

D'Amato, M.R. & Colombo, M. (1989). Serial learning with wild card items by monkeys (Cebus apella): implications for knowledge of ordinal position. Journal of Comparative Psychology, 103, 3, 252-261.

D'Amato, M.R. & Colombo, M. (1990). The symbolic distance effect in monkeys (Cebus apella). Animal Learning and Behaviour, 18, 2, 133-140.

De Lillo, C. & Visalberghi, E. (1994). Transfer index and mediational learning in tufted capuchin monkeys (Cebus apella). International Journal of Primatology.

Dickinson, T. (in preparation). Phd Thesis University of Edinburgh.

Darwin, C. (1871). The Descent of Man and Selection in Relation to Sex. London: Murray.

Diamond, A. (1985). Development of the ability of use recall to guide action, as indicated by infants performance on AB. Child Development, 56, 868-883.

Drozdall, J.G. Jr., & Flavell, J.H. (1975). A developmental study of logical search behavior. Child Development, 46, 389-393.

Fischer, K.W. & Jennings, S. (1981). The emergence of representation in search: understanding the hider as an independent agent. Developmental Review, 1, 18-30.

Flavell, J.H. & Wellman, H.M. (1977). Metamemory. In: R.V. Kail & J.W. Hagen (Eds.), Perspectives on the Development of Memory and Cognition. Hillsdale, N.J.: LEA.

Gallistel, R.G. (1990). The Organization of Learning. Cambridge, Massachusetts: MIT Press.

Gardner, R.A. & Gardner, B.T. (1969). Teaching sign language to a chimpanzee. Science, 165, 664-672.

Gardner, B.T. & Gardner, R.A. (1971). Two-Way communication with an infant chimpanzee. In: A. Schrier & F. Stollniz (Eds.) Behavior of Non-Human Primates, Vol. 4, pp. 117-184. New York: Academic

Gardner, B.T. & Gardner, R.A. (1974). Comparing the early utterances of child and chimpanzee. In A. Pick (Ed.) Minnesota Symposia on Child Psychology, Vol. 8, 3-23. University of Minnesota Press.

Gardner, R.A. & Gardner, B.T. (1978). Comparative psychology and language acquisition. Annals of New York Academy of Sciences, 309, 37-76.

Gluck, M. A. & Bower, G.H. (1988). From conditioning to category learning: An adaptive network model. Journal of Experimental Psychology: General, 117, 227-247.

Gould, S.J. (1977). Ontogeny and Phylogeny. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Hamilton, G.V. (1911). A study of trial and error reactions by mammals. Journal of Animal Behaviour, 1, 33-66.

Haake, R.J. & Somerville, S.C. (1985). Development of logical search skills in infancy. Developmental Psychology, 21, 1, 176-186.

Harlow, H.F. (1949). The formation of learning sets. Psychological Review, 56, 51-65.

Harlow, H.F. (1959). Learning set and error factor theory. In: S. Koch (Ed.), Psychology: A Study of a Science, Vol. 2: pp. 492-537. New York: McGraw-Hill.

Harris, M.R. (1988). Phd Thesis, Universtity of Edinburgh.

Harris, M.R. & McGonigle, B.O. (1994). A model of transitive choice. Quarterly Journal of Experimental Psychology (in press).

Herrnstein, R.J. (1984). Object categories and discriminative stimuli. In: R.L. Roitblat, T.G. Bever, H.S. Terrace (Eds.), Animal Cognition, Hillsdale, N.J.: LEA, pp. 233-262.

Hilgard, E.R. and Bower, G.H. (1975). Theories of Learning. Englewood Cliffs, N.J.: Prentice-Hall.

Hunter, W.S. (1914). The delayed reaction in animals and children. Behavior Monographs, 2, Vol.6.

Inhelder, B. & Piaget, J. (1964). The Early Growth of Logic in The Child. London: Routledge and Kegan Paul.

