13

THE GROWTH OF COGNITIVE STRUCTURE IN MONKEYS AND MEN

Brendan McGonigle and Margaret Chalmers
University of Edinburgh

BACKGROUND: AN UNBRIDGEABLE GAP IN THE STUDY OF COGNITIVE FUNCTIONING

There is a widespread view that the sorts of animal learning mechanisms most frequently studied in the laboratory are inductively too weak and unproductive to generate the kinds of behaviours expressed in higher order forms of human cognitive and linguistic adaptation [(Chomsky, 1980; Fodor & Pylyshyn, 1988; Piaget, 1971). One reason for this (Harlow, 1949) is that investigations are rarely followed through from one learning episode to another to assess the cumulative benefits (if any) as a function of the agent's task and life history. Yet the course of human development is protracted, and even sophisticated adult subjects frequently show dramatic changes in strategy when confronted with many problems of the same type, detecting pattern and structural invariance in some (e.g.(Wood, 1978), using analogies to bridge problems of a different surface structure (Gentner, 1983), and devising progressively economic, data reducing procedures to secure success with the least investment in resource (Anderson, 1990; McGonigle & Chalmers, 1986, 1998; McGonigle & Chalmers, 2001). Whilst Harlow's pioneering work on the learning set (LS) using primates is an exception, indicating the vast potential for accelerated learning, his generative claim that the LS leaves the organism free to attack problems of a new hierarchy of difficulty has never been properly realised due to the fact that all problems in conventional LS studies are of the same (simple binary) type and level of difficulty. Worse still, LS tasks conventionally involve arbitrarily chosen pairings of stimuli so that it is impossible in these circumstances for subjects to devise any rules of relation or stimulus classifier system which would connect one learning episode with another, despite the fact that relational codification and learning arguably constitutes a quite different type of adaptation with a different inductive and generative profile (Bryant, 1974; Dusek & Eichenbaum, 1997; Gazzaniga, Ivry, & Mangun, 1998; McGonigle & Chalmers, 1996; McGonigle & Jones, 1978; Wills & Mackintosh, 1998).

In the domain of human cognition, on the other hand, experimental paradigms heavily emphasise language based tasks often as their only means to access (or so it is perceived) complex language-like abstractions which can be operated only by symbolically competent agents (Russell, 1983). Thus thinking and reasoning tasks attempt to probe cognitive representation by giving subjects texts or verbal inference-drawing tasks from input such as "All of the bee-keepers are artists; some of the chemists are bee-keepers" (Johnson-Laird, 1983). Children have not been spared deep linguistically based probing either as in Piaget's celebrated tests of transitive reasoning, for example; "Edith is fairer than Lilli; Edith is darker than Suzanne, who is the fairest/darkest" and in his assessment of class inclusion competences in such tests as "are there more flowers or more daffodils ?" (Inhelder & Piaget, 1964). Beyond success, furthermore, justification of the answer is often a part of assessment as well (Karmiloff-Smith, 1992; Piaget, 1928), if only to help determine how explicitly the subject was aware of the **necessity** of the conclusion at an explicit propositional level of comprehension, yielding further insights into whether the answer was driven by the demands of a modal (logical) understanding rather than by empirical knowledge alone (Smith, 1993).

As a consequence, human cognitive psychology with its emphasis on abstract representational devices has left psychological and cognitive neuroscience with a huge credibility gap between bottom-up, knee-jerk portrayals of intelligence, and the sorts of organisation which support the highest achievements of man as portrayed in the laboratories of cognitive psychologists. When characterised as abstract, symbolic, and domain independent (see 'proof' based theories of representation (Fodor, 1983)), such architectures of the mind seem often to have no credible roots in earlier phylogenetic or ontogenetic adaptations. Whether concerned with representational format (see e.g. the analogue *versus* propositional issue (Pylyshyn, 1981)), structural organisation or problem solving (Anderson, 1990; Johnson-Laird, 1983), the currency of explanation starts from beyond the point at which most behaviour based learning accounts leave off, i.e. the currency of symbol manipulation.

The consequences for both cognitive theory and application are far-reaching. For if language is the only serious instrument of probing 'higher' intellectual functioning, where does this leave the psychology of the non-human, the developing human and the cognitively impaired? If impotent without “semiotic instruments and the like ", as Piaget (1971) puts it, to assess representational factors, a first casualty would be the investigation of cognitive functioning in non-humans and with it any prospect of a neurosciences model based on non-human research which could provide useful insights into high level deficits in executive memory and related subsystems—the dementias, for example. Such in-principle limitations would also apply to developmental approaches—especially those which are themselves informed by non-human comparisons (Chalmers & McGonigle, 1997; McGonigle & Chalmers, 1996, 1998; Terrace & McGonigle, 1994) and those focusing on transitions in human cognitive growth from its earliest stages (Fischer & Canfield, 1986). Finally, the clinician attempting to evaluate subtle cognitive deficits where language disorder itself may be a confounding factor; for example a syndrome such as Fragile X would be left 'high and dry' without adequate substitutes for language (and other symbol) based evaluations (Chalmers, 1998)

In this chapter, we provide a characterisation of cognitive systems from a comparative and developmental perspective, reviewing research that has led to our current focus on the dynamics of self organisation and the self-selection of procedures which provide the agent with progressively enhanced adaptive power with greater economy of resource (Anderson, 1990). Embracing many of the core phenomena such as linear and hierarchical organisation perceived by many as important indicants of high level brain organisation our aim has been to help establish a common currency of tasks and measures beyond the scope of knee jerk level intelligence, yet without pre-supposing linguistic competence.

In particular, we shall review a corpus of work on transitivity, seriation, and classification designed to evaluate both the power and primacy of relational competences, seen initially by James (1891) as central to an evolutionary shift from ‘habit of mind' mechanisms to the association of ideas by similarity (or the perception of relations (Boakes, 1984)). Whilst James saw the shift as a basic difference between human minds and those of ‘brutes’, Herrnstein (1989) provides a more accurate contemporary view:

> "abstract relations differentiate most sharply among animals. It has been noted that we see the largest gaps in comparative performance at the level of abstract relations... once a species has a foothold at the level of abstract relations, the possibilities are unbounded".

This switch of emphasis is also timely. For many now doubt that traditional learning approaches based on association principles are either the whole story or indeed any story at all (Gallistel, 1995). As Gazzaniga et al put it, ”.... when one is trying to understand how the brain enables learning, one must realise that there may be several mechanisms, not just one.” (Gazzaniga et al., 1998), p. 521.

Given the context of the forum which motivated this volume, furthermore, we shall also highlight some of the more immediate neurosciences implications and applications which now follow from our stance.

COGNITIVE STRUCTURE: RELATIONAL MECHANISMS AND THEIR ORIGINS

Relationally based organisation is undoubtedly a key player in supporting the structural properties which lie at the core of human cognitive organisation. As Hummel and Holyoak (2001) point out, “human thinking is structure sensitive.... in the sense that we can represent and reason about abstract relationships, appreciating the similarities and differences between the idea that “John loves Mary” and the idea that "Mary loves John" (see Fodor and Pylyshyn, 1988)”. In addition, a formidable weapon which makes human cognition both powerful and economic, is the human brain’s ability to exploit analogy (Gentner, 1983; Luger, 1994) using a common relational structure to bind otherwise related objects and events.

However, switching emphasis to relational rather than associative competences confronts the investigator of non humans with some serious, historically entrenched problems both conceptual and procedural, all of which have had a blighting influence on decades of comparative research. First is the long-standing assumption already noted that it is only at the abstract level of representation that such systematic, productive aspects of relational comprehension operate. Linked to this has been the notion that language may provide not only an essential window on cognition but is a key casual instrument in the determination of core cognitive matters. Indeed, Fodor and Pylyshyn (1988) claim that the systematicity of thought *follows from* the systematicity of language. Certainly, the most palpable evidence of relational competence in humans is found in the vocabulary of relational connectives such as comparatives and scalars (we **point** to objects; we **declare** relationships!). So it is perhaps unsurprising that relational differences between objects expressed by terms such as ‘bigger’ and ‘smaller’ once attracted the view that language itself provides the bridge or the mediating device which make such achievements possible.

Two factors reinforced this view. The first was that many demonstrations designed to demonstrate relational competence in non-humans, and based on a one-step transposition paradigm (Reese, 1968) were also interpretable in simple stimulus generalisation terms (Spence, 1937). The second factor inhibiting the development of relationally based learning theories was the emergence of language based ‘mediation’ theories to account for the development of such expertise in human cognitive development. Even within a behaviourist stance, some appeal to linguistic factors was made necessary because older children and adult subjects transposed relationally on what were known as 'far' tests i.e. to stimulus values outside the critical transfer range of Spence’s generalisation model. Advocated by Kuenne (Kuenne, 1946) and others, the idea was that a relevant linguistic label could come to act as an internal conditioned response (R_m) to the stimulus, itself then acting as mediated stimulus (S_m) for the final observable choice. The final chain is no longer S - R, but S - R_m - S_m - R. Kuenne argued that there were 'levels' of verbal behaviour itself in the context of discrimination learning. Only at later levels, she argued, do children's verbal responses come to control their choice behaviour, performing a role more like dimensional abstraction, even though, as Spence's student, she did not actually invoke such an overtly 'mentalistic' interpretation.

In discussing Kuenne's position, Bryant (Bryant, 1974) drew attention to the tautology noted by Lashley (1929). "The main trouble with the hypothesis that children begin to take in and use relations to help them solve problems because they learn the appropriate comparative terms like 'larger' is that it leaves unanswered the very awkward question of how they learned the meaning of these words in the first place." (page). A further problem for the mediation hypotheses was in determining whether or not language *per se* changes the 'level' of stimulus analysis. As late as 1972, Fein observed that there were at least four types of possible stimulus analysis which transposition research had failed to differentiate. A stimulus field, she argued, could be analysed as a succession of e.g. small and large objects, or as an array of objects, one of which is e.g. the large in the training pair, the small in the test pair. Higher order 'relational' types of analysis would be scalar in nature, i.e. a series of objects, in which e.g. A is bigger than B; B bigger than C. In a further refinement, such scalars may carry only ordinal information without representing the magnitude of differences between items thus scaled. A more advanced form would carry metrical information of the amount by which the items differed. As for ‘levels’, the logic of Kuenne's mediational analysis **had** to suggest that the controlling function of language was based on a grasp of the **scalar** nature of the dimension, necessary to give contrastive meaning to verbal labels such as ‘big’ and ‘small’. But if achieved by behavioural rather than linguistic regulation, the determiners of such relational achievements remain submerged under the dominance of associative learning stances. Yet if behavioural, it should be possible to show how changes in levels of stimulus analysis proceeds without benefit of linguistic codes.

PARADIGMATIC CHANGE IN TRANSPOSITION RESEARCH: ELIMINATING THE ONE STEP METHOD

Two sets of investigations, one with monkeys (McGonigle & Jones, 1978) and one with children (Lawrenson & Bryant, 1972), put some of these fundamental hypothesis to the test. In Lawrenson and Bryant ‘s study, four and six year old children were trained on one of two types of double discrimination task. In one case they were to choose the same relation such as 'bigger than'; in the other the same absolute size. Lawrenson and Bryant gave half their subjects a 'far' condition as one of the trained pairs (the absolute group were required to take the stimulus nearest in size to the original trained value). Under these conditions, there was no 'far' effect for the relational group, both the near and far training conditions were equally easy for both ages of child **and** both were substantially easier to learn than the absolute condition.

By finding such successful ‘far’ transposition under direct training conditions, Lawrenson and Bryant had neatly demonstrated a lower competence bound for relational responding in children than previous investigators relying on spontaneous transfer. The relative difficulty of maintaining an 'absolute' response under these same training conditions, moreover, now clearly suggested a more primitive **ontological** status for relational as opposed to 'absolute' codification. However, given that the subjects were human and were capable of covert if not explicit linguistically based augmentation of the test conditions, it was left to studies such as McGonigle and Jones (1978) to break the link with language in relational encoding which couldn’t be explained away as a stimulus generalisation effect.

Begun in 1969, the Edinburgh programme on relational learning also took a training stance. Informed partially by experiments on cats which showed that equivalence tests seriously underestimated positive transfer when compared to savings measures based on re-acquisition procedures, McGonigle and Jones ran a series of studies one of which we shall summarise here. Applying a training procedure which derived from Harlow’s LS procedure, but critically featuring conditions where the training stimuli varied relationally in size or brightness, McGonigle and Jones (1978) trained one group of squirrel monkeys to choose objects on a relational basis, and another group to choose on an absolute one. In the size studies, initial training stimuli were three white polystyrene cubes, which differed in size from one another by just over 1 cm.[2]. The three cubes were presented as pairs. These are designated BC and CD (as two further cubes were added later), where B is the largest, C the middle-sized and D the smallest. For the relational group, choice of item, B in BC and C in CD (i.e. the larger in each case) was rewarded with a peanut; for the absolute group choice of C in both cases was rewarded.

Whilst initial learning was relatively rapid for both groups, significant differences began to emerge with task changes. The first and most obvious one was the lack of stable retention shown by the ‘absolute’ stimulus condition. For example, removing the visible context within which the objects were viewed (the objects glowed with luminous paint) adversely affected this condition only. Whilst context, including egocentric based reference frames based on the sight of the hand and other means of calibrating size would be removed by darkness, a relational code based on object-object differences alone proved robust enough to allow the relationally trained group to transfer both from light to dark and dark to light conditions respectively.

