

eScholarship

International Journal of Comparative Psychology

Title

A Behavior-Based Fractionation of Cognitive Competence with Clinical Applications: A Comparative Approach

Permalink

<https://escholarship.org/uc/item/6cr1s9kx>

Journal

International Journal of Comparative Psychology, 15(2)

ISSN

0889-3675

Authors

McGonigle, Brendan
Chalmers, Margaret

Publication Date

2002

DOI

10.46867/C44W2G

Copyright Information

Copyright 2002 by the author(s). This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

A Behavior-Based Fractionation of Cognitive Competence with Clinical Applications: A Comparative Approach

Brendan McGonigle and Margaret Chalmers
University of Edinburgh, United Kingdom

We describe experimental techniques based on serial ordering tasks using touch screens, designed to assess memory and high level cognitive organization in human and nonhuman subjects. We demonstrate the applicability of these techniques to a wide range of cognitive competence and executive functioning, illustrating some promising new applications in areas of cognitive dysfunction in humans, specifically Fragile X syndrome and Autistic Spectrum Disorder in children, and Alzheimer's disease and bipolar disorder in adults. We conclude that these techniques have implications both for work in areas of cognitive remediation, as well as in the promotion of animal models of human cognitive function and executive control.

Whilst many cognitive models (Morton, 1980) have been applied in neuropsychology (Shallice, 1988), their traditional dependence on language makes them relatively inaccessible to the comparative psychologist. One major implication of this is a possible limit on the generality of findings from experimental neuroscience using nonhuman subjects to human clinical domains (Morris, 2001). Urgently required in this sphere are purely behavior based tasks that establish both a common currency of comparative measurement and (critically) reflect homologous cognitive processes in humans and nonhumans alike (Conway & Christiansen, 2001; Hauser, Chomsky, & Fitch, 2002; McGonigle, Ravenscroft, & Chalmers, 2002). Confined by convention mainly to associative learning paradigms and simpler forms of adaptation, and buttressed by the argument that these reflect universal learning mechanisms for all (Pearce, 1987), such paradigms have lacked credibility as reflecting the causal antecedents of high level human cognitive and linguistic achievements. Instead, such associative processes are now viewed as representing only one of several, different types of learning (Gazzaniga, Ivry, & Mangun, 1998), each specialized for a different adaptive role. From a de-

The first author acknowledges support from NATO (International Collaborative Grants) with Professor Herb Terrace, Columbia University, New York, the Medical Research Council and the Economic and Social Research Council for their support of much of the comparative and developmental research reported in this paper, Dr Alan Young, Consultant, Newcastle University Department of Psychiatry and Neuroscience, Royal Victoria Infirmary, for financial support of software developments, and to Professor Richard Morris, FRS, co-founder of the Laboratory for cognitive neuroscience for timely practical help and encouragement. The second author wishes to acknowledge the University of Edinburgh's Small Project grant and Development Trust scheme in support of the initial software developments, and the British Academy for a small project grant to support additional software developed in 2001/2002, implemented by Lorenzo Vigentini. She would also like to acknowledge substantial undergraduate involvement in the testing of autistic and Fragile X affected children: V. Bhatt, F. Clark, J. Dean, E. Drury, J. Gray, C. Haworth, A. Hodgson, A. House, J. Pearson, J. Neill, A. Scott, K. Steel. Finally, we are both extremely grateful for helpful edits and constructive criticisms from 2 anonymous referees. Correspondence concerning this article may be addressed to Brendan McGonigle, Laboratory for Cognitive neuroscience, Center for neuroscience, University of Edinburgh, United Kingdom, Appleton Tower, George Square, Edinburgh EH 8 9QJ, United Kingdom (b.mcgonigle@holyrood.ed.ac.uk).

sign stance, furthermore, this system characterization makes better sense. As Pinker (1998, p. 497) puts it:

It is ...important to think of the brain as a family of systems engineered to solve the kinds of problems the organism faced in its evolutionary history...rather than hoping to explain intelligence exclusively with very crude general mechanisms like forming associative bonds.

In line with this view, cognitively oriented comparative research now reveals evidence of a variety of putatively high level functions in nonhumans such as relational learning (McGonigle & Jones, 1978), transitivity (Boysen et al., 1993; McGonigle & Chalmers, 1977, 1992), declarative memory (Dusek & Eichenbaum, 1997; McGonigle & Chalmers, 2002), seriation and hierarchical classification (Johnson-Pynn et al., 1999; McGonigle, Chalmers, & Dickinson, 2003), numerosity and ordering (Brannon & Terrace, 1998) and episodic-like memory (Morris, 2001). Yet the problems of convergence between human and animal research programs in a neuroscience context remain, if only because many of the tests of human cognition that inform clinical assessments are essentially linguistic. Here, analyses of the cognitive processes that lie at the heart of many pathological deficits, such as working and declarative memory (Squire, 1993; Tulving & Markowitsch, 1998) and executive function (Pennington & Ozonoff, 1996) barricade cognition within a symbol-manipulating system where the constituents of the tasks are often symbolic or linguistic as in digit, word and sentence recall tasks (Baddeley, 1974; Tulving & Donaldson, 1972) or their implementation depends on linguistic instruction as in e.g. the Wisconsin Card Sort Task or Self-Ordered Pointing Task (SOPT) (Petrides & Milner, 1982).

Psychometric tests fare little better. Widely used in clinical evaluations of cognitive dysfunction in children and adults, test batteries such as the Stanford-Binet (Terman & Merrill, 1960), WISC-R (Wechsler, 1974) and the WAIS-R (Wechsler, 1981), are also heavily dependent on language based items. Even when test batteries are biased toward nonverbal assessments as with the Kaufman-ABC test instrument for children, the administration of the battery usually requires considerable linguistic instruction.

In summary, behavior based approaches to cognition that eschew language are clearly required if a comparative-informed psychology of cognition is to be successful within a clinical context, but must overcome two main problems. The first is taxonomic: Which nonverbal behaviors can be said to reflect similar complex memorial and cognitive functions as studied in language based tasks? The second is implementational: How (if any are posited) can these be made tractable for research with nonhumans without diluting or dumbing down the target behaviors to the point of making their resemblance with human cognitive achievements merely a superficial one?

A Behavior- Based Taxonomy: Natural Fractionation of Cognitive Systems in Human Development

Confronted with this dilemma, we turned to human cognitive developmental research as a source of inspiration (Chalmers & McGonigle, 1997; McGonigle & Chalmers, 1996, 1998, 2001). In particular, an important window on the “embryol-

ogy of the mind” (Piaget, 1971), has been provided by the extensive use of nonverbal methods with young children. Here competences such as classification, reasoning, and ordering skills are exposed in the course of human development as a series of subskills. Charted from early infancy (Younger, 1993), skills such as object categorization can reveal a natural fractionation of the full blown cognitive system as expressed in adults (Collins & Quillian, 1969). In object sorting tasks (Hayne, Greco-Vitorito, & Rovee-Collier, 1993; Inhelder & Piaget, 1964; Reznick & Kagan, 1983; Sugarman, 1983; Vygotsky, 1962) young children begin to sort and classify in ways that can be seen to develop both qualitatively and quantitatively. At first, forming small groups of objects on the basis of arbitrary or spatial relations, children move on to using perceptual features as a basis for sorting (Inhelder & Piaget, 1964; Langer, 1994; Vygotsky, 1962). The way that children classify from then on shows a change in their comprehension of classificatory structures, from disjoint classes, with no shared features (e.g., rocks and penguins) to reciprocal ones where different classes share properties (e.g., ostriches and penguins) and finally ones which are hierarchically organized with both shared and nonshared properties (e.g., birds and ostriches). All but the last of these have been demonstrated by children sorting actual objects (Inhelder & Piaget, 1964; Sugarman, 1983; Vygotsky, 1962). However, their comprehension of full-blown hierarchical class inclusion principles is traditionally indicated by their answers to questions such as “are there more flowers than daffodils?” Now expressed at the linguistic level, success reflects the knowledge that daffodils are only a subset of the superordinate “flowers” (Inhelder & Piaget, 1964).