Karmiloff-Smith, A. (1992). Beyond Modularity: A Developmental Perspective on Cognitive Science. Cambridge, Massachusetts: MIT Press.

Koehler, W. (1917/1976). L'intelligenza nelle Scimmie Antropomorfe (Italian translation of "The Mentality of Apes"). Firenze: Giunti.

Koehler, W. (1918/1938). Simple structural functions in the chimpanzee and in the chicken. Translated from German and condensed in: W.D. Ellis (Ed.), A Source Book of Gestalt Psychology, New York: Harcourt, Brace & World.

Kobasigawa, A. (1977). Retrieval strategies in the development of memory. In: R.V. Kail Jr. and Hagen J.W., Perspectives on the Development of Memory and Cognition, Lawrence Erlbaum Associates, Hillsdale N.J., pp. 177-201.

Krebs, J.R. (1981). Foraging. In: D. McFarland (Ed.) The Oxford Companion of Animal Behaviour. Oxford: Oxford University Press, pp. 214-217

Lashley, K.S. (1951). The problem of serial order in behavior. In L.A. Jeffries (Ed.), Cerebral Mechanisms in Behavior. New York: Wiley.

Levine, M. (1965). Hypothesis Behavior. In: M. Schrier, H.F. Harlow and F. Stollnitz (Eds.) Behaviour of Nonhuman Primates. Vol 1, pp. 97-128. New York: Academic Press.

Luger, G.F. & Stern, C. (1990). Expert Systems and the abductive circle. The University of New Mexico. Technical Report No. CS90-17.

Luria, A.R. (1987) Reductionism in psychology. In: R.L. Gregory (Ed.), The Oxford Companion to the Mind, Oxford: Oxford University Press, pp. 675-676.

Lyons, J. (1991). Chomsky. London: Fontana

Mackintosh, N.J. (1974). The Psychology of Animal Learning. London: Academic Press.

McGonigle, B.O. (1980). Sign, symbol and syntax in the language of apes. Nature, 286, 761-762.

McGonigle, B.O. (1984). Intelligence from the stand-point of comparative psychology. Paper presented at the Thyssen Philosophy Group. Matlock, September.

McGonigle, B.O. (1987a). Non verbal thinking by animals ? Nature, 325, 110-112.

McGonigle, B.O. (1987b). Inference and seriation in child and monkey: a decomposition of primate cognitive skill. Report in Medical Research Council of G.B.

McGonigle, B.O. (1989). An analysis of serial ordering skills in children. Project Grant R0002, 1989-1993, Economic & Social Research Council of G.B.

McGonigle, B.O. (1991). Incrementing intelligent systems by design. In: J. Meyer & S.W. Wilson (Eds.), Proceedings of the First International Conference on Simulation of Adaptive Behavior. Cambridge, Massachusetts: MIT Press, pp. 525-531.

McGonigle, B.O. (1994). Seriation Paradigm n. 4. Laboratory for Cognive Neuroscience Memo.

McGonigle, B.O. & Chalmers, M. (1977). Are monkeys logical? Nature, 267, 694-696.

McGonigle, B.O. & Chalmers, M. (1986). Representations and strategies during inference. In: T. Myers, K. Brown & B.O. McGonigle (Eds.), Reasoning and Discourse processing, pp. 141-164. London: Academic Press.

McGonigle, B.O. & Chalmers, M. (1992). Monkeys are Rational. The Quarterly Journal of Experimental Psychology. Section B. Comparative and Physiological Psychology. 45b, 3, 189-228.

McGonigle, B.O., De Lillo, C. & Dickinson, A. (1992). A comparative and developmental analysis of serially motivated organization in young children and Cebus apella. Paper presented at the V European Conference on Developmental Psychology. Seville, 6-9 Sept. 1992.

McGonigle, B.O. & Jaswall, V. (1993) Categorical Seriation in Children. (manuscript in preparation).

McGonigle, B.O. & Jones, B.T. (1978). Levels of stimulus processing by the squirrel monkey. Perception, 7, 635-659.