Transferred to triads derived from the training set, furthermore, the relational group immediately chose largest, showing an elementary form of perceptual transitivity. The fact too that monkeys in the relational condition could predict on the *first trial* of any new problem which object should be selected (as distinct from forming a non-specific LS as absolute stimulus learners did), combined with the general robustness of their performance in retaining what they had learnt, led McGonigle and Jones (1978) to claim that relational encoding was a *design primitive* and quite separate in profile and thus not derived from, associative, absolute stimulus learning. Just as crucially, such relational codes did not depend on language; instead McGonigle and Jones had found a set of precursors to linguistic comparatives. But what sort of precursors? Was the simple binary, asymmetrical relational code (bigger than) as illustrated here merely the tip of an iceberg obscuring a whole network or system of relationships which have yet to be exhibited by new behaviour based research methods? If so exposed, moreover, would such relationships operate as an (installed) ensemble with strict rules of interdependency, reflecting a logical hierarchy of relational types? As in the domain of transitive relationships, for example, where, within a formal symbol system, A>B, B>C implies A>C ? Or in the principled, linear seriation of size? Or in hierarchical acts of classification based on the perception of similarities and differences? Or as staged expressions of relational competences derived from minimalist design primitives and forged though environmental interaction? Or as punctate competences only, operating pragmatically and unsystematically to suit local task circumstances?

These questions are of considerable contemporary interest, not only in the light of rapidly growing research on the role of the hippocampus and relational coding (Dusek & Eichenbaum, 1997) (Morris, Garrud, Rawlins, & O'Keefe, 1982), but also because the move to ground linguistic phenomena in the pragmatics of real world adaptation has also grown apace (H. H. Clark, 1973; Jackendoff, 1983). Even the most symbol driven of symbolic representationalists such as Fodor and Pylyshyn (Fodor & Pylyshyn, 1988) have been moved to suggest that a rat learning a black versus a white discrimination would be hooked into a more extensive relational system, involving, for example, a 'darker than' relation and its inverse.

Yet no criteria are provided by these authors which would enable a distinction between punctate from non-punctate relational code use in non-humans. This is hardly surprising. In animal learning, it is rare to find both experiments and procedures which enable a **concurrency** of operation of an ensemble of relational codes in the same subject at the same time. Yet this is a condition which crucially has to be met to help evaluate the role of any one relational code in the operation of another. With humans, although inevitably confounded by linguistic factors, the trajectory of acquisition of antonymic terms by children (big/bigger; small/smaller), for example, indicates a psychological hierarchy through the apparent dependency between relational codes which in a formal (logical) system would have equal status. Illustrated by the lag shown (E. Clark, 1973; Donaldson & Wales, 1970) between the comprehension and production of the unmarked (e.g. bigger) *versus* the marked term (e.g. smaller) in young children, a natural syntax of rule acquisition is suggested as Heidenheimer () has argued, where children who first understand, big(ness), next derive an intermediate stage of understanding 'not-big', before acquiring an understanding of the inverse (smaller than), an interpretation endorsed following experiments on mental comparisons using linguistic instructions by McGonigle and Chalmers (McGonigle & Chalmers, 1984).

That language is merely the expression of this developing competence rather than its cause, moreover is suggested by Clark (Clark, 1970; H. H. Clark, 1973) and others as based on the structure of the environment, rather than deriving from (deep) structures of language itself. Persisting in adult performance throughout the life cycle as shown by a consistent asymmetry in the speed with which adult human subjects make decisions following comparative questions featuring marked versus unmarked terms (Clark, 1969), such 'marking' effects were once thought to have a lexical origin. However, in a seminal paper Clark (H. H. Clark, 1973) rejected his earlier linguistically derived 'deep-structural' argument, arguing instead that such effects reflected instead a fundamental asymmetry in the organisation of the perceptual world—the P-space. In spatial locative terms, Clark argues, there is an egocentrically defined privilege for the upward and

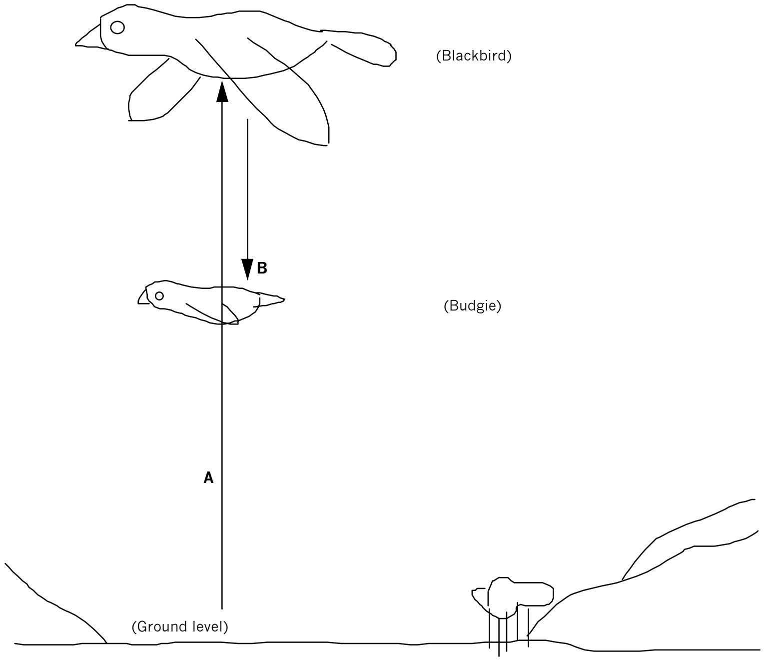


Figure 1. The perceptual origins of lexical marking (after Clark, 1973). The ground plane provides a natural reference for the primary direction A (higher). An object located along this dimension provides a secondary reference point for the derivative (marked) direction B (lower).

forwards direction as in above (below), over (under), or in front of (behind). A natural reference point for such directions is the ground—see Figure 1. However size and saliency of objects can also determine their status as natural reference points; objects on the horizon, the tallest or highest object in the visual field can form natural reference points for the judgment of size.

ORIGINS: LEXICAL MARKING AND THE P SPACE

This example, coupled with the trajectory profile for the growth of relational comprehension, neatly demonstrates the biological answer to the symbol grounding problem (Harnad, 1987). Biological systems do not make ‘abstraction’ mistakes! Instead, it would seem that ecological factors extrinsic to subsequent relational coding of sets of objects anchor the ensemble to the world from the outset.

Preliminary Criteria for Relational Systematicity

These related sets of phenomena now need to be taken to the 'proof 'stage if human psychological data on relational competence is to be used as indexical of systematic rule use by monkeys. Crucially too, as the human data is linguistico-relational, non-human competences in this sphere would unambiguously indicate a non-linguistic origin for them all. As for indicators, we shall take antonymic asymmetry as a prime 'marker' for dependency and system growth. And following that, a core question is the trajectory of relational acquisition. Does it have a natural syntax? If the rule for 'smaller than' derives from a negation of 'bigness', how is this expressed behaviourally? Beyond this case of possible rule dependency, furthermore, would a 'middleness' rule emerge, for example, only in subjects who had first acquired relational rules of both 'bigger' and 'smaller than'? Theoretical matters aside, how can assays of concurrent multiple rule, and thus possible system evaluation be achieved in the same subject and in the same test context without linguistic instruction? We address these matters in the next section.

RELATIONAL RULE SYSTEMS IN THE MAKING: PRIMITIVES AND DERIVATIONS

Conventional methods of training and testing animals are inappropriate to the requirement to assess concurrent relational use in the same subject. To mimic the effects of linguistic instructions, a conditional code was needed. We (McGonigle, 1987; McGonigle & Chalmers, 1980, 1986) decided to use a colour conditional code requiring subjects to select the 'larger' of two objects if (say) all were black, and the 'smaller' of a pair if all were white.

A. Lexical Marking Examined

Four squirrel monkeys (Brown, Blue, Green and White) from McGonigle and Jones (1978) were re-trained on pairwise size discriminations, using a colour conditional procedure (see Figure 2). The stimuli were wooden blocks, the same sizes as before (the sides of each face ranging from approximately 1.2 cm to 6.2 cm. with a 1.2 cm. interval difference). There were two such sets; one was painted white, the other black. For two monkeys, white denoted 'take larger', black denoted 'take smaller', and vice-versa for the other two. They were presented with the four adjacent pairs AB, BC, CD and DE in randomly presented trial blocks of five (first) and then ten, until they were 80% correct on each. In the following phase, all ten pairs deriving from the set were presented in

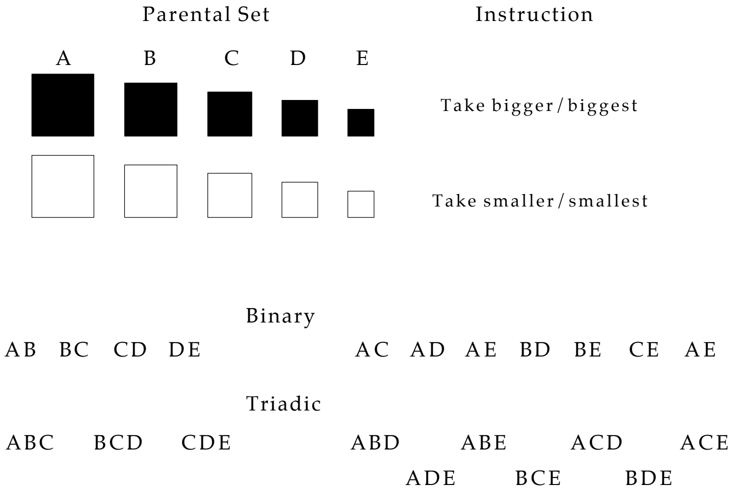


Figure 2. A colour conditional size rule training task used with monkeys (McGonigle & Chalmers, 1980, 1996).

randomised trial blocks of ten, until an 80% criterion was reached on each pair. Then the second, new instruction was learned in the same way using the other colour of stimuli.

Based on the finding from classic psychophysics that size discrimination difficulty can be measured using subjects' reaction time, we incorporated a psychophysical method of this sort, first to ensure that whatever the effects of instruction, that our RT measures would accurately reflect the relative discriminability of the pairwise combinations (10) which derived from the training sets given.

In the penultimate phase, trial blocks of 10 were given in which the instructions were alternated across blocks, until the monkeys could sustain high levels of accurate choice. Finally the instructions were randomised within trial blocks so that monkeys could predict neither instruction nor pair of objects on any trial.

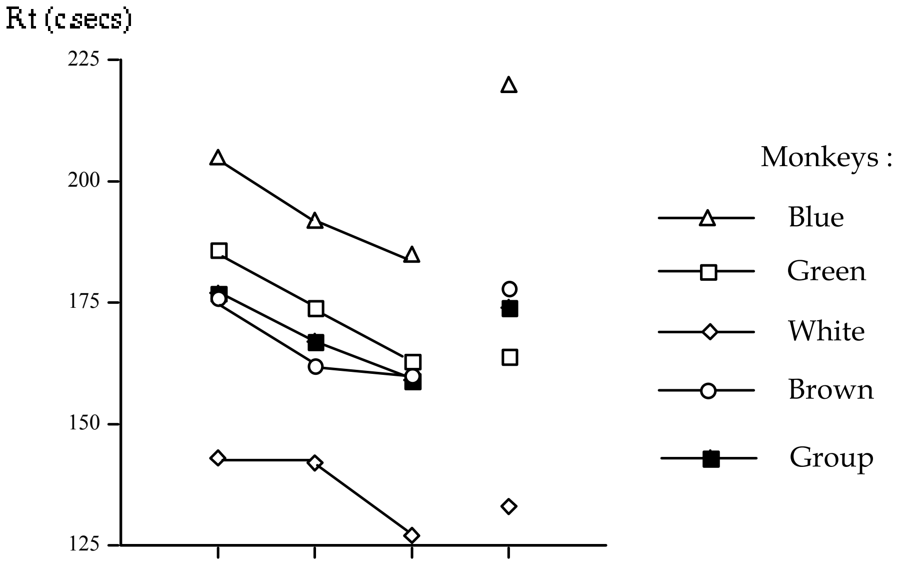


Figure 3. Psychophysical distance effects in the RTs for binary size comparisons obtained for individual monkeys, where distance is measured in steps (1 step represents the adjacent pairs).

Results

Acquisition

All monkeys completed training. Re-training on their original larger rule was relatively rapid, averaging 98 trials per pair on the initial phase. The 'smaller than' relation was learned even faster: only 39 trials were required on average. Stable performance involving both instructions in alternation was acquired with few errors and sustained at over 90% levels of accuracy for both instructions and for all comparisons.

RT Assay

A minimum of ten decision times recorded per instruction for each of the ten pairs during the final phase of testing was analysed. Analysis is based on correct choices only. There was a significant effect of psychophysical distance (measured as steps) for each subject as Figure 3 shows.

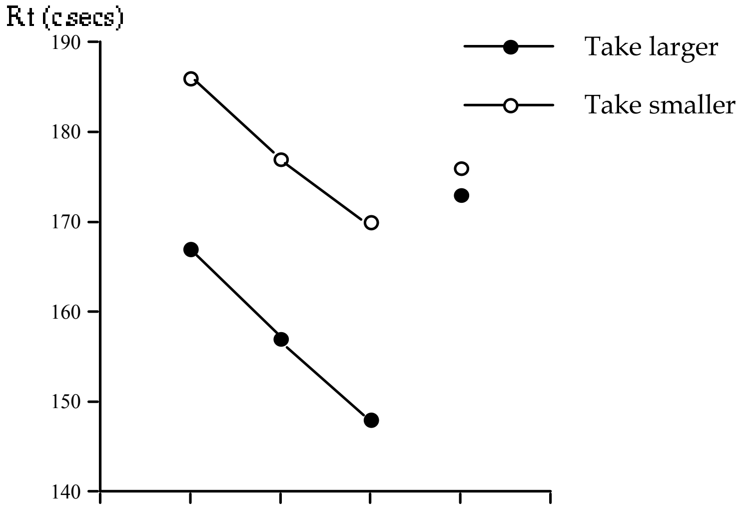


Figure 4. A marking effect in the RT data from monkeys.