In this way, behavioral measures of classification illuminate most of the development of cognitive hierarchical organizational so necessary to linguistic and mathematical skills (Conway & Christiansen, 2001; Hauser et al., 2002). A similar window on the development of important collateral linear organization has been provided by size seriation (Inhelder & Piaget, 1964; Piaget & Szeminska, 1952). Interpreted by Piaget and others as an important precursor of abstract reasoning as measured by linguistic tests of transitive inference (if A is longer than B and B longer than C, which is the longest?), size seriation is still one of the most robust indicators of cognitive growth up the age of 6 years or so (Kingma, 1984). Yet, as in the case of classification, success is not an all-or-none phenomenon; a number of substages of cognitive growth have been identified using object based tasks requiring the ordering of rods of varying sizes into a linear series. In Piaget and Inhelder's classic size seriation task featuring 10 objects (Inhelder & Piaget, 1964), children in the pre-school age range are likely to succeed only with a small subsets of the total test pool, gradually extending their range to larger sets (Kingma, 1984; Young, 1976). Success apart, the procedures that used underlying success also change from trial-and-error based corrections by 5/6 year olds, to the final operational or executive stage achieved around 6/7 years of age where children show principled selection of test objects, starting with for example, the largest object, then continuing in strict descending order of size (Inhelder & Piaget, 1964; Kingma, 1984; Leiser & Gillieron, 1990).

In short, behavior based tasks have been found successful as indicators of both cognitive achievement and cognitive style. In addition, such high level cognitive functioning is not all-or-none, but emerges instead from a layered set of sub-

skills many of which can be assessed without presuming linguistic competence on the part of the subject.

Extending Developmental Tests to Animals

Whilst the cognitive pedigree of behavior based indicators of cognitive growth is unassailable and strongly prescriptive for the comparative psychology of cognition, few tests of this sort can be implemented using the traditional apparatus and procedures of the animal learning laboratory. Many of the latter are based on 2-choice discrimination learning tasks. Adapted for studies of animal reasoning and representation with the additional use of reaction time measures, (McGonigle & Chalmers, 1977, 1986, 1992; McGonigle & Jones, 1978), there is no doubt that these procedures have an important role to play. A persistent feature of many core human developmental tasks, however, is their high response demands, which are integral with the competence under investigation. In classification tasks, these require subjects, for example, to sort, manipulate, and place objects in piles, or, in seriation tasks, to place rods of differing heights in a neat row. Yet monkeys, and even apes, are poor candidates for such tasks as conventionally given to children. Getting monkeys to copy a model series, let alone select blocks from a jumbled set and place them in a row is not a realistic objective (although an easier form of seriation with nested cups has been performed successfully by monkeys and apes where the physical containment relationships between the cups provides an important and useful constraint for the subject; Johnson-Pynn et al., 1999). Otherwise, current unstructured free-sorting tasks, as used to assess classification in monkeys and apes (Langer, 1994; Poti, Langer, & Savage-Rumbaugh, 1990), have almost certainly underspecified the nonhuman primate competence in this domain (McGonigle et al., 2003; Parker & McKinney, 1999).

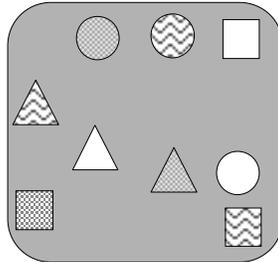
To circumvent the manipulation problem whilst testing subjects under conditions which meaningfully convey the point of the task, we turned to serial ordering tasks, changing the test medium to touch sensitive screens (McGonigle & Chalmers, 1993, 1996; McGonigle et al., in press; McGonigle, De Lillo, & Dickinson, 1992, 1994)

Executive Functioning in Human and Nonhuman Primates: Sequencing Using Touchscreens

Whilst seriation and classification abilities are conventionally demonstrated by the placement of objects into rows or collections, equivalent tasks using touch sensitive screens can be administered by computer without the need for physical manipulation. Here it is necessary only to touch icons displayed on the screen. Whilst each touch is simple to achieve, the *sequence* of such touches is the key to mapping behaviors examined under these conditions with those obtained using conventional tasks. For classification, the items presented can vary in several dimensions (see Figure 1A), offering opportunities to classify by shape by touching, say, all the squares first followed by all the circles followed by all the triangles; for size seriation, the items vary systematically in size alone (see Figure 1B). For purposes of assessing arbitrary list learning, the items may simply consist of a set of unrelated objects (Figure 1C) and for spatial search, would feature identical items. Using

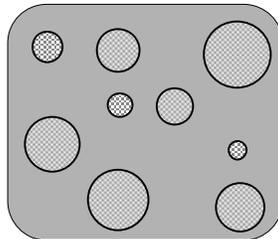
these types of stimuli, we have developed two main types of ordering tasks in research with children (Chalmers & McGonigle, 2003; McGonigle & Chalmers, 1993, 1996; McGonigle et al., 1992) and capuchin monkeys (*Cebus apella*) (McGonigle & Chalmers, 1993, 1998; McGonigle et al., 2003; McGonigle et al., 1992, 1994). In supervised ordering tasks, the subject is explicitly taught a given

A. a classification task :



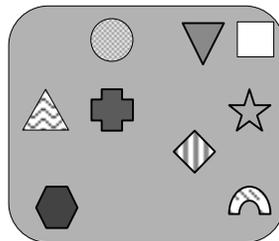
Elements can be touched in order of shape or color

B. a seriation task



Elements can be touched in order of size

c. an arbitrary list task :



Elements have no intrinsic relation to one another other than spatial proximity

sequence. In unsupervised, free ordering tasks, subjects are free to sequence the test items in any order they choose. We summarize these 2 main techniques below, together with the quantitative and qualitative measures that derive from them as a prelude to describing some of our recent clinical applications based on these developments.

Figure 1. Three types of touchscreen task. The learning may be supervised, i.e. all icons must be touched in a specified order, or it may be unsupervised (free search), where the only requirement is that every icon is touched once only.

Interface Between Pure and Applied Research

Supervised Ordering

In this type of sequencing task, subjects must comply with a requirement to order icons as specified by the control program. For classification, the program determines the order in which different shape classes should be selected; for size seriation, the order of sizes. A successful ordering is followed by a reward signal; for monkeys this is a peanut; for human subjects it takes the form of auditory and visual feedback designed to amuse children, or in the case of older subjects, a well-done sign might be flashed on the screen. An ordering mistake in the course of sequencing is conventionally followed by the screen going blank and, after a delay, the icons are presented again in a different spatial configuration. At this point the subject must start the sequence all over again. In all cases, however, a registration signal is given to the subject following each touch. This is to show that the touch (whether correct or incorrect) has been logged by the machine. Our main measures in this context, are trials and errors to criterion, and timing measures, which usefully indicate phrasing and other chunking strategy subjects may use (McGonigle & Chalmers, 1998). Timing measures also clearly indicate the amount of time a subject takes to interrogate the test field before starting the sequence—a measure of forward planning on the part of successful subjects at least — as initial response latency generally reflects the size of the test set (McGonigle & Chalmers, 1998; Terrace & McGonigle, 1994).