McGonigle, B.O. and Neapolitan, D. (1993). Quarterly Journal of Comparative Psychology (under review).

Menzel, E.W. Jr. (1974). A group of young chimpanzees in a one-acre field, In: A.M. Schrier & F. Stollnitz (Eds.), Behavior of Nonhuman Primates. New York: Academic, pp. 83-153.

Miller, G.A. (1956). The magical number seven plus or minus two. Psychological Review, 63, 81-97.

Miyashita, Y. (1988). Neural correlate of visual associative long-term-memory in the primate temporal cortex. Nature, 335, 817-820.

Moyer, R.S. (1973). Comparing objects in memory: evidence suggesting an internal psychophysics. Perception and Psychophysics. 13 (2), 180-184.

Moynahan, E.D. (1973). The development of knowledge concerning the effect of categorisation upon free recall. Child Development, 44, 238-246.

Murray, A.E. & Mishkin, M. (1987). Experimental studies of memory in monkeys. National Forum, Spring, 33-37.

Natale, F. (1989). Pattern of Object Manipulation. In: F. Antinucci (Ed.), Cognitive Structure and Development in Non-Human Primates. Hillsdale, N.J.: LEA, pp. 145-161.

Olton, S.O. (1982) Spatially organized behaviors in animals: behavioral and neurological studies. In M. Potegal (Ed.), Spatial abilities. Development and physiological foundations. New York: Academic Press, pp. 335-360.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. Memory and Cognition, 3, 6, 635-647.

Patterson, F. (1978). Conversations with a gorilla. National Geographic, 154, 4, 438-465.

Peirce, C.S. (1943). The Philosophical Writings of Peirce. New York: Dover Publications.

Piaget, J. (1955). The Construction of Reality in the Child. London: Routledge and Kegan Paul.

Povinelli, D.J., Parks, K.A. and Novak, M.A. (1991). Do rhesus monkeys (Macaca mulatta) Attribute knowledge and ignorance to others?. Journal of Comparative Psychology, 105, 4, 318-325.

Premack, D. (1976). Intelligence in Ape and Man. Hillsdale , N.J.: Lawrence Erlbaum.

Reese, H.W. (1964). Discrimination learning set in rhesus monkeys. Psychological Bulletin, 61, 321.

Reese, H.M. (1968). Perception of Stimulus Relations. New York: Academic.

Restle, F. (1958). Toward a quantitative description of learning set data. Psychological Review, 65, 77.

Romanes, G.J. (1882). Animal Intelligence. London: Kegan Paul.

Romanes, G.J. (1884). Mental Evolution in Animals. New York: Appleton.

Rumbaugh, D.M. (ed.) (1977). Language Learning by a Chimpanzee: The LANA Project. New York: Academic.

Savage-Rumbaugh, E.S. (1986). Ape Language: From Conditioned Response to Symbol. New York: Columbia University Press.

Salatas, H. & Flavell, J.H. (1976). Behavioral and metamnemonic indicators of strategic behaviors under remember instructions in first grade. Child Development, 47, 81-89.

Siegel, S. & Castellan, N.J. Jr. (1988). Nonparametric Statistics for the Behavioral Sciences (Second Edition). New York: McGraw-Hill.

Skinner, B.F. (1938). The Behavior of Organisms: An Experimental Analysis. New York: Appleton Century Crofts.

Skinner, B.F. (1953). Science and Human Behaviour. New York: Macmillan.

Skinner, B.F. (1956). A case history in scientific method. American Psychologist, 11, 221-233.

Skinner, B.F. (1957). Verbal Behavior. New York: Appleton Century Crofts

Skinner, B.F. (1974). About behaviourism. New York: Knopf (reprinted in Great Britain by Penguin Books, 1993).

Sophian, C. & Wellman, H. (1987). The development of indirect search strategies. British Journal of Developmental Psychology, 5, 9-18.