As Figure 3 indicates, all monkeys show a significant inverse linear relationship between ordinal separation for all but the last item. Apart from the dramatic upturn on the last pairing (based on the two end items A and E which children also show (McGonigle & Chalmers, 1984), monkeys show profiles even at the level of the individual which are similar to those found with human subjects in memory experiments (Moyer, 1973).

Second as Figure 4 shows, there was a highly significant ‘lexical’ marking effect at the level of reaction times for the group as a whole and for each individual there was a significant effect of instruction in favour of the ‘unmarked’ instruction. This effect was unaffected either by practice or by psychophysical salience which did not interact with these variables, indicating that the marking effect is an access effect.

Third, there was an interaction between instructions and the 'ends’ of the continuum found also in the memory search data for similar sorts of task with adult humans (Banks, 1977). This was indicated in the monkey data by an analysis in which the data are plotted *qua* distance functions, but from each end of the series respectively as illustrated in Figure 5. Here we can now assess the distance effects from end points only. As can be seen, the 'take larger' instruction produces a much more pronounced distance effect when plotted from the small end. Whilst the smaller than instruction also produces a more marked distance effect from the small end, they do not decrease so dramatically as distance increases. In short, the marking effect is not only an instructional effect but one which is related to a fundamental series asymmetry favoring the 'big' end of the physical continuum.

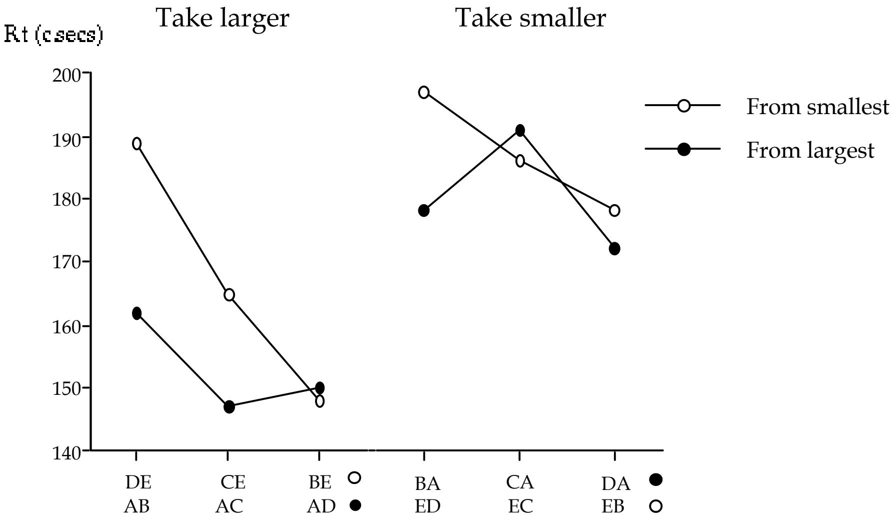


Figure 5. Asymmetrical interaction of instruction with distance effect in the RT data from monkeys' size comparisons.

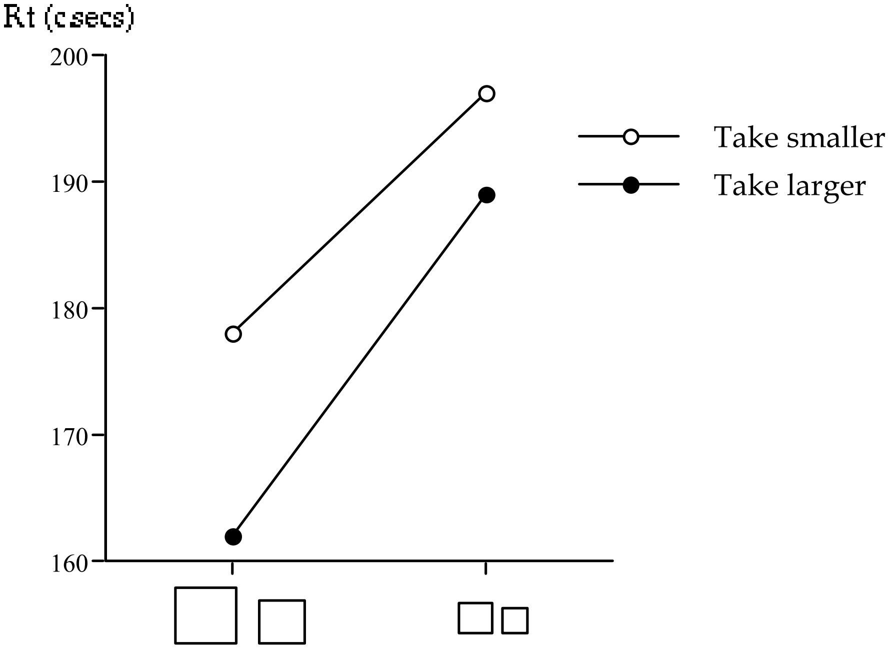


Figure 6. "Ends" by instruction asymmetry in the RT data from monkeys.

This is most clearly represented in an excerpt (Figure 6) from the data used above and which represents the variation caused by the ordinal position of test items (now taken from the extreme ends of the set).

Conclusions: 'Lexical' Marking

Without relying on acquisition data, but performatives instead, and based on steady state phases of performance at that, we can map both the macro and the micropattern of simian performance unto that of the human to an extensive degree. Crucially, we have shown not only that monkeys can operate two rules of relation concurrently in the same experiment, but that there is evidence of the marking effect used in human research to suggest a dependency between the operation of one rule and the operation of its inverse. The marking effect is clear and is undiminished either by practice or psychophysical salience. Showing impressive and parallel linear functions at the macro level, however, the micro-assays show how one end of the series anchors all decisions, such that the further away test stimuli are from the big end of the set, the slower choice is for both instructions.

We interpret these effects of instruction in the monkey data therefore, as strong evidence of asymmetrical, and unidirectional processing of the size dimension, similar to those we have obtained with children as old as nine years in memory experiments when asked comparative size questions about real world objects (McGonigle and Chalmers, 1984). Nevertheless, the implementation by the monkey of the 'smaller than' instruction was accurate and consistent. Encouraged, therefore, by the success of the conditional learning procedures, together with the gain in information obtained when more than one relation could be tested using a concurrent procedure, we next extended the method to 3 rules (middle) then to a set of 5 rules, where every size in the set we used had to have an independent code.

B. Introducing Middle-Sized

First, triads were presented, derived from the training set used above, and all subjects tested on bigger/biggest; smaller/smallest using the same instruction codes. As in the McGonigle and Jones (1978) study, monkeys generalised to the triads readily, showing once again elementary perceptual transitivity for both directions of instruction. Then a third instruction was introduced. The colour red now signified the operation of the 'middle' rule. Training on middle sized was lengthy. That monkeys required around 100 trials per triad before reaching criterion, shows that it was relatively costly for the monkeys to acquire this internal rule. However, and crucially, the monkeys were all successful, suggesting that their

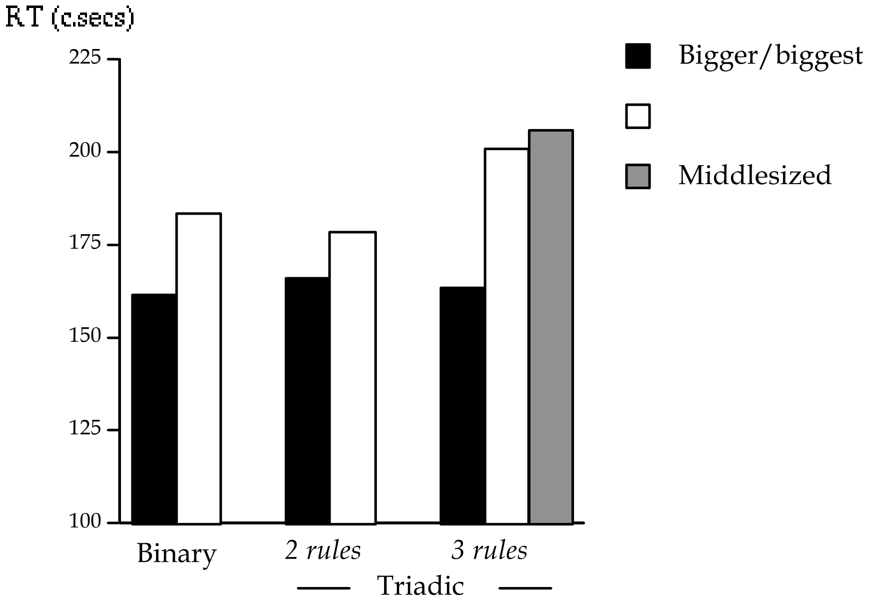


Figure 7. Mean RTs for monkeys as a function of instruction and the number of response categories required by a size comparison task.

previous failure in the McGonigle and Jones (1978) study was because they had lacked prior training on both comparatives ‘bigger’ and ‘smaller’.

As before, a steady state phase was then introduced for purposes of measuring RTs in the context of correct performance. Figure 7 below shows these for the triplet instruction set and compares them with those obtained under double instructions from the earlier two stimulus phase and the later triadic phases.

The first finding to emerge from the comparisons depicted in Figure 7 is that whereas the triadic condition *per se* did not itself produce any increase in mean RT, the triple instruction did have an effect on RT selective to the smallest and middle-sized rules. In the latter condition, each of these showed a significant increase on their previous level. By virtue of this fact, the marking effect also increased under the triple instruction condition indicating that it is rule rather than stimulus competition which is a key factor here. Consistent, moreover, with the strong role of the ‘big end’ of the continuum, this rule competition seems to be between ‘middle’ and ‘small’ instructions only, leaving the ‘larger/largest’ instruction unaffected.

Conclusions: Middle-Sized

So far, the syntax of size rule learning by monkey is fully consistent with memory based encoding and retrieval models from human adult cognition. In their experiments on human memory and internal psychophysics, Moyer and Bayer (Moyer & Bayer, 1976) also found that, within a given size range, faster RTs were recorded at the large end of the continuum (even) where the original perceived differences are less psychophysically salient than at the small end. The memorial process, they argued, requires that the subject searches from the end of the continuum suggested by the instruction (in their case take larger) and begins a self-terminating search until one of the two items is found.

It is also has features in common with Parkman's model for digit (size) comparisons (Parkman, 1971). This suggests that when asked which is the larger of two digits, the human subject will mentally scan upward from zero until one of the digits is found. The larger digit is then deemed to be the 'other' one and is retrieved accordingly. This model was devised to explain the 'min' and 'split' effects in mental comparison experiments, where the Split effect refers to the distance effect (after Moyer (1973), as described above) and the Min effect refers to the fact that the nearer the smaller of the two items is to the small anchor, the faster the RT. For Parkman, the Split effect is seen to arise because of the statistical likelihood of one of the items being nearer to the Min as distance increases, and the Min effect itself derives from the fact that numbers are learned from zero upwards, providing the system with a fundamental directionality and a fixed starting point for any comparison. Whatever the particular search mechanisms, therefore, both models suggest, in common with Clark's p-space predictions, that binary judgments depend upon the establishment of a primary direction of relational codification.

Still consistent with Parkman's search model is the next stage of rule acquisition (smaller than) for, on this upwards or scanning model, the secondary direction is computed as a derivative only after the forward scanning has taken place. Consequently all 'smaller than' comparisons will be slower than all 'larger than' ones, where the scan procedure first finds the larger item and uses it as a reference point to locate the smaller one in a downward search.

The emergence of such processing effects is clearly evident in the monkey RT data. Thus applying Parkman's model to the question of how a series of ordinal values are acquired within a set, enables the prediction that specific rule differentiation might proceed somewhat paradoxically from the secondary reference point as it is only in the derivative downward direction that items are specifically tagged with respect to the reference point; in the upward direction, they attract an undifferentiated label 'larger than'. This

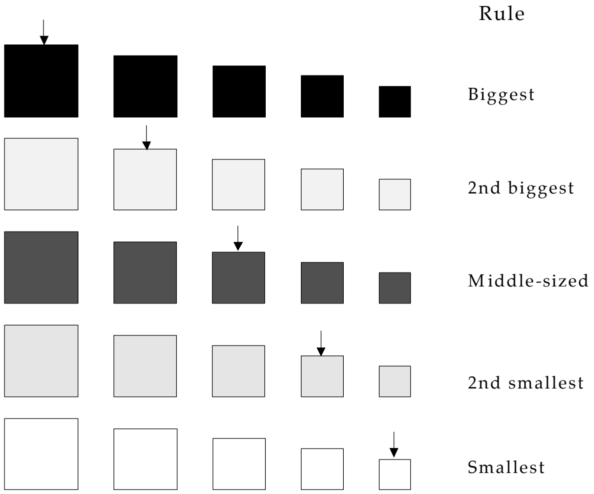


Figure 8. A colour conditional paradigm for training monkeys on five size rules (Chalmers & McGonigle, 1994; McGonigle & Chalmers, 1996).

model is well fitted to the monkey data by the fact that the introduction of middle-sized caused an apparent increase in the uncertainty attaching to 'smallest'. If this denotes the emergence of smallest as a secondary reference point (for computing middle-sized), then it should also follow that the next rule to be acquired would be the one adjacent to this. In a three item set, this would be middle-sized; in a five item set, however, it would be second smallest, not middle-sized. The next stage of our study allowed us to test this prediction.