Many of our findings from supervised seriation have been reported elsewhere (see McGonigle & Chalmers, 2002, for a review). Suffice it here to report that the techniques have been implemented with success with both children (McGonigle & Chalmers, 1993, 1996, 1998) and monkeys (McGonigle et al., in press). Acquisition measures together with final levels of performance achieved indicate, moreover, that the tasks are highly sensitive to developmental stage as well as phylogenetic status (Terrace & McGonigle, 1994). Finally, as we have argued elsewhere (McGonigle & Chalmers, 1998, 2001, 2002; McGonigle et al., in press), the relative ease with which size seriation and classification have been trained in children and animals under supervised conditions suggests that these forms of cognitive organization are economic for human and nonhuman primates. That is, linear organization by a single dimension such as size, and grouping or classifying by similarity are data *reducing* procedures that minimize cognitive resource—an adaptive principle claimed to be a core feature of human adult cognition (Anderson, 1990). As well as suggesting that supervised procedures on these tasks may offer a window on cognitive dysfunction in an applied context, the further implication already under test by us with normal children and animals (McGonigle & Chalmers, 1998; McGonigle, Ravenscroft, & Chalmers, 2001), is that subjects ought to be able to devise their own strategies to enable economic search without any supervision at all. This led to the second main paradigm we have now applied in a clinical context.

Free Search: Selfinitiated Strategies for Sequencing

Here, explicit training is eliminated in favor of allowing subjects to devise their own means of solution, akin to the sorts of (verbal) tests of subjective memory organization as given by Tulving (Tulving & Donaldson, 1972), in free-recall tasks, in which subjects who have first heard a list of items are free to produce their own order in the recall test. Our free search tasks (McGonigle & Chalmers, 2001, 2002; McGonigle et al., 2001) simply require exhaustive searches of all icons displayed on the screen; the order of icon selection is up to the subject. However, to prevent the use of a simple spatial strategy, the position of all icons in the test array changes following the selection of each icon in the set (see Figure 2). This means that the subject must remember the visual attributes of each test item chosen if they are not to lose track of where they are in the sequence. When this has been achieved, the task terminates with a feedback signal. In contrast to the free recall task of Tulving, our subjects do not need to have a literal recall of all the items. This enables us to evaluate a broader, and arguably a more powerful, range of strategies beyond one of, for example, repeating the same order of test items in successive recall episodes. Whilst the latter procedure is effective, if limited to the actual items used in test, a sorting or categorical strategy by contrast would generalize over a wide range of entirely novel tasks. Crucially, however, it also allows us to detect, in ways not possible under explicit training, what strategies may be used by participants with cognitive dysfunction.

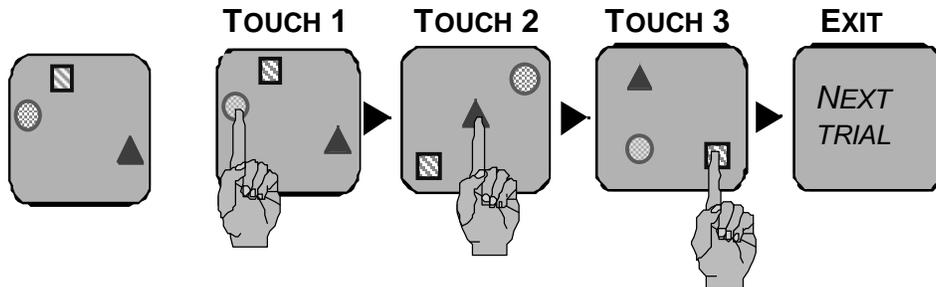


Figure 2. Search based on object properties. To ensure that sequencing is based on the visual properties of the stimuli, they move around after each touch.

Quantitative Measures. Our main quantitative measure is sequencing efficiency; if there are 6 icons on display, then 6 responses, one to each of the individual test items is sufficient to search the set exhaustively. However, if a subject, through reiterations, records a score of 8 responses for the 6 item test, then surplus touches (in this case 2) reflect relative inefficiency. Computed as a simple efficiency ratio (ER), the maximally efficient score of 6 is unity (1.0); if 8 touches have been recorded, it is 0.75, etc. To weight efficiency measures to reflect levels of task difficulty, the ER is multiplied by the number of test icons; this we call the adjusted efficiency ratio or AER (McGonigle et al., 2001). Thus a maximally efficient ER score of 1.0 in a 6 item set produces an AER score of 6.0, in an 8 item set it produces an AER score of 8.0, etc.

Qualitative Measures and Strategic Utility. Whilst (relative) efficiency can be computed objectively as described above, this measure does not in itself indicate *how* the subject is carrying out the task. Given that the responses of the subject are serial, we can compute the ordinal distance functions of items in a sequence, using a Euclidean metric (McGonigle et al., 2001). For example, in a test set of 6 icons composed of 3 categories A, B, C, an order such as AAABBBCCC has a Euclidean Distance (ED) score of 0, as all exemplars of one category are selected before the exemplars of the second, etc. In contrast, an order (with the same level of efficiency) such as ABACCBACB shows considerable ordinal violations of an adjacency principle. There is a total of 4 ordinal steps separating each selection of exemplars from class A, for example, and so the full sequence has a cumulative ED score of 11 and a mean ED of 1.8 (2 for A, 2.5 for B and 1 for C).

We shall now review several applications of our research methods and measures in current assessments of Fragile X syndrome, autism, and the clinical assessments of Alzheimer's and bipolar disorders.

Clinical Applications in Progress

Fragile X Syndrome and Autism

Fragile X syndrome is a neurogenetic disease and the most prevalent known form of developmental disability, affecting an estimated 1 in 2,500. Autistic Spectrum Disorder (ASD) have an as yet unknown etiology but a high and increasing prevalence rate--recently estimated as many as 11 in 1,000 and thought to be rising (Gillberg, Steffenburg, & Schaumann, 1991). Both syndromes show a wide range of dysfunction. Sex-linked, Fragile X syndrome is likely to show a more severe phenotypic expression in males as a double X chromosome effectively dilutes the effect of the mutation in females; severe mental retardation is consequently more likely in males. Taking both genders into account, Fragile X syndrome thus has a phenotypic expression of cognitive function that varies from very mild to severely retarded: 100% of boys and 50% of females with full the mutation have some form of intellectual impairment ranging from moderate to severe (Bregman et al., 1987). For ASD, around 70 % of autistic children are thought to have IQs in the subnormal range (Diagnostic and Statistical Manual of Mental Disorders, 1980) and there is now a widespread belief that there are several subtypes contained within the spectrum, at least some of which are related to very low levels of verbal and nonverbal intelligence (Beglinger & Smith, 2001).

Problems of Cognitive Assessment. Fragile X affected children are selectively impaired on sequential tests such as the digit span subtest of the Wechsler (Kemper et al., 1986), bead memory subtest of the Stanford-Binet (Freund & Reiss, 1991), and picture sequence memory tests in the Kaufman-ABC (Dykens, 1995). Whilst indicating an area for further exploration within this syndrome, conventional IQ tests often fail to properly capture either the extent or the precise nature of the sequential dysfunction in affected individuals (Chalmers, 1997). With characteristically poor or non-existent verbal skills, many low functioning children are often untestable on standard clinical intelligence assessments and their subtests, making low scores difficult to interpret (Tager-Flusberg, 1999).