Somerville, S.C. & Capuani-Shumaker, A. (1984). Logical searches of young children in hiding and finding tasks. British Journal of Developmental Psychology, 2, 315-328.

Spence, K.W. (1942). The basis of solution by chimpanzee of the intermediate size problem. Journal of Experimental Psychology, 31, 257-271.

Spence, K.W. (1951). Theoretical interpretations of learning. In: S.S. Stevens (Ed.), Handbook of Experimental Psychology. New York: Wiley.

Spinozzi, G. & Natale, F. (1989). Classification. In: F. Antinucci (Ed.), Cognitive Structure and Development in Non-Human Primates. Hillsdale, N.J.: LEA, pp. 163-187.

Sternberg, S., Knoll, R.L. & Wright, C.E. (1982). Control of rapid action sequences in speech and typewriting. Talk given by Saul Sternberg at Bell Laboratories, February 9.

Stillings, N.A., Feinstein, M.H., Garfield, J.L., Rissland E.L., Rosembaum, D.A., Weisler, S.E. & Baker-Ward, L. (1987). Cognitive Science: An Introduction. Cambridge, Massachusetts: MIT Press.

St Johnston (1993). Phd Thesis, University of Edinburgh.

Straub, R.O. & Terrace, H.S. (1981). Generalization of serial learning in the pigeon. Animal Learning and Behavior, 9, 4, 454-468.

Sugarman, S. (1983). Children's early thought: Developments in classification. Cambridge: Cambridge University Press.

Swartz, K.B., Chen, S. & Terrace, H.S. (1991). Serial learning by rhesus monkeys I: Acquisition and retention of multiple four items lists. Journal of Experimental Psychology: Animal Behavior Processes, 17, 396-410.

Terrace, H.S. (1979). Nim. New York: Knopf.

Terrace, H.S. (1984). Simultaneous chaining: the problem it poses for traditional chaining theory. In: M.L. Commons, R.J. Herrnstein & A.R. Wagner (Eds.), Quantitative Analyses of Behaviour: Discrimination processes. Cambridge, Massachusetts: Ballinger Publishing Company.

Terrace, H.S. (1985). In the beginning was the name. American Psychologist, 40, 1011-1028.

Terrace, H.S. (1987). Chunking by a pigeon in a serial learning task. Nature, 325, 149-151.

Terrace, H.S. (1991). Chunking during serial learning by a pigeon: I basic evidence. Journal of Experimental Psychology: Animal Behavior Processes, 17, 1, 81-93

Terrace, H.S. & Chen, S. (1991a). Chunking during serial learning by a pigeon: II integrity of a chunk on a new list, Journal of Experimental Psychology: Animal Behavior Processes, 17, 1, 94-106.

Terrace, H.S. & Chen, S. (1991b). Chunking during serial learning by a pigeon: III what are the necessary conditions for establishing a chunk ? Journal of Experimental Psychology: Animal Behavior Processes, 17, 1, 107-118.

Terrace, H.S. & McGonigle, B.O. (1994). Memory and Representation of Serial Order by Children, Monkeys and Pigeons,. Current Directions in Psychological Science, A.R. Gallistel (Ed.) (in press).

Terrace, H.S., Petitto, L.A., Sanders, R.J. & Bever, T.J. (1979). Can an ape create a sentence ? Science, 206, 891-902.

Terrace, H.S., Petitto L.A., Sanders, R.J. & Bever, T.J. (1980). On the grammatical capacities of apes. In: K. Nelsdon (Ed.), Children Language, Vol 2, 371-495. New York: Gardner Press.

Terrace, H.S., Straub, R.O., Bever, T.G. & Seideberg, M.S. (1977). Representation of a sequence by a pigeon. Bulletin of the Psychonomic Society, 10, 269.

Thorndike, E.L. (1911). Animal Intelligence. New York: MacMillan.

Trabasso, T., Riley, C. & Wilson, E. (1975). The representation of linear and spatial strategies in reasoning: a developmental study. In: R.J. Falmagne (Ed.), Reasoning: Representation and process in children and adults. Hillsdale N.J.: Erlbaum.