C. Five Size Rule Learning

Now using a very large WGTA that enabled the subject to view five objects simultaneously, monkeys were adapted to new procedures using five differently coloured sets each of five sizes (Chalmers, 1994; McGonigle & Chalmers, 1996). Three new colours were added to the colour conditional discrimination task, producing the five by five design as shown in Figure 8. As before, on a given trial all stimuli were of one colour and only one size rule rewarded. The spatial configuration of the items on the tray varied randomly from trial to trial.

Training was conducted using initially trial blocks of 5 trials per rule, in random alternation with the constraint that no rule was repeated until the set of five had been exhausted. Five such trial blocks were typically run within a single session. Training continued until a performance criterion was met of 80% correct across 50 trials (10 per rule), with no one rule at or below 60%. This was repeated with trial blocks of two trials per rule. Finally the five colours were presented in single random alternation across trials, but no colour was repeated until the full set had been exhausted. Two different criteria were applied at this stage. The first was the same (80%) one as before. Training continued after this point, however, to establish a stable level of performance against which subsequent transfer behaviour could be measured and required that each and every rule could be operated (simultaneously) at the 95% confidence level. This was set at 70% correct overall across 16 consecutive sessions (80 observations per rule), with no rule less than 60% overall. Finally, an overtraining phase was given to establish, what, if any, further improvement could be sustained over a second set of 16 consecutive sessions.

To ensure that subjects were learning to make set-referenced ordinal judgments rather than ones based on absolute size, we introduced a second size range, overlapping with the first and tested size rule transfer.

Results

All subjects learnt all five rules, and operated these concurrently in the final phase of the experiment. Figure 9 illustrates the choice profiles obtained and compares them with those we obtained from four year old children following similar training. However, in the case of the children, all five size rules were trained concurrently from the outset. Unlike the initial end-point asymmetry we found under binary and triadic conditions of testing, the data now show **both** end points superior to the learning of codes denoting internal set values, especially following the attainment of criterion levels of training. Of equal interest, is the ranking of the internal ordinal rules as shown in Figure 9. From this figure, it can be seen that all monkeys now show a skewed choice distribution in favour of the 'small' end-point, from the earliest phase of acquisition. This item was not only the most accurately identified; it also attracted the fewest incorrect choices during performance on other rules. The biggest item was generally the second most accurately identified throughout, but it attracted many more incorrect choices than the smallest item, principally from rule competition with the second biggest stimulus rule. Of the three inner items—and crucial to our predictions concerning rule acquisition—second smallest was the best identified from the outset, followed by middle, followed by second biggest. Overall improvement with practice in the monkey is better characterized as a change in the pattern of error as a function of ordinal position, such that error distributions become more symmetrically distributed across the series.

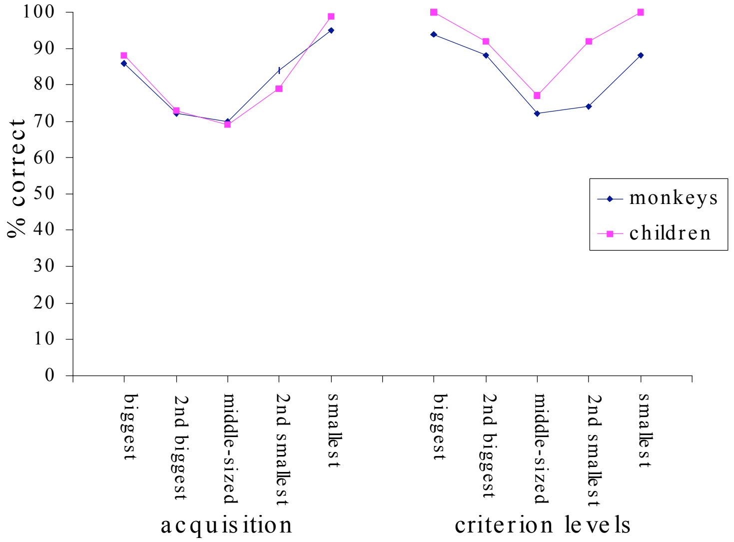


Figure 9. Pattern of size rule acquisition for monkeys and four year old children; data were taken from trial blocks during the later phases of acquisition and when the subjects had reached criterion.

In the monkey, furthermore, whereas choosing the smallest stimulus during binary tests was accompanied by relatively high costs, as reflected in the decision times (Figure 7), these costs now seem to shift to items in the centre of the set when multiple codes must be computed. This is illustrated in Figure 10.

Transferred to a second size range, designed to ensure that subjects computed 'set relative' as distinct from absolute size values, all subjects were able to sustain high levels of performance and recovered criterion levels within four sessions. The strong generalization of codes is shown in Figure 11.

Conclusions

There is now strong warrant for the belief that monkeys are adept not only at size relational codification, but that they can use such multiple relationships concurrently and systematically, following a profile of acquisition and, later, performance which suggests a very similar one to that expressed both in human development and adult human use of linguistic comparatives. That the codes operate in both sets indicates, furthermore, that

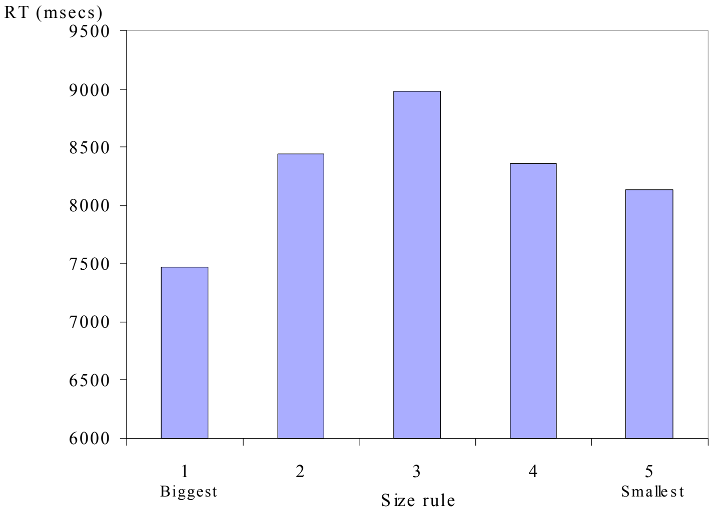


Figure 10. Reaction time distributions for monkeys during steady-state performance on five size rules.

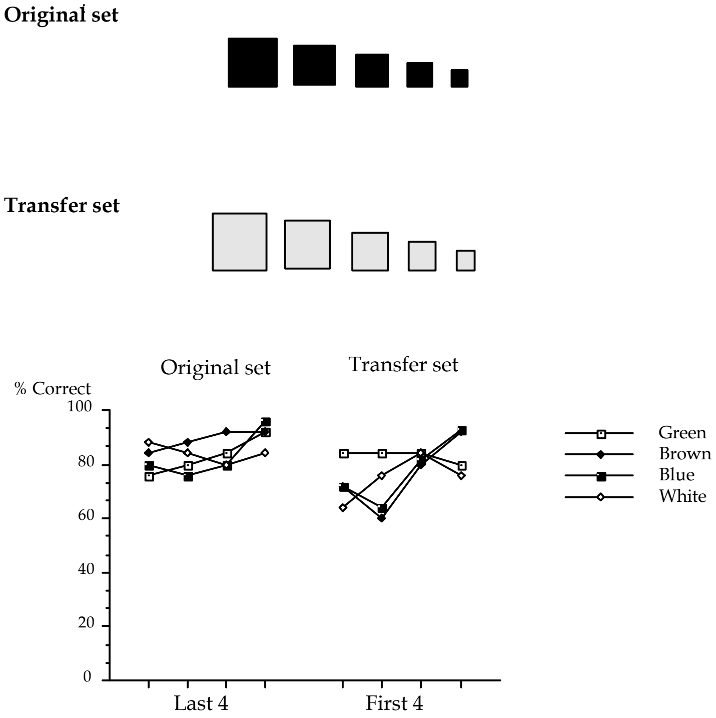


Figure 11. Transfer condition run following acquisition of five size rules and results for all monkeys.

the monkey can learn to determine at least four distinct relational codes in a five item set—the 5th and ‘middle ‘one may only be a default procedure. In this context, it is clear that the 'assembly logic’ which enables the individuation of a full set of items is not based on a co-ordinate logic which constructs a third position from two binary codes. Instead, the empirical evidence dictates that a second anchor based on the 'other' end of the range provides a (secondary) basis from which further ordinal specification takes place—see also (H. H. Clark, 1973; Parkman, 1971).

Now a symmetrical profile, the selective uptake of the relational rules, and the syntax of their acquisition indicates a **trajectory** based change in coding as the subjects moved from single to binary to a three valued and then to a five valued set of relational codes. Of course it could be argued that, initially at any rate, monkeys were taught one relational rule at a time; so some serial uncovering of the whole system was inevitable, i.e. that the rules existed all the time, only their ‘existence’ proof depended on staged procedures. This is a difficult point to answer in almost any domain of learning as the pragmatics of testing often requires a staged or parsed testing procedure first inducing the subject into something easy and tractable before moving to more complex matters. Nonetheless, we would argue, that the selective lag in acquisition of different rules, even under conditions of concurrent testing (five of them), followed by the durable symmetrical profiles recorded post acquisition with RT data, suggests a principled computational logic through which the whole apparatus is seeded from minimalist, inherently asymmetrical relational connections, then altered in a dynamic reflecting greater demands for object individuation and calibration requiring both ends of the stimulus range. The consequence of this transition appears to be a migration from strong directional asymmetry based on one end-anchor to a more symmetrical two-anchor based search procedure.

Certainly the RT-based ‘signatures’ for performance when subjects operate with only two or even three relational rules is different than in cases where the same rules operate within a context where all ordinal codes must be individuated. This provides a 'look-up' table where the degree of series symmetry may usefully indicate the level of ordinal code specification within the set. And this suggests in turn a diagnostic use for such signatures in circumstances where it is difficult to disambiguate different possible control mechanisms underwriting choice. We now apply them to the assessment of the mechanisms that underwrite relational transitivity: one of the most central competences to be investigated in the domain of cognitive growth and human development.

TRANSITIVE CHOICE

In the search for high level cognitive mechanisms requiring complex relational calculation, there is hardly a more seductive task than the linear inference task (Halford, 1993). At core, it provides the subjects with minimally two logical arguments as in Piaget's celebrated version, 'Edith is fairer than Lilli, Edith is darker than Susanne' (Piaget, 1928). For Piaget, a critical solution procedure indexical of abstract though was in the subject's co-ordination of the relationship between Edith is fairer than and Edith is darker than so that Edith could be located ordinally as a middle value between Lilli and Suzanne. When children failed (as they invariably did below the age of eleven or so), Piaget attributed to a failure to seriate mentally and to create three ordinal positions, the superlatives, biggest and smallest together with the new middle position, from the binary codes as given in the premises.

However, in more recent developments, memory factors have been implicated as an important restriction on performance in such tasks and a variety of procedures have been developed to ensure that young subjects remember the connecting premises at the time of transitivity testing. In this way, memory failures can be distinguished from a failure to integrate the pairwise information into a series.

Adapting the classic reasoning task for use with young children, Bryant and Trabasso (1971) developed what is known as a 5-term series task into a training paradigm. Formally represented by the predicates A>B/B<A; B>C/C<B; C>D/D<C; D>E/E<D followed by pairwise tests such as B? D, in this version, the child saw a pair of rods protruding from a box, where one was e.g. yellow (A), the other blue (B). To control for memory failure, the child was trained to remember that e.g. the blue one was bigger than the yellow and that yellow was smaller than blue. The child was similarly trained on the three other connecting pairs from the series without ever seeing the sticks varying in size.

As is always the case in transitivity tasks (Youniss & Murray, 1970), care had to be taken that simple end-point labeling strategies were not the basis for solution. The crucial comparison afforded by using five rather than three item therefore was B v D where both these items have been double coded as larger and smaller than. In the first report using this procedure, Bryant and Trabasso produced evidence that that when memory can be eliminated as factor, then children make transitive choices very well, even on the crucial B D comparison. They also supported the view that children were integrating the pairs by showing that the level of transitive choice obtained could be predicted on the basis of the Cartesian product of remembering both connecting pairs.