ASD is also increasingly viewed as a disorder that can affect sequential memory and executive control (Ozonoff, Pennington, & Rogers, 1991; Zelazo et al., 2001), and, accordingly has been described as a frontal syndrome (Pennington & Ozonoff, 1996). However, conclusions regarding executive functioning in ASD have been made hard to achieve through the inconsistent sampling of different levels of mental functioning across the spectrum. Early reports of cognitive skills were usually based on children with verbal and/or nonverbal IQs in the subnormal range (Hermelin & O'Connor, 1970), but there has an increasing tendency to restrict experiments to children with a normal verbal IQ (see, e.g. Burack et al., 2001). One reason for this is that in ASD, investigations of executive functions are often carried out with the aim of clarifying the source of the social dysfunction in autism, which, in turn, is classically measured by a number of verbally administered false belief and social imagination tasks (Happé, 1994, 2001). Few, if any, single studies of executive skills in autism are conducted with participants across the full range of mental functioning found within the spectrum. An important challenge in designing such studies, therefore, is the adequate measurement of performance falling below developmental norms for the age of child being tested.

First Applications of Touchscreen Techniques. The general applicability of our comparative techniques to syndromes where language may be delayed or nonexistent in some participants lies in their essentially nonverbal nature. For ASD furthermore, they make no demand on social cooperation. In contradistinction to one-off clinical tests, furthermore, we can systematically discover the upper bounds of success for each participant within a learning context, as well as identify reasons for failure within individuals, using error analysis. This allows us to make two important distinctions that could apply in a clinical context. First, in the context of difficulty or failure, there is an important difference between restrictions on efficiency that can be traced to ineffective strategies for planning and organization, and those indexed by for example, high levels of reiterations (which could indicate severe memory problems), or high levels of perseverative touching (possibly indicating serious behavioral inhibition deficits). The first source of difficulty is similar to that found in early development, and is more likely to denote retardation of a skill than its absence. The second sort reflects more pathological deviations from a normal developmental trajectory-either in the form of a fundamental memory or information-processing deficit, or in the form of a behavioral or motor abnormality, such as perseverative responding. This distinction could help identify which low-functioning participants are affected by a more widespread neuropathology than that due to the syndrome itself (Joseph, 1999). Successful performance as indexed by efficiency can also be subdivided into that based on generally adaptive strategies, as measured by a good subjective organization score, and success based on idiosyncratic and brittle solutions (such as rote memorizing) that are sometimes thought to be characteristic, for example, of autistic individuals (Frith, 1989).

Preliminary Study: Finessing Procedures for Low-Functioning Subjects. We first applied our techniques to sequential learning in Fragile X syndrome, using a moderately affected fifteen year old Fragile X girl, GY (Chalmers, 1998). Low in verbal competence, our participant had failed to perform on simple tests of sequential memory within the Kaufman-ABC battery, apparently failing to

understand the instructions. Our preliminary study with GY employed a spatial free search task in which identical shapes placed randomly on the screen were to be touched once and once only, and two simple supervised serial learning tasks. One of these supervised tasks involved training on an arbitrary color sequence (red, then green, then blue, etc.), the other on a monotonic size sequence where squares were to be touched from smallest to biggest. All task requirements were made explicit through auditory feedback; a bleep followed a correct touch in the supervised tasks whilst a buzz followed an incorrect touch, and auditory and graphic feedback was used on all tasks indicating a good trial (a smiling face and a “well done G---“ voice-over). She settled well into all the tasks, was compliant with the general requirements to look at the screen and touch the icons. Results clearly showed, however, that whilst capable of controlling a spatial search of up to 9 icons, she was unable to learn a very simple sequence of two colors (or two sizes), as Figure 3 illustrates. GY’s learning on the shape based sequences was thus pathological and seemed relatively unmodifiable through explicit training.

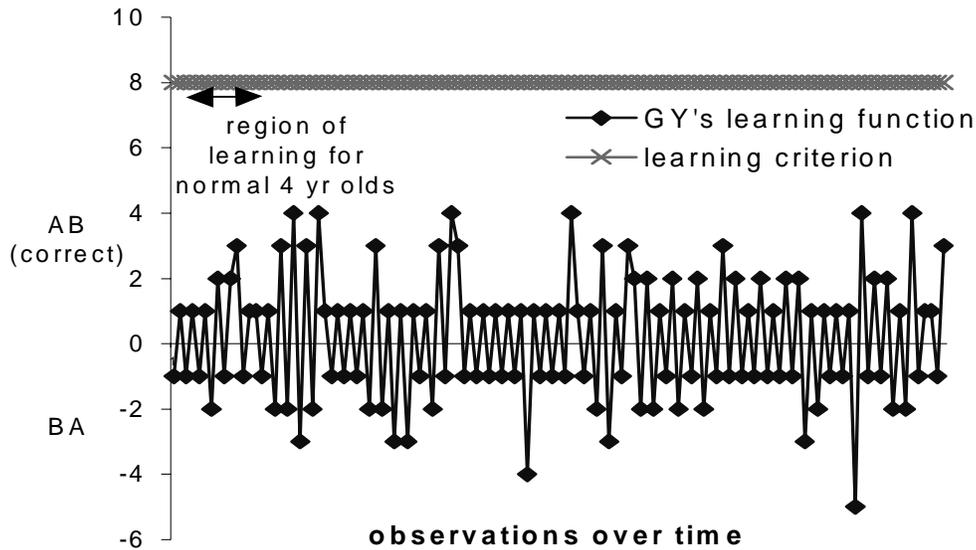


Figure 3. An illustration of zero learning in over 100 trials on a simple (supervised) two-item sequence by a moderately affected Fragile X female, where AB is the correct sequence of colors (or sizes)

Game Based Developments of the Software. The poor performance and restricted learning capability of GY was dramatic, but possibly based on a failure to engage in the point and purpose of the supervised sequential task. Our next pilot attempts were designed to improve motivation to succeed, using tasks that were more like games (i.e., introduced more advanced graphic feedback throughout the task). For example, one game involved frogs going to a party. Positive reinforcement constituted a splash and “ribbit-ribbit” sound and a picture of a smiling frog; incorrect trials were negatively reinforced by withdrawing this feedback and replacing it simply with a low “mmm” sound. However, it was soon clear that low func-

tioning Fragile X children (now represented by two severely affected males) and a low functioning autistic pilot participant were all failing to detect the contingency between a correct (nonreiterative sequence) and a game based reward (e.g., some happy frogs on the screen!). The main study we report below involved a new software development in which the graphic feedback was made more integral with the task. That is, instead of providing differential reinforcement after the touch (and after the stimuli had disappeared from the screen), we provided more immediate and more reinforcing feedback following good (nonreiterative and nonperseverative) touches. This we did by animating the shape icons, and turning them into footballer characters at the moment they were touched, but leaving incorrectly touched icons inanimate, thus selectively reinforcing efficient and nonperseverative touching as vividly as possible. Figure 4 shows these task features as introduced through new software, commissioned by us and implemented by Lorenzo Vigentini.