Visalberghi, E. (1990). Tool use in Cebus. Folia Primatologica, 54, 146-154.

Visalberghi, E. & Fragatszy, D. (1990). Do monkeys ape? In S. Parker & K. Gibson (Eds.), "Language and intelligence in monkeys and apes. New York: Cambridge University Press, pp. 247-273.

Wallmann, J. (1992). Aping Language. Cambridge: Cambridge University Press.

Warren, J.M. (1965). Primate Learning in Comparative Perspective. In: M. Schrier, H.F. Harlow and F. Stollnitz (Eds.) Behaviour of Nonhuman Primates. Vol 1. New York: Academic Press, pp. 249-281.

Wasserman, E.A. (1984). Animal Intelligence: Understanding the minds of animals through their behavioral "ambassadors". In H.L. Roitblat, T.G. Bever, & H.S. Terrace (Eds.), Animal Cognition. Hillsdale, NJ: Erlbaum, pp. 45-60.

Wasserman, E.A. (1993). Comparative Cognition: Beginning the second century of the study of Animal Intelligence. Psychological Bulletin, 113, 2, 211-228

Watson, J.B. (1913). Psychology as the behaviourist views it. Psychological Review, 20, 158-177.

Watson, J.B. (1914). Behaviour: an Introduction to Comparative Psychology. New York: Henry Holt.

Wellman, H.M., Somerville, S.C. & Haake, R.J. (1979). Development of search procedures in real-life spatial environments. Developmental Psychology, 15, 5, 530-542.

Winer, B.J. (1971). Statistical Principles in Experimental Design. 2nd edition. New York: McGraw-Hill.

Wright, C.E. (1990). Controlling sequential motor activity. In: D.N. Osherson, S.M. Kosslyn & J.M. Hollerbach (Eds.), An Invitation to Cognitive Science, Vol. 2: Visual Cognition and Action. Cambridge Massachusetts: MIT Press.

Wright, P.C. (1985). Phd Thesis, University of Edinburgh.

Yerkes, R.M. (1917). Methods of exhibiting reactive tendencies characteristic of ontogenetic and phylogenetic states. Journal of Animal Behavior, 7, 11-28.

Yerkes, R.M. & Yerkes, A.W. (1929). The great apes: a study of anthropoid life. New Haven: Yale University.

# APPENDIX A

## SOME ADDITIONAL DATA AVAILABLE IN THE LITERATURE ON SDE IN MONKEYS

Following McGonigle and Chalmers (1977; 1986) and in the attempt to identify the basis of possible representation of linear order in monkeys, D'Amato and Colombo (1990) tried to elicit the Symbolic Distance Effect (SDE) in capuchin monkeys (Cebus apella).

D'Amato and Colombo assumed that if the SDE was found in monkeys performing a serial learning task, then the effect could have been taken as evidence for the fact that their performance was based on a spatial representational device.

In a first experiment, D'Amato and Colombo (1990) trained their subjects to perform an arbitrarious sequence of responses, using the conventional forward procedure described in Chapter II. Then, they tested the monkeys on all the possible pairings of the five terms and registered their response time for pairs characterized by different distances between the two stimuli (such as, AB vs AC vs AD vs AE). The monkeys failed to show a Symbolic Distance Effect.

Interestingly, the effect appeared in their second experiment, where monkeys were trained on a task derived from the study of McGonigle and Chalmers on transitivity (McGonigle and Chalmers, 1977). In this second experiment the monkeys were trained on couples of adjacent items (AB, BC, CD, DE) and then tested, as usual, on all the possible pairings of the five terms.