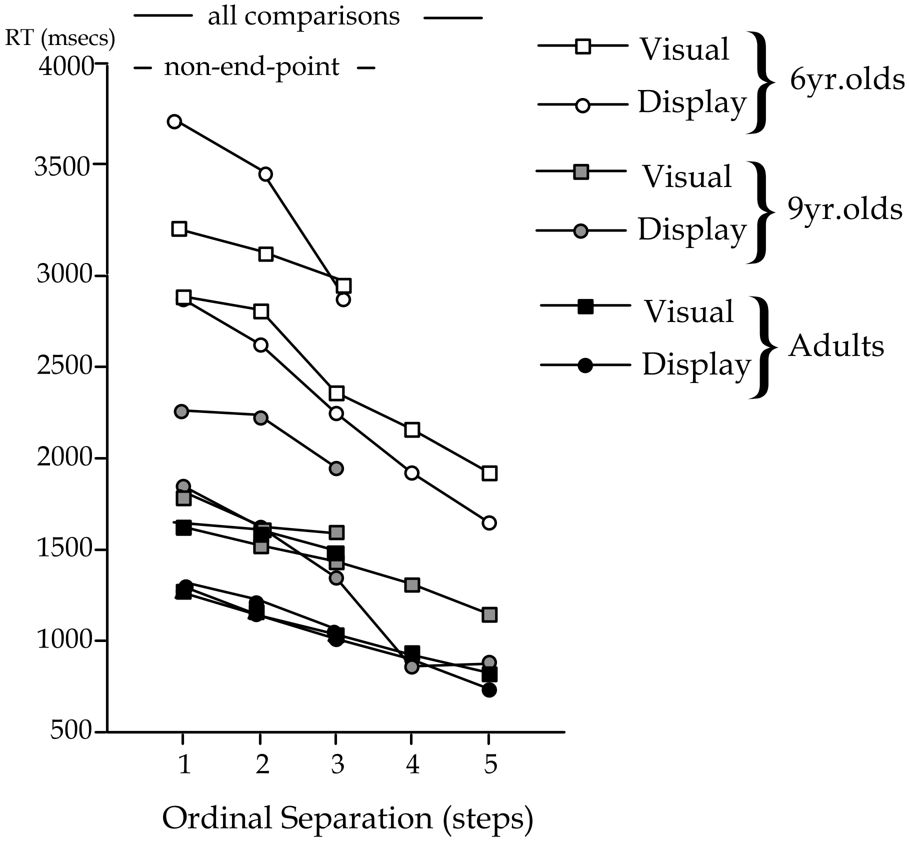


Figure 12. A sample of the SDEs (Symbolic Distance Effect) reported by Trabasso (1977) for different ages and conditions.

In a further development, Trabasso extended the task to a six term one, testing subjects of six and nine years of age as well as adults (Trabasso, 1977; Trabasso, Riley, & Wilson, 1975). In this ground-breaking series of experiments Trabasso and his colleagues also varied the medium of presentation from the quasi-concrete versions used by Bryant and Trabasso to directly perceptual ones where the subjects could either see the size values during training or could see a display of items representing the full series being tested.

With virtually all subjects and all conditions, Trabasso et al. found that transitivity of choice was accompanied by the Symbolic Distance Effect interpreted in the domain of internal psychophysics, as we have reviewed, as indicating a mental series representation of the items in a linear order. A sample of their findings are depicted in Figure 12.

MONKEY TRANSITIVITY

Given that the more recent research on transitivity in children suggested that the lower bound of a transitivity/seriation competence might be much earlier than was at first believed, there were implications for non-humans which we were the first to implement with squirrel monkeys (McGonigle & Chalmers, 1977); replicated in part with chimpanzees by Gillan (1981), pigeons (Fersen, Wynne, Delius, & Staddon, 1991) and rats (Roberts & Phelps, 1994) and now extended into the neuroscience of memory by Eichenbaum and his associates (Dusek & Eichenbaum, 1997). Our experiments with squirrel monkeys (*Saimiri scuireus*) (McGonigle & Chalmers, 1977, 1992) were based on similar conditions of training and test to those of Bryant & Trabasso (1971). Instead of sticks, however, we used coloured tins which differed in weight (not a mediate property of objects) and trained and tested as illustrated in Figure 13. Only two weight values were used throughout, however, so that no simple material scale could be used based on a weight series mapped onto the colour values. In our second study (McGonigle & Chalmers, 1992), RTs were measured during binary tests.

Training

Trained initially to a 90% accuracy on a random sequence of training pairs, all subjects received an overtraining phase on the '92 study in anticipation of extended use of RT data for an analysis of the SDE. This ensured that a minimum level of 90% success could be maintained to each pair over a block of 80 trials of post acquisition criterion testing.

Transitivity and High Retention Levels

The results of both the 1977 and the 1992 study show that choice transitivity was highly significant, and for all monkeys, including the non-end anchored comparisons B v D. High levels of retention were also recorded for training pairs, which was 99.5% correct (McGonigle & Chalmers, 1992).

THE FIRST SYMBOLIC DISTANCE EFFECTS IN MONKEYS

Uniquely, each individual was examined for evidence of an SDE. For each 100 trial test block, subjects were subjected to analysis of variance to see there was any significant variation across test pairs. If there was no such

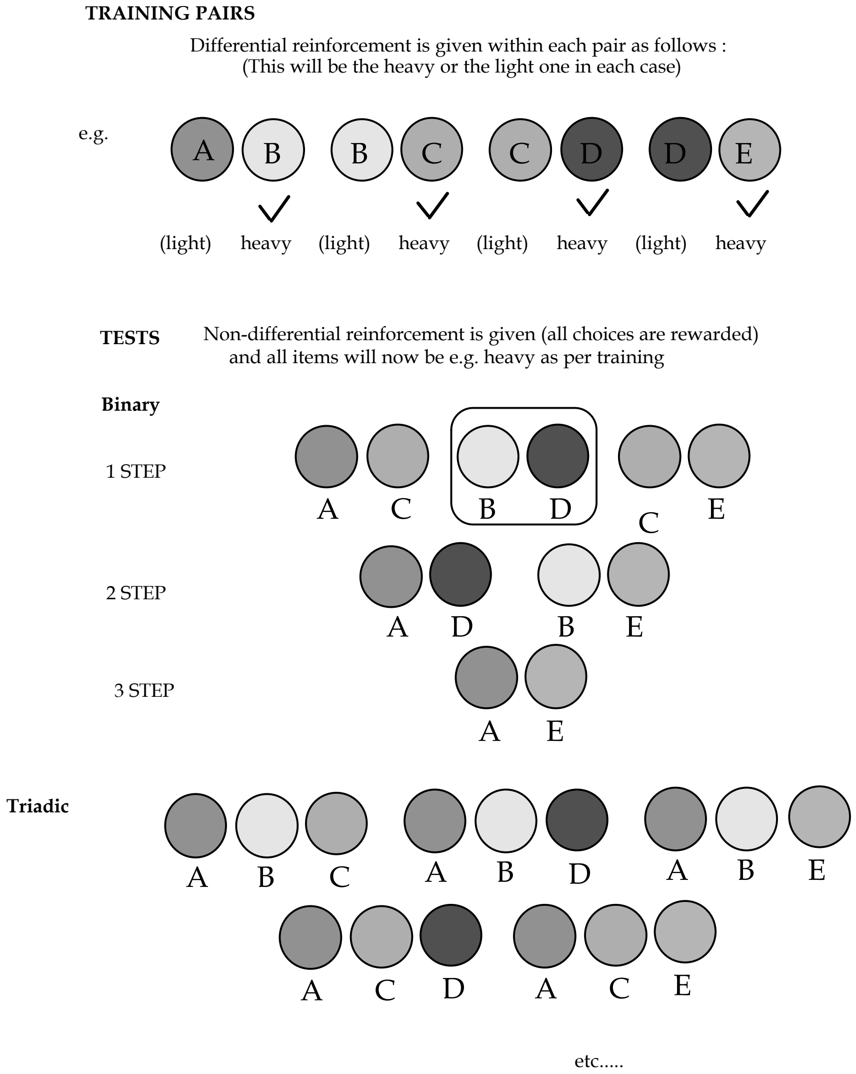


Figure 13. The transitivity paradigm used with monkeys by McGonigle & Chalmers (1977; 1992).

variation, monkeys went on to a second 100 trial test module. Similarly if pair variation was not accompanied by a significant effect of ordinal separation, or if any effect of ordinal separation was not significantly linear, the monkeys were maintained on testing, for a further module of ten (by ten) observations, and so on.

All monkeys recorded an SDE by the third ten by ten test module: one (Blue) showed the effect by the end of the first module, two (Green and

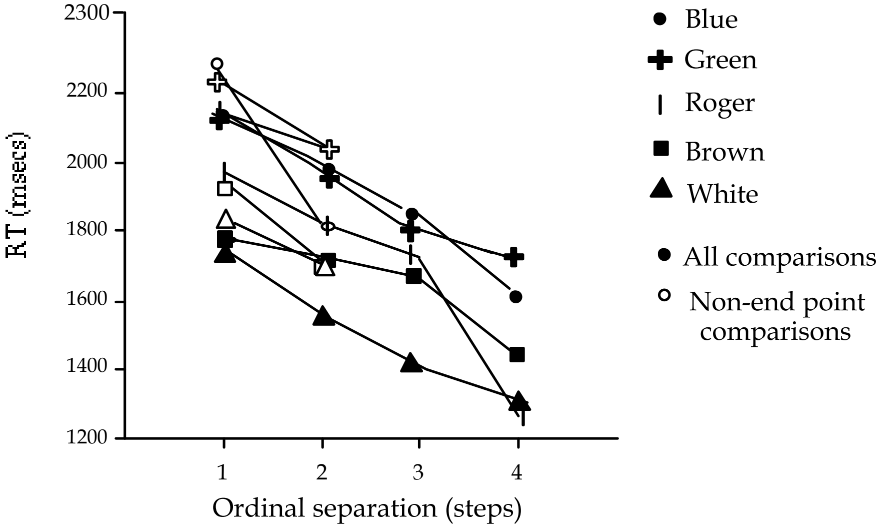


Figure 14. Symbolic Distance Effects obtained for individual monkeys.

Roger) by the end of the second and two (Brown and White) by the end of the third. No other significant RT variation was recorded. As with the marking effects we report earlier, practice did not abolish the effect. Instead, it sharpened the SDE in individuals most probably due to the reduction in 'rogue' variance or noise. All effects were also accompanied by a Symbolic Distance Effect for the non-end point comparisons which was highly significant for the group as a whole as Figure 14 illustrates.

Distance Effects and Their Implications

In a long theoretical analysis of his team's research on the SDE, Trabasso (1977) entertains many different models all of which make similar predictions but which have very different processing implications. Trabasso tends, however, to favour an imagery account whereby a spatial layout is used 'in the mind's eye', then scanned in memory. Based on 'think aloud' protocols where some subjects at least report the use of spatial paralogical device enabling test items to be read off as if physically present, Trabasso (1977) envisages the read-off a serial search from one end of the series:

> "...suppose that the child represents the coloured sticks as an ordered list (1,2,3,4,5,6). When a question on the relationship between a pair of sticks is posed, he begins by scanning the list from the end mentioned in the question. If "longer," then he starts with 6; if "shorter," he starts with 1. Suppose the question is on the

> relationship between Stick 2 and Stick 4. If the question contains "shorter?" then he must scan two members, 1 and 2, to find 2, and he can terminate at this point. If the question contains "longer", he will scan three members, 6, 5 and 4, to find 4. On average 2 ½ members are scanned. For a question on Sticks 2 and 3, however, the respective number of members scanned are two and four, or an average of three members." (page 346/347).

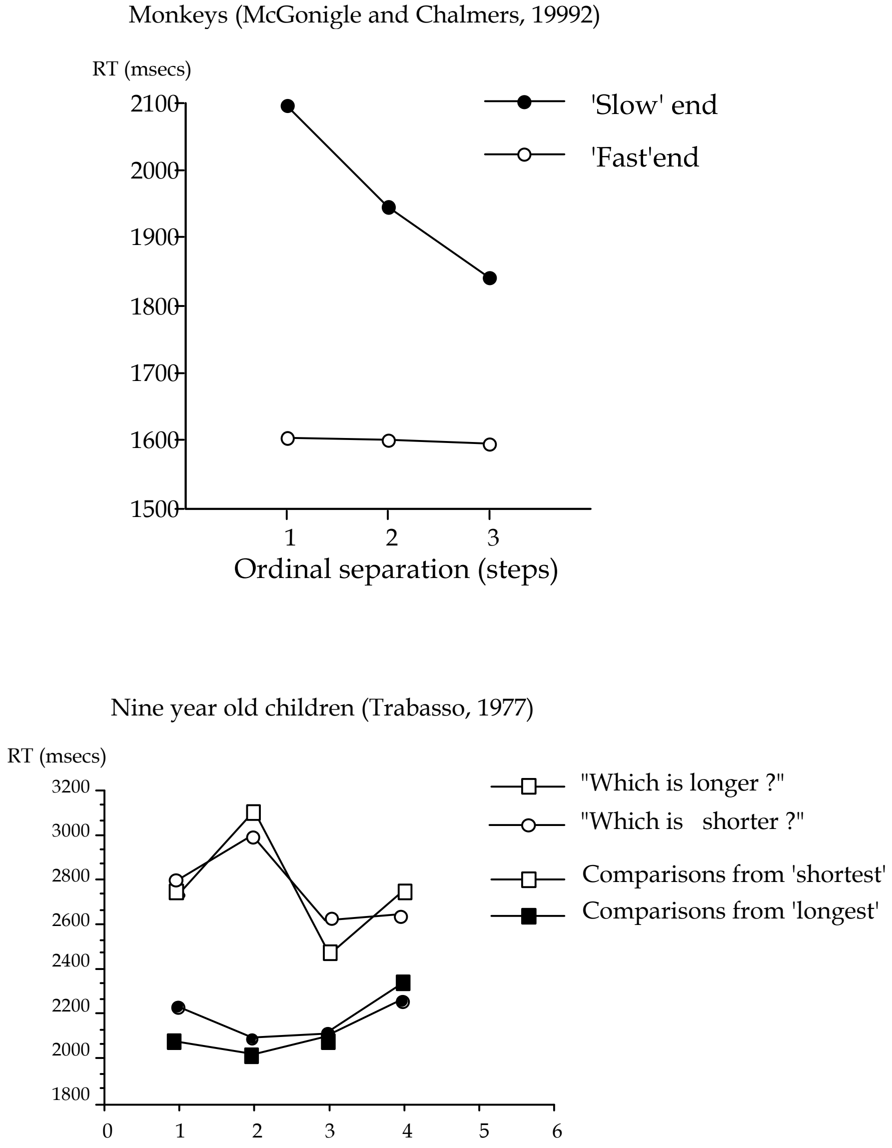
A modification of that view which he then proposes, because it produces a better fit with the empirical data, is that the search is self-terminating after only one item is found. Both versions of the search model, however, are based on the idea that RT reflects distance from an end-point rather than inter-item distance *per se*. This is very compatible with the Parkman's search model for digit comparisons and the mechanisms suggested above for ordinal rule learning by monkey.[1] However, the SDE by itself does not arbitrate between 'read-off' models and search models. If search, furthermore, is it with—or without—benefit of symmetrical series representation that would allow **both** ends of the series to be used as reference points?