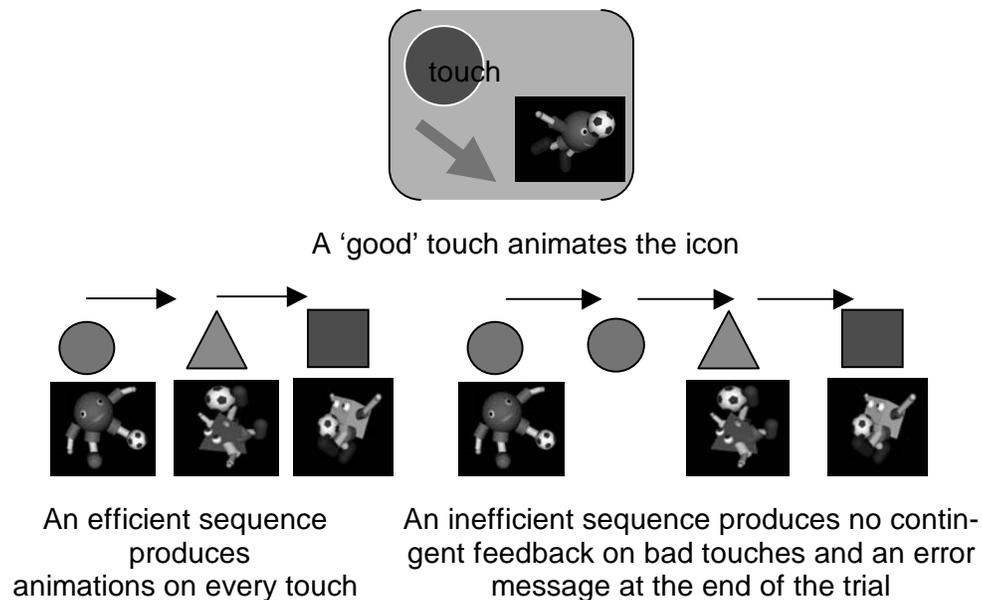


Figure 4. The game based enhancement produced by the CAST software to help support efficient searching

Fractionating Cognitive Deficit within a Single Task. Finessed from the pilot investigations and now incorporating our new game based features, we carried out an investigation using 14 autistic children between the ages of 7 and 14 years (as well as the 2 Fragile X boys from the pilot study). The autistic children represented a wide range of functioning, with some diagnosed as low functioning and with a mental age of around four years; others high functioning and in mainstream schools. As we were using the full spectrum of ability, we now had to confront the issue of measuring cognitive functioning and fractionating cognitive deficit within a single task. As previous research restricted to non-retarded participants between the ages of 7 and 18 years had indicated that there is "intact working memory in autism" (Ozonoff & Strayer, 2001, p.257), it was important in our study to clarify for which participants this was true, and if true, whether just in quantitative terms (measured

as efficiency of search) or, in addition, in qualitative terms (measured in terms of strategic type).

The study we then carried out was a free search task, designed to allow fractionation across the range of cognitive functioning found in Fragile X syndrome and ASD, into age-appropriate and age-subnormal levels of performance. This was implemented by incrementing the task, level by level from 2 icons up to a total of 13. The study was a 2-tier one in which the first levels of the task simply required adequate working memory of three discrete visual items on the screen (e.g., green square, red circle, blue star). From level 4 (4 icons) onwards, however, the stimulus properties began to recur (e.g., now featuring, say, a red square), allowing the participant the use of a grouping principle (e.g., red square, red circle...) in dealing with the increasing working memory demands of the task. This design provided us with a task based fractionation of skill in which gross problems of sequential memory per se would be indicated by a failure to proceed beyond a few items, whilst more subtle problems of executive control at higher overall levels of efficiency could be detected in terms of a failure to impose an effective form of organization. Failure to proceed beyond level 4 was used as a measure of low-scoring performance and would be age-subnormal for any child within our sample.

Results: Quantitative Measures. Using level reached and efficiency measures within each level, we found a bimodal distribution, where 7 of the children were found able to perform with maximum efficiency as measured by AER scores, up to levels equivalent to those achieved by chronologically aged matched peers (i.e., in the range of 7 -13 icons) whilst the other 7 were significantly inferior, measured both by level achieved and mean ER. The level of performance across the spectrum was generally, but not invariably, related to level of functioning as indicated by clinical diagnosis. The upper limit on the sequences achieved by the low scoring subgroup was slightly below that found at around the 3–4 year old age level in normally developing children (an average of 3 icons), and efficiency scores were low. The two Fragile X boys performed at levels similar to the lowest functioning autistic children. Figure 5 shows a scatter plot of level reached by all participants.

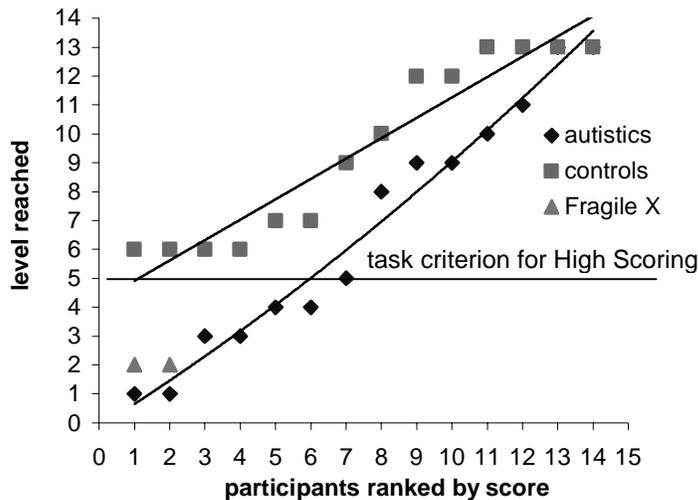


Figure 5. Performance recorded as maximum level of efficient search achieved by autistic children from 6 to 14 years and an age-matched control sample.

Whilst such a bimodal distribution of autistic participants might be expected in theory, in practice it is extremely rare for children from the full autistic spectrum to be compared on a common currency of measurement within a single study. In our study, we could be confident that no participant's scores needed to be removed from the database due to difficulties in interpreting failure; performance in low scoring participants was assessed during test and retest episodes where we could be confident that the final level achieved represented a genuine plateau on performance. We were also able to identify two further subgroups within the low scoring participants. The Fragile X children and two of the lowest scoring ASD children appeared to show a deep and pathological deficit in the ability to use visual features information to control search as measured by high levels of reiterative errors. We discounted any explanation of this based on general noncompliance, as all of these children were, in fact, able to improve their performance when the stimuli did not move around after every touch, but remained static. Whilst some appeared to be reverting to a purely spatial strategy, others began to organize the sequence by color, shape or size, suggesting that the deficit does not lie along a clean break between spatial and object-based memory but along certain spatio-temporal aspects of visual processing that require further analyses (Chalmers & Vigentini 2003). Whilst we could thus eliminate a general problem of short-term memory in the very low scoring children, their performance was more pathological than the other low-scoring subgroup, who, whilst developmentally retarded in terms of their executive memory, were at least recordable as on a normal developmental trajectory in terms both of efficiency and strategy use. Such findings leave open the possibility that the autistic spectrum can include severely affected individuals who should be regarded at least as on the same continuum of executive functioning with higher-scoring children with the same syndrome.

Qualitative Measures: Subtle Differences in High Scoring Autistics.

While many claim that there are no fundamental working memory impairments in autism (Ozonoff & Strayer, 2001), there is a persistent claim that memory organization in such individuals is nevertheless deviant from the neurotypical age matched peers (e.g., Boucher & Lewis, 1989). In our study, we found that efficiency in our high scoring group was at normal levels, whilst qualitative assessments based on clustering using the ED metric described above, produced a significant difference between autistics and age-matched controls. These showed that fewer high scoring autistics than age matched control subjects used consistent spontaneous classification to support their search. As Figure 6 illustrates, this was due to three of the autistic children in particular, and, as also illustrated, these three were the most restricted in terms of upper level reached to criterion. Coupled with the finding that there were lower initial inspection times for the autistic participants in this high-scoring subgroup than for the age matched controls, an implication (and one we are now testing using size seriation tests) is that at least some high functioning autistic children fail to plan an executive task according to the most effective strategy.