In order to account for the differences in the results of their two experiments, D'Amato and Colombo were forced to assume that the subjects had at their disposal two different sorts of processes to deal with the serial organization of behaviour. One is the classical associations chain, which they claim would account for serial learning, and the other is the spatial paralogical device which can explain the appearance of the Symbolic Distance Effect when subjects are trained on pairs of conditional discrimination (mutuated from an ordered series). They had also to accommodate this latter interpretation to their previous accounts of the results from serial learning experiments and those using wild cards (see chapter II, section 2.2).

An interesting conclusion which can be drawn from this study is that capuchin monkeys seem to develop more sophisticated forms of representations in situations where they have to find by themselves the relationship between pairs of items belonging to a non-arbitrarily ordered sequence. By contrast, in a situation where the sequence is arbitrarily defined by the experimenters and rigidly thought step by step, they do not spontaneously devise such a type of organization.

It is also interesting to note that D'Amato and Colombo obtained the Symbolic Distance Effect only after the second of two training/testing cycles. The authors interpreted this finding as follows:

"It might be possible that the 9 training and 10 testing sessions of their first cycle might non have been sufficient to completely divert these subjects from the associative chain mode of processing, which was the rationale for the second training/test cycle" (D'Amato and Colombo, 1990, p. 137).

Thus, overall, it seems not only that items which can be organized in meaningful (non-arbitrary) sequences allow the subjects to construct sophisticated forms of representation of their serial production, but a protracted exposure to the task is required in order to change the form of representation and the processes which act upon them.

# APPENDIX B

## THE RANDOM PROBABILITY MODEL

(quoted by permission from St Johnston, 1993).

Assuming that each stimulus is chosen at random at each point in the trial, what is the probability that the agent will receive the reward (i.e. have exhausted the set of stimuli) on turn T ? The key to working this out is to forget that the trial ends when the set is exhausted and assume that the trial just continues on and on.

A(N, M): The event that in N goes the agent will have touched all M stimuli.

If we know this, then we can work out the probability that it gets the reward on turn T, because this is simply p(A (T, M) - p(A (T - 1, M)), the probability that it has touched them all by turn T minus the probability that it has touched them all already by turn T - 1.

To solve this consider the number of sequences which the agent can make to satisfy A(N, M).

The total number of possible sequences is $M^N$, but we have to subtract off this number of sequences which don't involve touching all M stimuli. Any sequence will touch precisely i stimuli for some i in 1, 2, ...., M. Let us call the number of stimuli a sequence involves (touches) its index.

So the index of

x x x

is 1, while the index of

x y

is 2.

Now, fix the length of the sequence (N above). From all sequences of length N, we want to find those with index M. This is all sequence of length N less those with an index of M - 1, M - 2,..., 3, 2, or 1. (Since every sequence must have some index in 1,..., M).

But the number of sequences of length N with index M - 1 is simply the number of sequences satisfying A(N, M - 1), and so on. Well, that's almost true. We need to allow for the fact that if we have M stimuli then here are many different sets of M - 1 stimuli the sequence could have involved. In general if we consider A(N, m) there are $\binom{M}{m}$ sets of stimuli. Since no sequence with index m can be in any two of the sets of m from M stimuli (because ALL of the stimuli have to be included in the sequence) we count no sequence twice in this analysis.

So, the total number of sequences with index m we have to subtract is $\binom{M}{m}$ A(N, m).

So, summing over all the indexes we have to take away from the original number of sequences, we find

$$A(N, M) = M^N - \sum_{M=1}^{M-1} \binom{M}{m} A(N, m).$$

This is the total number of sequences with index M.

The probability of having won by turn N with M stimuli (assuming we carry on playing after we win) is simply

$$\frac{A(N, M)}{M^N} = p(N, M)$$

and after a little bit of maths this boils down to the following recurrence relation:

$p(N, M) = 0$ for $M > N$

$p(N, 1) = 1$ for $N > 0$

$$p(N, M) = 1 - \sum_{M=1}^{M-1} \binom{M}{m} \frac{m^N}{M^N} p(N, m) \text{ otherwise.}$$

As mentioned before, the probability of winning on turn T is simply $p(T, M) - p(T - 1, M)$.