Which Signature: Imaginal Read-Off, or Serial Search?

Given the unquestionable importance of RT data as a means of arbitrating between competing theoretical accounts, McGonigle and Chalmers (1992) undertook an extensive sub-analysis of the SDE in monkeys. With an new affluent data base per subject, McGonigle and Chalmers (1992) were able to evaluate selective directional effects, and compare the RT signatures obtained with those as reported earlier in the size relational conditions. As with our relational data, we now assessed the role of different ends of the series in the SDE, plotting decision times from each pole separately. This revealed for both the group as a whole, and for each subject bar one (who showed a symmetrical effect—the same one, White, who showed symmetry in the binary size condition) a highly **asymmetric** profile as the wedge function illustrated in Figure 15 reveals. The RT data were found to assort into two populations. One, a population of fast scores, was insensitive to distance effects. The second population was derived from slower RTs and these did show a strong distance function and internal ranking as a function of proximity to the slow end-point.

---

[1] Parkman's model suggests that zero is the reference point for digits; for size, the monkey and child data suggest it is the largest stimulus in the field.



Monkeys (McGonigle and Chalmers, 19992)

Nine year old children (Trabasso, 1977)

Figure 15. The SDE for monkeys and nine year old children (Trabasso, 1977) plotted separately from each end-point of the series.

These data are certainly counter-indicative of a 'read-off' from a Euclidean mental array and suggest, instead, that subjects' decision times are determined largely by proximity of the test items to a single privileged anchor point. As monkeys, are, however, effectively only 'asked' to use one type of comparative (heavier **or** lighter denotes the direction of training), it is difficult to tell whether or not they would show symmetrical distance effects suggesting the use of a secondary reference point. However, a re-analysis of data from Trabasso's child subjects suggests that children as old as nine (in contradistinction to adults) do **not** in fact show any 'cross-over' when the instruction changes from longer to shorter. As Figure 15 shows, there is similar evidence for series asymmetry in the child data, independently of the question asked.

Conclusions

Using our own size rule ‘signatures’ described earlier as a basis for interpretation of the levels of control implied here, we can now locate the transitivity mechanisms deployed by monkeys somewhere between an ‘early’ stage of open set control, lacking the closed, or bounded set symmetry of processing these same subjects expressed early in the five rule training and a more principled search procedure allowing search from both ends. That RTs may be ranked as a function of proximity to the 'slow' end of the series—giving rise to the SDE—suggests that monkeys are using a conjunction of a relational tag (e.g. heavy) and its negation (not heavy). Comparisons involving the ' not heavy' end-point or items associated with this anchor through proximity (Breslow, 1981), add to the RT through the operation of negation. The more likely it is that a decision can be made without negation—by using the positive, (e.g. heavy) reference point or one close to it, the faster the decision. This would conform to Heidenhemer's model for the use of antonymic terms early in development, and to one we have devised for quite independent studies of children's mental comparisons (McGonigle & Chalmers, 1984). In the present context, such a model was give a formal warrant in a production systems model devised by Harris and McGonigle (1994) and informed by our squirrel monkey data. Based on production systems each item (in the transitivity series A- to E+) was represented as a condition-action rule such as 'if E then select E' and negation-based rules such as 'if A then avoid A'. Each rule had a place in an overall stack. Thus 'if E then E (select)', could be followed in the stack by 'if A avoid A', etc. An example would be

1. If E select E
2. If A avoid A
3. if D select D
4. If C select C

A crucial property of this model is that *all* stacks that perform correctly on the adjacent pairs also perform correctly on the remote pairs without making further assumptions. This contrasts with other formal models such as the rank tensor product model of Halford (Halford, Wilson, & Phillips, 1998), in which a transitive decision about a transitive pairing is higher-level and more complex than a decisions concerning adjacent pairings.

A main implication of these converging analysis both formal and empirical is that transitive choice, for primates at least, may occur early within a trajectory, using primitive forms of relational judgment and that these leave open many possible new developments, not the least of which is the migration to finite set, symmetrical ends inward processing using two comparatives rather than one and its negation.

Triads

Some confirmation of this within the transitivity paradigm has been provided by our use of triads. As in the size conditions described above, Chalmers and McGonigle (1984) and McGonigle and Chalmers (1977, 1992) followed binary tests of transitivity with triadic tests derived permutatively from the same training set. Whilst choices were significantly transitive under these circumstances for both monkeys and children, there was an initial drop in the strength of the transitive profile for both species as Figure 16 illustrates

Triads, where there is no perceivable solution, do, of course, require more finely tuned codes than do binary choices. This is formally shown in the Harris and McGonigle (1994) model which show that certain triads, at least, require a deeper search to disambiguate the ‘winning’ item. This further constrains the set of rule stacks which can sustain high levels of transitive choice in triadic situations. Take the case of the rule stacks 'if E, select E' followed by 'if A, avoid A' and *vice-versa*. In either case, the first rule reduces overall uncertainty by 40% by accounting for 40% of all pairs; the second rule by a further 30%. In the case of the triads, however, the 'avoid' rule leaves the residual uncertainty concerning which of the remaining two items to select in every triad. Thus, 'Avoid A' now has 0% uncertainty reduction, therefore, whether in first or in second place in the stack. The first useful constraint that would thus become apparent within triads is the need to proceed monotonically on the basis of a **positive** selection strategy. Stacking 'If E, select E', followed by 'if D, select D' and so on, produces the highest rate of information gain measured as uncertainty reduction. In the binary case, there would be no such advantage for a monotonically ordered search (McGonigle & Chalmers, 1998).

That response competition might motivate the beginnings of what might be genuine linear search was apparent from spontaneous changes in the monkeys' triadic response distributions as a function of test exposure.

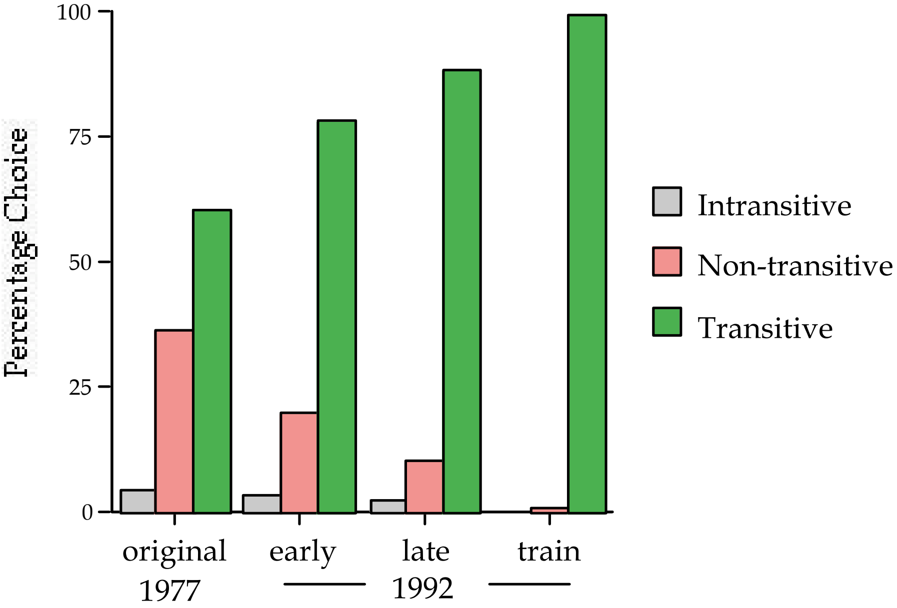


Figure 16. The spontaneous drift in triadic choice distributions for monkeys towards 100% transitivity.

Although there was an initial drop in overall levels of transitivity, even though these remained at a highly significant level, McGonigle and Chalmers (1992) found that with protracted exposure to triadic tests, monkeys achieved a level of 92%, revealing a shift in choice behaviour such that later triads revealed a level of transitive choices almost as high as in the binary case. This shift is illustrated in Figure 16, and occurred without differential reinforcement.

Triads and Seriation

The crucial advance brought about by establishing a bounded set with two reference points is, as we have seen from relational rule learning the formation of a scale on which objects can be differentiated not just by direction of comparison (bigger versus smaller), but also by establishing a finite range within which stimuli can be both calibrated and searched from both ends in accordance with the response demands of the task. In the case of object relations, it was necessary with the primate to make calibration explicit by requiring multiple responses to be sustained concurrently within the same set of values.

In the case of transitivity, however, only inferences can be drawn about the underlying calibration of the test series. And the signatures from the RT assays suggest that transitive choice is the result of low level relational encoding dominated by a salient reference point. Whilst the triadic conditions increase item competition and induce more fine tuning, they neither motivate, nor do they require extensive ordinal codification or the extensive search that a five term series could **in principle** allow.

Conclusions

Given that young children show similar patterns of binary and triadic transitive choice to those of monkeys (Chalmers & McGonigle, 1984) and similar RT profiles during test (Trabasso, 1977), the conclusion which we must inevitably draw from this is that a core rational mechanism demands rather less ordinal computation and explicit search than was once thought. Given the adaptive importance of the mechanism, furthermore, this is perhaps as it should be! It is initially, on this view, a 'small world' phenomenon adapted for small sets of alternatives, a limited and bounded rational mechanism where the ranking of object codes is based at best on their derived distance from one or two stable reference markers, and the search is non-exhaustive, terminating when the target item is found. Such mechanisms may indeed form the basis of judgments in real space involving deictic codification of the relative position of environmental features with respect to a common fixed source. Could this be the reason perhaps that Dusek and Eichenbaum (1997) find hippocampal substrates underwriting transitive choice in rodents in ways that Morris (1982) and others do for spatial codification? The 'landmark' logic may be similar.

Seriation

If transitivity initially reflects a limited search based ordering procedure, as we suggest, where ranking of items internal to the set derives from their relative distance from a common end anchor, it is perhaps not surprising that explicit tests of seriation which require the individuation of every item in the set *vis-a-vis* every other one of the set, emerges later in human development than the earliest demonstrations of transitivity. Achieved normatively by children around seven years, this classic seriation task as devised by Piaget (Piaget & Szeminska, 1952), requires subjects to explicitly order a large number of objects (8-10) differing in size, placing them in row. Here no inference is required that ordering of all the items has occurred as the behaviour is overt and public—especially when the child orders in a principled way (selecting and placing the rods in a monotonic sequence without trial and error), as in what Piaget terms 'operational' seriation.

In such a sequencing task, however, using a finite or bounded set procedure to achieve full ordinal specification by computing for example, biggest, second biggest, third smallest, etc. as a preamble to seriating them would be a cumbersome and almost certainly inaccurate procedure for sets larger than a few items—and the classic version conventionally employs 8 - 10. Instead, it would seem much more likely that the executive demands of serial ordering items within a single production (as distinct from successive relational judgments made in simple choice situations), demand much more streamlined mechanisms designed to support rapid serial control as in piano playing or typing (Lashley, 1951). However behaviours of this sort cannot be studied readily using conventional choice mechanisms as in transitivity where behaviour 'halts' following a single decision to choose one object over another. Yet conventional seriation tasks are impossible to administer to non-humans given the serious manipulative restrictions imposed by their motor control systems. These embargo the adoption of the 'select and seriate' tasks used by Piaget. To eschew these problems, we turned to touch screen based versions of seriation, developing a family of both supervised (trained) and free search based ordering mechanisms that enable us to establish a common currency of measurement in comparisons between human and non-human species.

TOUCH SCREENS AND EXECUTIVE MEMORY PARADIGMS

With touchscreens, we can now help deliver on a long-standing ambition of Lashley (Lashley, 1951), when he urged the study of serial ordering as a key to understanding the nervous system and a key player in adaptive behaviours. For few behaviours, whether at the level of language or at the level of action, make much adaptive sense without evidence of serial control, be it based on simple finite state transitions or higher order dependencies as in word strings.

However, until recently, the learning of serial orders has rarely been demonstrated (Terrace & McGonigle, 1994; Terrace, 1987; McGonigle, 1987). In simple cases of serial behaviour occurring in natural contexts, instinct rules have been posited; in complex linguistic constructions, generative grammars have a strong influence. As for child development, of which the development of seriation is one of its most robust indicants, the dynamic constructivist position of Piaget has viewed learning as a poor relation—the consequence and not the cause of cognitive growth.