Of significance for the claim that autistic individuals sometimes show superior rote memory (Frith, 1989), we found that one (high functioning and high scoring) child in our sample completed the task by remembering the (arbitrary) se-

quence used to generate the stimuli as the task incremented from level to level, noting the new item at each level of task incrementation and adding it on to the end of the list. High in efficiency, this subject failed to demonstrate any classification ability, using instead a brittle strategy that would be maladaptive under other circumstances. Indeed the costs of using this form of verbatim memory were suggested by exceptionally long latencies recorded by this participant.

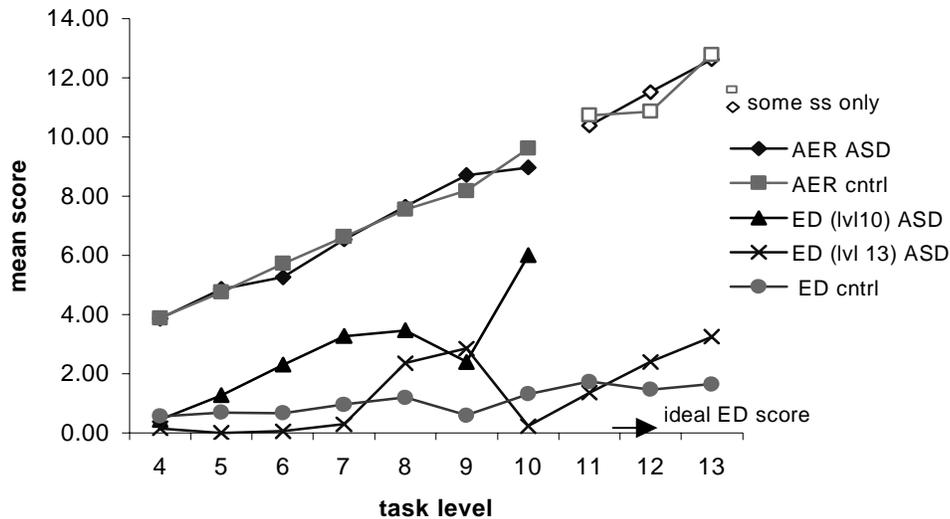


Figure 6. Sequencing efficiency (AER) scores for high-scoring autistic children and age-matched controls compared with strategic organization scores (ED), showing lower organization in the autistic sample, especially for a subgroup would not proceed beyond level 9.

Summary. Now specifically adjusted to support testing of children with Fragile X syndrome and autism, our techniques have been shown suitable for assessing executive functioning in both these syndromes. We have found that children with highly disparate behavioral and intellectual manifestations of the same syndrome can be compared on a common currency of assessment, and where low achievement does not mean 'not scorable'. The combination of qualitative and quantitative assessments, furthermore, has enabled interpretation of success and failure for all subjects at an *individual* level, and thus offers a new methodology that could help inform the issue of how many possible cognitive behavioral sub-types subsume these syndromes - and thus also better inform FMRI and other neuro-anatomical investigations in this area.

Alzheimer's Disease, Bipolar Disorders and Executive Memory

Alzheimer's disease (AD) initially affects the temporal lobe, and, accounting for 50-60% of cases of dementia in people aged over 65 years, is the most prevalent cause of abnormal cognitive decline. The main cognitive deficit is in episodic memory, but other explicit and implicit memory components are also thought to be affected. There is a current need for better forms of assessment both in terms of early screening when cognitive deficits are minor and could lead to easier drug

treatments, as well as in terms of techniques which can track the course of the disorder from early expression to later decline using the same tests. As AD is a multifactorial one, qualitative as well as quantitative measures are needed to portray the various memory and attentional components that may selectively decline in AD as compared with normal memory loss due to aging per se. In particular, the potential loss of effective strategies may have a profound effect on the robustness of information stored by low level encoding methods. Yet the identification of strategies is itself a problem endemic in this area (Lowe & Rabbit, 1998).

As our tasks and our methods of measurement have been designed for both quantitative and strategic assessments, we have begun to assess late Alzheimer's cases. In a preliminary study the first author and Elena Cook, a medical student working under his direct supervision, examined four Alzheimer's patients aged over 75 years, attending a day care center (Cook, 2001). The subjects varied in severity; one was severely affected, the other three moderately so. Stimuli were derived from a set of pictorial stimuli depicting examples from three taxonomic classes (vehicles, animals, plants), as used with nursery children in a previous study (McGonigle et al., 2001). This allowed direct comparison with the performance of very young subjects. The tasks took around 20 min per subject per session and were unpaced.

The free search task was used, which changed the spatial layout of the test icons after each successive touch. To keep patients from becoming unduly discouraged, the maximum number of responses allowed per sequence was capped to a maximum of 4 touches (excluding repeat touches to the same icon) over the number of test icons. Patients were given registration feedback in the form of a bleep heard after each touch to show that it had been recorded by the machine. After each sequence fulfilling the exhaustive search criterion, patients were shown a 'congratulations' sign to the sound of a fanfare. Following a failure to search all test items within the number of touches allocated, a sign on the screen read "unlucky this time..." Following this test phase, the tasks were repeated but hints were given by the tester to help determine if subjects could utilize strategies when being told "look for a way that the objects may fit into categories".

The first finding is that the subjects were compliant overall in coping with the test procedures. However, memory failures surfaced rapidly in both quantitative and qualitative ways. Overall, the length of sequence that could be completed to criterion hovered around 6 test items, a figure well below the 8 or 9 item sequences achieved by 4 year old children. Within sequences completed, moreover, search was not efficient; an ER averaged 0.76 on first trial attempts per problem level. Significantly, in this context, evidence for spontaneous classification was sparse, with an overall average of only 25% of all sequences reflecting categorical use. (This figure declined further when alternative stimuli were used in succeeding tests.) Overall, the number of reiterative errors was so high as to make computation of ED irrelevant. Perseverative behavior in the form of repeat touches to the same test icon was also a feature, with upwards of 6 successive repeat touches being recorded before a subject chose another test icon.

Despite the lack of strong categorical use in spontaneous tests, the results of the augmented tests that provided hints showed a strong contrast with approximately 70% of sequences showing evidence of categorical use. This suggests that spontaneous failures may stem, in part at least, to the lack of an explicit strategy to compensate for other, perhaps more automatic ones, lost in AD. It also suggests, in

common with other evidence, that remediation procedures along these lines may be effective in compensating for some of the deterioration due to this disease (Clare et al., 1999).

Still to be fully analyzed, the overall pattern of results are encouraging us to refine our tests, incorporating some new features to help conduct early and longitudinal screening of mild cognitive deficit in adults of 55 years and over, in collaboration with Dr John Gray, consultant clinical neuropsychologist attached to the laboratory. In these cases, a percentage of the sample will convert to Alzheimer's each year. It is hypothesized that early screening and good markers of the syndrome may be beneficial in enabling drug therapies to maximize the length of remission these currently provide, now spanning from 18 months to 2 years.

Another clinical application comes from work on bipolar affective disorder. This is a recurrent, cyclical disorder that is characterized by recurrent episodes of depression and mania. Positive volumetric MRI findings have mainly been reported in the frontal lobe, medial temporal lobe and striatum (Thompson et al., 2002). Cognitive deficits are also implicated in this condition and failures of executive control are suspected.