A key feature of our new procedures in this context, therefore is that we can train our subjects to seriate, manipulating key independent variables such as string length, and the stimulus characteristics of the test set. These

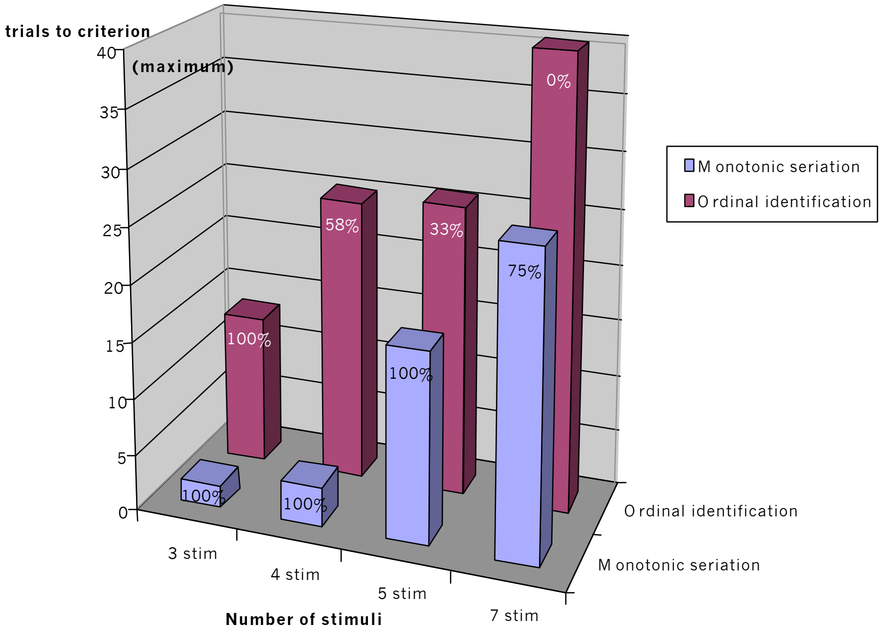


Figure 17. The relationship between learning ordinal identification for five sizes and learning to monotonically order those sizes by five year old children. The percentages on the bars represent the number of children who succeeded in reaching criterion within 40 trials (per series and per size rule).

techniques also allow subjects free search on the touch screen, enabling them to devise their own means of keeping a track of the items they have already selected. Our first experiments of this sort were of the supervised variety carried out with children of five and seven years of age (Chalmers & McGonigle, 1997). Required, for example, to touch five items (rectangles) arranged in a random linear array on the screen in the order from biggest to smallest, children as young as five were able to learn to do so using error free search within a single session or so. Whilst their executive control is clearly inferior to that of seven year olds under these circumstances, this skill is nevertheless much more easily trained in young subjects than the concurrent learning of set-referenced ordinal codes for each and every item in the set as measured by both by a matching to sample procedure and by the five rule colour conditional procedure described for actual objects earlier. This relationship is illustrated in Figure 17.

Interpreting these findings in terms of an economy-based model of serial organization (McGonigle & Chalmers, 1998; McGonigle & Chalmers, 2001), we have argued that linear monotonic size seriation is an example of a highly adaptive form of on-line memory organization, where the principled iteration of an elementary code allows a large number of items to be searched uni-directionally and rapidly on-line—without requiring the costly off-line codes of ordinal position that locate each item within a set bounded at both ends of the range.

LINEAR AND HIERARCHICAL CONTROL IN EXECUTIVE MEMORY

One of the general implications of this is that higher order cognitive organisation in humans reflects an optimising procedure where solutions are selected both to minimize memory cost and to maximise adaptive utility (Anderson, 1990; McGonigle & Chalmers, 1998). Linear search procedures aside, classification procedures which segment the serial production into manageable chunks or units should also be selected on this view for their data reducing memory effects if these can be shown to have utility value.

In the final study we report here, based on seven *Cebus apella,* we have combined both linear size and class based ordering procedures in the same experiment run concurrently To induce concurrent classification and linear size seriation we used a novel design which first trained the subject to order three different shapes. Each shape stood for a category of icons which ultimately varied in size. Thus, from the first simple string, we generated longer sequences featuring multiple exemplars of each shape category, also varying in size. At first only three levels of size were used throughout the series. Thus the same three sizes recurred from category to category, and the subjects had to order both the categories on a shape basis, and the exemplars within each on a size basis. The rationale here was to assess the extent to which the subject could combine linear size and class based information to order a full set of nine items. Here, shape information could support only three ordinal positions, and size could only support three. So to sustain the nine item sequence, subjects had to combine shape and size information, nested hierarchically such that shape provided an ordinal segment (of three items per segment) individuated by size seriation, so that nine independent values were possible (see Figure 18).

Finally, we required the *Cebus* to order all nine items along a linear size series of nine different sizes. This was done under two main conditions run concurrently. In one all sizes came from either category A or B or C. In another, the linear series was overlaid with three categories with the first three sizes of the series going to category A, the next three to category B, and the final three to category C. The rationale here was to determine the extent to which adding category differences to a linear size series led to

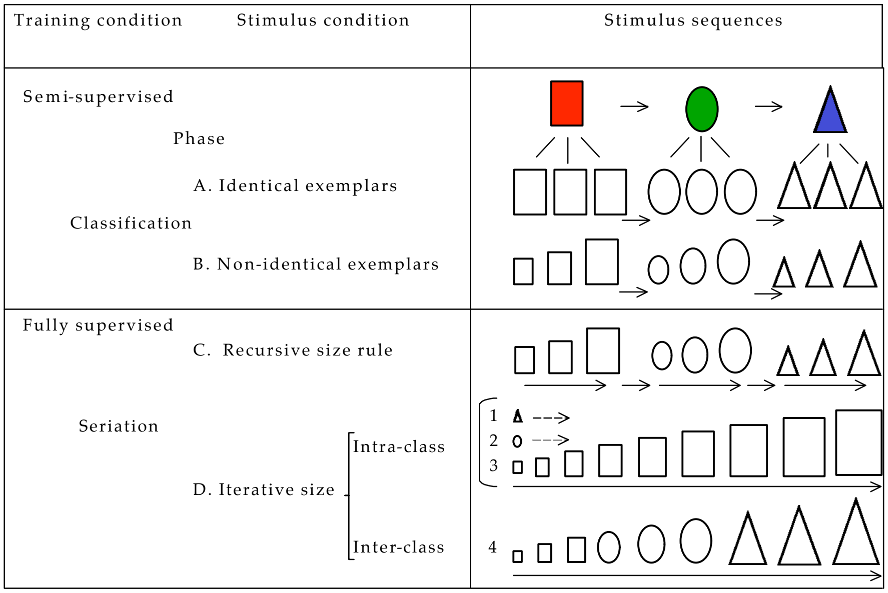


Figure 18. The design of a longitudinal study of monkey classification and seriation carried out with *Cebus apella*.

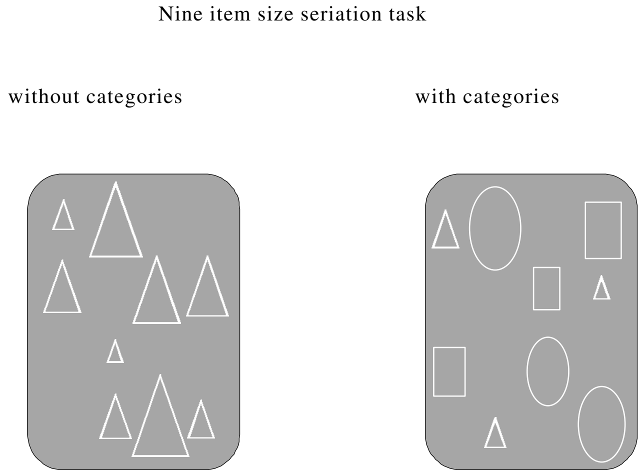


Figure 19. Examples of nine-item seriation tasks given to monkeys as they appear on the touchscreen.

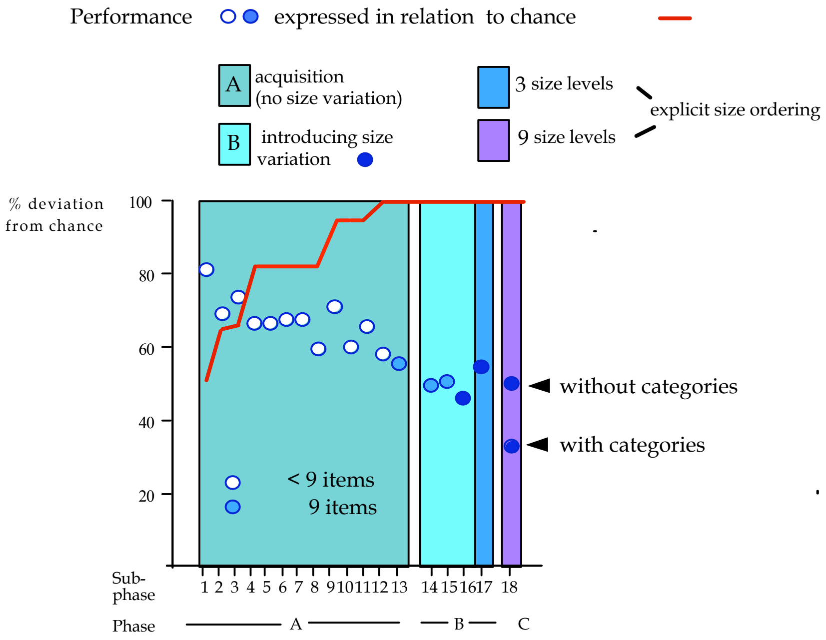


Figure 20. Learning-to-learn profile from *Cebus apella* as they proceeded through stages of the nine-item classification and seriation study. Errors committed during each sub-stage of training are expressed as a deviation from the errors expected on chance assumptions. The final points show the comparison between nine item size seriation with, and without, opportunities for classification.

improved performance. Figure 19 shows the two types of sequence deployed as they appear on the screen.

Results

Over a lengthy series of stages, in an experiment lasting several years, monkeys achieved nine item seriation. As Figure 20 shows, monkeys adapted to progressive increases both in sequence length and the levels of specification demanded of the constituents of the sequences. Early stages of sequence production could be controlled, furthermore, by coarse chunks of three classes each containing identical exemplars and individuated only by their relative position to one another within the test array. Later sequences, however, required the differentiation and then principled ordering of both exemplars and categories.

In this way, the tasks can be ordered in hierarchies of difficulty. Apart from greater constituent specification in later tasks, the progressive increase in the number of constituents *per se* renders later problems subject to serious combinatorial (explosion) effects. That linear nine item size seriation is also solved, therefore, with no disproportionate increases in error, as Figure 20 also shows, indicates to us at least the high utility value of unidirectional size coding where a series can be generated, and successor items anticipated prospectively with a low cost search procedure. A further demonstration of this is provided by the enhanced utility value which categorical seriation has over the purely linear case. Figure 20 shows that at the final nine item size seriation stage, performance is uniformly better on the multiple classification condition, although both sorts of seriation return high levels of performance.

Conclusions

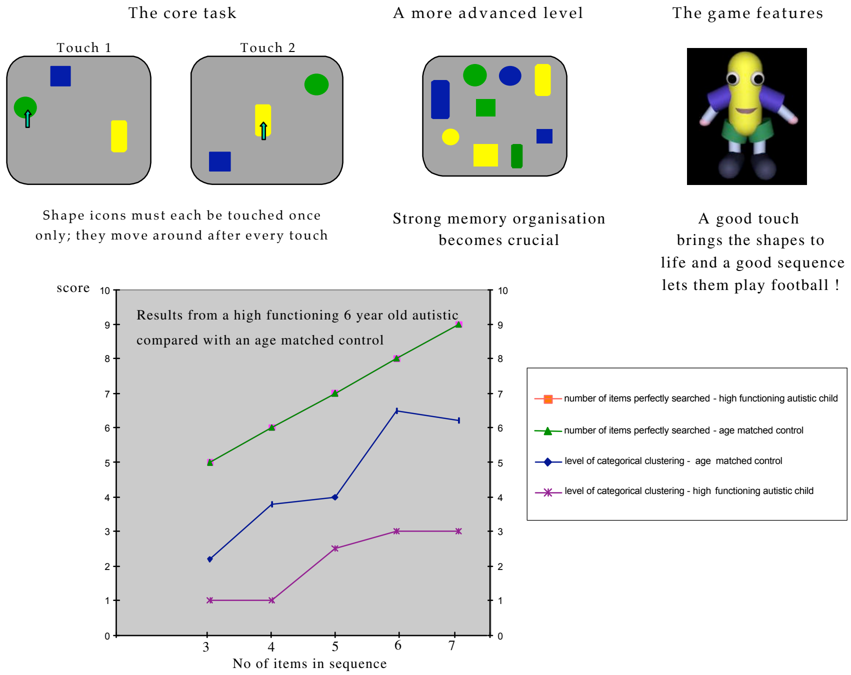
Nine item size seriation is regarded in human development as a benchmark of cognitive growth. Whilst the methods reported here required that monkeys were trained to do what children appear to do spontaneously at around the age of seven, their attainment of this goal is nevertheless the first of its kind to be reported, and at an operational level executed as a principled selection procedure on every successful trial. As such, it represents the most complex form of organising multiple elements in executive memory in the non-human literature. The most recent claims (Brannon & Terrace, 1998) for nine item ordering (based on numerosity) by two rhesus monkeys whilst fully consistent with these findings are based on data obtained from overlapping sequences. The objective measure of the utility of classification within a sequencing task, furthermore, gives strong warrant for the belief that the classifier system is being actively used.

Now that the objective utilities of different sorts of search procedure can be evaluated over and above successful performance, we are now addressing the question of the extent to which subjects themselves can detect and discriminate these relative utilities and self regulate executive control in the context of unsupervised free search tasks. As such tasks do not presume the way in which such strategies might be expressed (by imposing a 'right' answer on the subject), these techniques are particularly suitable in certain clinical applications as we review briefly below.

CLINICAL APPLICATIONS AND ANIMAL MODELS

There now seems little to constrain comparisons between non-primate and primate cognitive functioning at almost all levels of functioning. Executive control, spontaneous strategic elaboration, classification, phrasing, search routines and the use of state based externalised feedback to further reduce memory load (McGonigle & Chalmers, 2001)—are all now possible. In addition, the measurement of performance in real time enables on-line and off-line distinctions which have not previously been possible using more traditional forms of assessment. For human clinical conditions too, where cognitive dysfunction is under review, impairment can now be evaluated using a rich possible degradation space which enables the clinician to distinguish between meaningful (relational) and associative memory, active versus passive memory deficits, and the relative failure of patients to construct appropriate strategies in the light of progressive and well calibrated task demands which impose an ever increasing memorial burden on the subject. In the context of Alzheimer's research, for example, recent work at Edinburgh and Newcastle Infirmary has been heavily reliant on our free search tasks as described above in helping determine the sorts of strategies, if any, patients with dementia can still furnish to minimise memory loss. In addition, with a mix of supervised and unsupervised (free search) procedures, we are in a position to evaluate the extent to which strategies can be re-acquired or taught *de novo*—a remediation issue in this domain.

Another application of this work to which we have recently turned our attention concerns the characterisation of cognitive deficit associated with autism and the sex-chromosome linked condition, Fragile X syndrome. The cognitive deficit associated with such disorders are usually established through psychometric testing (especially using the Kaufman Assessment Battery for Children), and in the case of Fragile X are already thought to be as sequential in nature (Dykens, Hodapp, & Leckman, 1994). But with such testing devices the cognitive diagnosis of the condition is unlikely to advance further. Apart from the shallow, one-off nature of psychometric test items, they also confound linguistic and educational retardation with core deficits which might cause such retardation in the first place. It seemed to us that methods such as those we have described remove many of these problems at a stroke. Highly motivating and essentially non-verbal in nature they require low levels of manual dexterity, minimal communication with the experimenter, presume on no particular taught or world knowledge, and can be geared to very reduced levels of competence such as a 2-item sequence. But above all, in being capable of fractionation into component skills, they allow us to specify learning pathologies with a level of precision not available from standard techniques (Chalmers, 1998). An example is given in Figure 21, depicting a free search task as given recently by us to autistic children (Chalmers, In preparation) and children with Fragile X (Chalmers, In preparation), together with a learning profile from a high functioning autistic child, revealing that, whilst sequential learning was as good as that found with an age matched unaffected child, the signatures for categorical clustering were absent in the autistic child.


The core task

A more advanced level

The game features

Shape icons must each be touched once only; they move around after every touch

Strong memory organisation becomes crucial

A good touch brings the shapes to life and a good sequence lets them play football !

Results from a high functioning 6 year old autistic compared with an age matched control

Figure 21. Example of a free search paradigm and results from an autistic subject showing perfect sequence learning for up to 7 items, but without the degree of organisational control found in a normal age matched control.

The obverse has also been found using our tasks, where an apparent failure to learn to make an exhaustive search through a set of four items by a Fragile X subject revealed a 'hidden' ability to show organisational constraints in sections of his performance similar to those exhibited by a normal (mental) age matched control subject. In both these cases we can now pursue the implications of spared and/or absent competences by explicit training or further task variations. For example, our autistic subject is to be re-tested, using new highly flexible testing programs on a task that requires exhaustive on-line search as before, but that will disallow (by varying the items from trial to trial), the characteristic autistic strategy of rote memorising a specific list of items. This should reveal, what, if any, volitional control such an autistic subject might show when confronted with visual input that demands flexibility within short-term memory, by tailored experimental techniques still relatively rare in the area of childhood cognitive dysfunction.

CONCLUSIONS AND THE FUTURE

We have offered a sample of experiments from within our programme because we believe that it is time for a new agenda to help make possible a full behaviour-based analysis of cognitive complexity without requiring language.

In this biomedical context, we are only too aware that a persistent problem in the neurosciences concerned with human cognitive function and its degradation as in the amnesias, is to find animal (species) and paradigms which will enable the neuroscientist to explore brain-behaviour relationships in animals crucially homologous with higher order functions in man. Although similarities in design logic as assessed at the behavioural level do not necessarily imply a similar form of implementation at the ‘wetware’ levels, it does at least create a climate of competence fractionation and evaluation enabling neuroscientists to asks more accurate questions of those adaptive competences brains (of whatever species) have evolved to deliver.

Certainly a rich space of measurement enables us now to converge comparative research with clinical and neuroscience research areas, where a major problem has been to establish task scenarios which enable both long term assays of performance longitudinally and which keep the subject's interest. With our latest generation of software we can both create intrinsically interesting interactive tasks which keep even very young subjects of 2.5 years engaged, and for long periods. And we have tested primate subjects for years in longitudinal studies in these test environments. Now being developed as an intelligent tutoring system which automatically adjusts task type, task difficulty and training conditions based on samples of a subject's ongoing performance, our new system is conceived both as a laboratory instrument for cognitive research and as a portable prototype to support a wide variety of clinical applications.

REFERENCES

Anderson, J., R. (1990). *The adaptive character of thought*. Hillsdale: Lawrence Erlbaum.

Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. II). New York: Academic Press.

Boakes, R. (1984). *From Darwin to behaviorism. Psychology and the minds of animals*. Cambridge: Cambridge University Press.

Brannon, E. M., & Terrace, H. (1998). Ordering of the numerosities 1-9 by monkeys. *Science,* **282**, 746-749.

Breslow, L. (1981). Re-evaluation of the literature on the development of transitive inference. *Psychological Bulletin,* **89**(2), 325-351.

Bryant, P. E. (1974). *Perception and understanding in young children*. London: Methuen.

Chalmers, M. (1998). *Identifying core causes of complex cognitive deficit in Fragile X syndrome*. Paper presented at the 8th. Annual meeting of the Society for the Study of Behavioural Phenotypes, Johns Hopkins University , Baltimore.

Chalmers, M., & McGonigle, B. (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. (1997). Capturing dynamic structuralism in the laboratory. In L.Smith & J.Dockrell & P.Tomlinson (Eds.), *Piaget, Vygotsky and Beyond* (pp. 183-200). London: Routledge.

Chomsky, N. (1980). *Rules and representations*. Oxford: Blackwell.

Clark, E. (1973). What's in a word? In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review,* **76**, 387-404.

Clark, H. H. (1970). The primitive nature of children's relational concepts. In J. R. Hayes (Ed.), *Cognition and the development of language*.

Clark, H. H. (1973). Space, time, semantics and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (Vol. New York, pp. 27-63): Academic Press.

Donaldson, M., & Wales, R. J. (1970). On the acquisition of some relational terms. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the national academy of science, USA,* **94**, 7109-7114.

Dykens, E. M., Hodapp, R. M., & Leckman, J. F. (1994). *Behaviour and development in Fragile X Syndrome*. Thousand Oaks: Sage.

Fersen, L. v., Wynne, C. D. L., Delius, J. D., & Staddon, J. E. R. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal BehaviourPprocesses,* **17**, 334-341.

Fischer, K. W., & Canfield, R. L. (1986). The ambiguity of stage and structure in behavior: person and environment in the development of psychological structures. In I.Levin (Ed.), *Stage and Structure: re-opening the debate* (pp. 246-267). Norwood, N.J.: Ablex.

Fodor, J. A. (1983). *Modularity of Mind*. Cambridge: MIT Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture. A critical analysis. *Cognition,* **28**, 3-71.

Gallistel, C. R. (1995). The replacement of general purpose theories with adaptive specializations. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 1255-1267). Cambridge: MIT Press.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (1998). *Cognitive neuroscience*. *The biology of the mind*. New York: Norton.

Gentner, D. (1983). Structure-mapping : A theoretical framework for analogy. *Cognitive Science,* **7**, 155-170.

Halford, G. (1993). *Children's understanding*. *The development of mental models*. Hillsdale: Lawrence Erlbaum Associates.

Halford, G., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity : Implications for comparative, developmental and cognitive psychology. *Behavioral and Brain Sciences,* **21**.

Harlow, H. (1949). The formation of learning sets. *Psychological Review,* **56**, 51-65.

Harnad, S. (1987). *Categorical perception; the groundwork of cognition*. Cambridge: Cambridge University Press.

Hummel, J. E., & Holyoak, K. J. (2001). A process model of human transitive inference. In M. Gattis (Ed.), *Spatial schemas and abstract thought*. London: MIT Press.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child*. London: Routledge and Kegan-Paul.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge: MIT Press.

James, W. (1891). *The principles of psychology*. *Vol II*. London: Macmillan.

Johnson-Laird, P. N. (1983). *Mental models*. Harvard University Press.

Karmiloff-Smith, A. (1992). *Beyond modularity*. London: MIT Press.

Kuenne, M. R. (1946). Experimental investigation of the relation of language to transposition. *Journal of Experimental Psychology,* **36**, 471-490.

Lashley, K. S. (1951). The problem of serial order in behaviour. In L.A.Jeffries (Ed.), *Cerebral mechanisms in behaviour*. New York: John Wiley.

Lawrenson, W., & Bryant, P. E. (1972). Absolute and relative codes in young children. *Journal of Child Psychology and Psychiatry,* **13**, 25-35.

Luger, G. (1994). *Cognitive Science*. *The science of intelligent systems*. San Diego: Academic Press.

McGonigle, B. (1987). Non-verbal thinking by animals. *Nature,* **325**, 110-112.

McGonigle, B., & Chalmers, M. (1977). Are monkeys logical ? *Nature,* **267**, 694-697.

McGonigle, B., & Chalmers, M. (1980). On the genesis of relational terms : A comparative study of monkeys and human children. *Antroplogia Contemporanea,* **3**, 236.

McGonigle, B., & Chalmers, M. (1984). The selective impact of question form and input mode on the symbolic distance effect in children. *Journal of Experimental Child psychology,* **37**, 525-554.

McGonigle, B., & Chalmers, M. (1986). Representations and strategies during inference. In T.Myers & K.Brown (Eds.), *Reasoning and Discourse Processes*. London: Academic Press.

McGonigle, B., & Chalmers, M. (1992). Monkeys are rational ! *The Quarterly Journal of Experimental Psychology,* **45B**(3), 189-228.

McGonigle, B., & Chalmers, M. (1996). The ontology of order. In L. Smith (Ed.), *Critical readings on Piaget*. London: Routledge.

McGonigle, B., & Chalmers, M. (1998). Rationality as optimized cognitive self-regulation. In M. Oaksford & N.Chater (Eds.), *Rational models of cognition* (pp. 501-529).

McGonigle, B., & Chalmers, M. (2001). Spatial representation as cause and effect : Circular causality comes to cognition. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 247-277). London: MIT Press.

McGonigle, B., & Jones, B. (1978). Levels of stimulus processing by the squirrel monkey: Relative and absolute judgments compared. *Perception,* **7**, 635-659.

Morris, R. G. M., Garrud, P., Rawlins, J. N. P., & O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature,* **297**, 681-683.

Moyer, R. S. (1973). Comparing objects in memory : Evidence suggesting an internal psychophysics. *Perception and Psychophysics,* **13**, 180-184.

Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology,* **8**, 228-246.

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgements. *Journal of Experimental Psychology,* **91**, 191-605.

Piaget, J. (1928). *Judgement and reasoning in the child*. London: Routledge and Kegan Paul.

Piaget, J. (1971). *Biology and Knowledge*. Edinburgh: Edinburgh University Press.

Piaget, J., & Szeminska, A. (1952). *The child's conception of number*. London: Routledge and Kegan Paul.

Pylyshyn, Z. W. (1981). The imagery debate. Analogue media versus tacit knowledge. *Psychological Review.*

Reese, H. W. (1968). *The perception of stimulus relations*. New York: Academic Press.

Roberts, W. A., & Phelps, M. T. (1994). Transitive inference in rats: A test of the spatial coding hypothesis. *Psychological Science,* **5**, 368-374.

Russell. (1983). The child's appreciation of necessary truth and necessary falseness of propositions. *British Journal of Psychology,* **73**, 256-266.

Smith, L. (1993). *Necessary knowledge. Piagetian perspectives on constructivism*. Hove: Lawrence Erlbaum.

Spence, K. W. (1937). The differential response in animals varying within a single dimension. *Psychological Review,* **44**, 430-444.

Terrace, H., & McGonigle, B. (1994). Memory and representation of serial order by children. monkeys and pigeons. *Current directions in psychological science*.

Terrace, H. S. (1987). Chunking by a pigeon in a serial learning task. *Nature,* **325**, 149-151.

Trabasso, T. (1977). The role of memory as a system in making transitive inferences. In R. V. k. a. J. W. Hagen (Ed.), *Perspectives on the development of memory and cognition*. Hillsdale: Erlbaum.

Trabasso, T., Riley, C. A., & Wilson, F. G. (1975). The representation of linear order and spatial strategies in reasoning. In R. Falmagne (Ed.), *Reasoning, representation and process*. Hillsdale: Erlbaum.

Wills, S., & Mackintosh, N. J. (1998). Peak shift on an artificial dimension. *the quarterly journal of experimental psychology,* **51B**(1), 1-31.

Wood, D. J. (1978). The nature and development of strategies. In G.Underwood (Ed.), *Strategies for Information Processing*.: Academic Press.

Youniss, J., & Murray, J. P. (1970). Transitive inference with non-transitive solutions controlled. *Developmental Psychology,* **2.2**, 169-175.