To address this issue, a recently completed study has been conducted by a team at the school of Neurosciences and Psychiatry, Royal Victoria Infirmary, Newcastle upon Tyne. Using a version of our free search task and implemented by our own software, Thompson et al (2002) have examined impaired working memory in 20 patients with DSM-IV confirmed bipolar affective disorder, compared with 20 (DSM-IV) screened controls. A CANTAB pattern recognition task was also run on each session. In the Newcastle implementation of our task, arrays of abstract designs were presented and the subject was asked to choose them in any order they wished. Task difficulty was varied with set size levels of 4, 6, 8 and 10, and 3 trials were administered at each level to assess practice effects. Stimuli were conserved within sets, but changed from set level to set level. Unlike our own procedures, however, the total number of responses permitted per subject per trial was strictly limited to the size of the test set so that, for example, if four stimuli, then only four responses were permitted, and the trial was concluded. This study was constrained in the use of measures available in an attempt to keep the task as analogous as possible to a non-computer assisted, verbally administered one called the self ordered pointing task (SOPT). The latter is based on presenting subjects with a series of sheets or cards, and has an established clinical provenance, but different origin. A possible consequence of this reduction in measurement possibilities may be the finding that no significant interaction between group and set size was statistically significant even though errors for both groups rose with set size. Had subjects been allowed to continue within each task beyond the ceiling imposed by set size, and/or had the procedures included a paced condition, it is interesting to speculate just how much more vivid the differences between patients and control subjects might have been. The issue of quantitative differences apart, the exclusive use of arbitrarily composed test arrays precluded assessments of categorical strategies with either group.

Nevertheless, even implemented as a modified SOPT task, the results of the computer based version showed that patients made significantly more errors than controls, suggesting that the deficit is characterized as due to an inability to monitor the contents of working memory. Perseverative errors did not account for this differ-

ence; instead, patients tended to reiterate choices to stimuli chosen earlier in the sequence. The deficit was selective to the executive task, furthermore, as the results of the CANTAB test showed that item recognition performance was not significantly different from that of controls.

Clinical Summary and New Applications from the Program

Although developed within a comparative program, our touchscreen based assessments of executive function have been successfully applied across a spectrum of clinical conditions, even for difficult subjects such as low functioning autistics, and advanced AD patients. Based on a tailoring of interactive feedback to sustain subjects' attention and commitment, our procedures have enabled a learning based, rather than a snapshot based, analysis of individual performance, assessed in both quantitative and qualitative terms. Accordingly, the techniques we describe could contribute to improved diagnostic and remedial programs in these areas.

As for animal models, we seek to develop further the relationship between comparative research and its cognitive neuroscience applications. Our initial, behavior based versions of human inference tasks (McGonigle & Chalmers, 1977, 1992), attempting to achieve this goal, have already had a neuroscience based application in the fractionation of associative and declarative memory in the hippocampus (Dusek & Eichenbaum, 1997). Now exploring executive competences of the sort we have described, which strongly implicate homologous neural substrates in human and nonhumans alike (Conway & Christiansen, 2001; McGonigle et al., 2003), will further this agenda, and will, we believe, offer enhanced clinical neuroscience benefits as a consequence.

References

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale: Lawrence Erlbaum.
- Baddeley, A. D. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.
- Beglinger, L. J., & Smith, T. H. (2001). A review of sub-typing in autism and proposed dimensional classification model. *Journal of Autism and Developmental Disorders*, **31**, 411-422.
- Boucher, J., & Lewis, V. (1989). Memory impairments and communication in relatively able autistic children. *Journal of Child Psychology and Psychiatry*, **30**, 99-122.
- Boysen, S. T., Bernson, G. G., Shreyer, T. A., & Quigley. (1993). Processing of ordinality and transitivity by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, **107**, 208-215.
- Brannon, E. M., & Terrace, H. (1998). Ordering of the numerosities 1-9 by monkeys. *Science*, **282**, 746-749.
- Bregman, J. D., Dykens, E., Watson, M., Ort, S. I., & Leckman, J. F. (1987). Fragile X syndrome : variability of phenotypic expression. *Journal of the American Academy of Child and Adolescent Psychiatry*, **26**, 463-471.
- Burack, J. A., Charman, T., Yirmiya, N., & Zelazo, P. R. (2001). *The development of autism: Perspectives from theory and research*. Mahwah, NJ: Erlbaum.
- Chalmers, M. (1997). *Exploring the cognitive deficit associated with Fragile X syndrome : A case study*. Paper presented at the Society for the Study of Behavioural Phenotypes 7th. Annual Meeting, Cambridge University, Cambridge, United Kingdom.
- Chalmers, M. (1998). *Identifying Core causes of Complex Cognitive Deficit in Fragile X Syndrome*. Paper presented at the Society for the Study of Behavioural Phenotypes 8th. Annual Meeting, Johns Hopkins University, Baltimore, USA.

- 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.
- Chalmers, M., & McGonigle, B. (2003). *Seriation as a window on the development of executive control*. Manuscript in preparation.
- Chalmers, M., & Vigentini, L. (2003). *Sequential processing by autistic children: A learning-based approach to individual assessment*. Manuscript submitted for publication.
- Clare, L., Wilson, B. A., Breen, K., & Hodges, J. R. (1999). Errorless learning of face-name associations in early Alzheimer's disease. *Neurocase*, **5**, 37-46.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **9**, 240-247.
- Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in nonhuman primates. *Trends in Cognitive Sciences*, **5**, 539-546.
- Cook, E. (2001). *Assessment of strategic memory deficits*. Unpublished Undergraduate Dissertation for Final Honours, Edinburgh University, United Kingdom.
- Diagnostic and statistical manual of mental disorders* (1980). (3rd ed.). Washington, DC: American Psychiatric Association.
- 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. (1995). Adaptive behavior in males with fragile X syndrome. *Mental Retardation and Developmental Disabilities*, **1**, 281- 285.
- Freund, L. S., & Reiss, A. L. (1991). Cognitive profiles associated with the fra(X) syndrome in males and females. *American Journal of Medical Genetics*, **38**, 542-547.
- Frith, U. (1989). *Autism. Explaining the enigma*. Oxford: Blackwell.
- Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (1998). *Cognitive neuroscience. The biology of the mind*. New York: Norton.
- Gillberg, C., Steffenburg, S., & Schaumann, H. (1991). Is autism more common now than 10 years ago? *British Journal of Psychiatry*, **158**, 403-409.
- Happe, F. (1994). *Autism: An introduction to psychological theory*. London: University College London Press.
- Happe, F. (2001). Social and nonsocial development in autism: Where are the links ? In P. R. Zelazo (Ed.), *The development of autism. Perspectives from theory and research* (pp. 237-253). Mahwah, NJ: Erlbaum.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language : What is it, Who has it, and how did it evolve ? *Science*, **298**, 1569-1579.
- Hayne, H., Greco-Vitorito, C., & Rovee-Collier, C. (1993). Forming contextual categorizations in infancy. *Cognitive Development*, **8**, 63-82.
- Hermelin, B., & O'Connor, N. (1970). *Experiments with autistic children*. Oxford, UK: Pergamon Press.
- Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child*. London: Routledge and Kegan Paul.
- Johnson-Pynn, J., Frigaszy, D. M., Hirsch, E. M., Brakke, K. E., & Greenfield, P. M. (1999). Strategies used to combine seriated cups by chimpanzees (*Pan troglodytes*), Bonobos (*Pan paniscus*) and capuchins (*Cebus apella*). *Journal of Comparative Psychology*, **113**, 137-148.
- Joseph, R. (1999). Neuropsychological frameworks for understanding autism. *International Review of Psychiatry*, **11**, 309-325.
- Kemper, M. B., Hagerman, R. J., Ahmad, R. S., & Mariner, R. (1986). Cognitive profiles and the spectrum of clinical manifestations in heterozygous fra(X) females. *American Journal of Medical Genetics*, **23**, 139-156.
- Kingma, J. (1984). Task sensitivity and the sequence of development in seriation, ordinal correspondence, and cardination. *Genetic Psychology Monographs*, **110**, 181 -205.
- Langer, J. (1994). From acting to understanding : The comparative development of meaning. In W. E. Overton & D. S. Palermo (Eds.), *The nature and ontogenesis of meaning* (pp. 191-213). Hillsdale, NJ: Erlbaum.
- Leiser, D., & Gillieron, C. (1990). *Cognitive science and genetic epistemology*. New York: Plenum.
- Lowe, C., & Rabbit, P. (1998). Test/Re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Neuropsychologia*, **36(9)**, 915-923.
- McGonigle, B., & Chalmers, M. (1977). Are monkeys logical ? *Nature*, **267**, 694-697.

- McGonigle, B., & Chalmers, M. (1986). Representations and strategies during inference. In T. Myers & K. Brown & B. McGonigle (Eds.), *Reasoning and discourse processes* (pp. 141-164). London: Academic Press.
- McGonigle, B., & Chalmers, M. (1992). Monkeys are rational! *Quarterly Journal of Experimental Psychology*, **45(B)**, 189-228.
- McGonigle, B., & Chalmers, M. (1993). *An experimental analysis of ordering skills in children. Final project report on ESRC grant*. London: British Lending Library.
- McGonigle, B., & Chalmers, M. (1996). The ontology of order. In L. Smith (Ed.), *Critical readings on Piaget* (pp. 279-311). London: Routledge.
- McGonigle, B., & Chalmers, M. (1998). Rationality as optimised cognitive self-regulation. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 501-534). Oxford, UK: Oxford University Press.
- 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., & Chalmers, M. (2002). The growth of cognitive structure in monkeys and men. In S. B. Fountain & M. D. Bunsey & J. H. Danks & M. K. McBeath (Eds.), *Animal cognition and sequential behaviour : Behavioral, biological and computational perspectives* (pp. 269-314). Boston, MA: Kluwer Academic.
- McGonigle, B., Chalmers, M., & Dickinson, A. (in press). Concurrent disjoint and reciprocal classification by *Cebus apella* in serial ordering tasks : evidence for hierarchical organization. *Animal Cognition*.
- McGonigle, B., De Lillo, C., & Dickinson, A. R. (1992). *Serial order induced search in children and monkeys*. Paper presented at the 5th. European Conference on Developmental Psychology, Seville, Spain.
- McGonigle, B., De Lillo, C., & Dickinson, A. R. (1994). *Classification to order: A comparative analysis of categorical seriation in monkey and man*. Paper presented at the XVth. Congress of the International Primatological Society, Bali, Indonesia.
- McGonigle, B., & Jones, B. (1978). Levels of stimulus processing by the squirrel monkey: Relative and absolute judgments compared. *Perception*, **7**, 635-659.
- McGonigle, B., Ravenscroft, J., & Chalmers, M. (2001). *Self organised memory strategies during unsupervised visuo-spatial search tasks by monkeys (Cebus apella) and very young children*. Paper presented at the 8th International Symposium on Imagery and Cognition, St. Malo, France.
- McGonigle, B., Ravenscroft, J., & Chalmers, M. (2002). *Assessing the utility of classification in executive memory tasks by very young children*. Manuscript submitted for publication.
- Morris, R. G. (2001). Episodic-like memory in animals: psychological criteria, neural mechanisms and the value of episodic-like tasks to investigate animal models of neurodegenerative disease. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 1453-1465.
- Morton, J. (1980). The logogen model and orthographic structure. In U. Frith (Ed.), *Cognitive approaches in spelling*. London: Academic Press.
- Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry*, **32**, 1081-1105.
- Ozonoff, S., & Strayer, D. L. (2001). Further evidence of intact working memory in autism. *Journal of Autism and Developmental Disorders*, **31**, 257-263.
- Parker, S., & McKinney, M. (1999). *The evolution of cognitive development in monkeys, apes and humans*. Baltimore, MD: Johns Hopkins University Press.
- Pearce, J. M. (1987). *An introduction to animal cognition*. Hove: Erlbaum.
- Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental pathology. *Journal of Child Psychology and Psychiatry*, **37**, 51-87.
- Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal lobe lesions in man. *Neuropsychologia*, **20**, 249-262.
- 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.

- Pinker, S. (1998). An interview with Stephen Pinker, Ph.D. In M. S. Gazzaniga, R. B. Ivry, & G. R. Mangun (Eds.), *Cognitive neuroscience. The biology of the mind* (pp. 496-497). New York: Norton.
- Poti, P., Langer, J., & Savage-Rumbaugh, S. (1990). Spontaneous logicomathematical constructions by chimpanzees. *Animal Cognition*, **2**, 147-156.
- Reznick, J. S., & Kagan, J. (1983). Category detection in infancy. In L. P. Lipsitt (Ed.), *Advances in infancy research*, Vol. 2 (pp. 79-111). Norwood, NJ: Ablex.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, UK: Cambridge University Press.
- Squire, L. R. (1993). The organization of declarative and nondeclarative memory. In T. Ono, L. R. Squire, M. E. Raichle, D. I. Perret, & M. Fukuda (Eds.), *Brain mechanisms of perception and memory: from neuron to behavior* (pp. 219-227). New York: Oxford University Press.
- Sugarman, S. (1983). *Children's early thoughts: Developments in classification*. Cambridge, UK: Cambridge University Press.
- Tager-Flusberg, H. (1999). An introduction to research on neurodevelopmental disorders. In H. Tager-Flusberg (Ed.), *Neurodevelopmental disorders* (pp. 3-24). Cambridge, MA: MIT Press.
- Terman, L. M., & Merrill, M. A. (1960). *Stanford-Binet intelligence scale, manual for the third revision*. Boston, MA: Houghton-Mifflin.
- Terrace, H. S., & McGonigle, B. (1994). Memory and representation of serial order by children, monkeys and pigeons. *Current Directions in Psychological Science*, **3**, 180-185.
- Thompson, J. M., Gray, J. M., Hughes, J. H., Watson, S., Ferrier, I. N., & Young, A. H. (2002). *Impaired working memory monitoring on the Self-Ordered Pointing Task in euthymic bipolar patients*. Manuscript submitted for publication.
- Tulving, E., & Donaldson, W. (Eds.). (1972). *Organization of memory*. New York: Academic Press.
- Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: role of the hippocampus. *Hippocampus*, **8**, 198-204.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Wechsler, D. (1974). *Manual for the Wechsler intelligence scale for children - revised*. New York: The Psychological Corporation.
- Wechsler, D. (1981). *Manual for the Wechsler intelligence scale - revised*. New York: The Psychological Corporation.
- Young, R. M. (1976). *Children's seriation behaviour - An artificial intelligence analysis of a Piagetian task*. Basel, Switzerland: Birkhauser.
- Younger, B. A. (1993). Understanding category members as "the same sort of thing": Explicit categorization in ten-month-old infants. *Child Development*, **64**, 309-320.
- Zelazo, P. D., Burack, J. A., Boseovski, J. J., Jacques, S., & Frye, D. (2001). A cognitive complexity and control framework for the study of autism. In J. A. Burack & T. Charman & N. Yirmiya & P. D. Zelazo (Eds.), *The development of autism : perspectives from theory and research* (pp. 195-218). Mahwah, NJ: Lawrence Erlbaum.

Received August 22, 2002.

First revision received November 22, 2003.

Second revision received April 4, 2003.

Accepted April 16, 2003.